Synthese 191 (2014)
pp. 1627–49

Erratum corrected
in this version.

KNOWABILITY AS POTENTIAL KNOWLEDGE

ANDRÉ FUHRMANN

Abstract. The thesis that every truth is knowable is usually glossed by decomposing knowability into possibility and knowledge. Under elementary assumptions about possibility and knowledge, considered as modal operators, the thesis collapses the distinction between truth and knowledge (as shown by the so-called Fitch-argument). We show that there is a more plausible interpretation of knowability—one that does not decompose the notion in the usual way—to which the Fitch-argument does not apply. We call this the potential knowledge-interpretation of knowability. We compare our interpretation with the rephrasal of knowability proposed by Edgington and Rabinowicz and Segerberg, inserting an actuality-operator. This proposal shares some key features with ours but suffers from requiring specific transworld-knowledge. We observe that potential knowledge involves no transworld-knowledge. We describe the logic of potential knowledge by providing models for interpreting the new operator. Finally we show that the knowability thesis can be added to elementary conditions on potential knowledge without collapsing modal distinctions.

Keywords. Belief revision, Fitch, knowability, knowledge, modality, possibility.

> This seem to me all that you can clearly mean when you say that truth pre-exists to knowledge. It is knowledge anticipated, knowledge in the form of possibility merely.
>
> William James, *The Meaning of Truth*[1]

1. Introduction

It is a truism—so it seems (but see below)—that nothing can be known unless it is true. It is decidedly not a truism that everything true must be knowable. In fact that latter thesis, placing truth under a particularly strong epistemic condition, has never been able to enlist many followers. We may put the thesis very shortly thus:

Ver. $$A \to \text{Knowable}\, A,$$

where the label Ver may stand for (weak) verificationism or verifiability. This way of succinctly putting the thesis has at least the advantage of showing clearly that much depends on how we understand "knowability". (Much, not everything: The fate of Ver may also depend on the range of the implicit quantification over sentences or propositions; see e.g. [18].) As a first attempt we may try to decompose knowability into possibility and knowledge:

PossK. $$A \to \Diamond \mathsf{K} A$$

As is well known, in this form the thesis falls prey to the Fitch-argument.[2] It is a natural response, therefore, to find a formulation of the knowability thesis other than PossK. The proposal presented here is of this type. We propose to read Ver not in the sense of PossK. We propose not to decompose—not, at any rate, in the way it is usually done.[3]

The importance of the Fitch-argument does not so much rest on its claim to refuting in a few lines a thesis that has arguably and very seriously been endorsed by a number of important philosophers (from Kant over James to Dummett). Rather, the Fitch-argument constitutes a serious paradox. It leads from a plausible premiss to an implausible conclusion by apparently coercive reasoning. The plausible premise is that there are unknown truths. The implausible conclusion is that the premise *logically* collides with the knowability thesis. Some sort of collision one would expect. Given that unknown truths are abundant and that for many of them we have not even

[1] From the chapter "A Dialogue", [10, p. 295].

[2] The argument appeared first in print in Frederic Fitch's [6] from 1963 after having been suggested to Fitch by a then anonymous referee in 1945. Only recently was it discovered that the referee was Alonzo Church. See the recent collection [16] with material on the history of the argument.

[3] In a technical sense, our approach may also be seen as decomposing. But we do not decompose into a quantification over worlds and the assessment of a knowledge state therein.

the slightest idea of how we may investigate the matter, the extreme optimism expressed by the verifiability thesis about the possible extent of our knowledge must seem ill-founded. Foes of Ver argue that the thesis has indeed little to recommend it. They typically do not expect, however, to win as easily as the Fitch-argument would have it. This all too effortless victory is the scandal of reason encapsulated in the argument.

The plan of the paper is as follows. In the next section we collect some requisites on any plausible notion of knowability and show that *possible knowledge* cannot be a candidate analysis of knowability. In section 3 we investigate the prospects of analysing knowability in terms of *possible knowledge of actual facts*, a proposal due to Dorothy Edgington [3][4] and spelt out by Wlodek Rabinowicz and Krister Segerberg [14] (see also [12]). Although the analysis satisfies the requisites established in section 2, its decompositional character still generates triviality arguments, essentially of the kind pointed out by Timothy Williamson [19][20]. In section 4 we outline a theory of potential knowledge as knowledge in a range of potential states of knowledge and argue that potential knowledge satisfies the conditions on knowability. In section 5 we use some central observations from the theory of belief change to explain how the range of potential knowledge states can be fixed such that the analysis of knowability in terms of potential knowledge is safe from the kind of objection brought forward against the rival theory of section 3. In section 6 models of potential knowledge are described and in section 7 we show that the thesis that every truth is potentially known is immune to the Fitch-argument.

