Notre Dame Journal of Formal Logic
Volume 40, Number 1, Winter 1999

Which Modal Logic is the Right One?

JOHN P. BURGESS

Abstract The question, "Which modal logic is the right one for logical necessity?," divides into two questions, one about model-theoretic validity, the other about proof-theoretic demonstrability. The arguments of Halldén and others that the right validity argument is **S5**, and the right demonstrability logic includes **S4**, are reviewed, and certain common objections are argued to be fallacious. A new argument, based on work of Słupecki and Bryll, is presented for the claim that the right demonstrability logic must be contained in **S5**, and a more speculative argument for the claim that it does not include **S4.2** is also presented.

1 The question Which if any of the many systems of modal logic in the literature is it whose theorems are all and only the right general laws of necessity? That depends on what kind of necessity is in question, so I should begin by making distinctions.

A first distinction that must be noted is between *metaphysical necessity* or *inevitability*, 'what could not have been otherwise', and *logical necessity* or *tautology*, 'what it is self-contradictory to say is otherwise'. The stock example to distinguish the two is this: 'Water is a compound and not an element.' Water could not have been anything other than what it is, a compound of hydrogen and oxygen; but there is no self-contradiction in saying, as was often said, that water is one of four elements along with earth and air and fire.

The logic of inevitability might be called *mood logic*, by analogy with tense logic. For the one aims to do for the distinction between the indicative 'it is the case that . . . ' and the subjunctive 'it could have been the case that . . . ' something like what the other does for the distinction between the present 'it is the case that . . . ' and the future 'it will be the case that . . . ' or the past 'it was the case that . . . '. The logic of tautology might be called *endometalogic*, since it attempts to treat within the object-language notions that classical logic treats only in the metalanguage. However, it hardly deserves a name, since it immediately splits up into two subjects.

For a second distinction must be made between two senses of tautology. On the one hand, there is *model-theoretic logical necessity* or *validity*, the nonexistence of

Received April 16, 1998

a falsifying interpretation, 'being true by logical form alone'. On the other hand, there is *proof-theoretic logical necessity* or *demonstrability*, the existence of a verifying derivation, 'being recognizable as true by logical considerations alone'. Likewise, there is a distinction between two notions of contradiction, model-theoretic satisfiability and proof-theoretic consistency, and between two notions of implication, model-theoretic consequence and proof-theoretic deducibility. There would be at least a conceptual distinction even if logic were understood narrowly as first-order logic, where the model-theoretic and proof-theoretic notions coincide in extension by the Gödel Completeness Theorem. There may be a difference in extension between them when logic is understood more broadly, for instance, if it is taken to include higher-order logic and the mathematics that goes with it.

Logicians often call model-theoretic and proof-theoretic necessity *semantic* and *syntactic* logical necessity. However, there is a conflict between this usage and the older usage of linguists on which, roughly speaking, "semantic" means "pertaining to meaning" and "syntactic" means "pertaining to grammar." There is a conflict between the two usages of "semantics," especially, because there is or may be a gap between mathematical modeling and intended meaning. In any case, shorter labels than 'the logic of semantic logical necessity' and 'the logic of syntactic logical necessity' would be useful. One might use *proplasmatic* logic and *apodictic* logic, from the Greek for model and proof. But it may be more suggestive to use *validity logic* and *demonstrability logic*, by analogy with *provability logic*. The analogy between provability logic and demonstrability logic is especially close, the one being concerned with what a theory can prove, the other with what we can demonstrate, the "can" in each pertaining to ability in principle, regardless of practical limitations.

The question Which is the right system of tense logic? is not one for the logician: the logician can indicate how this or that or the other system corresponds to this or that or the other theory of the nature of time, but which is the right theory of the nature of time is a question for the physicist. Similarly, the question Which is the right system of mood logic? would seem to be one not for the logician, but for the metaphysician. By contrast, the question Which is the right system of validity or demonstrability logic? cannot be passed off by logic to some other discipline.

The question Which is the right validity logic? has been answered at the sentential level, which is the only level that will be considered here: it is the system known as **S5**. This result is essentially established already in Carnap [4].

The question Which is the right demonstrability logic? again at the sentential level, goes back to the earliest days of modern modal logic. For though the founder of the subject, C. I. Lewis, did not clearly distinguish among metaphysical, model-theoretic logical, and proof-theoretic logical modalities, still he did always write of necessitation as implication, and did often write of implication as deducibility, so that it is reasonable to conclude that by necessity he primarily meant tautology, by which in turn he primarily meant demonstrability. No one today, however, takes seriously his suggestion that the right logic for this notion might be the feeble **S1** or the bizarre **S3**. To the extent that there is any consensus or plurality view among logicians today, I take the view to be that the right demonstrability logic is **S4**. (Even in "relevance" or "relevant" logic, where **S4** cannot be literally accepted, since the classical sentential logic it is based on is rejected, still it seems to be a consensus or plurality view that the right logic should be "**S4**-like".) The *locus classicus* for such a view is a paper from the proceedings of a famous 1962 Helsinki conference on modal logic, Halldén [7].

