MODELS FOR THE LOGIC OF POSSIBLE PROOFS

BY

LEON HORSTEN

Abstract: The present paper investigates the logical structure of possible proofs. We present and philosophically motivate a class of possible proof models that describes in some detail the modal-epistemic propositional logical structure of possible proofs. This class of models is then recursively axiomatized.

1. Introduction

Pure epistemic logic, as pioneered by Hintikka in the sixties,[1] has not been a uniquely successful discipline. This was due to two factors. On the one hand, there was a disturbing lack of philosophical clarity about the concept of knowledge. It was not at all clear what the necessary and sufficient conditions are for knowing a proposition. On the other hand, it was realized quickly that whatever the exact outcome would be of attempts to improve this situation, pure epistemic logic was going to be such a weak logic as to be almost trivial from a proof-theoretical point of view. Also, it became clear that because of their proof-theoretic weakness, it would be difficult to construct an interesting formal semantics for the resulting calculi.

Investigating the *interaction* between epistemic and modal concepts turned out to be more fruitful. From a philosophical point of view, Kripke[2] made considerable progress through his subtle analysis of the relations between modal concepts and certain *particular* epistemic concepts. From a more technical point of view, Shapiro[3] and others showed that if we take *absolute provability* as our primitive notion, we obtain a nontrivial logic. Shapiro argued that in the context of first-order arithmetic, this notion obeys the *S*4 laws of modal logic.

Pacific Philosophical Quarterly 81 (2000) 49–66 0279–0750/00/0100–0000

In previous work,[4] I have analyzed Shapiro's notion of absolute provability into, on the one hand, a modal notion (possibility), and on the other hand, a particular epistemic notion (having a proof). An attempt was made to describe the logical laws governing their interaction. This led to a logical calculus for which a possible worlds semantics was given. But for several reasons, this logical calculus and its accompanying semantics are not very satisfactory. The overall complaint is that the system still is too much under the spell of the familiar modal systems.[5] The notion of absolute provability, now regarded as a *complex* notion, is still governed by the familiar *S*4 laws.[6] But when one attempts to extend this modal-epistemic system to higher order settings (e.g. to type theory), it becomes apparent that a fundamental principle of normal modal logics does not hold for the complex notion of absolute provability. This can be seen as follows.

Example 1: Assume a modal-epistemic type theoretical language $\mathcal{L}$.[7] Let $\Diamond$ be the familiar possibility operator, and let P be a purely epistemic operator which stands for "X has a proof that," where we keep the prover X fixed across possible situations. Let A be a description in $\mathcal{L}$ of a subset of the set of the natural numbers N. Then we can express in $\mathcal{L}$ that there is a possible situation in which X *has* proved that {2, 4} is the collection of numbers that she has proved to belong to A. This sentence can be expressed in $\mathcal{L}$ in the following way:

$$\Diamond P \forall x\, [P(x \in A) \leftrightarrow (x = 2 \vee x = 4)].$$

We can also express in $\mathcal{L}$ that there is a possible situation in which X has proved that {2, 4, 6} is the collection of numbers that she has proved to belong to A. One way to express this is the following:

$$\Diamond P \forall x\, [P(x \in A) \leftrightarrow (x = 2 \vee x = 4 \vee x = 6)].$$

These two sentences can both be true. Suppose A is the term

$$\{x \mid \exists y \in \mathrm{N}\ (x = y \times 2)\}.$$

In some situation a, X may have proved only the numbers 2 and 4 to be even. And by reflection on what she has proved she may in situation a have realized this. Then situation a verifies the first sentence. For similar reasons there can be a situation b verifying the second sentence. If we assume the familiar aggregation principle

$$\Diamond PA \rightarrow (\Diamond PB \rightarrow \Diamond P\ (A \wedge B)),$$

then we can infer to

$$\Diamond P \begin{bmatrix} \forall x\,[P\,(x \in A) \leftrightarrow (x = 2 \vee x = 4)] \wedge \\ \forall x\,[P\,(x \in A) \leftrightarrow (x = 2 \vee x = 4 \vee x = 6)] \end{bmatrix}.$$

But this sentence can never be true. So the seemingly innocuous aggregation principle must after all be invalid.

The principle $\Diamond PA \rightarrow (\Diamond PB \rightarrow \Diamond P\,(A \wedge B))$ is provable in the systems of Shapiro (1985), Flagg (1986), and Horsten (1994). This does not lead to trouble because they are working in a more *restrictive* language: either it is first-order, or it is higher-order, but the absolute provability operator is not explicitly analyzed. One of the aims of the present paper is to construct a propositional modal-epistemic logic which is more *stable*, so that even in highly expressive languages, and in the context of (formalizations of) strong mathematical theories, its propositional laws are preserved. Ideally, one would like the logic to be *abolutely safe* in this respect.

Anderson[8] has made a more cautious attempt to describe in a propositional context some of the logical principles governing the interaction between the concepts of empirical knowledge, a priori knowledge, and necessity. He was skeptical of aggregation principles like the one discussed above, and did not include them in his system.[9] Dropping the above aggregation principle leads us away from the paradigm of *normal* modal logic. But there is a second defect of *both* Horsten (1994) and Anderson (1993) which, when properly addressed, leads us even further away from the structures of the familiar modal systems. For *every* proposition q, the sentence $\Diamond\neg Pq$ is a logical truth. In a sense, this principle, which we call the *density principle*, is the *opposite* of logical omniscience. One would expect it to be a theorem of modal-epistemic propositional logic; it is in fact independent of the systems proposed in Anderson (1993) and Horsten (1994). But systems which have $\Diamond\neg Pq$ as a theorem for all q differ from the more traditional modal systems in that the "eraser translation" (i.e. the translation that erases all occurrences of intensional operators) is no longer conservative over the underlying logic. For one thing, this means that it will usually be more difficult to prove consistency of such systems.

