A NEW APPROACH TO FORMALIZATION OF A LOGIC OF KNOWLEDGE AND BELIEF*

Kathleen Johnson Wu

§ 1. Introduction.

There are no logical reasons why someone should know any of the logical consequences of what he knows. Yet, there seems to be general agreement among those interested in formalizing the logic of knowledge that, if a sufficiently powerful system is to be developed, it must be assumed that everyone knows at least some of the logical consequences of what he knows. No consensus has been reached on exactly which logical consequences he must know, however. And, as a result, various systems have been proposed, but none widely accepted. The same is true for the logic of belief.

In this paper a different approach is taken. No extralogical assumptions are involved in the system proposed which consists of natural deduction rules for an S4 alethic modal logic together with rules for knowledge and belief as well as some other closely related notions. What makes the system interesting, besides the notions formalized, is 1) the ease with which straightforward extensions of the system can be shown to contain other systems, e.g. Hintikka's, and 2) the fact that it provides a method for dealing with Moore's Paradox, the Prediction Paradox and other similar problems.

§ 2. The language. The symbols are the following

(1) *Sentential connectives*: — (not), & (and), v (or), ⊃ (only if), ≡ (if and only if).

(2) *Alethic modal operators*: □ (it is logically necessary that), ◇ (it is logically possible that).

(3) *Epistemic relations*: K (knows that), K^0 (is open to consider evidence that), $\dot{K}$ (implicitly knows that), K^0 (is open to know that).

(4) *Doxastic relations*: B (believes that), B^0 (is open to believe that), B (implicitly believes that), B^0 (is open to consistently believe that).

(5) *Auxiliary signs:* Parentheses.

(6) *Names of individuals*: a, b, c, ... As syntactical variables that have these symbols as their values, I use the letters: *a, b, c,* ...

(7) *Atomic sentences*: p, q, r, p_1, q_1, r_1, ... As syntactical variables that have as their values these symbols and those strings of symbols defined as sentences in (10) below, I use the letters: p, q, r, p_1, q_1, r_1, ...

Special symbol formations are as follows:

(8) *Epistemic operators:* "K_a," "K^0_a," "$\dot{K}_a$," "$\dot{K}^0_a$." [1]

(9) *Doxastic operators*: "B_a," "B^0_a," "$\dot{B}_a$" "$\dot{B}^0_a$."

(10) *Sentences*: The class of sentences is defined as follows:

(a) All atomic sentences (listed in (7) above) are sentences.

(b) "$-p$," "$p \& q$," "$p \vee q$," "$p \supset q$," and "$p \equiv q$" are sentences.

(c) "$\Box p$" and "$\Diamond p$" are sentences.

(d) "$K_a p$," "$K^0_a p$," "$\dot{K}_a p$," and "$\dot{K}^0_a p$" are sentences.

(e) "$B_a p$," "$B^0_a p$," "$\dot{B}_a p$," and "$\dot{B}^0_a p$" are sentences.

(f) The only sentences are those that are such in virtue of (a) — (e).

§ 3. Proof method and rules.

The reader is assumed to be familiar with the proof method, the rules of reiteration and the rules of introduction and elimination for the sentential connectives and the alethic modal operators. They are essentially the same as Fitch uses to develop his two-valued S4 propositional alethic modal system in [5]. A detailed explanation of the method and rules is given in [2]. Only the rules for the epistemic and doxastic operators will be given here.

The symbol 'K' for knowledge is to be understood as expressing a relation between an individual and a proposition such that if this relation obtains between an individual and a proposition, it follows that the individual has conclusive evidence for the proposition and that the proposition is, therefore, true. It need not follow that the individual knows any of the logical consequences of the proposition. The rule for knowledge elimination below states, in effect, that if someone knows a proposition, that proposition is true.

Knowledge elimination ("kn elim"). p is a direct consequence (hereafter abbreviated as "d.c.") of "$K_a p$."

The symbol 'K^0' for openness to consider evidence is to be understood as expressing a relation that obtains between an individual and any proposition the negate of which he does not know. That is, someone is open to consider evidence that a proposition is true if and only if it is not the case that he knows that it is false. This is what is, in effect, stated by the following pair of introduction and elimination rules

Openness to consider evidence introduction ("op con ev int"). "$K^0_a p$" is a d.c. of "$— K_a — p$."