2. Knowability and possible knowledge

How many guitar strings did Keith Richards replace during his career? We do not know and will never know. But although he almost certainly did not, Richards *may* have kept a record of his string replacements. There is nothing in the notion of a string-replacement that could keep Richards from counting instances of them. So if he did, he would know. Suppose that n is in fact the total number of replacements up to a certain point in time t. It is knowable for Richards that strings have been replaced n times up to t. And if he has replaced strings n times, then it cannot be knowable that he replaced them some other number of times. In other words, knowability entails truth:

$$\text{Knowable}\, A \to A. \tag{1}$$

Now, for any number m (up to a natural threshold), it is *possible to know* that Richards replaced strings m times (up to t). For each m there is a possible world in which there have been m replacements and in which Richards has reliably counted them. Thus, for example, there is a possible world in which Richards has been so gentle on his strings that he never had to replace any ($m = 0$) and in which Richards knows this. So it is possible to know that Richards never replaced a string although, as a matter of fact, he replaced many. In other words, it is not in general the case that possible knowledge does entail truth:

$$\Diamond \mathsf{K}A \to A, \tag{2}$$

It follows from the truth of (1) and the falsity of (2) that knowability and possible knowledge are distinct notions.

Does the one notion imply the other? If $\Diamond\mathsf{K}$ implied knowability, i.e.

$$\Diamond \mathsf{K}A \to \text{Knowable } A, \tag{3}$$

then by (1) we would generally have (2)—which, as we have just observed, we do not. As to the other direction, i.e.

$$\text{Knowable } A \to \Diamond \mathsf{K}A, \tag{4}$$

suppose that we endorsed Ver. Then PossK would follow and thus the Fitch-argument would apply to refute Ver. Any account of knowability that escapes fitching must therefore be such as to render also (4) invalid. In particular, the required notion of knowability ought to render it knowable that true P is unknown (i.e. that Fitch-sentences $P \wedge \neg\mathsf{K}P$ are knowable) while—for familiar reasons—it remains impossible to know that P is an unknown truth (i.e. $\neg\Diamond\mathsf{K}(P \wedge \neg\mathsf{K}P)$).

A sound account of knowability compatible with the knowability thesis Ver must thus break the connection between knowability and possible knowledge in both directions. It must be such that

K1. knowability entails truth,
K2. knowability does not follow from possible knowledge, and
K3. knowability does not entail possible knowledge.

Below we shall offer an account of knowability as *potential knowledge* (as opposed to possible knowledge) that meets these conditions. But before we turn to potential knowledge, we shall first consider an already known candidate explication of knowability that satisfies the conditions above. Unfortunately this explication fails to distance itself from possible knowledge in one important respect. As a consequence it eventually confronts the same sort of problem as PossK does. We shall see that apart from K1-3 another adequacy condition for knowability is needed.

3. Knowability as possible knowledge of actual facts

The reason why possible knowledge cannot generally satisfy (1), i.e. imply truth, is simple. A proposition $\Diamond\mathsf{K}A$ can be true in a world a ("here") in virtue of A being true and known in a different world b ("there"). But if A is true and known there, then it is true here that A is possibly known even though A may be false here:

$$\begin{array}{ccc} a \bullet & \longrightarrow & \bullet\, b \\ \neg A & & A \\ \Diamond\mathsf{K}A & & \mathsf{K}A \end{array}$$

The diamond drives us away from a and there is nothing following the diamond that makes us return. We can force such a return by inserting an actuality operator: It is possible to know that A is true here, where "here" refers to some distinguished world—the actual world—at which truth simpliciter is evaluated. Thus, introducing an operator [0] for actuality, we can interpret "A is knowable" as "it is possible to know that actually A" ($\Diamond\mathsf{K}[0]A$). A full semantic analysis of the actuality-operator requires double-indexing (just as the completely analogous now-operator in tense logic). Suppressing one of the required indices and simplifying, the analysis will have this feature:

(5) $$a \models [0]A \;\Rightarrow\; 0 \models A$$

(assuming that 0 is designated as the actual world). This rules out counterexamples to (1) as just described. Given the standard analysis of actuality it follows indeed immediately that every world will satisfy the schema

(6) $$\Diamond\mathsf{K}[0]A \to [0]A.$$

If we define validity as truth in all models and truth in a model as truth in a distinguished ("actual") world 0, then

$$\Diamond\mathsf{K}[0]A \to A$$

is valid in this sense and so condition K1 on any adequate account of knowability is satisfied.[4] Condition K2 is also satisfied. For, if A is a proposition false in 0 but true in some other possible world, then it is possible to know that A without it being possible to know that actually A, i.e. that A is true in 0.

The converse of schema (6), i.e.

ActK. $$[0]A \to \Diamond\mathsf{K}[0]A.$$

is now a salient candidate for formulating the knowability thesis Ver: Of every actual truth it is possible to know that it is actually true.

[4] Truth in all models at the 0-point is what is called *weak* validity in [14]. Strong validity would be truth in all models at all points. Both notions of validity have their use and there is little point in discussing which is the right one in the present context.