In the present paper the *semantic view of theories*[10] will be resolutely adopted. In our case, this means that we adopt the semantic conception of *epistemological* theories. In this approach, an epistemological theory is seen first and foremost as a collection of structures that intend to model knowledge situations. Of course, this does not exclude the possibility that in particular cases (e.g. in some propositional logical settings) complete axiomatizations of nontrivial classes of knowledge situations can be obtained.

In the artificial intelligence literature on knowledge representation, the semantic approach is not uncommon. In contrast, philosophical logicians usually adopt the *syntactical method*. They mostly treat epistemological theories in the first place as axiomatically presented formal systems.

In the present paper we want to model a particular class of epistemic objects, namely the class of *possible proofs*. We will philosophically motivate and mathematically describe a class of structures, which are called *possible proof models*. Subsequently this class will be recursively axiomatized. Whereas Shapiro (1985) and Anderson (1993) do not present a formal semantics, Horsten (1994) does give a model-theoretic semantics for modal-epistemic logic. But the set of models given there does not describe the structure of possible proofs in sufficient detail. The present paper aims at constructing a semantics which describes the structure of possible proofs in a more fine-grained way.

2. *The modal-epistemic propositional structure of possible proofs*

Proofs are taken to be *a priori demonstrations*. We construct a theory of the demonstrable a priori. And we are following Shapiro (1985) in that the *absolute* or *intuitive* notion of proof is investigated, not proof in this or that formal system. Moreover, unrestricted use of *classical* logic is allowed in the construction of proofs.

A *proof* is a finite list of sentence-tokens, each of which is created in one of two ways:

1. by writing down an axiom;
2. by a rule of inference, on the basis of previous lines of the proof.

Proofs can be proofs of (classical) logical and of mathematical truths. Proofs can also be of modal truths, or about the notion of proof itself. One might, e.g., try to prove that if φ is provable, and ψ is provable, then $\varphi \wedge \psi$ is provable.[11] And, to be sure, proofs can be about countless other things (probability, truth, etc.). Our prover is likely to work in an informal language – the mathematical fragment of English, say. But we are at present only interested in the *modal-epistemic propositional logical structure* of possible proofs. So we represent possible proofs in a formal language, which will be called $\mathcal{L}_{\Diamond P}$. This language contains an infinite supply of proposition letters p, q, r, . . . , the usual propositional logical connectives, and the intensional operators P and $\Diamond$. Informally, we think of the propositional variables p, q, r, . . . as standing for mutually distinct sentences which are *propositionally atomic*, i.e. sentences which are not of the form $\varphi * \psi$ (for $* \in \{\wedge, \vee, \rightarrow\}$), $\neg\varphi$, $P\varphi$ or $\Diamond\varphi$. This means that the sentences of $\mathcal{L}_{\Diamond P}$ are intended to represent only the modal-epistemic propositional structure of informal sentences – we are hereby suppressing the predicate and higher order logical structure of the informal language.

The operator $\Diamond$ of course expresses the familiar notion of possibility. What does the epistemic operator P express? As was said above, proofs

consist of sentence-tokens, which are human constructions. So we will take a "personal" perspective here, and consider an epistemic agent (call her X), and reason about what she proves in different possible situations. Then Pq expresses that X has (in the situation under consideration) a proof of q. So the prover proves sentences of $\mathcal{L}_{\Diamond P}$, which themselves may contain the notion "X has a proof that." Hence these possible proofs exhibit a *self-referential structure*. The prover is engaged proving theorems using principles of *auto-epistemic logic*.

The possible situations that are quantified over by the modal operator $\Diamond$ are called *proof situations*. They are thought of not as possible world-histories but as *time-slices* of world-histories (or possible worlds-at-a-time, if you like). These time-slices do not all belong to one and the same world-history. That means that we are here, for the sake of simplicity, in effect lumping together a temporal and a modal component. It is left for future work to explicitly disentangle the two by including, beside the epistemic operator P and the modal operator $\Diamond$, also a temporal operator in the language.

In the models that will be constructed, the collection of proved sentences varies from proof situation to proof situation. If the prover has written down a proof in some proof situation, then she may go on to extend her proof – thus creating a *new* proof situation. But given that we are quantifying also over time-slices of different world-histories, there may be proof situations in which the prover has constructed a different proof altogether.

In proof situations belonging to different world-histories different sentences may even count as axioms. This can be due to the conventional element inherent in all theories. For instance, in one situation the Well-ordering Principle may be taken as an axiom, and Zorn's Lemma proved from it, whereas in another situation it is the other way round. But it may also be due to the fact that the "intuitive" capacities of the epistemic agent vary from possible world to possible world. It may be, for instance, that the Riemann Hypothesis in analysis will for ever be out of our epistemic reach. Suppose that it is in fact independent of even the strongest analytic principles that we will ever adopt. If, in addition, the conjecture is true, then it cannot be a priori excluded that had our mathematical intuition of the real number structure been stronger,[12] we would have been able to see that it was an irreducible truth: an axiom of analysis.