While the argument for the soundness of **S4** as a demonstrability logic given there seems as compelling as an "informally rigorous" argument can be, there is no real argument for completeness, which remains an open question. It therefore remains conceivable that the right logic is something stronger than **S4**: that it is something intermediate between **S4** and **S5**, such as **S4.2** or **S4.3**; or that it is something stronger than **S4** but incomparable with **S5**, such as the logic called **Grz** after Grzegorczyk [6] and the logic that ought to be called **McK** after McKinsey [12]. (In the literature it has heretofore been misleadingly called **S4.1**, though it is not intermediate between **S4** and **S5**.)

The issues are sufficiently illustrated by the cases of the distinctive axioms of **S4.2** and of **McK**, which are equivalent respectively to $\sim(\Box\sim\Box p \wedge \Box\sim\Box\sim p)$, the principle that "nothing is both demonstrably not demonstrably true and demonstrably not demonstrably false," and to $\Box\sim\Box p \vee \Box\sim\Box\sim p$, the principle that "everything is either demonstrably not demonstrably true or demonstrably not demonstrably false." Halldén rightly says of the latter—what he could also have said of the former—that it is not an intuitively plausible principle when the box $\Box$ is meant as demonstrability. But to say this is to do something less than to give an "informally rigorous" argument for the claim that either principle outright fails as a general law, let alone for the claim that any principle not a theorem of **S4** does so.

The question Which is the right provability logic? has been answered, and though results are often stated for a single theory, classical, first-order arithmetic, many hold for all true theories satisfying certain minimum requirements of strength. Actually, one must distinguish the question Which logic gives all and only those principles about provability all whose instances are provable by the theory in question? from the question Which gives all and only those principles that are valid (or demonstable by us)? The answer to the former question is given by a system **GL**, and to the latter question by a system **GLS**. Both differ from the Lewis systems, **S4** and **S5**, by lacking the law $\Box(\Box p \rightarrow p)$. The failure of this law is, roughly speaking, the content of the Gödel Incompleteness Theorems. The standard reference is, of course, Boolos [2].

Below, in Section 2, I will recall the case for the soundness of **S5** as a validity logic and of **S4** as a demonstrability logic. In Section 3, I will recall the Carnapian case for the completeness of **S5**. In Section 4, I will indicate the minimal requirements of strength which are assumed in provability logic, and which I will be assuming in demonstrability logic also, and attempt to clarify the relationship between the two logics. In Section 5, I will present a case against **McK** as a demonstrability logic; and it will generalize to a case against any system not contained in **S5**, such as **Grz**. Finally, in Section 6, I will present a case against **S4.2**; and this will, of course, also constitute a case against any stronger system, such as **S4.3**. But the case of weaker systems intermediate between **S4** and **S5** will be left open and with it the general question.

2 Soundness A key consequence of the step of treating modality in the object language, treating $\Box$ as a one-place connective on a par with $\sim$, is that *iterated* modalities, modalities embedded inside modalities, as in $\Box\sim\Box p$, are allowed. By contrast, when “valid” is expressed only by a word of the metalanguage, applicable only to formulas of the object language, there can be no question of iterations such as ‘it is valid that it is not valid that . . . ’. All the modal systems most commonly considered in the literature agree with classical metalogic, in the sense that if where classical metalogic has a law, for example, ‘a valid conclusion is a consequence of any premise’, these systems will have a corresponding law, in the example $\Box\, p \to (q \Longrightarrow p)$. Agreeing as they all do with classical logic, these systems agree with each other for formulas without iterated modalities. What distinguishes **S5** is that it has laws that make every iterated formula more or less trivially equivalent to an uniterated formula.

