ON S4 AS THE "LOGIC OF JUSTIFIED KNOWLEDGE"*

by

JAVIER LEGRIS

Abstract

Some years ago W. Lenzen and F. v. Kutschera argued that S4.2 and S4.4 (but not S4) are the "logics of knowledge". In this paper I give good reasons to consider S4 as the "logic of justified knowledge". From the basis of a doxastic system for the notion of justified belief and the definition of knowledge as "true justified belief" I construct an epistemic counterpart to S4, and I show semantically that at most this system is derivable from these assumptions. In this sense the original conjecture of J. Hintikka about S4 as the exact tantamount to the logic of "real" knowledge becomes defensible.

In his paper (1969) J. Hintikka conjectured modal S4 to be the logic of knowledge. F. v. Kutschera in (1976) and W. Lenzen in (1979) and (1980) rejected this idea. Lenzen constructed two epistemic systems corresponding to modal S4.4 and S4.2 on the basis of knowledge as true conviction. In this paper I discuss both systems and - above all - present another one based on the notion of knowledge as justified true belief, which comes to be a counterpart to modal S4. Thus, in a certain sense, Hintikka's conjecture becomes defensible.

* The author wishes to thank the anonymous referees of an earlier draft of this paper for helpful comments and corrections.
J. Hintikka conjectured in his paper (1969) that the logic of "real" knowledge (that is, not mere true opinion) "is exactly tantamount to S4" (p. 83) when the modal operator of necessity is replaced by the epistemic operator $K$ of knowledge. This conjecture relied only upon intuitive grounds and it was formulated without a deep elucidation of this notion of "knowledge". Now, following an old tradition beginning with Plato (in this Thaetetus) "knowledge" can be defined as "true belief". Symbolically:

$$DK1: KA =_df NA \land A,$$

where $Np$ means that a person believes (in the strong sense of "being convinced of") $p$. By means of this definition epistemic logic comes to be reducible to doxastic logic (i.e. the logic of belief) plus classical logic.

In a doxastic propositional language including the classical connectives negation ($\neg$), conditional ($\supset$), and the doxastic operator $N$, the system $D$ is constructed. Its axioms are sentences of the language of the following forms:

D0: $A$, if $A$ is a classical tautology;

D1: $NA \supset \neg N \neg A$;

D2: $N(A \supset B) \supset (NA \supset NB)$;

D3: $NA \supset NNA$;
D4: \(\neg N \vardash N \neg N A\);

and it includes the rule of *modus ponens* and the rule

\[\text{RD: } \vdash N A, \text{ if } \vdash A.\]

At the present time \(D\) is regarded as the standard system for belief in the sense of conviction (cf. KUTSCHERA 1976, p. 92). Axiom D1 expresses the consistency of someone’s beliefs. D2 and RD express the “normality” of the system. (The persons are assumed to have “logical omnibelief”. ) D3 is a doxastic counterpart to the characteristic axiom for the modal S4, and D4 for S5. However, the system does not correspond to S5 because of the lack of axioms of the form

\[N A \vdash A,\]

which does not seem plausible in a doxastic interpretation. (Generally, we do no expect our convictions to be true.) In defense of D2 and RD the *implicit* aspect of the notion of belief can be argued. A person believes *implicitly* the classical tautologies and the logical consequences of his own beliefs (cf. LENZEN 1978, p. 63). In what follows I take for granted this system as “the logic of conviction”.

A possible-world-semantics for \(D\) can be constructed. Intuitively, one can assume a set \(I\) of possible worlds and a subset \(S_D\) of \(I\), consisting of that worlds, which are believed by an arbitrary person \(a\) in the world \(i\) as possible. For this reason I shall talk about subsets of \(I\) instead of the usual “accessibility relation” between worlds of \(I\).
A \textbf{D-model} is a triple $M = <I, S_D, V>$ satisfying the following conditions:

1) $I$ is a non-empty set.

2) For every $i \in I$, $S_D, \subseteq I$, so that
   da) $S_{D, i}$ is non empty.
   db) $S_{D, j} = S_{D, i}$, provided that $j \in S_{D, i}$.

