Logique & Analyse **129-130** (1990), 3-20

KNOWING, KNOWLEDGE, KNOWN

David BERLINSKI

1. Knowledge with an Indefinite Subject(1)

It is clear that

1.1 *A knows P* $\supset$ *P is known*,

but does

1.2. *P is known* $\supset$ $(\exists x)$*(x knows P)?*

In mathematics,(2) the subject of knowledge is often indefinite; knowledge is *distributed* over agents.(3)

Suppose that

1.3 *A knows P* & $\sim$*(A knows Q)*

and

(1) A word about the point of view that informs this paper: Epistemic logic, broadly conceived, covers three programs. There is, in the first place, the effort to determine a set of valid epistemic sentences. It is an effort that is easily displaced by metamathematical speculations about the system that results. There is, in the second place, the effort to determine the logical form of epistemic sentences. And finally there is the effort to explore the structure of an epistemic space whose points are equivalent sets of epistemic sentences. I regard the first effort as uninteresting and the second as hopeless. It is this last program that is of interest here. George N. Schlesinger's *The Range of Epistemic Logic* (Humanities Press, New Jersey: 1985) is written from a comparable point of view. An overview of related work is provided by Joseph Halpern's IBM Research Report, 'Reasoning about Knowledge,' RJ 8068 (74007), April, 1991.

(2) And not only in mathematics. Epistemic distribution is in evidence whenever historians talk, say, of what is known by a given culture (or era).

(3) The distinction between lumped and distributed parameters in control theory is similar to the distinction between singular and distributed epistemic subjects.

1.4 B *knows* Q & $\sim$(B *knows* P).

Suppose further that

1.5 A *knows* (B *knows* Q),

and that

1.6 B *knows* (A *knows* P).

It follows from 1.1 that

1.7 P *is known* & Q *is known*.

The epistemic principle

1.8 P *is known* & Q *is known* $\supset$ $\{P$ & $Q\}$ *is known*

is obviously false; 1.9 is weaker and more plausible:

1.9 P *is known* & Q *is known* $\supset$ $\{P$ & $Q\}$ *is known* $\equiv$ $\exists x \exists y$[x *knows* P & y *knows* Q & x *knows* (y *knows* Q) & y *knows* (x *knows* P)].

Note that 1.9 does *not* imply

1.10 x *knows* (y *knows* P) $\supset$ x *knows* P

even though

1.11 x *knows* (x *knows* P) $\supset$ x *knows* P.

It follows from 1.1 and 1.3 that

1.12 P *is known*;

and from 1.1 and 1.4 that

1.13 Q *is known*.

It follows from 1.7 and 1.9 that

1.14 *$\{P$ & $Q\}$ is known.*

Yet from 1.3 and 1.4 it follows that

1.15 $\sim\exists x(x$ *knows* $\{P$ & $Q\})$.

1.14 and 1.15 specify a property of *separability*. The epistemic attitudes of mathematicians toward proofs generated by computer searches are often separable. The Appel-Hakens proof that every planar map is four colorable provides an example.([4]) The practical problems involved in checking complex and very lengthy computer programs also require that mathematicians and computer scientists adopt separable epistemic attitudes toward their work.([5])

2. *Gettier Paradoxes*

On many views

2.1 *A knows that P $\equiv$ P & A believes P & A believes P justifiably*

([4]) In addition to cases in which the subjects of knowledge are indefinite, there are cases as well in which the *object* of knowledge is indefinite. The question posed by Joseph Rotman: "When does one "know" a group?" suggests that there is an epistemic sense in which knowledge takes as its objects entities that are complex but non-propositional. See. Joseph Rotman, *The Theory of Groups* (Alyn & Bacon, Boston: 1973), p. 9. Rotman suggests that an agent A knows **G**, where **G** is a group, if and only if given any other group **H**, A can determine whether **G** and **H** are isomorphic. This formulation becomes precise (and interesting) if the phrase 'can determine whether' is replaced by 'can prove that.'