The first step in establishing **S5** as the right logic of validity is to establish the soundness of **S5** as a validity logic: to establish that every theorem of **S5** is correct as a general law about validity, or what comes to the same thing, that every axiom of **S5** is thus correct, and that every rule of **S5** preserves such correctness. This is completely unproblematic for the nonmodal axioms and rules, which are simply those of the sentential component of classical, nonmodal sentential logic. Moreover, though **S5** is usually formulated with a specifically modal rule allowing $\Box A$ to be taken as a theorem whenever A is a theorem, this rule can be dispensed with in favor of adding $\Box A$ as an axiom whenever A is an axiom of the usual formulation, which is to say, whenever A is either a classical, nonmodal axiom or one of the following modal axioms:

1. $\Box p \to p$
2. $\Box(p \to q) \to (\Box p \to \Box q)$
3. $\Box p \to \Box\Box p$
4. $\sim\Box p \to \Box \sim\Box p$

Again for the classical, nonmodal axioms this is completely unproblematic, while in making a case for—which is to say, in demonstrating—any one of (1) – (4), one will at the same time be making a case for its demonstrability, and a fortiori for its validity.

Thus the problem of establishing the soundness of **S5** for validity logic reduces to that of establishing the correctness of (1) – (4), and similarly the problem of establishing the soundness of **S4** for demonstrability logic reduces to that of establishing the correctness of (1) – (3). Indeed, for (1) – (3) correctness for the box as validity and for the box as demonstrability can be established by more or less parallel arguments. Consider (1), for example. The arguments are simply the parallel ones that whatever is true by logical form alone must be true, and that whatever can be recognized to be true by logical considerations alone must be true. But indeed, I need not enlarge on the case for (1) – (3), which is adequately made by Halldén.

It remains to consider the distinctive axiom 4, which, of course, is being proposed only as a general law of validity, not of demonstrability. Here the main point is just as follows. Consider any particular instance of (4):

5. If it is not true by logical form alone that π, then it is true by logical form alone that it is not true by logical form alone that π.

Suppose that the antecedent of (5) is true, which is to say that the following is false:

6. It is true by logical form alone that π.

Since (6) is false, there must be some ψ of the same logical form as π that is false. Now consider anything else of the same logical form as (6). It will look like the following, wherein ρ has the same logical form as π:

7. It is true by logical form alone that ρ.

But then ψ also has the same logical form as ρ, and since ψ is false, (7) is false. In other words, anything of the same logical form as (6) is false, and hence the following is true:

8. It is true by logical form alone that it is not true by logical form alone that π.

Thus the consequent of (5) is true, as required.

3 Completeness for validity logic In the case of provability logic, as expounded in [2], and of intuitionistic logic, as expounded in Burgess [3], once the candidate logic **S** has been identified, the argument that it is the right one consists of three parts: soundness, *formal* completeness, and *material* completeness. That is, it is shown that every theorem of **S** is acceptable as a general law under the intended interpretation; a class Σ of mathematical models is identified and it is shown that (every theorem of **S** and) no nontheorem of **S** comes out true in all models of class; and it is shown that any formula that comes out untrue in some model of class Σ is unacceptable as a general law under the intended interpretation. This last step, bridging the gap between "semantics" in the logicians' sense and "semantics" in something more like the linguists' sense, is due in the case of provability logic to Solovay, and in the case of intuitionistic logic to Kreisel, who coined the phrase "informal rigor" in this connection. In both cases, the last step is the most difficult. The situation is rather similar in the case of validity logic.

Beginning with formal completeness for the case of validity logic (soundness having already been discussed), the kind of models now standard in modal logic first became widely known through another talk at the 1962 Helsinki conference, this one by Kripke; another kind of model near to those standardly used became widely known through yet another talk at the same conference, this one by Hintikka. A *frame* model M, as in Kripke [9], consists of two parts, a *frame* and a *valuation*. The frame consists of a set W of indices, a two-place relation R on it, and a designated member w_0 of it. A valuation V is a specification for each x in W and each sentence letter $p, q, r, \ldots$ of whether or not the sentence letter counts as true in that index. The notion of truth in an index is extended to compound formulas by recursion:

$\sim A$ is true at x if and only if A is not true at x.
$A \wedge B$ is true at x if and only if A is true at x and B is true at x.
$\Box A$ is true at x if and only if A is true at y for every y in W such that Rxy.

It is permitted to have two indices x and y at which exactly the same set of formulas is true, and such *duplication* is often important. A formula counts as holding in M if it is true at w_0, and as being valid in a class K of frames if it holds in every model whose frame is in that class.

The proposal in Hintikka [8] is less purely "semantic" or model-theoretic: it is still "syntactic" or proof-theoretic in that, whereas it has a relation R, what this relation relates are not abstract indices, but sets of formulas, and so one does not have duplication. But as it happens, in connection with the system **S5**, differences between the approaches of Kripke and Hintikka, such as permitting or forbidding duplication, are unimportant, and so for that matter is the main similarity between the two approaches, the presence of a relation R. For while the theorems of **S5** can be characterized as the formulas valid for the class of reflexive, transitive, and symmetric frames, this characterization reduces, by a series of steps too familiar to bear repetition here, to a much simpler one.