This way of formulating Ver is recommended by the further observation that the Fitch-argument is now blocked in a natural way. Here is the Fitch-argument. It starts, on the left branch, with the assumption of an unknown truth and confronts it with an observation derived on the right branch from elementary modal (RN) and epistemic (KT) principles.

$$\dfrac{\dfrac{\dfrac{P \wedge \neg\mathsf{K}P}{\Diamond\mathsf{K}(P \wedge \neg\mathsf{K}P)}\,\text{PossK}}{\Diamond(\mathsf{K}P \wedge \mathsf{K}\neg\mathsf{K}P)}\,\text{K} \qquad \dfrac{\dfrac{\dfrac{}{\mathsf{K}\neg\mathsf{K}P \to \neg\mathsf{K}P}\,\text{KT}}{\Box(\mathsf{K}\neg\mathsf{K}P \to \neg\mathsf{K}P)}\,\text{RN}}{\neg\Diamond(\mathsf{K}P \wedge \mathsf{K}\neg\mathsf{K}P)}\,\equiv}{\dfrac{\bot}{P \to \mathsf{K}P}}$$

Now, from the hypothesis that there is an *actually* unknown truth P, i.e.

$$[0](P \wedge \neg\mathsf{K}P),$$

the Fitch-argument gets *via* $\Diamond\mathsf{K}[0](P \wedge \neg\mathsf{K}P)$ (by ActK) no further than this (by K and the schema $\mathsf{K}[0]\mathsf{K}A \to [0]\mathsf{K}A$):

$$\Diamond(\mathsf{K}[0]P \wedge [0]\neg\mathsf{K}P) :$$

It is possible to know an actually unknown truth. (Take a "bird's eye view" of the actual world: You "see" both that P is true there and that this goes unnoticed.) This seems exactly what the proponent of Ver should want to say about actually unknown truths. There is not the slightest hint of paradox.

Of course, the argument is effectively blocked only under the condition that K3 is satisfied, i.e. that $\Diamond\mathsf{K}[0](P\wedge\neg\mathsf{K}P)$ does not imply $\Diamond\mathsf{K}(P\wedge\neg\mathsf{K}P)$. Rabinowicz and Segerberg [14] show that given enough care in the semantic analysis of the actuality-operator (here double-indexing plays an essential rôle), the condition can indeed be made to hold.

Discomfort sets in when noticing that the consequent of ActK requires that we have knowledge in some possible world b about facts in some other possible world a (not necessarily the actual world):

$$(\dagger) \qquad b \models \mathsf{K}[a]B.$$

Of course, in certain cases we do have such knowledge. For example, we know in the actual world that in all possible worlds 2+2=4. We also know that in some possible world other than ours, Gödel never proved anything and that there are no possible worlds in which water is not H_2O. So transworld-knowledge is *per se* not problematic—or so we may assume for the present discussion. What may be problematic is *specific* (rather than general, i.e. logical or semantical) transworld-knowledge: knowing in one world that in a *specific* different world some contingent fact obtains. For, the specific world needs to be specified and this proves to be difficult, as Williamson [20] has shown.

The world a in (†) cannot be specified indexically. For, we need to specify a in b, and saying "here" (or "actually") in b refers to b, not to a. Thus a must be picked out descriptively.

Let A be a descriptive specification of a. We look now at the proposal that (†) is true in virtue of the existence of a specification A of a, a claim which we shall write down thus:

$$b \models \mathsf{K}[A]B. \tag{‡}$$

For present purposes we need not know much about specifications. Specifications are descriptions. The interpretation $i([A])$ of $[A]$ will thus be a set of worlds. Not much will depend on whether specifications can be partial, thus picking out more than one world, or whether they must be maximal, i.e. singling out a unique world. Let us assume first that specifications are allowed to be partial.

A description A specifies a world a just in case a is in $i([A])$. So the proposal under consideration is this:

(∗) (†) is true iff $\exists A$ such that (‡) and $a \in i([A])$.

For the arguments below, it suffices to assume that the right-hand-side of (∗) is sufficient for (†).

All worlds specified by a given description satisfy that description. Thus $[A]A$ should be universally true. Moreover, we have the following closure principle: If all worlds described by A satisfy B and B cannot be satisfied without satisfying the condition C, then all worlds satisfying the description A must satisfy the condition C.

There may be distinct but equivalent specifications: propositions A and A' that apply to precisely the same worlds. In that case it is true (for all C) that in all worlds $[A]C \leftrightarrow [A']C$, a fact which we abbreviate thus: $[A] \equiv [A']$.

If B is true under the specification $[A]$, then whatever is true under the conjunctive specification $[A \wedge B]$ must already be true under $[A]$ alone:

$$[A]B \rightarrow ([A \wedge B]C \rightarrow [A]C). \tag{7}$$

The converse of the consequent holds under the same antecedent-condition: Suppose that all worlds specified in a given world by the description A satisfy the condition B. Then the conjunctive specification $[A \wedge B]$ cannot be more restrictive than $[A]$, for $[A]$ alone already restricts the selected worlds to B-worlds. Thus we have

$$[A]B \rightarrow ([A]C \rightarrow [A \wedge B]C) \tag{8}$$

and further, by joining (7) and (8),

I. $$[A]B \rightarrow [A] \equiv [A \wedge B].$$

In other words, under the supposition that $[A]B$ holds in a given world, A and $A \wedge B$ are equivalent specifications in that world.

If the knowledge that B in a in (†) is knowledge *of* the world a that B obtains, i.e. knowledge *de re*, then the truth of the knowledge claim must not depend on how a is specified. A necessary condition for knowledge in (‡) being *de re* is that the schema

II. $$[A] \equiv [A'] \rightarrow (\mathsf{K}[A]B \rightarrow \mathsf{K}[A']B)$$

be valid.