([5]) Knowledge that is distributed among epistemic agents is not at all the same thing as *common* knowledge. Knowledge is held in common if everyone in an epistemic community knows roughly the same thing. Common knowledge has been the subject of investigation in the computer science and aritificial intelligence communities. And there are, of course, interesting connections between distributed and common knowledge, notions which in some sense are dual. See J.Y. Halpern and Y. Moses, 'Knowledge and Common Knowledge in a Distributed Environment,' *Journal of the ACM*, Volume 37(3), pp. 549-587.

has the force of an irrefragable assumption.(6)
It is false that

2.2 *A believes P & (P ⊃ Q) ⊃ A believes Q,*

but

2.3 *A believes P & A knows (P ⊃ Q) ⊃ A believes Q*

is plausible;(7) 2.3 in turn entails

2.4 *A believes P & A believes (P ⊃ Q) ⊃ A believes Q*

in virtue of 2.1 itself.

By the *operator of appearances*, I mean the variable-binding, sententia operator *it appears to x that P*. It is possible that a proposition be false anc appear true; or that it be true and appear false. These are semantic obser- vations. Two metaphysical theses now follow:

2.5 ◇[(∀x) ~ {(*it appears to x that P*) ⊃ *P*}];

but equally

2.6 ◇[~ {*P* ⊃ (∃x)(*it appears to x that P*)}].

2.5 and 2.6 are in a sense principles of realism.(8) They are followed ir

(6) The fact that 2.1 is intuitively compelling is itself not without significance for philoso phy, a discipline historically weak in stable intellectual structures. On the other hand, th decision to treat 2.1 as a *definition* may well be a mistake.

(7) Indeed, on the assumption that *A knows (P ⊃ Q) ⊃ (A knows P ⊃ A knows Q)*, 2.: expresses no more than a triviality. On the assumption, note. The proposition that *A believe. P & (A believes P ⊃ A believes Q) ⊃ A believes Q* is, of course, an instance of a law o logic.

(8) Although a determined irrealist could with ingenuity defend both assertions. Susa Haack has recently argued that the proposition (∀*P*) ◇ (*A believes P*) expresses a version o failibilism, the doctrine that "we are capable of believing any proposition irrespective o whether it is true or false." Susan Haack, 'Failibilism and Necessity,' *Synthese*, Volume 41 1979. The point is discussed in Schlesinger, *Op. Cit.*, p. 26. The words quoted are Schle singer's as well. Failibilism will be true only if there are no sentences that compel appearan

turn by a principle of *credulity*:

2.7 *it appears to A that P* $\supset$ *A believes P.*

It is 2.7 that draws a connection between perception and belief. Absent some such connection, it would be difficult to understand how beliefs are ever formed on the basis of appearances, and thus to understand how science is itself possible.

Let

2.8 $J_A[\Xi, \xi]$

indicate that the finite set of sentences Ξ justifies the finite set of sentences ξ to A. ξ may be either simple or complex. If ξ is complex, its justification must be recursive. Simple sentences must either be self-justifying or justified by the structure of sets of simple sentences in Ξ. There is no reason to believe that the classification of such sets of sentences is apt to be any easier than the classification of finite simple groups.

Quite apart from these vexed issues, there is the question of the atttitude an agent must bear toward justification in order to be justified in believing a proposition. The thesis that 2.8 is true if and only if

2.9 *A knows* $J_A[\Xi, \xi]$

leads to circularity in the analysis of 2.1; but equally, the thesis that 2.8 is true if and only if

2.10 *A believes* $J_A[\Xi, \xi]$

commences a third-man spiral if 2.10 must itself be justified. If an agent's attitude toward justification is neither knowledge nor justified belief, it is difficult to know what else it might be.

Suppose now that

2.11 *it appears to A that Fa*;

ces. See section 4, especially 4.3.

it follows from 2.7 that

2.12 *A believes Fa.*

Suppose further that

2.13 *A believes* $(Fa \supset \exists xFx)$.

It follows from 2.4 that

2.14 *A believes* $\exists xFx$

Whatever the definition of justification, it might be argued,

2.15 $J_A[\Xi, \{\exists xFx\}]$

must be a consequence of a clause in the definition of justification if for some *a*

2.16 $J_A[\Xi, \{Fa\}]$

and

2.17 *A believes Fa.*

Assume that in virtue of 2.17 and 2.16, 2.15 is in fact true. This is an *assumption*, note.