Consider, for any k, the formulas involving only the sentential variables or atomic formulas $p_1, \ldots, p_k$. Then for such formulas, a model may be taken to consist simply of a nonempty subset W of the set of rows of the truth table $p_1, \ldots, p_k$, with one such row w_0 designated. The notion of truth at a row x in the truth table is defined for compound formulas by a recursion in which the first two clauses are exactly the same as above, while the third reads as follows:

$\Box A$ is true at x if and only if A is true at y for every y in W.

A formula counts as holding in such a model M if it is true at w_0. The theorems of **S5** may be characterized as the formulas that hold in all such models.

Turning to material completeness for the case of validity logic, it may be well to begin by considering an objection to the modeling just described that has been independently advanced by several writers. One of them put it as follows:

> What is needed for logical necessity of a sentence p in a world w_0 is more than its truth in each one of some arbitrarily selected set of alternatives to w_0. What is needed is its truth in each *logically possible* world. However, in Kripke semantics it is not required that all such worlds are among the alternatives to a given one.

It is then suggested that one should adopt not the standard model theory, nor the simplification thereof described above, but rather a deviant model theory, which after simplification amounts to just this, that the only model admitted is the one consisting of *all* rows of the truth table.

There is a fallacy or confusion here. What is wanted is that the technical notion of coming out true in all models should correspond to the intuitive notion of coming out true under all interpretations, or all substitutions of specific $\pi_1, \ldots, \pi_k$ for the variables $p_1, \ldots, p_k$. Since, for instance, among all the many substitutions available there are ones in which the π_1 substituted for p_1 *is the same as* the π_2 substituted for p_2, so that it is impossible for π_1 and π_2 to have different truth values, there must correspondingly be among the models one available where the only rows of the truth table present are those for which the value given to p_1 *is the same as* the value given to p_2.

The confusion in the objection becomes apparent when one notes that in the deviant model theory suggested, $\sim\Box\sim(p_1\wedge\sim p_2)$ counts as valid, whereas, of course, $\sim\Box\sim(p_1\wedge\sim p_1)$ does not, so that the standard rule of substitution fails. But the rule of substitution must hold so long as one adheres to the standard conception of the role of the variables $p_1, \ldots, p_k$, according to which arbitrarily selected $\pi_1, \ldots, \pi_k$ may be substituted for them. Indeed, the deviant model theory corresponds to a deviant conception on which *independent* $\pi_1, \ldots, \pi_k$ *must be substituted for distinct* $p_1, \ldots, p_k$. The confusion is more deeply confounded when it is suggested that the difference between the standard and deviant model theories somehow corresponds to a difference between nonlogical and logical notions of necessity. For what is at issue are, to repeat, differences in conceptions of the role of variables, not in conceptions of the nature of necessity.

Yet, confused as it is, the objection does serve to call attention to an important question. Each substitution of specific $\pi_1, \ldots, \pi_k$ for $p_1, \ldots, p_k$ determines a nonempty set of rows of the truth table, consisting of all and only those rows x such that it is not impossible by the logical forms of the π_i alone for them to have the truth values x assigned to the corresponding p_i. The question is, is it the case that for any arbitrarily selected nonempty subset W of the set of rows of the truth table, there are specific statements, specific $\pi_1, \ldots, \pi_k$, that determine, in the manner just described, exactly that subset? In other words, if a formula is not a theorem of **S5**, and therefore fails in some standard model, is there some specific instance in which it fails? An affirmative answer to this question is precisely what is needed to establish the material completeness of **S5** as validity logic.

It is a reasonable assumption, and one presumably made by the critics alluded to, that there exist indefinitely many $\alpha, \beta, \gamma, \ldots$ that are independent in the sense that any conjunction of some of them with the negations of the rest of them is possible, in the relevant sense of possibility. For instance, if $\alpha, \beta, \gamma, \ldots$ are of simple subject-predicate form with distinct subjects and predicates in each, they will be thus independent. Given this assumption, an affirmative answer to the foregoing question is forthcoming. As this result has in effect already been expounded several times in the literature, in [4] and Makinson [10] and Thomason [16], there should be no need for me to do more than give an illustrative example here. Indeed, a simple one, involving just three sentence letters $p, q, r,$ should suffice.

Consider the set W containing just the three rows in which two of p, q, r are true and the other false. Call the one where r alone is false x, the one where q alone is false y, and the one where p alone is false z. What is to be established is that given independent $\alpha, \beta, \ldots$, there are truth-functional compounds π, ψ, ρ thereof that might be substituted for $p, q, r,$ for which the three rows indicated represent all and only the combinations of truth values that are not false by logical form alone.