Given I and II, we can reinstate a version of the Fitch-result by showing that $[A]B$ entails $\mathsf{K}[A]B$:

(A) $$\frac{\dfrac{\dfrac{\overline{[A \wedge B]A \wedge B} \qquad \overline{A \wedge B \rightarrow B}}{[A \wedge B]B}\,\text{Closure}}{\mathsf{K}[A \wedge B]B}\,\text{RN} \qquad\qquad \dfrac{[A]B}{[A] \equiv [A \wedge B]}\,\text{I}}{\mathsf{K}[A]B}\,\text{II}$$

(RN is the necessitation rule for the knowledge operator.) Thus (‡) cannot be interpreted *de re*. Can the proposal escape fitching by waiving the *de re*-condition II?

To refute the proposal ($*$) we do in fact not need condition II, effectively restricting the interpretation of (†) and (‡) to *de re* readings. There is a very plausible third condition on specifications that can replace II to obtain an equally unpleasant result. The new condition is this: If a world a is not among the worlds picked out by a specification, then a does not satisfy the specification. Put contrapositively and more succinctly:

III. $$a \models A \;\Rightarrow\; a \in i([A])$$

Given III, the proposal ($*$) again yields the Fitch-result in the following form: Every proposition true at a world is at that world known to be true.

(B) $$\frac{\dfrac{\overline{[A]A}}{a \models \mathsf{K}[A]A}\,\text{RN} \qquad \dfrac{a \models A}{a \in i([A])}\,\text{III}}{a \models \mathsf{K}[a]A}\,(*)$$

Let us finally consider, as promised, a variant of the above proposal according to which specifications must be maximal, i.e. single out a unique world. This condition on specifications strengthens the right-hand-side of the proposal (*) and thus weakens the relevant direction of (*) to

($**$) If $\exists A$ such that (‡) $b \models \mathsf{K}[A]B$ and $i([A]) = \{a\}$, then (†) $b \models \mathsf{K}[a]B$ is true.

Under this new understanding of what a specification is, conditions I and II remain in place and so does the argument (A). In argument (B) we can use (∗∗) instead of (∗), if III can be strengthened to

IV. $$a \models A \;\Rightarrow\; i([A]) = \{a\}$$

This new condition is indeed the rightful heir of III under the changed conception of specifications. If a satisfies the maximal specification $[A]$, then, by uniqueness, $\forall b$ such that b satisfies the specification $[A]$, $a = b$. So, given that we accepted III under the more liberal conception of specification, we must accept IV under the stricter conception. The argument (B) is thus easily converted into a variant to the same effect.

Remark. Knowledge of counterfactuals seems to furnish straightforwardly knowledge across possible worlds. Peter knows that if he had not looked at the window, he would not have seen that bird. So Peter seems to actually know something about a non-actual world, namely one in which he did not look at the window. That non-actual world is specified by the antecedent of the counterfactual. So actual knowledge of a counterfactual is knowledge of the typically non-actual antecedent-worlds.

Based on this observation Dorothy Edgington [3] has been taken to make a proposal which fits under the schema above as follows: Read $[A]B$ in (‡) as a counterfactual with A as its antecedent and B as its consequent and let $i([A])$ in (∗) be $c(a, |A|)$, the set of closest-to-a antecedent-worlds, as in the standard, Lewisian semantics for counterfactual conditionals. (In retrospect of Edgington [4] it is not altogether clear whether the proposal should be ascribed to Edgington, as both [14] and [20] do.) Under this more specific interpretation of what it means to specify a range of worlds, condition I expresses a principle of cautious monotony (in one direction; the other direction expresses cautious cut), condition II renders the antecedent position of a conditional inside the scope of knowledge extensional (thus giving the required sense of a *de re* specification of worlds), and condition III corresponds to the weak centering condition on choice functions, requiring that if a satisfies the condition A, then a is to be included in $c(a, |A|)$. Under this reading I and III are standard features of the logic of counterfactual conditionals and II expresses a necessary condition on a *de re*-circumscription of antecedent worlds. Thus the two arguments above apply. I take Williamson [20] and Rabinowicz and Segerberg [14] as having made essentially these observations.

In [20] (but not in[14]) Edgington's proposal is reconstructed under the assumption that specifications be maximal. This has the effect that the relevant counterfactuals are of the kind modelled by the Stalnaker-semantics, i.e. making the selection function always singleton-valued. So strong centering, $c(a, |A|) = \{a\}$ ($\forall A$ true in a), is required, which corresponds to our condition IV. (*End of remark.*)

The problem of specific transworld-knowledge afflicting ActK arises from the fact that ActK tries to force a return to the reference-world (here: the actual world) inside the scope of the possibility-operator. ActK shares with

PossK the decomposition of knowability in possibility and knowledge and thereby falls again victim to fitching. While PossK glossed knowability as knowledge in some possible world, ActK opted for knowledge in some possible world of actual facts. The reference to what is actually the case is well-taken. But the reference is ill-placed within the scope of knowledge in non-actual worlds. Knowability should better not require epistemic leaps across worlds—it should be a purely this-worldly affair. Thus we have the following fourth condition on any sound account of knowability:

K4. What is knowable about a world a must be determined in a itself.