Suppose further that

2.18 $\sim Fa \ \& \ \exists xFx$.

It follows directly that

2.19 $\sim Fa$;

and it follows from 2.5 that 2.19 and 2.11 are at least consistent.

2.20 *A knows that* $\exists xFx$.

now follows from 2.14 (belief), 2.15 (justification), 2.18 (truth), and 2.1.(9)

3. *Primary Paradoxes*

Suppose that

3.1 *it appears to A that P.*

It follows from 2.7 that

3.2 *A believes P.*

Suppose further that

3.3 $\sim$(*A believes Q*)

and that

3.4 *A believes* $P \supset P \vee Q$.

It follows from 3.2 and 2.4 that

3.5 *A believes* $P \vee Q$.

Disjunctions are complex statements; their justification is recursive:

(9) See Keith Lehrer, *Theory of Knowledge* (Westview Press, Boulder & San Francisco: 1990), pp. 16-17. 2.20 is paradoxical only to the extent that intuition upholds its denial; no formal paradox is forthcoming, a circumstance that may well suggest that what is at issue in this area of epistemology is less the adequacy of a definition than the definition of adequacy. My own view, which I lack the space to argue here, is that 2.1 and like statements are not so much true as *typically* true. They resemble the assertion that straight lines in two-dimensional space are likely to intersect. The thesis is set out in my *Prediction and Infection in Dynamical Systems* (forthcoming from the Princeton University Press). The endeavor to so purify 2.1 as to eliminate all counter-examples seems to me as misguided as the correlative effort to define the derivative D[*f*] of a real valued function *f* in such a way that D[*f*] is never 0.

3.6 $J_A[\Xi, \{Fa \vee Fb\}]$

if

3.7 $J_A[\Xi, \{Fa\}]$ *or* $J_A[\Xi, \{Fb\}]$

and

3.8 *A believes* $\{Fa \vee Fb\}$.

Assume that in virtue of 3.8 and 3.7, 3.6 is in fact true. This, too, is an *assumption*.

Suppose now that

3.9 $\sim P \;\&\; Q$.

It follows that

3.10 $\sim P$;

it follows from 2.5 that 3.10 and 3.1 are consistent.

3.11 *A knows* $(P \vee Q)$

follows from 3.5 (belief), 3.6 (justification), and 3.9 (truth).

Gettier paradoxes may be reduced to primary paradoxes on the assumption that

3.12 $\exists x Fx$

is equivalent to

3.13 $Fa \vee Fb \vee Fc \vee \ldots \vee Fk$, where {a, b, c, ... , k} are all of the elements in a domain D.(10)

4. *Appearance, Adherence, Compunction*

The sentence

4.1 *it appears to A that P*

is *indicative* (of the truth) if

4.2 it appears to A that $P \supset P$;

compelled (by the truth) if(11)

4.3 $P \supset$ *it appears to A that P*;

and *miscompelled* (by the truth) if:

4.4 $\exists Q(Q \,\&\, Q \supset$ *it appears to A that P*);

It is in general false that

4.5. *A believes* $P \supset P$;

but

4.6 *A believes* (*it appears to A that P*) $\supset$ *it appears to A that P*

is rather more persuasive in virtue of the authority an agent brings to his

(10) Arguments showing that certain epistemic paradoxes are, under given assumptions, variants of one another are important to the extent that they point toward the distant goal of *classifying* epistemic singularities. Sard's theorem says, roughly, that the set of singular points of a continuously differentiable real valued function has measure 0. Something similar is wanted in epistemology.

(11) See note 8.

own experiences.[12]

Suppose that

4.7 *A believes (it appears to A that P).*

Suppose also that 4.1 is not compelled:

4.8 $\sim$(*P* $\supset$ *it appears to A that P*).

It follows that

4.9 *P* & $\sim$(*it appears to A that P*);

and hence that

4.10 $\sim$(*it appears to A that P*).