To find the required compounds, one first finds three auxiliary compounds ξ, υ, ζ, that are pairwise exclusive and jointly exhaustive, meaning that the conjunction of any two must be false by logical form alone, while the disjunction of all three must be true by logical form alone. Setting $\xi = \alpha$ and $\upsilon = \sim\alpha \wedge \beta$ and $\zeta = \sim\alpha \wedge \sim\beta$ will do. One next lets the auxiliaries ξ, υ, ζ correspond to the rows x, y, z, and takes as the substitute for a given one of p, q, r the disjunction of the auxiliaries corresponding to the rows in which it is true. Thus the substitute π for p should be $\xi \vee \upsilon$ or $\alpha \vee (\sim\alpha \wedge \beta)$, which simplifies to $\alpha \vee \beta$. It can be worked out that the substitutes ψ and ρ for q and r simplify to $\alpha \vee \sim\beta$ and $\sim\alpha$, respectively. And it can then be worked out that exactly two of the three, $\alpha \vee \beta$ and $\alpha \vee \sim\beta$ and $\sim\alpha$, must be true, and that given the independence of α and β it may be any two of the three, as required.

Before leaving the topic of validity logic it may be mentioned that the fact that **S5** is indeed the right logic can be confirmed in a different way. After soundness is established in order to show that no stronger system than **S5** is acceptable, one would appeal to the result of Scroggs [13], according to which the only extensions of **S5** are finitely many-valued logics. One would then argue that no finitely many-valued logic can be correct for semantic logical necessity (given the same reasonable assumption as above, that there are indefinitely distinct independent statements).

4 Demonstrability and provability A word must be said to clarify the relationship between demonstrability and provability logics and to dispel a puzzle about that relationship.

The minimal assumptions of strength needed for provability logic are three. They may be formulated either as assumptions on the notion of proof-for-the-theory, or as assumptions on the set of theorems provable. On the formulation in terms of proofs, the first assumption would be that whether something is or is not a proof-in-the-theory is decidable, which by Church's Thesis implies that the relation of proof-in-the-theory to theorem proved is recursive. The second assumption would be that the rules of classical first-order logic may be used in proofs-in-the-theory. The third assumption would be that certain basic, finite, combinatorial modes of reasoning—whose exact scope need not be discussed here, except to say that, since we want to get the Second Incompleteness Theorem, the scope needs to be somewhat wider than it would need to be if we only wanted to get the First Incompleteness Theorem—may be used in proofs-in-the-theory.

On the formulation in terms of theorems, it would first be assumed that the set of theorems provable is recursively enumerable. It would second be assumed that the set of theorems provable is closed under the rules of classical first-order logic. And it would third be assumed that the set of theorems provable includes certain basic, finite, combinatorial results. Clearly, the list of assumptions on the theory stated earlier yields the list of assumptions on the set of theorems just stated. And conversely, by Craig's Lemma, any set of conclusions satisfying this latter list of assumptions coincides with the set of conclusions provable in some theory satisfying the former list of assumptions.

In demonstrability logic, at least as I will be considering it here, it is to be assumed that whether something constitutes a demonstration of a given conclusion is decidable, that the rules of classical first-order logic may be used in demonstrations, and that certain basic, finite, combinatorial modes of reasoning may be used in demonstrations. By what has already been said, it follows that the set of conclusions we can demonstrate coincides with the set of conclusions that can be proved in some theory of the kind to which provability logic applies. And yet, provability logic and demonstrability logic are supposed to be different, in that by what has been said in earlier sections, (1) below is false, while (2) below is true.

1. It can be proved in such-and-such theory **T** that if something can be proved in such-and-such theory **T**, then its negation cannot also be.
2. It can be demonstrated by us that if something can be demonstrated by use, then its negation cannot also be.

What may be puzzling is how it can be that (1) fails while (2) holds and as already indicated the above-stated assumptions commit one to the truth of something of the following form:

3. What can be demonstrated by us coincides with what can be proved in such-and-such theory **T**.

Indeed, a notorious objection to (3) above, associated with the names of Lucas and Penrose, claims that it, together with the true (2) above, yields the false (1) above.

The solution to the puzzle is to point out that this objection commits the fallacy of assuming that co-extensive terms, such as 'what we can demonstrate' and 'what such-and-such theory **T** can prove' can be substituted without change of truth value everywhere, even in intensional contexts, such as 'we can demonstrate that . . . ' or 'such-and-such theory **T** can prove that…'. To get (1) above from (2) above one would need something stronger than (3) above, namely, the following:

4. We can demonstrate (3).

The solution of the puzzle is that (3) does not yield (4). (By analogy, those familiar with the work of Feferman and Tait on the standpoints of the predicativists and the finitists, thinkers who owing to their philosophical prejudices cannot demonstrate all that we can, will recall that what is demonstrable from the standpoint of one or the other of these *ism*s can indeed be exactly captured by a theory, though the *ist*s themselves cannot recognize as much.)