In an attempt to satisfy this condition we shall investigate in what follows the prospects of a notion of knowability as a kind of actual knowledge: *actual potential knowledge*.

4. Potential knowledge and possible knowledge

It is tempting to analyse ...bility-concepts as *composita* involving the concept of possibility. That such a procedure always succeeds is a risky and therefore substantial thesis. There are many examples where the apparently natural analysis looses its temptation on closer examination.

It is certainly possible that I finish a triathlon in less than 8 hours tomorrow. It suffices to focus on a possible world where I am in very much better shape than I actually am and where I do finish in less than 8 hours. But equally certainly I am quite incapable of even finishing a triathlon at all. For me a triathlon is not "finishable". Likewise, Thomas Mann was unable to converse in Chinese. But it is of course possible that he did—he even may have written the *Zauberberg* originally in Chinese. Sometimes we describe such situations by saying that a certain option is, though possible, not feasible. Being capable of ϕing or able to ϕ is not the same as it being possible that one ϕs. The latter does not entail the former.[5]

[5] In these examples we understand possibility in the widest, i.e. logically weakest sense, sometimes called metaphysical possibility. This is the sense that is at issue in discussing the merits of the PossK-version of the knowability principle *vis-à-vis* the Fitch-argument. Fara [5] argues to essentially the same conclusion as to the relation between knowability and the possibility that one knows. But he does so on the basis of a stronger intermediate step: "Sometimes one might have the capacity to do so-and-so even though it is not possible that one does so-and-so in *any* sense of possible obtained by restriction of metaphysical possibility" (p. 68). Although this thesis is certainly interesting, it goes well beyond what is required for decoupling the knowability thesis Ver from its interpretation PossK in terms of possible knowledge.

Based on the observation that Ver does not entail PossK, Fara points out in the final part of [5] that knowability may be interpreted as a capacity to know without inviting the Fitch-argument nor the transworld-knowledge challenge discussed in the last section. The general strategy is thus the same as in the present paper. Fara does not wish to commit himself on the basis and nature of that capacity. In particular, there is no attempt at a semantical analysis of knowability claims under the capacity-reading. Sometimes such neutrality can be prudent. But in this case it remains unclear in which sense capacity can be construed as a sentential operator as suggested in the final pages of [5]. Moreover, we do not know enough about the aimed at notion of a capacity-to-know to confidently judge that it is immune to fitching and to transworld-embarassments. What seems plausible enough though, is the thesis that some way of cashing out that notion is safe in this sense. One such way is explained here, though it is one that—contrary to Fara's suggestion—does not decompose knowability into a knowledge- and a capacity-operator.

More to the present point, it is possible to believe that I finish a triathlon in less than 8 hours tomorrow. Any possible world where I did,[6] serves as a truth-maker for the proposition. It is incredible, however, that I even participate in a triathlon tomorrow, not to mention that I finish with a record time. Credibility is not possible belief.

Someone, X, who finds the proposition that pigs fly incredible need not wish to rule out that there are possible circumstances in which he believes pigs to fly. In a possible world inhabited by winged pigs, X would typically believe that pigs fly. X can acknowledge such a possibility and will, in this sense, assent to the claim that it is possible for him to believe that pigs fly. It does not seem odd for X to maintain in the same context that he finds flying pigs incredible, thus asserting

(9) It is possible that I believe pigs to fly but I find it incredible that they do.

If possible belief implied believability, then (9) would express an outright contradiction. But it does not—or so we are plausibly supposing. Thus, there must be at least one natural interpretation of believability such that it is not entailed by possible belief.

What has been said so far does not rule out that assertions like (9) may be rendered consistent by identifying in the ability-expression a notion of possibility that differs from the one in which the possibility of belief is claimed. The appearance of contradiction may then be understood as resulting from an equivocation on 'possibility'. Thus one might gloss (9) by saying that although it is metaphysically possible that I believe that pigs fly it is in some other (here, presumably, epistemic) sense not possible that I do. But, first, this is not the only option available, as we shall see. Second, if [5] (see footnote 5) is right, then the option may indeed not be available in cases of interest here. Third, this way of treating -bilities is of no avail in resolving the Fitch-paradox. For of whatever kind the modality in PossK may be, nothing stops us from instantiating PossK with a Fitch-sentence in which possibility is of the same kind. So if we explore a different approach to knowability as it features in Fitch-sentences, then simplicity speaks in favour of exploring the benefits of that approach with respect to other -bilities.

How then may we understand the right conjunct of (9) under the assumption that (9) is consistent? When X claims it to be incredible that pigs fly, he asserts that, for all he believes, pigs do not fly and that he will continue to believe that pigs do not fly whatever evidence may reasonably be expected to turn up. X offers a bet, as it were, on a range of belief changes induced by evidence on the matter of flying pigs. This range is delimited by his expectations in and with respect to the actual world. Speakers sometimes emphasise the point by inserting an "actually", as in "I find it actually incredible that ...", "I am actually incapable of ...", or "it is actually not a feasible option that ...". We will shortly expand on this idea.

[6] I.e. where I believed that I finished the event in less than 8 hours because I shaped up to perfom the feat or because I was taken in by some illusion to the effect.