3) For every $i \in I$, $V_i$ is a function satisfying the classical truth conditions for $\neg$ and $\supset$, and:

   $V_i(NA) = T$ iff for every $j \in S_{D, i}$, $V_j(A) = T$.

The standard relation of accessibility can be defined for \textbf{D}-models in the following way:

$$(dR): iRj =_{df} \{ <i, j>: j \in S_{D, i} \}$$

(where $iRj$ means that the world $j$ is \textit{accessible from} the world $i$). Satisfaction, validity and \textbf{D}-validity are defined in the standard way. According to Kripke's methods, soundness and completeness of the system can be proved. In the case of completeness the task is to construct canonical models from maximal consistent sets (s. v.g. KUTSCHERA 1976, pp. 38 ff).

W. Lenzen showed in his paper (1979) that a special epistemic
system results from $D$ and the definition $DK1$, its main feature
being to correspond to the modal $S4.4$ (s. SOBOCINSKI 1964).

This epistemic system, here called $E1$, consists accordingly of
the following axioms:

E0: $A$, if $A$ is a classical tautology;

E1: $KA \supset A$;

E2: $K(A \supset B) \supset (KA \supset KB)$;

E3: $KA \supset KKA$;

E4: $A \supset (\neg K \neg KA \supset KA)$;

and the rules of modus ponens and

RK: $\vdash KA$, if $\vdash A$.

Lenzen characterized this system as the logic of the “radical
knower” because it relies upon the definition of knowledge as true
conviction, i.e. a “radical” notion of knowledge (cf. 1979, p. 38).
What he proved is, in first place, the derivability of $E1$ from $D +$
$DK1$. It is easy to see that $E2$ and $E3$ follow from $D2$ and $D3$ (with
$DK1$ and classical logic). $E1$ is a straightforward consequence of
$DK1$. Finally $E4$ follows from $D4$, $D3$ and $D2$. To prove in what
sense $E1$ corresponds exactly to $D$ and $DK1$ Lenzen made use of
the following equivalence, provable in $D + DK1$:
NK: \( NA \equiv \neg K \neg KA, \)

which allows the definition of conviction in \( \textbf{E1} \) in terms of knowledge, i.e as

\[ \text{DN: } NA =_{df} \neg K \neg KA. \]

In this way the following proposition holds:

**PROPOSITION 2.1:** \( \textbf{E1} + \text{DN} \) is deductively equivalent to \( \textbf{D} + \text{DK1} \).

On one direction, the derivability of \( \textbf{E1} \) from \( \textbf{D} \) is routine matter, and \( \text{DN} \) is deduced directly from \( \textbf{D} \). On the other direction, from \( \textbf{E1} \), \( \textbf{E4} \) and \( \text{DN} \), \( \text{DK1} \) follows. From \( \textbf{E5} \), \( \textbf{D1} \) results. \( \textbf{E5} \), \( \text{NK} \) and \( \textbf{E2} \) imply \( \textbf{D3} \). \( \textbf{D4} \) results from \( \textbf{E3} \) and \( \text{DN} \).

Now, the semantic counterpart to \( \text{DK1} \) is

\[ (\text{dk1}): S_K; =_{df} S_D; \cup \{i\}. \]

Intuitively, one might think of \( S_K; \) as the smallest set of worlds of those about which a person knows in \( i \), that one of them is the real world. Trivially the sets \( S_K; \) satisfy the condition

\[ \neg ka) \ i \in S_K; \]

(corresponding to the reflexivity of the accessibility relation \( R \) in the standard semantics for modal logic), so that they are non-empty. Besides, they do not satisfy condition 2b) for \( \textbf{D} \)-models, but only
in one direction

kb) $S_{K_j} \subseteq S_{K_i}$, provided $j \in S_{K_i}$,

(corresponding to the transitivity of $R$); and the following special case

kc) $S_{K_i} \subseteq S_{K_j} \cup \{i\}$, provided that $j \in S_{K_i}$,

which corresponds to E4. These conditions make possible a definition of $E1$-models which make system $E1$ sound and complete. The proof follows the guidelines for $D$ with the corresponding changes; it remains only to show that kc) holds in the canonical model.