Contraposition on 4.6 and modus ponens via 4.10 yield

4.11 $\sim$*A believes (it appears to A that P)*,

contradicting 4.7. This is a simple (minded) doxastic paradox.[13]

An epistemic paradox now follows. Assume again that

4.12 A believes (*it appears to A that P*),

(12) 4.6 is relatively a weak epistemic principle; attempts to provide stronger versions are vexing in their unclarity. Thus consider the principle of infailibility that Schlesinger attributes to William Alston: *A believes that P* $\supset$ *P*, where *P* 'ranges over propositions describing A's current mental state.' Schlesinger, *Op. Cit.*, p. 10. But whatever the proposition *P*, if A *does* believe *P*, *P* describes at least one of A's mental states, namely the state of believing *P*. No other proposition describes the state nearly so well; indeed, no other proposition describes the state at all. The principle of infailibility thus comes perilously close to the affirmation that whatever is believed is true. No better is Alston's correlative principle of omniscience. See William Alston, 'Varieties of Priviledged Access,' *American Philosophical Quarterly*, Volume 8, 1971.

(13) An obvious inconsistency may be recovered directly from 4.6, 4.7 and 4.10. 4.11 is interesting in that it indicates an unsuspected conflict in belief.

where P is a simple sentence of the form Fa.

No scheme for the justification of simple sentences, it must be admitted, commands wide and general assent. Still, it seems plausible to suggest that

4.13 $J_A[\Xi, \{P\}]$

must be a consequence of a clause in any reasonable definition of justification if

4.14 *it appears to A that P*

and

4.15 *it appears to A that P* $\supset$ *P*.

Few other strategies for the justification of simple sentences suggest themselves.

Given 4.12,

4.16 *it appears to A that P*

follows via 4.6; and

4.17 *A believes P*

follows from 2.7 via 4.16.

Assume now that 4.1 is miscompelled:

4.18 $\exists Q(Q \mathbin{\&} Q \supset$ *it appears to A that P*)

but indicative:

4.19 *it appears to A that P* $\supset$ *P*.

It follows by modus ponens from 4.16 and 4.19 that

4.20 *P*;

hence that

4.21 *A knows that P*,

which follows from 4.20(truth), 4.17 (belief), and 4.16 and 4.19 (justification). To the extent that intuition scruples at 4.21 in virtue of 4.18, this is a paradox of appearances.([14])

An appearance can be neither true nor false, but it may be mistaken. 4.1 is inadherent (to the truth) if it is neither indicative nor compelled:

4.22 $\sim$(*it appears to A that P* $\supset$ *P*) & $\sim$(*P* $\supset$ *it appears to A that P*).

But $P \supset Q \vee Q \supset P$ is a law of logic; 4.22, an instance of its negation, and so a contradiction. Evidently 4.1 is inadherent (to the truth) only if

4.23 *it appears to A that* [$\sim$(*it appears to A that P* $\supset$ *P*) & $\sim$(*P* $\supset$ *it appears to A that P*)].

Assume now that

4.24 *A believes* (*it appears to A that P*).

Given 4.23, the appearance corresponding to 4.24 is apt to seem incoherent, with 4.1 itself *miscued* (by the truth). This impression is expressed by a principle of coherence:

4.25 *it appears to A that* [$\sim$(*it appears to A that P* $\supset$ *P*) & $\sim$(*P* $\supset$ *it appears to A that P*)] $\supset$ $\sim$(*it appears to A that P*).

The operator *it appears to A that P* admits of iteration.([15]) Iteration on 4.1, for example, yields

4.26 *it appears to A that Q*;

where Q is itself 4.1. (In fact, 2.5 and 2.6 have iterated analogues.) It is tempting to draw a connection between inadherence and iteration, as in a

([14]) Readers familiar with the literature will no doubt be able to invent a scenario to exemplify this schemata.

([15]) This might suggest that from a technical point of view, epistemic operators are best studied as an aspect of the theory of algebraic operators.

principle of second order appearances:

4.27 *A believes (it appears to A that P) & ~(it appears to A that P) ⊃ it appears to A that [it appears to A that P].*

An inadherent appearance thus receives an interpretation in terms of *its* appearance. This is an attractive conclusion inasmuch as it is compatible with 4.23 and explains the fact that (at 4.24) A believes that it is *P* that expresses the content of an appearance.

4.24 and modus ponens on 4.25 via 4.23 imply that

4.28 *it appears to A that [it appears to A that P]*

via modus ponens on 4.27 itself.