Openness to consider evidence elimination ("op con ev elim"). "$— K_a — p$" is a d.c. of "$K^0_a p$."

The symbol '$\dot{K}$' for implicit knowledge is to be understood as expressing a relation that obtains between an individual and any proposition that is a logical consequence of propositions he knows. That is, someone implicity knows a proposition if and only if the proposition is logically implied by what he knows. [2] This is what is, in effect, stated by the following pair of introduction and elimination rules:

Implicit knowledge introduction ("impl kn int"). "$\dot{K}_a p$" is a d.c. of the sentences "$K_a q_1 \& .. \& K_a q_n$" and "$\Box ((q_1 \& .. \& q_n) \supset p)$." Here, $n \geqslant 0$. In case $n = 0$, the sentence "$K_a q_1 \& .. K_a q_n$" is dropped and the sentence "$\Box ((q_1 \& .. \& q_n) \supset p)$ is changed to "$\Box p$."

Implicit knowledge elimination ("impl kn elim"). r is a d.c. of "$\dot{K}_a p$" and a regular subproof general with respect to $q_1, .., q_n$ having the sentences "$K_a q_1 \& .. \& K_a qn$" and "$\Box ((q_1 \& .. \& q_n) \supset p)$" as its only hypotheses and r among its items. Here $n \geqslant 0$ and it is assumed that neither $q_1, ..,$ nor q_n occur in r.

This rule may be expressed schematically [3] as follows:

1	$\dot{K}_a p$		hyp
2	$q_1, .., q_n$	$K_a q_1 \& .. \& K_a q_n$	hyp
3		$\Box ((q_1 \& .. \& q_n) \supset p)$	hyp
.		.	
.		.	
.		.	
i		r	
j	r		1, 2 - i, impl kn elim

By *a regular subproof general with respect to* $q_1, .., q_n$ is meant a regular subproof in which the method of deriving any item or some variation of that method must hold regardless of the length n of the sequence of sentences $q_1, .., q_n$ and regardless of what sentences $q_1, .., q_n$ are. To insure that a subproof is general in this respect it is understood that no sentence in which the sentence $q_1, ..,$ or q_n occurs can be reiteratet into the subproof.

The symbol '$\dot{K}^0$' for openness to know is to be understood as expressing a relation that obtains between an individual and any proposition the negate of which is not a logical consequence of that individual's knowing what he knows. That is, a person is open to know a proposition if it is consistent with what he knows as well as with his knowing what he knows. This is what is, in effect, stated by the following pair of introduction and elimination rules:

Openness to know introduction ("op kn int"). "$\dot{K}^0_a p$" is a d.c. of a categorical subproof general with respect to $q_1, \ldots, q_n$ having the sentence "$(K_a q_1 \& \ldots \& K_a q_n) \supset - \Box ((K_a q_1 \& \ldots \& K_a q_n) \supset - p)$" among its items. Here, again, it is assumed that $n \geqslant 0$.

This rule may be expressed schematically as follows:

1	$q_1, \ldots, q_n$	$(K_a q_1 \& \ldots \& K_a q_n) \supset$ $\Box((K_a q_1 \& \ldots \& K_a q_n) \supset - p)$	hyp
			1, op kn in
2	$\dot{K}^0_a p$		

Openness to know elimination ("op kn elim").
"$(K_a q_1 \& \ldots \& K_a q_n) \supset - \Box ((K_a q_1 \& \ldots \& K_a q_n) \supset - p)$" is a d.c. of "$\dot{K}^0_a p$." Here, $n \geqslant 0$. In case $n = 0$, the sentence "$K_a q_1 \& \ldots \& K_a q_n) \supset - \Box ((K_a q_1 \& \ldots \& K_a q_n) \supset - p)$" is changed to "$- \Box - p$."

The symbol 'B' for belief is to be understood as expressing a relation between an individual and a proposition such that if this relation obtains between an individual and a proposition, it follows that the individual is persuaded that the proposition, rather than its negate, is true, and, therefore, does not believe its negate. It need not follow, however, that the individual believes any of the logical consequences of the proposition. It is assumed that if an individual knows a proposition then he also believes it. This is what is in effect stated by the following rule:

Belief introduction ("be int"). "$B_a p$" is a d.c. of $K_a p$."