Between both notions of knowledge and belief, there exist many relations like the obvious "bridge-principle" (connecting knowledge and belief in the sense of conviction)

ED1: $KA \supset NA$,

that is, knowledge implies conviction, and the plausible

ED2: $NA \supset \neg N \neg KA$,

ED3: $NA \supset KNA$.

Someone who is convinced of the truth of $A$ cannot be convinced that she or he does not know $A$, and knows the own convictions.

By means of the last two principles, instead of the stronger definition $DK1$, Lenzen also showed the derivability of an epistemic
correlate of modal S4.2 from D and DK1, which I call here E2. Its characteristic axiom, instead of E4, is:

\[ \neg K \neg KA \supset K \neg K \neg A, \]

its semantical counterpart being:

\[ \text{(kd)} \text{ for every } i, j, k \in I, \text{ if } j \in S_{K_i} \text{ and } k \in S_{K_j}, \text{ so there is a } 1 \text{ with } 1 \in S_{K_j} \text{ and } 1 \in S_{K_k}. \]

Formulated in terms of (dR) kd) means that the accessibility relation is "convergent" (s. v.g. HUGHES and CRESSWELL 1984, p. 31).

In fact, Lenzen proved (in 1979, p. 45):

PROPOSITION 2.2: E2 + DN is deductively equivalent to D + ED2 + ED3.

For, on one direction, E2 + DN contains system D and ED2 + ED3 implies axiom E5. On the other direction, from ED2 the sentence \( Na \supset \neg N\neg KA \) is obtained, and from ED3 \( \neg K \neg KA \supset A \), so that DN can be introduced. Through DN and D axiom E5 results.

3

So far we have two epistemic systems based on belief as conviction. Now, Plato himself considered DK1 as an unsatisfactory definition of knowledge. He argued that a person is said to know a state of affairs when he or she has reasons for his true belief in
it, i.e. when his or her belief is justified (cf. PLATO, *Thaetetus*, 201c, where he uses the expression *meta logos* to express the justification requirement). Consequently, a new definition of knowledge is needed, namely

\[
\text{DK2: } KA =_{df} JA \land A,
\]

where \( JA \) means that a person is justified in believing \( A \). Now, the notion of epistemic justification is rather obscure. If "justification" is to be understood as *logical* foundation of own beliefs, the notion comes to be circular or implies a *regressio ad infinitum*, just as E. Gettier in (1963) pointed out. In order to avoid such problems a different characterization of epistemic justification was proposed in my (1989) and (1990). According to it, justification is characterized in terms of *intersubjective standards*. The belief of a person is justified if, and only if, he or she obtained it according to intersubjective standards of rationality. These intersubjective standards are patterns or criteria regarded by the person within an "epistemic community" as valid and reliable for characterising a belief as justified. Observation, reliable reports, deductive and inductive arguments, coherence with former knowledge or with generally accepted assumptions are, among others, procedures which are normally accepted by the persons as criteria or standards for the characterization of a belief as justified.

On the basis of this notion of epistemic justification I construct a system for justified belief, which I call \( J \) and consists of the following axioms:

\[
\text{J0: } A, \text{ if } A \text{ is a classical tautology};
\]
J1: JA ⊃ \neg J\neg A;

J2: J(A ⊃ B) ⊃ (JA ⊃ JB);

J3: JA ⊃ JJA;

J4: J(JA ⊃ A),

and the rule

RJ: ⊢ JA, if ⊢ A,

together with modus ponens. J1 avoids the justified belief in contradictions; J2 and RJ express the "normality" of the system, J3 is an iteration principle, and the idea behind J4 is the justified claim to conclude true propositions from justified beliefs. J differs from D mainly in the lack of a corresponding axiom for D4, namely

J4*: \neg JA ⊃ J\neg JA,

which does not seem reasonable. Suppose there is a person a who is convinced of proposition ρ, but whose conviction is not regarded as justified. However, a really believes that his conviction is justified, that is:

(⋆) Nρ ∧ \neg Jρ ∧ NJρ.

Now, it is beyond doubt that justified belief implies belief, that is
JD1: $JA \supset NA$.