Axioms are usually taken to be *necessary* truths. But it is not easy to see why they would *have* to be. For something to be a possible axiom, it could be argued, all that is required is that it be possible to know a priori and perhaps somehow immediately that it is true. Dedekind proposed to take as an axiom the proposition "I exist," and to deduce the laws of arithmetic from it.[13] If we leave aside the details of his admittedly questionable deduction, it is not clear what is objectionable about his proposal. In any case, without committing ourselves to any particular

contingent proposition being a possible axiom, *contingent axioms will in principle be allowed* in the models which will be constructed.

In our models, we will assume that elementary truths of propositional modal and epistemic logic (and their interaction) are always available for use in possible proofs, as well as two *special* rules of inference expressing forms of introspection. Specifically, the following operations are allowed in possible proofs:

- The prover can write down a classical tautology.
- The prover can apply Modus Ponens.
- The prover can write down a theorem of $S5$ modal logic. The agent is *not* in general allowed to use the *Necessitation Rule* in possible proofs because there can be contingent sentences in possible proofs (as we have seen above).
- The prover can use the epistemic law $PA \rightarrow A$, and its necessitation $\Box (PA \rightarrow A)$.
- If there is a possible situation in which the prover has proved A, and a possible situation in which the prover has proved B, then there is a possible situation in which the prover takes $\Diamond A \wedge \Diamond B$ as an axiom.
- From a line in a possible proof containing a sentence A, the prover may infer to PA ("I have proved A") on a subsequent line. This "positive introspection" rule is called the *rule of* P-*Introduction*.
- The prover can always come to realize, of a sencence A that she has not proved, *that* she has not proved A. So a possible proof can be extended by a line containing $\neg PA$. This "negative introspection" rule is called the *rule of* $\neg$P-*Introduction*.

The rules of P-Introduction and $\neg P$-Introduction force us to impose a *global restriction* on possible proofs:

> A sentence A can be written down on a line in a possible proof in accordance with the above rules only if $\neg PA$ does not follow from $PA_1, \ldots, PA_n$, where $A_1, \ldots, A_n$ are earlier lines of the possible proof.

Precise statements of these rules for introducing sentences in possible proofs will be given in the following sections. But it may be helpful at this point to give some illustrations of how the P-Introduction rule, the $\neg P$-Introduction rule and the global restriction on possible proofs function.

Example 2: Let there be given a proof situation a, where the prover has written down a proof of the following form:

1. p (Axiom)
2. $p \rightarrow p \vee q$ (Propositional Logic)
3. $p \vee q$ (Modus Ponens 1, 2)

Then there exists a proof situation b where the prover has extended her proof by adding one line, to become:

1. p (Axiom)
2. $p \rightarrow p \vee q$ (Propositional Logic)
3. $p \vee q$ (Modus Ponens 1, 2)
4. $P\,(p \vee q)$ (P-Introduction, 1–3)

The sentence on line 4 expresses that there exists a proof of $p \vee q$ in proof situation b.

Example 3: Assume again proof situation a of the previous example. Then there must be a proof situation c where the prover has extended her proof using the $\neg P$-Introduction rule:

1. p (Axiom)
2. $p \rightarrow p \vee q$ (Propositional Logic)
3. $p \vee q$ (Modus Ponens 1, 2)
4. $\neg Pq$ ($\neg P$-Introduction, 1–3)

Line 4 here expresses that there is no proof of q in proof situation c.

Example 4: Assume again proof situation a. Then the following extension of the possible proof in a does *not* count as a possible proof:

1. p (Axiom)
2. $p \rightarrow p \vee q$ (Propositional Logic)
3. $q \vee q$ (Modus Ponens 1, 2)
4. $\neg Pp$ ($\neg P$-Introduction, 1–3)

By writing down line 4, the prover violates that global restriction on possible proofs. For from Pp one can readily infer, using the epistemic law $PA \rightarrow A$, the $\neg P \neg Pp$.

Example 5: Assume proof situation c. Then the following one-line extension of the possible proof in c does *not* count as a possible proof:

1. p (Axiom)
2. $p \rightarrow p \vee q$ (Propositional Logic)
3. $p \vee q$ (Modus Ponens 1, 2)
4. $\neg Pq$ ($\neg P$-Introduction, 1–3)
5. q (Axiom)

By writing down line 5, the prover violates that global restriction on possible proofs. For $\neg Pq$ is easliy derived from $P \neg Pq$.

We maintain that the rules of P-Introduction and $\neg P$-Introduction are admissible rules in a priori demonstrations.[14] An inference to PA from a

proof of A does not depend on sense perception for its justification. So such an inference preserves a prioricity. Similarly, an inference to $\neg PB$ from the fact that the proof that has been constructed in a given situation does not prove B, does not depend on sense perception for its justification. So such inferences also preserve a prioricity.