The next rule states, in effect, that if someone believes a proposition, he does not believe that it is false.

Belief elimination ("be elim"). "$- B_a - p$" is a d.c. of "$B_a p$."

The symbol 'B^0' for openness to believe is to be understood as expressing a relation that obtains between an individual and any proposition the negate of which that individual does not believe. That is, an individual is open to believe that a proposition is true if and only if he does not believe that it is false. This is what is, in effect, stated by the following pair of introduction and elimination rules:

Openness to believe introduction ("op be int").
"$B^0_a p$" is a d.c. of "$-B_a - p$."

Openness to believe elimination ("op be elim").
"$-B_a - p$" is a d.c. of "$B^0_a p$."

The symbol '$\dot{B}$' for implicit belief is to be understood as expressing a relation between an individual and any proposition that is a logical consequence of propositions he believes. That is, an individual implicity believes a proposition if and only if it is logically implied by what he believes. (4) This is what is, in effect, stated by the following pair of introduction and elimination rules:

Implicit belief introduction ("impl be int"). "$\dot{B}_a p$" is a d.c. of the pair of sentences "$B_a q_1 \,\&\, .. \,\&\, B_a q_n$" and "$\Box\, ((q_1 \,\&\, .. \,\&\, q_n) \supset p)$." Here, $n \geqslant 0$. In case $n = 0$, the sentence "$B_a q_1 \,\&\, .. \,\&\, B_a q_n$" is dropped and the sentence "$\Box\, ((q_1 \,\&\, .. \,\&\, q_n) \supset p)$" is changed to "$\Box\, p$."

Implicit belief elimination ("impl be elim"). r is a d.c. of "$\dot{B}_a p$" and a subproof general with respect to $q_1, .. , q_n$ having the sentences "$B_a q_1 \,\&\, .. \,\&\, B_a q_n$" and "$\Box\, ((q_1 \,\&\, .. \,\&\, q_n) \supset p)$" as its only hypotheses and r among its items. It is assumed that $n \geqslant 0$ and that neither $q_1, .. ,$ nor q_n occur in r.

The symbol '$\dot{B}^0$' for openness to consistently believe is to be understood as expressing a relation that obtains between an individual and any proposition the negate of which is not logically implied by what he believes or by his believing what he believes. That is, an individual may consistently believe a proposition if it is consistent with his believing what he believes as well as with what he believes. This is what is, in effect, stated by the following pair of introduction and elimination rules:

Openness to consistently believe introduction ("op cons be int"). "$\dot{B}^0_a p$" is a d.c. of a categorical subproof general with respect to $q_1, .. , q_n$ having the sentences "$(B_a q_1 \,\&\, .. \,\&\, B_a q_n)$

$\supset - \Box ((B_a q_1 \,\&\, q_1 \,\&\, .. \,\&\, B_a q_n \,\&\, q_n) \supset - p)$" among its items. Here, again, it is assumed that $n \geqslant 0$.

This rule may be expressed schematically as follows:

1	$q_1, \ldots, q_n$	$(B_a q_1 \,\&\, .. \,\&\, B_a q_n) \supset$ $-\Box((B_a q_1 \,\&\, q_1 \,\&\, .. \,\&\, B_a q_n \,\&\, q_n) \supset -p)$	hyp
			1, op cons
2	$\dot{B}^0_a p$		be int

Openness to consistently believe elimination ("op cons be elim"). "$(B_a q_1 \,\&\, .. \,\&\, B_a q_n) \supset - \Box ((B_a q_1 \,\&\, q_1 \,\&\, .. \,\&\, B_a q_n \,\&\, q_n) \supset - p)$" is a d.c. of "$B^0 p$." Here, $n > 0$. In case $n = 0$, the sentence "$(B_a q_1 \,\&\, .. \,\&\, B_a q_n) \supset - \Box ((B_a q_1 \,\&\, q_1 \,\&\, .. \,\&\, B_a q_n \,\&\, q_n) \supset - p)$" is changed to "$- \Box - p$."

§ 4. Properties of the system.