In whatever way we may precisify the gloss on believability just offered, the sketch suffices to answer the question as to whether believability (in some such sense) implies possible belief (where the relevant notion of possibility is, as above, logical or metaphysical). Is it true (for any proposition A) that

(10) if A is believable for the subject X at time t (X_t),
then it is possible that X_t believes that A ?

Consider a proposition of the form

(11) P and X_t does not believe that P.

For example, let P be the proposition that Peter is at the party. X_t cannot rule out that his actual belief state at t may plausibly be replaced by a belief state in which he believes that P and continues to be aware of the fact that he actually, at t, does not believe that P. X_t is aware (or should be aware) of an *actually potential belief state* in which (11) holds. According to our gloss on believability, X_t may properly assert "I find it believable that Peter is at the party while believing that he is not." Thus

(12) it is believable for X_t that (11).

But X_t cannot possibly believe that (11), as (10) would have it given (12). For, in that case (letting B stand here for "X_t believes that") we would have

$$\Diamond\mathsf{B}(P \wedge \neg\mathsf{B}P).$$

From this one obtains by minimal principles governing belief[7]

$$\Diamond(\mathsf{B}P \wedge \neg\mathsf{B}P)$$

—which declares a contradiction possible. When uttered by X_t, (11) is of course a Moore-sentence.[8] We have thus observed that while I cannot believe Moore-sentences I may give them credibility. It follows that believability cannot imply possible belief. There is no logical connection between the two notions either way.

[7] The minimal principles are those one finds in the logic **KD4**, a common—though not uncontroversial—starter for doxastic logics. It extends the set of tautologies by the 4-schema $\Box A \rightarrow \Box\Box A$, the D-schema $\Box A \rightarrow \Diamond A$, and closes under Modus Ponens and Necessitation, $A/\Box A$.

[8] Moore-sentences are of the form "P and I do not believe that P." They may well be true but cannot be properly asserted. A standard way of accounting for this, at first sight paradoxical characteristic is by observing that assertion must be taken to express belief, whence the assertion of a Moore-sentence expresses contradictory content. Note that we may substitute knowledge for belief in a Moore-sentence, thus obtaining a Fitch-sentence. Just like Moore-sentences, asserting a Fitch-sentence sounds odd, a fact that can be explained by the thesis that knowledge is the norm of assertion: someone who asserts that P signals that he knows that P. On this explanation, the assertion of a Fitch-sentences would likewise have contradictory content. See Moore [13], where Moore's paradox first appeared in print and Williamson [20, ch. 11] for a statement of the knowledge account of assertion.

Before turning to the question as to how the range of actually potential belief states is fixed, let us briefly transfer the results obtained to the target notion of knowledge. Analogously to our gloss on believability we now interpret knowability as acceptance in some actually potential state of knowledge.

First, possible knowledge does not imply knowability. Take any contingently false proposition P. Then there is a possible world in which P is known. But since P is actually false, there is no knowledge state that is *actually* within reach in which P is accepted. So P is not knowable in the actual world.

Second, knowability does not imply possible knowledge. Take a Fitch-sentence, F=(P and X does not know that P). It is not possible that X knows that F (by the Fitch-argument). But F is knowable to X: The course of evidence in the actual world may plausibly run such that under improved evidential conditions X knows that P while recognising, i.e. knowing, that under the less favourable conditions X cannot count as knowing that P. Under the potential, favourable conditions F would be known to X and thus under the given, less favourable conditions F is knowable for X. Analogously to Moore-sentences, Fitch-sentences cannot be known but may be knowable.

We have used knowability in a sense that derives from quantification over a range of potential states of knowledge. A proposition is knowable for an agent X_t situated in a certain world just in case the knowledge of X_t in that world is such that it makes a state of knowledge in that world potentially available in which that proposition is accepted. It is now time to ask what makes a state of knowledge potentially available to an agent in a given world. In particular we shall need to ascertain that the properties of states available in a given world are fixed exclusively in that world. If we succeed in this task, we shall have shown that knowability in the sense of potential knowledge is an intra-world affair and thus need not take on the challenge to explain how specific interworld-knowledge is possible.

5. A lesson from belief revision

A belief state is not a mere repository of items called ‘beliefs’. In addition to hosting beliefs, a belief state must be equipped with a mechanism for processing evidence. Otherwise belief states would remain stationary, thereby loosing their main point: to be a serviceable representation of a changing world on the basis of changing evidence. Belief revision theory (as explained e.g. in [1] or in [11]) provides us with an abstract description of the dynamics of belief states. A knowledge state is, as it were, an elitist portion of a belief state. Notwithstanding the many theories as to what it takes for a belief to join the elite, essentially the same abstract picture of belief revision theory applies to the dynamics of knowledge states.

The picture is this: Given a piece of evidence impacting on a belief state K, it is not *logically* determined what the successor of K reacting to the evidence should be like. Suppose that X_t accepts in belief state K that A and that $A \to \neg B$, and that X_t is faced with evidence to the effect that B.

If X_t can deliberate about the matter—i.e. such that he is not forced into an inconsistent belief state—, he has the following choices:[9]

(a) reject the evidence and keep both A and $A \to \neg B$.
(b) accept B and reject A
(c) accept B and reject $A \to \neg B$
(d) accept B and reject both A and $A \to \neg B$.