The fallacy should become obvious on comparing (1) – (3) above with the following:

1′. The only man with such-and-such number **N** of hairs on his head knows that the only man with such-and-such number **N** of hairs on his head is gray haired.

2′. I know that I am gray haired.

3′. I am the only man with such-and-such number **N** of hairs on his head.

Clearly (3′) above by itself does not, with (2′) above, yield (1′) above. Rather, one would need the following stronger assumption.

4′. I know (3′).

But (3′) above does not yield (4′) above.

***5 Against* McK** Beginning with formal completeness for the case of demonstrability logic, soundness having already been discussed, the theorems of **S4** can be characterized as the formulas valid for the class of reflexive and transitive frames; and equally, they can be characterized as the formulas valid for the class of finite reflexive and transitive frames, a deeper result implying the decidability of the logic. An historical fact is worth mentioning, that the result just stated follows immediately from two results already in the literature two decades before the famous Helsinki conference. One of these, from McKinsey [11], characterizes the theorems of **S4** in terms of a class of finite models of a different kind, based not on frame structures but on algebraic structures of a certain kind. The other of these, from Birkhoff [1], connects finite algebraic structures of the kind in question with finite reflexive and transitive frame structures. (The former paper makes no mention of frames and the latter no mention of modal logic.) This history is worth mentioning among other reasons because the older algebraic modeling involved, which is sometimes not taught to students of the subject today, still has its uses even after the development of frame models, and I will be citing an instance later.

Turning to material completeness, no decisive results have yet been obtained, and my aim will be only to present, case by case, some partial results. To begin with, it is not only a reasonable assumption, as already said in an earlier section, that there exist indefinitely many $\alpha, \beta, \gamma, \ldots$ that are independent, but also a reasonable assumption that there exist indefinitely many $\alpha, \beta, \gamma, \ldots$ that are *demonstrably* independent. For instance, if $\alpha, \beta, \gamma, \ldots$ are *recognizably* of simple subject-predicate form with distinct subjects and predicates in each, they will be thus demonstrably independent. It follows that if D is any compound formed by negation and conjunction or disjunction from $p, q, r, \ldots$ such that D is a not a theorem of classical sentential logic, or in other words, such that D comes out false in some row of the pertinent truth table, then the result Δ of substituting $\alpha, \beta, \gamma, \ldots$ for $p, q, r, \ldots$ in D will be demonstrably not demonstrable, or demonstrably *indemonstrable*. Similarly, if D is such that $\sim D$ is not a theorem of classical sentential logic, or in other words, such that D comes out true in some row of the pertinent truth table, then Δ will be demonstrably not demonstrably false, or demonstrably *irrefutable*. Thus from any D such that neither it nor its negation is a theorem of classical sentential logic, we get a counterexample Δ to the **McK** axiom that nothing is both demonstrably indemonstrable and demonstrably irrefutable. This argument can be generalized to apply to any axiom that is not a theorem of **S5**.

Perhaps the easiest route to a generalization is to draw on the work of Słupecki and Bryll [15]. They pursue the old idea of the Polish school that a logic should have in addition to its axioms and rules of the ordinary kind, its axioms and rules of acceptance, indicating that certain formulas are acceptable as general laws, some axioms and rules of an opposite kind, axioms and rules of rejection, indicating that certain formulas are unacceptable as general laws. Just as a formula P is a theorem of the system, in symbols $\vdash P$, if there is a sequence of steps, each an axiom of acceptance or following from earlier ones by a rule of acceptance ending in P, so the goal would be to have for each formula Q that is not a theorem of the system, in symbols $\dashv Q$, a sequence of steps involving axioms and rules of rejection ending in Q.

For classical logic there would be the axiomatic rejection of the constant false $\dashv \bot$ and rules of rejection that are the reverse of the usual classical rules of acceptance: if $\dashv P'$ where P' is a substitution instance of P, then $\dashv P$, and if $\dashv Q$ and $\vdash P \supset Q$, then $\dashv P$. For any modal logic there would be also the rule of rejection that is the reverse of the usual modal rule of acceptance: if $\dashv \Box P$ then $\dashv P$. For each particular modal system additional rules of rejection would be needed. For **S5** Słupecki and Bryll show that just one additional rule suffices.

$$\text{if } \dashv P \rightarrow Q_1 \text{ and } \cdots \text{ and } \dashv P \rightarrow Q_N \text{ then } \dashv \Box P \rightarrow (\Box Q_1 \vee \cdots \vee \Box Q_N)$$

where P and the Q_i involve no modalities.