The principle according to which the existence of a one line possible proof consisting of the sentence $\Diamond A \wedge \Diamond B$ follows from the existence of a possible proof of A and a possible proof of B also deserves some comment. It can perhaps best be motivated as follows.[15] Observe first that, from the existence of a possible proof of A and a possible proof of B, the existence of a possible proof of $A \wedge B$ does *not* always follow. For it may be that the content of A and B entail that they cannot *both* be true in the *same* possible world, let alone their conjunction be proved there. For instance, it may be that $A = \neg PC$ and $B = PC$ for some C. But it seems that the content of two sentences A and B (such that there exists a possible proof of A and a possible proof of B) can exclude the possibility of being conjoined in one proof in this way *only because* at least one of the two sentences is *contingent*. Therefore even in such a situation, there seems to be nothing that can prevent the prover from taking the sentence $\Diamond A \wedge \Diamond B$ as her (sole) axiom. For one thing, both $\Diamond A$ and $\Diamond B$ will in the situation under consideration certainly be *true*. For A is true in some proof situation (since proved there) and B is true in some proof situation (since proved there).

In sections 4 and 5 a rigorous description is given of the class of models that was informally presented in this section. The aim of sections 6, 7, and 8, then, is to recursively axiomatize the set of modal-epistemic sentences that are true in every model of this class. But there is a philosophical problem. Whereas it is unreasonable to expect that all *rules* that the epistemic agent has available will be truth-preserving when added to the axiomatization of the class of possible proof-models,[16] it is reasonable to expect that all the *theorems* of the axiomatization should always be available to the epistemic agent. But if we add *them* to the demonstrative repertoire of the agent, then the axiomatization of the (slightly modified) class of models may change,[17] and so on. In the present paper, this difficulty is simply shelved. The price for this is that we cannot – and do not – claim that we have described *the* propositional logic of the demonstrable a priori.

3. *The system* S5P

Before giving a rigorous description of the class of possible proof models, we will in this section describe the system *S5P* of modal and epistemic truths to which the prover is guaranteed to have access in proof situations.

The main objective of this section is to show that $S5P$ is a *decidable* theory. This fact will be needed later on.

The language of $S5P$ is $\mathcal{L}_{\Diamond P}$. $S5P$ contains the axioms of the $S5$ system of modal logic, the principle $PA \rightarrow A$, and the necessitations of all these principles. Modus Ponens is the sole rule of inference. We have noted in the previous section that the Necessitation Rule is in general not a truth-preserving rule in possible proofs.

A model $M = \langle W, R, V \rangle$ for $\mathcal{L}_{\Diamond P}$ is like a Kripke model for modal propositional logic except that possible worlds are finite sets of sentences of $\mathcal{L}_{\Diamond P}$ and for all ϕ and all worlds w, $V(P\phi, w) = 1$ iff $P\phi \in w$.

Now consider the class $\mathcal{N}$ of models for $\mathcal{L}_{\Diamond P}$ such that for every world w, and for every sentence ϕ, if $V(P\phi, w) = 1$, then $V(\phi, w) = 1$, and such that R is an equivalence relation. We want to prove that $S5P$ is complete for $\mathcal{N}$ (soundness is obvious): for every ϕ which is not provable in $S5P$, we will construct a countermodel $N \in \mathcal{N}$. Moreover, this countermodel will be finite, which implies that $S5P$ is a decidable theory. We use a variant of a method due Kripke,[18] as presented by Boolos.[19] It is assumed that the reader is somewhat familiar with this method for proving completeness and decidability for systems of modal logic. A further adaptation of this method will in section 8 yield a completeness proof for our theory of possible proofs.

Lemma 1: $\mathcal{M}$ is complete for $S5P$.
Proof. (Sketch) The method for producing a countermodel for a nontheorem of $S5P$ is like the method for producing a countermodel for a nontheorem of the modal logic GL as described in Boolos (1993), chapter 10, except for the following:

(a) In the tree for $\neg\phi$, we have the following modal and epistemic rules:
(a1) if $\Box\psi$ occurs unchecked on a branch of the tree, then add ψ to the tree and check $\Box\psi$, and write both ψ and $\Box\psi$ in every window opened on the branch;
(a2) if $\neg\Box\psi$ occurs unchecked on the branch of the tree, then open a window in which you write $\neg\psi$, and check $\neg\Box\psi$;
(a3) if $P\psi$ occurs unchecked on a branch, then add ψ to the branch and check $P\psi$;
(b) In the resulting model N, W consists of the collection of open branches, where we identify worlds which make exactly the same sentences true. We let R be the total relation on W.

It is easily seen that the resulting countermodel N is an $S5P$-model. Moreover, since we are identifying worlds which make the same class of sentences true, the resulting model M will be finite. The reason for this is the following. In new windows we only write sentences of complexity at

most as high as that of the most complex formula on the branch on which the window is opened. So the most complex formula on the main branch imposes a finite upper bound on the number of windows with different formulas written on them. And since we are identifying worlds corresponding to windows with the same sentences written on them, there is a finite bound on the number of possible worlds.

To conclude the proof, we note that it can be shown inductively, in the usual way, that $N, w_0 \models \neg\phi$. ■

Corollary 2 $S5P$ is decidable.
Proof. Follows from Lemma 1. ■

4. *Possible proofs*

Now it is possible to give a precise formulation of what possible proofs are. *Possible proofs* are finite sequences of sentences of $\mathcal{L}_{\Diamond P}$ denoted as $\langle \phi_1, \ldots, \phi_n \rangle$, and subject to the following constraints.

There will be allowed sentences asserted on lines of such a proof and justified by *Axiom Introduction* (*AI*). Beside sentences introduced by *AI*, theorems of the system $S5P$ may be axiomatically introduced, and justified by *Logic Introduction* (*LI*).