From the point of view of modal logic the interesting operators are "$\dot{K}_a$," "$\dot{K}^0_a$," "$\dot{B}_a$," and "$\dot{B}^0_a$." "$\dot{K}_a$" has properties analogous to those of the operator for necessity in an alethic M system; whereas, "$\dot{K}^0_a$" has properties analogous to those of the operator for possibility in an alethic S4 system. This can be seen from the fact that the following can be proven as theorems (T) and derived rules (DR) of the system.

(T1) "$\dot{K}_a p \supset p$"

(T2) "$(\dot{K}_a(p \supset q)) \supset (\dot{K}_a p \supset \dot{K}_a q)$"

(DR1) "$\dot{K}_a p$" is a consequence of a strict subproof in which p is an item.

(T3) "$p \supset \dot{K}^0 p$"

(T4) "$(\dot{K}^0(p \vee q)) \supset (\dot{K}^0 p \vee \dot{K}^0 q)$"

(T5) "$\dot{K}^0 \dot{K}^0 p \supset \dot{K}^0 p$"

(DR2) "$- \dot{K}^0 - p$" is a consequence of a strict categorical subproof with p among its items.

"$\dot{B}_a$" has properties analogous to those of the operator for obligation in a deontic M system except that implicit belief does not imply possibility. "$\dot{B}^0_a$" has properties analogous to those of the operator for permission in a deontic S4 system. This can be seen from the fact that the following can be proven as theorems and derived rules of the system.

(T6) "$(\dot{B}_a(p \supset q)) \supset (\dot{B}_a p \supset \dot{B}_a q)$"

(DR3) "$\dot{B}_a p$" is a consequence of a strict categorical subproof with p among its items.

(T7) "$(\dot{B}^0(p \vee q)) \supset (\dot{B}^0 p \vee \dot{B}^0 q)$"

(T8) "$\dot{B}^0\dot{B}^0 p \supset \dot{B}^0 p$"

(DR4) "$- \dot{B}^0 - p$" is a consequence of a strict categorical subproof with p among its items.

If 'K,' 'K^0,' 'B,' and 'B^0' take the place of their dotted counterparts in the above theorems and derived rules, only (T1) and (T3) remain theorems. None of the following remain theorems, however, if '$\dot{K}$,' '$\dot{K}^0$,' '$\dot{B}$,' and '$\dot{B}^0$' take the place of their undotted counterparts.

(T9) "$B_a p \supset B^0_a p$"

(T10) "$B^0_a p \equiv - B_a - p$"

(T11) "$K^0_a p \equiv - K_a - p$"

Neither "$B_a p \equiv - B^0_a - p$" nor "$K_a p \equiv - K^0_a - p$" nor their dotted counterparts are theorems.

§ 5. Applications to Ideal Usage.

According to Hintikka's interpretation in *Knowledge and Belief* of the system he developes there, the results of the system are applicable to knowledge and belief statements only under the assumption that (1) everyone knows all the logical consequences of what he knows, (2) everyone believes all the logical consequences of what he believes, (3) everyone knows that he knows whatever he knows and (4) everyone believes that he believes whatever he believes. I have shown [see 2, pp 23-33, 6 and 7] that this interpretation of Hintikka's is inadequate and that in fact the results of his system are applicable only under the additional assumption that (5) everyone knows that (1) - (4), (6) everyone knows that (5), (7) everyone knows that (6), and so on.

The difference between Hintikka's system and a system which is applicable only under the assumption that (1) - (4) can be brought out nicely by considering two extensions, A and B, of my system. A is applicable only under the assumption that (1) - (4) and B, like Hintikka's, only under the assumption that (1), (2), (3), ...

A is formed by adding the following four rules for ideal usage.

Ideal usage$_1$ (i u$_1$). "$K_a p$" is a d.c. of "$\dot{K}_a p$"

Ideal usage$_2$ (i u$_2$). "$B_a p$" is a d.c. of "$\dot{B}_a p$"

Ideal usage$_3$ (i u$_3$). "$K_a K_a p$" is a d.c. of "$K_a p$"

Ideal usage$_4$ (i u$_4$). "$B_a B_a p$" is a d.c. of "$B_a p$"

Because these rules reflect extralogical assumptions, it is to be understood that no item of a strict subproof or item of an item of a strict subproof or so on, is an item in virtue of a rule of d.c. for ideal usage or any other kind of special usage. Adoption of the following four rules which are reminiscent, respectively, of Hintikka's (C. — K) (C. — P), (C. — B) and (C. — C) would have the same effect as the addition of the rules i u$_1$, i u$_2$, i u$_3$ and i u$_4$. In other words, the result would be a system equivalent to A.