If the subject wants to "doxastically go on", he will need to choose from the menu. Unless the choice is aleatoric, the subject must be assumed to take recourse to some sort of preference ranking. For the purpose of a theory of belief change, we must think of belief states as sets of beliefs equipped with some device that can effectuate the necessary choices.

The preference that dictates the choice must be in place before the evidence comes in. Otherwise one would have to accept the evidence as a variable for computing the preference. But then the subject could never opt for (a), i.e. not accepting the evidence. The requisite relation of epistemic preference that is in vigour in a belief state K must determine for each piece of evidence possibly impinging on K, how K is to change in response to that evidence. A belief state is thus a set of beliefs "surrounded" by a family of further belief states, one for each piece of evidence the subject may encounter.[10] The belief states surrounding a given belief state in this sense are potential in that they will be actualised by a change operation triggered by a certain piece of evidence. So the range of potential belief states for a given belief state K is a function of the subject's epistemic preferences which in turn is partially determined by the range of evidential events the subject is prepared to hypothetically consider.

At first glance, the picture of a belief state as coming equipped with a choice device that determines for each piece of evidence possibly to be encountered a unique successor state seems daunting. Although we are talking here of requirements for *rational* belief formation, it may be asked how these requirements can possibly serve as a model for the performance of actual subjects, i.e. such that the theory can have normative and explanatory merits. We should keep in mind, however, that the encoding of a preference relation can be more or less verbose. We frequently employ general rules by which we quickly compute particular instances of a relation. We may for example place all options based on miraculous evidence to the bottom of the scale or rank the value of evidence according to sources. Moreover, the picture carries no commitment to determining the sense of "possibly to be encountered evidence" in a way that is independent of the subject's choice

---

[9] The *caveat* is necessary because the agent may come to have contradictory beliefs—perhaps only for a short moment—by a mechanism beyond doxastic control, such as perception.—The picture we draw here includes the option that the evidence may be rejected (choice (a) in the menu). This option is not considered in classical belief revision theory ("AGM"). But AGM may be extended to cover the case, as in the theory of non-priorized belief revision; see [9].

[10] Essentially the same idea gave rise to the notion of a hyperproposition, as introduced in [7]. Hyperpropositions represent belief states together with their fallback positions in response to recalcitrant evidence. The fallbacks are determined by the choice that needs to be exercised in retracting beliefs.

function; "possible" evidence may just be evidence the subject is prepared to hypothetically consider, be it in terms of concrete evidential events or types of evidence such as originating from a reliable source or conforming to the laws of nature or to deeply entrenched expectations, etc. Thus understood it is not ruled out that the subject may encounter evidence for which he has no decision rule. Unless the subject accepts being stumped, he will have to make up a new rule on the fly. In our terms, he would thereby change to a new belief state, not by a change of belief but by a change of epistemic preference.

Given a belief state K situated in a world a, the range of potential belief states with respect to K as well as their properties are completely determined by the facts in a. These are facts about the epistemic subject accepting in a the beliefs in K and considering and responding to possible evidence according to K's choice function. A belief state H belongs to the the range of potential belief states associated with K just in case H can be reached from K by a series of belief changes that are legitimate from the viewpoint of K. Changes are legitimate according to K, if they respond to evidence that the subject is prepared to hypothetically consider in a in accordance with the subjects choice function in a. Thus, unlike possible belief, the notion of a potential belief carries no reference to worlds distinct from the world where potential beliefs are being assessed.[11]

States of knowledge are sub-states of belief states. There is, unfortunately, no litmus test for the subject to delineate the knowledge-state inside his corpus of beliefs. The way knowledge states are inscribed into belief states is intransparent to the subject. For the subject all beliefs are accepted as true and acquired in legitimate ways. They may differ, however, in their vulnerability to contrary evidence. Consider again the situation in which a belief state K containing A and $A \to \neg B$ is faced with evidence to the effect that B and thus has to choose from the options (a-d). If A and $A \to \neg B$ are both part of his knowledge state $K^!$ inside K, then he would be best served by opting for the first alternative, thus rejecting B. Only in that case would he not loose knowledge. But since, from the perspective of the subject, $K^!$ is hidden inside K, the subject cannot be blamed for making

[11] The credibility of the picture drawn here depends on resolving in one way or another an important issue concerning belief revision. We said that a belief state is a potential state for the agent ("within reach"), if it can be reached from the agent's state by a series of legitimate belief changes. Thus rational change functions have not only to determine a successor set of of beliefs but also a successor choice function. Only thus shall we get a belief state that can again respond to evidence in a non-trivial way—Rott [15] calls this "the problem of categorical matching" between the input and the output of a belief change function. The problem of finding rationality constraints for generating choice function that rule changes after changes is the topic of theories of iterated belief change; see [2] and also [17].