There are three rules of inference allowed in possible proofs. The first rule is *Modus Ponens* (*MP*). The second rule is the rule of *P-Introduction* (*PI*): it says that a sentence $P\phi$ may be written on a line of a possible proof if ϕ appears on an earlier line. The third rule is the rule of $\neg P$-*Introduction* ($\neg PI$): $\neg P\phi$ may be written on a line of a possible proof.

All these principles for introducing a sentence in a possible proof are subject to one *global restriction*:

> A sentence ϕ may not be written down on a line in a possible proof if $\neg P\phi$ can be derived in the system $S5P$ from sentences $P\phi_1, \ldots P\phi_n$, which are such that $\phi_1, \ldots, \phi_n$ appear earlier in the possible proof.

In section 2 we have given examples of how this global restriction can block the possibility of extending possible proofs in certain ways. The global restriction requires the prover to check each time before writing down a sentence ϕ, whether $\neg P\phi$ is derivable from $P\phi_1, \ldots, P\phi_n$, where $\phi_1, \ldots, \phi_n$ is the proof which she has already constructed. Now checking this may take quite some time, but it is always humanly possible to do this. For we have proved in the previous section that the global restriction specifies a *decidable* condition (Corollary 2).[20]

We say that $\langle \phi_1, \ldots, \phi_n \rangle$ is a 1-line $\mathcal{P}$-*extension* of $\langle \phi_1, \ldots, \phi_{n-1} \rangle$ if $\langle \phi_1, \ldots, \phi_n \rangle$ is obtained by adding ϕ_n to $\langle \phi_1, \ldots, \phi_{n-1} \rangle$ in accordance with

one of the principles for introducing lines in a possible proof. In a similar vein we define $\langle \phi_1, \ldots, \phi_{n-1} \rangle$ to be the 1-line $\mathcal{P}$-*contraction* of the possible proof $\langle \phi_1, \ldots, \phi_n \rangle$.

5. *Possible proof frames and possible proof models*

We now construct, using the notion of a possible proof, models for the language $\mathcal{L}_{\Diamond P}$. In the next section we will present a recursive axiomatization of this collection of models in $\mathcal{L}_{\Diamond P}$.

For a triple $\langle W, R, \Pi \rangle$ to be a *possible proof frame*, the following conditions have to be satisfied.

(1) W is a non-empty set of proof situations;
(2) R is an equivalence relation on W;
(3) Π is a function from proof situations to finite sequences of sentences of $\mathcal{L}_{\Diamond P}$;
(4) for every $w \in W$, if π is a 1-line $\mathcal{P}$-*extension* of $\Pi(w)$, then there is a $w' \in W$ such that wRw' and $\pi = \Pi(w')$;
(5) for every $w \in W$, if π is the 1-line $\mathcal{P}$-*contraction* of $\Pi(w)$, then there is a $w' \in W$ such that wRw' and $\pi = \Pi(w')$;
(6) for all w, $w' \in W$ such that wRw', if there are $\phi \in \Pi(w)$ and $\psi \in \Pi(w')$, then there is a $w'' \in W$ such that wRw'', $w'Rw''$, and $\Pi(w'')$ is a 1-line possible proof, consisting solely of the sentence $\Diamond\phi \wedge \Diamond\psi$, introduced by *AI*.

A *possible proof model* is a pair $\langle F, V \rangle$, where F is a possible proof frame $\langle W, R, \Pi \rangle$, and V is a valuation function for $\mathcal{L}_{\Diamond P}$ which obeys the usual evaluation rules of modal logic, together with the following conditions:

(1') For every $w \in W$, if $\phi \in \Pi(w)$, then $V(\phi, w) = 1$;
(2') for all ϕ and for all $w \in W$, $V(P\phi, w) = 1$ iff ϕ is a line of $\Pi(w)$.

The notions of truth in a possible proof model, validity and logical consequence are defined in the usual way.

6. *Axioms for possible proof models*

We now want to axiomatize in $\mathcal{L}_{\Diamond P}$ the class of possible proof models that was described in the previous section. We call our axiomatization *PP* ("possible proofs").

In what follows, I and J are parameters that range over finite sets of natural numbers. Here are the axioms and rules for *PP*:

Axiom 1 All theorems of $S5P$

Axiom 2 $(\Diamond P\phi \wedge \Diamond P\psi) \rightarrow \Diamond\left[P(\Diamond\phi \wedge \Diamond\psi) \wedge \bigwedge_{i\in I}\neg P\theta_i\right]$ with $\theta_i \neq \Diamond\phi \wedge \Diamond\psi$ for all $i \in I$.

Axiom 3 $\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j\right) \rightarrow \Diamond\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \wedge P\phi\right)$ with $\phi \in S5P$ and not: $\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \vdash_{S5P} \neg P\phi$

Axiom 4 $\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \wedge P(\phi \rightarrow \psi) \wedge P\phi\right) \rightarrow$ $\Diamond\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \wedge P(\phi \rightarrow \psi) \wedge P\phi \wedge P\psi\right)$ with not: $\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \wedge P(\phi \rightarrow \psi) \wedge P\phi \vdash_{S5P} \neg P\phi$

Axiom 5 $\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j\right) \rightarrow \Diamond\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \wedge PP\theta_k\right)$ with $k \in I$ and not: $\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \vdash_{S5P} \neg PP\theta_k$

Axiom 6 $\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j\right) \rightarrow \Diamond\left(\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \wedge P\neg P\gamma_l\right)$ with $l \in J$ and not: $\bigwedge_{i\in I}P\theta_i \wedge \bigwedge_{j\in J}\neg P\gamma_j \vdash_{S5P} \neg P\neg P\gamma_l$

Rule 1 Modus Ponens.