Then, we obtain from ($*$), JD1 and D1:

$\neg J\rho \land \neg J\neg J\rho$,

in contradiction with J4$*$.

The axiom J4 is intuitively sound. From the provable sentence in D

D5: $N(NA \supset A)$

and JD1 we obtain

JD2: $N(JA \supset A)$.

Since justified beliefs are regarded as reliable, JD2 is itself justified, so that J4 results.

Another relations between conviction and justified belief have also intuitive support:

JD3: $NA \supset JNA$;

JD4: $JA \supset NJA$.

The acceptability of JD2 relies upon the acknowledgment of the access to own beliefs as an intersubjective standard. Behind JD3 is the following assumption: If a person a rationally believes
a proposition $\rho$, $a$ did not satisfy the rationality criteria by chance but consciously.

On the contrary, the principle

\textbf{JD5*}: $NA \supset NJA$,

or its weaker version

\textbf{JD6*}: $NA \supset \neg N \neg JA$

are not acceptable. For I do not exclude the case of a person $a$, who is convinced of the truth of a proposition $\rho$, but $a$ is not convinced that he or she is justified in believing $\rho$. (That would be the case of convictions not obtained by means of intersubjective standards.) If they were sound, the equivalence between $N\rho$ and $\neg K \neg K \rho$ could be proved, so that E5 is obtained.

Now, a \textbf{J-model} is a triple $<S_J, V>$ like a D-model except for the conditions da)-db), which are to be replaced by

ja) = da) (with sets $S_J$ instead of $S_D$).

jb) $S_{J_j} \subset S_{J_i}$, provided $j \in S_J$,

and the condition

jc) $j \in S_{J_i}$, if $j \in S_{J_i}$.

Principle jb) reflects semantically axiom J3 and jc) axiom J4. Thus, the notion of J-model makes system J sound and complete,
as it is easy to prove.

Now, the definition DK2 reflects on the following semantical definition:

\[(d2): S_{K_i} = S_{J_i} \cup \{i\}.\]

According to this stipulation and the definition of J-model, sets \(S_{K_i}\) satisfy, on one hand, semantical conditions ka), kb) for \(E1\) models and an epistemic correlate of \(jc\), as it is easy to see. kb) is a version for \(jb\) and \(ka\) is obtained by \((d2)\). Naturally, they do no satisfy condition kd) for \(E2\) models. A simple counterexample is given by the finite model \(<I,S,V>\) defined as follows:
\(I = \{a,b,c,d,e\};\  Sa = I;\  Sb = \{b,d\};\  Sc = \{c,e\};\  Se = \{e\}.
\) Therefore, they do not satisfy condition kc) for \(E1\) models since kc) implies kd). On the other hand, the sets \(S_{J_i} \cup \{i\}\) satisfy trivially conditions ka) and kb).

Consequently, at most conditions ka) and kb) can be obtained from J-models and \((d2)\). (Besides, axioms E1, E2, E3 and RK derive from system J and DK1.) I call the epistemic system consisting only of E1-E3 and RK, \(E3\), which is an epistemic counterpart to S4. Finally, by means of sets \(S_{K_i}\), \(E3\)-models for E3 can be defined, which makes \(E3\) sound and complete.

Although the operator J does not have a definition in terms of \(K\) (like \(DN\) for \(N\)), it holds semantically: for every \(i \in I\), \(S_{K_i} - \{i\} \subseteq S_{J_i}\), but for some \(i \in I\), \(S_{J_i} = S_{K_i}\).

To sum up, it holds then semantically:

**Proposition 3.1:** For every sentence \(A\) of the epistemic language, \(A\) is valid in all \(E3\) models iff \(A\) is valid in all J-models
expanded with (dk2).

In other words E3 is the strongest epistemic system, which can be constructed from J + DK2, in the sense that at most E3 is derivable from J + DK2. Hence, it was shown that S4 is exactly the system that represents the notion of knowledge as justified belief.

References


—, (1980): *Glauben, Wissen und Wahrscheinlichkeit*, Wien-
N. York, Springer.


University of Buenos Aires
Buenos Aires, Argentina