In order to show that any nontheorem of **S5** should be rejected as a general law of demonstrability, it will suffice therefore to argue that the above rule of rejection is acceptable for demonstrability. And indeed, if the $P \rightarrow Q_i$ are unacceptable as general laws, there must for each be a row x_i of the truth table for the variables $p, q, \ldots$ involved on which P comes out true and Q_i comes out false. But then by Carnap's result there are specific π, ψ, ... that could be substituted for the variables $p, q, \ldots$ in P and the Q_i to give sentences Π and Ψ_i, such that the x_i represent all and only the possible combinations of truth values for the π, ψ, ... It follows that Π is an instance of a theorem of classical sentential logic, hence demonstrable, while each Ψ_i is demonstrably indemonstrable by the considerations of the preceding paragraph. Thus the following

$$\Box\Pi \rightarrow (\Box\Psi_1 \vee \cdots \vee \Box\Psi_N)$$

fails, and the formula of which it is an instance, namely, the following,

$$\Box P \rightarrow (\Box Q_1 \vee \cdots \vee \Box Q_N),$$

is not acceptable as a general law, as required.

***6 Against* S4.2** The **S4.2** principle says that everything is either demonstrably indemonstrable, or demonstrably irrefutable. An argument against the acceptability of this principle as a general law can be given.

In addition to assumptions about demonstrability listed in earlier sections, including the assumption that there are indefinitely many instances recognizable as being of simple subject-predicate form, I need the reasonable assumption that there is an instance recognizable as being of simple subject-verb-object form, with no further pertinent logical structure. That is, I assume there is a two-place predicate Φ that is recognizably a two-place predicate with no further pertinent logical structure, so that it is recognizable that all its pertinent logical structure is represented when it is represented by a simply two-place predicate variable F. Given such an example, for any compound Π formed from Φ using negation, conjunction or disjunction, and universal or existential quantification, the logical form of Π will recognizably be represented by a formula P of classical first-order logic formed from F using $\sim$, $\wedge$ or $\vee$, and $\forall$ or $\exists$. Let Λ be the set of all such compounds Π, and L the set of the corresponding formulas P.

Now suppose a formula P in L fails in no model of the kind used in classical first-order logic. Then P is a theorem of classical first-order logic by the Gödel Completeness Theorem; and by the assumption that the rules of classical first-order logic may be used in demonstrations, the corresponding Π in Λ will be demonstrable, and demonstrably so. Then by general laws represented by theorems of **S4**, ‘it is demonstrable that Π’ will be demonstrably irrefutable and not demonstrably indemonstrable.

Now suppose the formula P of L fails in some *finite* model of the kind used in classical first-order logic. Then basic, finite, combinatorial reasoning shows that it does so, and hence that it does not represent a correct general law; and by the assumption that basic, finite, combinatorial reasoning may be used in demonstrations, the corresponding Π in Λ will be indemonstrable, and demonstrably so. Then by general laws represented by theorems of **S4**, ‘it is demonstrable that Π’ will be demonstrably indemonstrable and not demonstrably irrefutable.

For any formula P of L, let $\Psi(P)$ be ‘it is demonstrable that Π’, where Π in Λ is the result of substituting Φ for F. Let X be the set of P such that $\Psi(P)$ is demonstrably irrefutable, and let Y be the set of P such that $\Psi(P)$ is demonstrably indemonstrable. What has been established so far is that if P has no countermodel, then P belongs to the difference set $X - Y$; while if P has a finite countermodel, then P belongs to the difference set $Y - X$. What the **S4.2** principle yields is that the union set $X \cup Y$ is all of L.

To complete the case against the **S4.2** principle, I must invoke the assumption that the set of demonstrable conclusions is recursively enumerable, from which it follows that the sets X and Y are also recursively enumerable. Then by the Reduction Theorem for recursively enumerable sets it follows that there are recursively enumerable sets X^* and Y^* satisfying the conditions that $X^* \subseteq X$ and $Y^* \subseteq Y$ and $X^* \cap Y^* = \varnothing$ and $X^* \cup Y^* = X \cup Y$. These conditions imply that $X - Y \subseteq X^*$ and $Y - X \subseteq Y^*$. What the **S4.2** principle yields is that X^* and Y^* are complements of each other in the recursive set L, which since both are recursively enumerable yields that both are recursive. What was established earlier yields that if P has no countermodel, then P belongs to X^*, while if P has a finite countermodel, then P belongs to Y^* and so does not belong to X^*. And now we have a contradiction, since by an elaborated version of Church's Theorem, there is no recursive set Z separating the formulas with no countermodels from those with finite countermodels.