Rule 2 Necessition for $\Box$.

Axiom 2 describes condition 6 on possible proof models. Axioms 3, 4, 5, 6 describe the principles that are allowed in possible proofs (*LI*, *MP*, *PI*, $\neg PI$, respectively). Observe that we do have the Necessitation Rule in the "home system." The reason for this is that *PP* is supposed *not* to prove any contingent sentences. *PP* is supposed to prove only *conceptual* truths about the interaction between the notion of possibility and the notion of proof.

Since theoremhood of $S5P$ is a decidable notion (Corollary 2 of section 3), *PP* is indeed a formal system: the class of its theorems is recursively enumerable. The choice of the axioms of *PP* is determined solely by what is needed to carry out the completeness proof. It may well be that a cleaner formulation of *PP* can be given, but we will not try to do so in this paper.

It is a theorem of *PP* – at least to the extent that $\mathcal{L}_{\Diamond P}$ is able to express it – that the agent might have proved nothing at all:

Proposition 3: $\vdash_{PP} \Diamond \left(\bigwedge_{i \in I} \neg P\theta_i \right)$, for any finite set I of natural numbers.
Proof. From Axiom 1, Axiom 2 and Axiom 3. ■

So the "density principle" that was presented and endorsed in the introduction is actually a theorem of *PP*.

7. *Soundness and consistency of* PP

The proof of the soundness of *PP* for possible proof models is proceeds as usual by induction on the length of proofs of *PP*. But clause (1') in the definition of possible proof models makes it a nontrivial question whether there *are* possible proof models. So the soundness of *PP* does not immediately imply the consistency of *PP*. We prove that *PP* is consistent by building a minimal model in which all axioms and rules of *PP* are satisfied.

Lemma 4: *PP* is consistent.
Proof. (Sketch) We build a minimal possible proof model

$$M_{\text{min}} = \langle \langle W_{\text{min}}, R_{\text{min}}, \Pi_{\text{min}} \rangle, V_{\text{min}} \rangle$$

in stages.

Stage 0. Introduce a proof situation w_0, and let $\Pi(w_0)$ be empty.
Stage n + 1. For each proof situation w that exists at stage n, if π is a 1-line $\mathcal{P}$-extension of $\Pi(w)$, then introduce a new proof situation w' such that $\pi = \Pi(w')$. And for every $\phi \in \Pi(w_1)$, $\psi \in \Pi(w_2)$, with w_1, w_2 existing at stage n, introduce a new proof situation w_3 such that $\Pi(w_3) = \langle \Diamond\phi \wedge \Diamond\psi \rangle$. W_{min} and Π_{min} are obtained by taking the minimal closure of this generation procedure. We let $V_{\text{min}}(w, p) = 0$ for all proposition letters p and for all $w \in W_{\text{min}}$, and we let R_{min} be the total relation on W_{min}.
Induction on the stages shows that for all $w \in W_{\text{min}}$, and for all ϕ, if $\phi \in \Pi_{\text{min}}(w)$, then $V_{\text{min}}(\phi, w) = 1$. So M_{min} is a possible proof model, and $P\psi \rightarrow \psi$ is true for each sentence ψ at each world. It follows from the construction of M_{min} that the other axioms of *PP* are satisfied everywhere in M_{min}, and that the rules of inference of *PP* are truth-preserving. ■

This allows us to show that objectionable modal-epistemic distribution principle that was discussed in the introduction is *not* provable in *PP*:

Proposition 5: $\Diamond P\phi \rightarrow (\Diamond P\psi \rightarrow \Diamond P(\phi \wedge \psi))$ is not a theorem of *PP*.
Proof. Let ϕ be of the form $P\theta$, with θ a theorem of *S5P*, and let $\psi = \neg P\theta$.

Then $\Diamond P\phi$ is provable in *PP* (using Axiom 3 and Axiom 5), and $\Diamond P\psi$ is provable in *PP* (Proposition 3). But by the consistency of *PP* (Lemma 4), $\Diamond P\ (\phi \wedge \psi)$ cannot be a theorem of *PP*. ■

8. *Completeness of* PP

We prove the completeness of *PP* with respect to $\mathcal{M}$ by means of a variation on the completeness proof of *S*5*P*. However, the resulting countermodel will now no longer be finite. So we do not immediately obtain decidability of *PP* as a corollary.

The rough idea is simple. As in the completeness theorem for *S*5*P*, we start with a proof situation w_0 at which we want to make $\neg\phi$ true, and we associate a possible proof $\Pi\ (w_0)$ with it. We also introduce new proof situations corresponding to sentences of the form $\neg\Box\psi$ that belong to w_0. Second, we construct a proof situation corresponding to the 1-line $\mathcal{P}$-contraction of $\Pi\ (w_0)$, associate the $\mathcal{P}$-contraction with it, construct a proof situation corresponding to *its* $\mathcal{P}$-contraction, and so on until we reach the empty proof. Again, we may have to introduce new proof situations corresponding to sentences of the form $\neg\Box\psi$ that appear in proof situations. Third, we construct proof situations corresponding to all possible 1-line $\mathcal{P}$-extensions of the possible proof which we now have, associate these 1-line $\mathcal{P}$-extensions with them, consider *their* 1-line extensions, and so on (ad infinitum). We let the valuation function of the resulting model be determined in the usual way at the various proof situations by the proposition letters that appear in the proof situation. Here are more details.