It follows from 2.7 that

4.29 *A believes (it appears to A that P).*

More generally,

4.30 $(I_n \supset x \text{ believes } I_{n-1})$,

where I_n is the nth iterate of *it appears to A that P.*

This suggests that inadherence has a simple interpretation as false belief. But, of course, if this is so then 4.6 must be rejected, inasmuch as 4.6 and 4.29 imply that *it appears to A that P*, while 4.23 and 4.25 that ~*(it appears to A that P).*([16])

5. *Appearance and Possibility*

The following is a principle often said to be useful in epistemic contexts:([17])

([16]) This is a conclusion I find congenial.

([17]) By Robert Nozick. among others. *Philosophical Explanations* (Harvard University Press: Cambridge, Massachusetts: 1981), chapter 3. See also Alvin Goldman, *Epistemology and Cognition* (Harvard University Press: Cambridge, Massachusetts: 1986), pp.45-46.

5.1 *If P were false, A would not believe P.*

Let $\{W_i\}$ be a set of possible worlds indexed by $i \in I$, and let

5.2 $D(W_i, W_j)$

be a metric on $\{W_i\}$. Let W_A be the actual world and let $\delta > 0$ be a real number. The set of worlds $\{W_i\}$ satisfying

5.3 $D(W_A, W_i) < \delta$

constitutes a δ-band B around W_A.

5.2 indicates that a metric exists; given 5.2, 5.3 serves to specify worlds that satisfy an inequality. Without a definition of distance,([18]) 5.2 and 5.3 remain entirely abstract. *Whatever* the definition of distance, the definition of the δ-band B at 5.3 is intended, on most explanations of counterfactuals, to specify possible worlds *similar* to a given world. Assume that such a δ-band B has been specified.

Let P_W mean that P is true at W. 5.1 admits of the following paraphrase:([19])

5.4 $(\forall W \in B)\{\sim P_W \supset \sim(A \textit{ believes } P)_W\}$.

The negation of 5.4, it is useful to recall, is

5.5 $(\exists W \in B)\sim\{\sim P_W \supset \sim(A \textit{ believes } P)_W\}$,

or, what comes to the same thing, as

5.6 $(\exists W \in B)\{\sim P_W \,\&\, (A \textit{ believes } P)_W\}$.

Suppose now that

5.7 *A believes Fa*

([18]) As in elementary analysis, when $\delta(x,y) = | x - y |$.

([19]) This is essentially the analysis proposed by David Lewis in *Counterfactuals* (Harvard University Press: Cambridge, Massachusetts: 1973).

5.8 *A believes* $\exists xFx$;

5.9 $\exists xFx$

and

5.10 $\sim Fa$.

From these assumptions, given 2.1, a paradox may easily be generated.

5.11 and 5.12 are weak enough to be compatible with virtually any interpretation of appearance or belief:

5.11 $(\forall W \in B)[(\textit{it appears to A that} \sim P)_W \supset \sim\{(\textit{it appears to A that } P)_W\}]$;

and

5.12 $(\forall W \in B)\{(\textit{A believes} \sim P)_W \supset \sim(\textit{A believes } P)_W\}$.[20]

Now let P be the sentence $\exists xFx$. And let

5.13 $(\forall W \in B)\{(\textit{it appears to A that} \sim P)_W \supset (\textit{A believes that} \sim P)_W\}$

be a modal version of 2.7.

Assume now 5.1 added to 2.1 as a condition of knowledge. Assume as well, given 5.7, 5.8, 5.9 and 5.10, that 5.1 is *false*. It follows from 5.6 that there exists a world W such that at W[21]

5.14 (*A believes that P*).

5.14 and contraposition on 5.12 imply again that at W

5.15 $\sim(\textit{A believes} \sim P)$;

and 5.15 and contraposition on 5.13 imply that at W

[20] Only 5.12 plays an inferential role in what follows.

[21] Subscripts have been dropped when the context makes clear their content; ditto quantifiers when they are specified in the sentence introducing a formula.

5.16 $\sim$(*it appears to A that* $\sim P$).

5.16 is a *diagnostic* sentence: It indicates *why* 5.14 is true, and correspondingly why 5.1 is false.