The foregoing argument applies to just one of the infinitely many formulas that are theorems of **S5** but not of **S4**. Can *all* such formulas be rejected? Clearly, if they can, some general argument, not a case by case examination of examples, will be needed to establish that fact. How might such an argument proceed? Well, rejection principles for **S4** have been formulated by Goranko [5], and simpler ones have been found by Skura [14], who works with finite algebraic models of the kind alluded to earlier. Skura requires two principles, the first being a slight variant of the rejection principle for **S5** considered earlier.

Unfortunately, Skura's second principle, though simpler than Goranko's principles, is complex enough that it is not very perspicuous, and it is not easy to argue why it should be acceptable for syntactic logical necessity. (Goranko and Skura do not themselves consider such questions.) Fortunately, Skura does not claim his rules are the simplest feasible, but on the contrary he explicitly poses it as an open question whether there are any simpler ones. It may be that this open question will have to be settled before one can settle the status of the conjecture that **S4** provides the answer to the question of which is the right demonstrability logic.

At present this question which, as I have said, goes back to the founders of modern modal logic, remains after most of a century still without a definitive answer.

REFERENCES

[1] Birkhoff, G., "Rings of sets," *Duke Mathematical Journal*, vol. 3 (1937), pp. 443–54. Zbl 0017.19403 5

[2] Boolos, G., "The logic of provability," *Cambridge University Press*, Cambridge, 1993. Zbl 0891.03004 MR 95c:03038 1, 3

[3] Burgess, J. P., "The completeness of intuitionistic propositional calculus for its intended interpretation," *Notre Dame Journal of Formal Logic*, vol. 22 (1981), pp. 17– 28. Zbl 0445.03030 MR 82g:03015 3

[4] Carnap, R., "Modalities and quantification," *The Journal of Symbolic Logic*, vol. 11 (1946), pp. 33–64. Zbl 0063.00713 MR 8,429c 1, 3

[5] Goranko, V., “Refutation systems in modal logic,” *Studia Logica*, vol. 53 (1994), pp. 299–324. Zbl 0803.03005 MR 95h:03042 6

[6] Grzgorczyk, A., “Some relational systems and the associated topological spaces,” *Fundamenta Mathematicæ*, vol. 60 (1967), pp. 223–31. 1

[7] Halldén, S., “A pragmatic approach to modal theory,” *Acta Philosophical Fennica*, vol. 16 (1963), pp. 53–64. 1

[8] Hintikka, J., “The modes of modality,” *Acta Philosophica Fennica*, vol. 16 (1963), pp. 65–82. Zbl 0135.00509 MR 29 3366 3

[9] Kripke, S. A., “Semantical considerations on modal logic,” *Acta Philosophica Fennica*, vol. 16 (1963), pp. 83–94. Zbl 0131.00602 MR 30 1035

[10] Makinson, D., “How meaningful are modal operators?,” *Australasian Journal of Philosophy*, vol. 44 (1966), pp. 331–37. 3

[11] McKinsey, J. C. C., “A solution of the decision problem for the Lewis systems **S2** and **S4**, with an application to topology,” *The Journal of Symbolic Logic*, vol. 6 (1941), pp. 117–34. MR 3,290d 5

[12] McKinsey, J. C. C., “On the syntactical construction of modal logic,” *The Journal of Symbolic Logic*, vol. 10 (1945), pp. 83–96. Zbl 0063.03865 MR 7,186b 1

[13] Scroggs, S. J., “Extensions of the Lewis system **S5**,” *The Journal of Symbolic Logic*, vol. 16 (1951), pp. 112–20. Zbl 0043.00804 MR 13,97a 3

[14] Skura, T., “A Łukasiewicz-style refutation system for the modal logic **S4**,” *Journal of Philosophical Logic*, vol. 24 (1995), pp. 573–82. Zbl 0847.03015 MR 96k:03057 6

[15] Słupecki, J. and Bryll, G., “Proof of the L-decidability of Lewis system **S5**,” *Studia Logica*, vol. 32 (1973), pp. 99–105. Zbl 0313.02012 MR 50 1845 5

[16] Thomason, S. K., “A new representation of **S5**,” *Notre Dame Journal of Formal Logic*, vol. 14 (1973), pp. 281–87. Zbl 0251.02027 MR 48 5825 3

Department of Philosophy
Princeton University
Princeton NJ 08544-1006
email: *jburgess@princeton.edu*