Theorem 6: *PP* is complete for $\mathcal{M}$.
Proof. (Sketch) Suppose ϕ is not a theorem of *PP*. The aim is to build a countermodel $M_{\neg\phi} = \langle\langle W_{\neg\phi}, R_{\neg\phi}, \Pi_{\neg\phi}\rangle, V_{\neg\phi}\rangle$ which makes $\neg\phi$ true. We build the countermodel in stages.

1. We start by writing down $\neg\phi$. Then we apply the rules for making an *S*5*P*-tree, as in the proof of the completeness theorem for *S*5*P*. [Each open branch will determine a proof situation.] Corresponding to each open branch *B*, we construct a possible proof out of the sentences ψ such that $P\psi$ occurs on *B*, in the following way: arrange all such ψ in an order of increasing complexity, yielding a sequence $\langle\psi_1, \ldots, \psi_n\rangle$. [This sequence will be the possible proof associated with the proof situation determined by *B*.] All sentences occurring in this sequence are taken to be obtained by *AI*.
2. We consider each 1-line contraction *C* of possible proofs that have already been obtained (in so far as they have not already been considered).

For each such C, we open a window in the branch corresponding to it, in which we write down $P\psi$ for every ψ in C. Then we again apply the rules for $S5P$-trees [to determine which other sentences are to be made true at the proof situation corresponding to this window], introducing new windows along the way if required by the rules.

3. We look at all 1-line extensions of proofs that have been obtained so far. For example, if $\langle \psi_1, \ldots, \psi_n, \psi_{n+1} \rangle$ is such a 1-line extension, obtained by *PI*, then we open a window on the open branch corresponding to $\langle \psi_1, \ldots, \psi_n \rangle$ in which we write $P\psi_1, \ldots, P\psi_n, P\psi_{n+1}$. Again we apply the rules for constructing $S5P$-trees. Similar for the other principles for extending possible proofs.

4. For any pair $\mathcal{P}_1, \mathcal{P}_2$ of possible proofs that have been obtained so far, and for any ψ_1, ψ_2, if ψ_1 occurs in $\mathcal{P}_1$ and ψ_2 occurs in $\mathcal{P}_2$, then we open a new window on the branch corresponding to $\mathcal{P}_1$ in which we write $P(\Diamond\psi_1 \wedge \Diamond\psi_2)$ (unless this has already been done at an earlier stage).

5. Repeat stages 1 through 4 until a minimal closure C is reached. $M_{\neg\phi}$ is determined by the minimal closure C. $W_{\neg\phi}$ is the set of all proof situations w_B such that B is an open branch of C, $R_{\neg\phi}$ is the total relation on $W_{\neg\phi}$, and for each $w_B \in W_{\neg\phi}$, $\Pi_{\neg\phi}(w_B)$ is the possible proof corresponding to the branch B. $V_{\neg\phi}$ is determined in the usual way at a proof situation w_B by the proposition letters that occur on B.

From the consistency of *PP* and from the axioms of *PP* it follows in the usual way that $V_{\neg\phi}$ is well-defined. $M_{\neg\phi}$ is constructed in such a way that all the axioms and rules of *PP* are satisfied. Moreover, for any proof situation w and for any ϕ, if $\phi \in \Pi_{\neg\phi}(w)$, then $P\phi$ occurs on the open branch corresponding to w, whereby also ϕ occurs on the branch corresponding to w, so that ϕ is true at w. Therefore $M_{\neg\phi}$ is a possible proof model.

To conclude, we must verify that the range of $\Pi_{\neg\phi}$ really consists of possible proofs, i.e. that the global restriction on writing down sentences in possible proofs is adhered to everywhere. Take any $\Pi_{\neg\phi}(w) = \langle \psi_1, \ldots, \psi_n, \psi_{n+1} \rangle$. We want to show that $\neg P\psi_{n+1}$ does not $S5P$-follow from

$$\{P\psi_1, \ldots, P\psi_n\} \cup \{\neg P\theta_i \mid \theta_i \notin \{\psi_1, \ldots, \psi_n, \psi_{n+1}\}\}.$$

But that must be the case, for consider the stage in the construction of $M_{\neg\phi}$ at which the branch corresponding to $\langle \psi_1, \ldots, \psi_n, \psi_{n+1} \rangle$ was constructed. An inductive argument shows that at that point we in effect constructed an $S5P$-model of

$$\{P\psi_1, \ldots, P\psi_n, P\psi_{n+1}\} \cup \{\neg P\theta_i \mid \theta_i \notin \{\psi_1, \ldots, \psi_n, \psi_{n+1}\}\}.$$

[In this inductive argument we need the assumption that $\psi_1, \ldots, \psi_n$ are arranged in an order of increasing complexity.]

This concludes the sketch of the completeness proof. ■

The method of this completeness proof is *flexible*. For instance, consider the class $\mathscr{M}'$ of models which are just like possible proof models, except that the rule for *P*-Introduction is *not* stipulated to be an admissible rule in possible proofs. The above completeness proof is easily adapted to show that the system which is just like *PP* except that it lacks Axiom 5 is (sound and) complete for $\mathscr{M}'$. Similar remarks apply to all the other principles for introducing sentences in possible proofs that were discussed, *except* for the principle of Axiom Introduction.