Ideal usage (i u$_5$). "$\dot{K}^0_a - p$" is a d.c. of "$- K_a p$"

Ideal usage (i u$_6$). "$K_a - p$" is a d.c. of "$- \dot{K}^0_a p$"

Ideal usage (i u$_7$). "$\dot{B}^0_a - p$" is a d.c. of "$- B_a p$"

Ideal usage (i u$_8$). "$B_a - p$" is a d.c. of "$- \dot{B}^0 p$"

In addition to those theorems and rules derivable in the basic system the following are derivable in A.

(AT1) "$(K_a(p \supset q)) \supset (K_a p \supset K_a q)$"

(AT2) "$K_a p \supset K_a K_a p$"

(ADR1) "$K_a p$" is a consequence of a strict subproof in which p is an item.

(AT3) "$K_a p \equiv - K^0_a - p$"

(AT4) "$\dot{K}\ p \equiv - \dot{K}^0 - p$"

(AT5) "$\dot{K}^0 p \equiv - \dot{K}\ - p$"

(AT6) "$(B_a\ (p \supset q)) \supset (B_a p \supset B_a q)$"

(AT7) "$B_a p \supset B_a B_a p$"

(ADR2) "$B_a p$" is a consequence of a strict subproof in which p is an item.

(AT8) "$B_a p \equiv - B^0_a - p$"

(AT9) "$\dot{\mathrm{B}}\, p \equiv -\dot{\mathrm{B}}^0 - p$"

(AT10) "$\dot{\mathrm{B}}^0 p \equiv -\dot{\mathrm{B}} - p$"

(AT11) "$\dot{\mathrm{B}}\, p \supset \dot{\mathrm{B}}^0 p$"

As those know who are familiar with Hintikka's system, it is S4. At first glance it might appear that A is also S4, but it is not, because (ADR1) and (ADR2) are restricted rules, as are for that matter (DR1), (DR2), (DR3) and (DR4) as a result of the addition of rules for ideal usage to the system. According to (ADR1) and (ADR2) if p is theorem, it does not follow that "$K_a p$" and "$B_a p$" are theorem unless p is a theorem of the basic system and not one in virtue of a rule of special usage. In other words A does not have a rule fully analogous to the Rule of Necessitation, which is at the basis of any S4 system.

With the addition of the following rule to A, B is formed. It is clearly S4.

Ideal usage$_9$ (i u$_9$). "$K_a p$" is a d.c. of a "K_a" categorical subproof having p among its items.

This rule may be expressed schematically as follows:

1	K_a	p	
2	$K_a p$		1, i u$_9$

By a "K_a" *categorical subproof* is understood a subproof into which no reiterate is allowed. It is to be understood, however, that an item of such a subproof may be an item in virtue of any of the rules of d.c. for ideal usage as well as any of the standard rules of d.c.

In addition to those theorems and rules derivable in A the following are derivable in B.

(BT1) "$K_b\, ((K_a\, (p \supset q)) \supset (K_a p \supset K_a q)))$"

(BT2) "$K_c K_b\, (K_a p \supset K_a K_a p)$"

(BT3) "$K_d K_c K_b\, ((B_a\, (p \supset q)) \supset (B_a p \supset B_a q)))$"

(BT4) "$B_e K_d B_c K_b\, (K_a p \supset K_a p)$"

(BDR1) "$B_a p$" is a consequence of a "B_a" categorical subproof having p among its items. A "B_a" categorical subproof is to be understood as satisfying the same conditions as a "K_a" categorical subproof.

The following is a proof of (BT2).

1	K_c	K_b		— K_ap	hyp
2				K_aK_ap	1, i u_3
3			$K_ap \supset K_aK_ap$		1-2, imp int
4		$K_b (K_ap \supset K_aK_ap)$			1-3, i u_9
5	$K_cK_b (K_ap \supset K_aK_ap)$				1-4, i u_9

§ 6. Applications to Actual Usage.