Suppose, on the other hand, that 5.10 is *true*, and not false. The epistemic context is *uninfected*. It follows that

5.18 *A knows P*,

is true. So too, then, is 5.1. Thus

5.19 $(\forall W \in B)\{\sim P_W \supset \sim(A \textit{ believes } P)_W\}$.

From contraposition on a variant of 5.13 (with P for $\sim$P), it follows that at *every* W at which P is false

5.20 $\sim$(*it appears to A that P*).

and thus that

5.21 $\sim P \supset \sim$(*it appears to A that P*).

Recall that the negation of 5.21 is

5.22 $(\exists W \in B)\{\sim P_W \,\&\, (\textit{it appears to A that } P)_W\}$.
From a slightly stronger version of 2.6, it follows that:([22])

5.23 $\Diamond\{\sim P \,\&\, (\textit{it appears to A that } P)\}$.

The modal paraphrase of 5.23 is just that

5.24 $(\exists W)\{\sim P_W \,\&\, (\textit{it appears to A that } P)_W\}$.

If
5.25 $W \in B$

([22]) 2.6 is expressed as an implication; 5.23 is a conjunction, but the choice is not essential to the argument.

it follows that 5.19 and 5.24 are inconsistent. It is somewhat surprising that an epistemic principle should have such strong metaphysical consequences.

Let B* be the set of possible worlds in which 5.22 is satisfied. For all worlds within B, a change in the truth of a proposition necessarily involves a change in its appearance. Not so for worlds within B*. It is possible to ask whether

5.26 *B is closer to* W_A *than B**

at all.

The obvious implications I leave to others.

APPENDIX

Epistemic or Doxastic Principles

The following principles have been investigated in this paper.

1.1 *A knows P ⊃ P is known.*
1.2. *P is known ⊃ ∃x(x knows P).*
1.8 *P is known & Q is known ⊃ {P & Q} is known.*
1.9 *P is known & Q is known ⊃ {P & Q} is known ≡ ∃x∃y[x knows P & y knows Q & x knows (y knows Q) & y knows (x knows P)].*
2.1 *A knows that P ≡ P & A believes P & A believes P justifiably.*
2.2 *A believes P & (P ⊃ Q) ⊃ A believes Q.*
2.3 *A believes P & A knows (P ⊃ Q) ⊃ A believes Q.*
2.5 ◇[(∀x) ~ {(*it appears to x that P*) ⊃ *P*}].
2.6 ◇[~ {*P* ⊃ (∃x)(*it appears to x that P*)}].
2.7 *it appears to A that P ⊃ A believes P.*
4.2 *it appears to A that P ⊃ P.*
4.3 *P ⊃ it appears to A that P.*
4.4 ∃*Q*(*Q & Q ⊃ it appears to A that P*).
4.5. *A believes P ⊃ P.*
4.6 *A believes (it appears to A that P) ⊃ it appears to A that P.*
4.22 ~ (*it appears to A that P ⊃ P*) & ~ (*P ⊃ it appears to A that P*).
4.23 *it appears to A that* [~ (*it appears to A that P ⊃ P*) & ~ (*P ⊃ it appears to A that P*)].

4.25 *it appears to A that* [$\sim$(*it appears to A that P* $\supset$ *P*) & $\sim$(*P* $\supset$ *it appears to A that P*)] $\supset$ $\sim$(*it appears to A that P*).

4.27 *A believes* (*it appears to A that P*) & $\sim$(*it appears to A that P*) $\supset$ *it appears to A that* [*it appears to A that P*].

4.30 ($I_n \supset$ *x believes* I_{n-1}).

5.1 *If P were false, A would not believe P.*

5.11 ($\forall W \in B$)[(*it appears to A that* $\sim P)_W \supset \sim$\{(*it appears to A that* $P)_W$\}].

5.12 ($\forall W \in B$)\{(*A believes* $\sim P)_W \supset \sim$(*A believes* $P)_W$\}.

5.13 ($\forall W \in B$)\{(*it appears to A that* $\sim P)_W \supset$ (*A believes that* $\sim P)_W$\}.

5.23 $\Diamond$\{$\sim P$ & (*it appears to A that P*)\}.