9. Conclusion

The aim of this paper was to get a grip on the propositional logic of possible proofs. An attempt was made to give a fairly detailed description of the structure of possible proofs and the interrelations between possible proofs, and to describe the modal-epistemic propositional sentences that are valid in all the resulting structures.

It was argued that the possible proof models presented here are in a sense more *realistic* than most of the models for epistemic logic that are in existence today. We think here especially of the models of Hintikka (1962) and its variants, and those of Horsten (1994). For instance, we have seen that the objectionable schematic principle

$$\Diamond PA \rightarrow (\Diamond PB \rightarrow \Diamond P(A \wedge B))$$

is not valid in all possible proof models. And the schematic principle $\Diamond\neg PA$ is valid in all possible proof models.

The possible proof models described here are also fairly *fine-grained*. Specifically, the *modal-epistemic propositional logical sequential structure* of possible proofs is explicitly exhibited in the possible proof models, which makes the possible proof models more fine-grained than those of Horsten (1994). Abstraction is made only of the predicate logical and higher-order structure of possible proofs.

Nevertheless, there are philosophical problems that remain unaddressed. As was briefly mentioned in section 2, there is a discrepancy between the system *PP* and the modal and epistemic principles that are guaranteed to be available in possible proofs (namely, the principles of the system *S*5*P*). If the theorems of *PP* are knowably valid principles of modal-epistemic logic, then should not the possible prover always be allowed to use any of *them* instead of only the theorems of the weaker system *S*5*P*? This is a problem that should be addressed in future research about the logic of possible proofs. Also, explicitly disentangling in the formal language the modal and temporal components that are implicit in the semantics may give us useful information about the structure of possible proofs. At the

same time, it should be kept in mind that with each step in the direction of extending the expressive power of the language, the structure of the models and of the formal systems becomes more complicated.

It was noted in passing in the previous sections that there are also purely technical questions that remain. It is not known at present how to axiomatize the class of models that results when axioms in possible proofs are required to be *necessary*. Whether *PP* is decidable is an open question. And one wonders whether intuitionistic propositional logic is interpretable in *PP* under a variant of the Gödel translation from intuitionistic to (intensional) classical logic.[21]

Institute of Philosophy
University of Leuven
e-mail: Leon.Horsten@hiw.kuleuven.ac.be

NOTES

[1] Hintikka, J. *Knowledge and Belief* (Ithaca: Cornell University Press, 1962).

[2] See, e.g., the discussion of the existence of contingent a priori and necessary a posteriori propositions in Kripke, S. *Naming and Necessity* (Harvard: Harvard University Press, 1980).

[3] Shapiro, S. "Epistemic and Intuitionistic Arithmetic," in S. Shapiro (ed), *Intensional Mathematics* (Amsterdam: North-Holland, 1985).

[4] Horsten, L. "Modal-Epistemic Variants of Shapiro's System of Epistemic Arithmetic," *Notre Dame Journal of Formal Logic* 35, 1994.

[5] This has been the curse of epistemic logic since its inception.

[6] Except for the fact that it was recognized that the reflexivity axiom has to be slightly weakened.

[7] More precisely, take the language of Epistemic Type Theory, as described in Flagg, R. "Integrating Intuitionistic and Epistemic Type Theory," *Annals of Pure and Applied Logic* 32, 1986. The only modification that we need to make to this language for our purposes is that instead of the absolute provability operator of Flagg's language, $\mathcal{L}$ has a modal operator $\Diamond$ and an epistemic operator P.

[8] Anderson, C. A. "Towards a Logic of A Priori Knowledge," *Philosophical Topics* 21, 1993.

[9] He gives a counterexample based on self-reference that is very similar to the aggregation principle of example 1. However, the results of Anderson (1993) and Horsten (1994) are not *directly* comparable, since they investigate slightly different notions.

[10] We draw this notion from the philosophy of science. See, for instance, Suppe, F. *The Structure of Scientific Theories*. Second Edition (Chicago: University of Illinois Press, 1977), pp. 221–30.

[11] Note that we have cast doubt on this principle in the previous section.

[12] Kripke has noted that it is difficult to say how far this stretching is allowed to go.

[13] Dedekind, R. "The Nature and Meaning of Numbers," in Dedekind, R. *Essays on the Theory of Numbers [1901]* (Mineola: Dover, 1963).

[14] For an analysis of the conditions that a piece of knowledge has to satisfy in order to count as *a priori* knowledge, see Burge, T. "Content Preservation," *Philosophical Review* 102, 1993.

[15] I realize that the following argument does not establish the soundness of this aggregation principle beyond all possible doubt.

[16] The "contingent rules" of P-Introduction and $\neg P$-Introduction, for instance, will not be truth-preserving when added to the axiomatization.

[17] It would be interesting to investigate under which conditions this process reaches a fixed point.

[18] Kripke, S. "A Completeness Theorem in Modal Logic," *Journal of Symbolic Logic* 24, 1959.

[19] Boolos, G. *The Logic of Provability* (Cambridge: Cambridge University Press, 1993), chapter 10.

[20] We ignore complexity considerations here.

[21] I wish to thank the participants in the conference on *A Priori Knowledge* at the University of Southern California, and especially Tony Anderson for valuable comments on an earlier version of this paper.