Most formalizations of the logic of knowledge and belief are developed with the purpose of explaining the way knowledge and belief statements are actually used. Indeed, Hintikka considers it a serious drawback that the results of his system in [4] are applicable to such statements

> only insofar as our world approximates,
> or can be made to approximate an "epistemically perfect world." [8, p. 2]

As a solution Hintikka has suggested restricting his system in such a way that if *q* is a logical consequence of *p* and a person knows (believes) that *p*, it need not follow that he also knows (believes) that *q* unless "$p \supset q$" is a surface tautology at the depth of *p* [9]. Even with this restriction, however Hintikka admits that his system will still involve some idealization, because it will be applicable only under the assumption that everyone "fully understands" what he knows and believes and not everyone does.

The problem with Hintikka's proposed solution is that regardless of how weak an extralogical assumption may be, it is possible that it is not warranted in a particular case. The advantage of the basic system proposed in this paper is that it is not restricted in its application by any such assumption, yet, in those cases in which special assumptions are needed, they can be made. For instance, once Hintikka works out the details of his restricted system it should not be difficult to see how to develop, as in § 4, an extension of the basic system proposed in this paper which would like Hintikka's be applicable only under the assumption that everyone "fully understands" what he knows and believes. This would be done by adopting special rules of usage to correspond to Hintikka's extralogical assumptions. Without the addition of such rules, however, the basic system is well-suited, indeed in my view better suited, to deal with problems the correct analysis of which depends on making clear particular assumptions involved rather than on appealing to general standards of deductive behavior. Moore's Paradox and the Prediction Paradox, both of which are dealt with in another paper, appear to be two such problems.

University of Alabama

NOTES

* This paper is an outgrowth of [1], [2] and [3]. I am indebted to Professor Frederic B. Fitch for many helpful comments. I also wish to thank Professor Hugues Leblanc for his remarks on an earlier draft of this paper.

(1) Double quotation marks are used in the same way that Jaakko Hintikka uses them in [4]. They are placed around an expression to refer to any expression which results from it by replacing each syntactical variable in it (like the italic letters '*a*' and '*p*') by some expression that variable refers to.

(2) In [2] a stronger notion of implicit knowledge is formalized. According to it an individual implicitly knows a proposition if and only if the proposition is logically implied by what he knows or by his knowing what he knows. The noton formalized here seems closer to what is ordinarily understood by "implicit knowledge."

(3) Rules are expressed schematically and proofs carried out using atomic sentences and names of individuals. It is to be understood, however, that they hold equally as well if the atomic sentences used are replaced by any other sentences and the names of individuals are replaced by any other names of individuals.

(4) In [2] a stronger notion of implicit belief is formalized. According to it an individual implicitly believes a proposition if and only if proposition is logically implied by what he believes or by hs believing what he believes. The notion formalized here seems closer to what is ordinarily understood by "implicit belief."

REFERENCES

[1] JOHNSON, Kathleen G. "A Formalization of the Logic of Knowledge and Belief," (Abstract) *The Journal of Symbolic Logic*, vol. 33 (1968) pp. 641-642.

[2] WU, Kathleen Gibbs Johnson. "A New Formalization of the Logic of Knowledge and Belief," a dissertation presented for the degree of Doctor of Philosophy in Yale University, 1970. Ann Arbor, Michigan: University Microfilm.

[3] WU, Kathleen Johnson. "The Question of Logical Omniscience in Epistemic Logic," (Abstract) *The Journal of Symbolic Logic*, vol. 37 (1972) pp. 438-439.

[4] HINTIKKA, Jaakko. *Knowledge and Belief*. Ithaca, New York: Cornell University Press, 1962.

[5] FITCH, Frederic B. 'Natural Deduction Rules for Obligation," *American Philosophical Quarterly*, vol. 3 (1963) pp. 1-12.

[6] WU, Kathleen G. Johnson. "Hintikka and Defensibility," *Ajatus*, vol. 32 (1970) pp. 25-31.

[7] WU, Kathleen Johnson. "Hintikka and Defensibility: Some Further Remarks," *Journal of Philosophical Logic*, vol. 1 (1972) pp. 259-261.

[8] HINTIKKA, Jaakko. «"Knowing Oneself» and Other Problems in Epistemic Logic,» *Theoria*, vol. 32 (1966) pp. 1-13.

[9] HINTIKKA, Jaakko. "Knowledge, Belief and Logical Consequence," *Ajatus*, vol. 32 (1970) pp. 32-47.