Our use of the notion of a potential belief state owes much to the work of Isaac Levi; see in particular [11]. Levi, however, requires strong structural constraints on the family of potential belief states generated by a given state. (Such families should, according to Levi, form complete atomic Boolean algebras, so as to serve as adequate tools in inquiry.) Levi's constraints are well worth considering, given the ways in which he makes use of potential belief states. Our purposes here are more modest and thus we can afford to remain non-committal with respect to conditions that turn no wheel in the present context.

one of the other choices. To the contrary, he should be applauded for choosing in accordance with his epistemic preferences, even if that meant loosing knowledge. Only such choices can be considered legitimate that are consistent with the preference ranking currently in vigour. If under no course of hypothetical evidence the subject is prepared to believe that B, then there is no potential—in the sense of legitimate—belief state in which he accepts that B. *A fortiori*, there can be no potential knowledge state in which B is accepted; so B is unknowable. If B happens to be true, then B would, in this case, be an unknowable truth for the subject—but see the next paragraph. If, conversely, the subject can change his beliefs by legitimate moves in response to evidence such that he would eventually know that B, then B is potentially known (knowable) in the original state of knowledge. Thus a proposition is knowable (potentially known) for a subject just in case there is a potential knowledge state for that subject in which he accepts the proposition in question (i.e. a potential belief state in which the proposition is known). In the next section we shall sketch what potential knowledge essentially comes down to.

Before we do so, we shall have to address a question that may have occurred to the reader in the last paragraph. Have we not just argued that there are unknowable truths—perhaps even that *no* truth is knowable? Whatever truth A we take, can there not be epistemic preferences to the effect that any evidence favouring A is rejected, so that a subject holding such preferences can never legitimately adopt a belief state whose inscribed body of knowledge contains A? If so, then A is true while there is no knowledge state available that would endorse A. So if we gloss the knowability of A as the availability of a knowledge state affirming A and allow for the stubborn rejection of evidence, then Ver, under this interpretation, must fail.

Contrapositively: If Ver is to be true (under the availability of relevant states of knowledge interpretation), then subjects must not adopt epistemic preferences that block the proper appreciation of evidence. Thus the effect of Ver is to place a normative constraint on the range of available states of knowledge, or, more simply, on the notion of knowability. Conventional wisdom has it that ought implies can. So the question of interest here is whether Ver *consistently* conditions the notion of knowability in this way. This is all what the discussion of the Fitch-paradox is about: whether the knowability principle is a consistent add-on to principles governing possibility and knowledge.[12] The answer, as we shall see, is a clear Yes.

[12] Lest not forget that the conditions on knowledge, assumed in the argument, have a normative character too.

6. The logic of potential knowledge

Let us consider a language that includes apart from a standard knowledge operator K a further unary sentential operator $\langle\mathsf{K}\rangle$ representing potential knowledge. We may think of $\langle\mathsf{K}\rangle$ as a composite expression, consisting of a sentential operator K and a *modal modifier* $\langle - \rangle$ which maps K to another sentential operator. Thus knowability can be decomposed—but not in the usual way.[13]

Given a set ATM of atomic formulæ, FML is the set of formulæ generated from ATM by closing under the connectives. To provide a semantic analysis of potential knowledge, we look at structures of the type

$$(W, S, \leq)$$

(called *hyperrelational frames* in [8]), where

W is a nonempty set of worlds,
S is a binary relation in $W \times W$ (modelling knowledge), and
$\leq$ is a binary relation in $W^2 \times W^2$, i.e. a relation between relations.

We take $S_a := \{b \in W : Sab\}$ as representing the knowledge of a subject at a world a, i.e. a knowledge state—except that S_a alone does not include a component representing epistemic preference in response to evidence as explained above. This component is indirectly represented by the closure of S_a under $\leq$,

$$\leq (S_a) := \{S'_a : S \leq S'\}.$$

This is the family of potential knowledge states (in a) from the perspective of S_a. As suggested by the informal picture drawn in the previous section, we may think of a potential knowledge state as one that can be reached from S_a in response to hypothetical evidence by way of a change as determined by the subject's choice function.[14]

A *model* adds to a structure a valuation, mapping atomic formulæ at worlds into truth-values,

$$v : \text{ATM} \times W \to \{0, 1\}.$$

Valuations are extended to a satisfaction relation $\models \subseteq W \times \text{FML}$ in the familiar way for Boolean connectives, starting with $a \models P$ iff $v(P, a) = 1$. The interesting clauses are these:

K. $\quad a \models \mathsf{K}A$ iff $\forall b : Sab \Rightarrow b \models A$,

$\langle\mathsf{K}\rangle$. $\quad a \models \langle\mathsf{K}\rangle A$ iff $\exists S' : S \leq S'$ & $(\forall b : S'ab \Rightarrow b \models A)$.

We may gloss the clause for $\langle\mathsf{K}\rangle$ thus: A is potentially known (in a world a) just in case there exists a potential state (in a) in which A is known. In what follows, we shall sometimes abbreviate $\forall b : Sab \Rightarrow b \models A$ to $S_a \models A$.

[13] For more on modal modifiers see the companion paper [8].
[14] This way of generating legitimate transitions is not represented in our formal models, since the language to be modelled is not expressive enough to refer to such apparatus.

Truth in a model is truth at all points in the model. *Validity* is truth in all models. The *theory* of structures of the type just describes, $Th(\mathbb{S})$, is the class of all formulæ valid in arbitrary models on arbitrary frames of the type. $Th(\mathbb{S})$ is thus the basic *Logic of Potential Knowledge*.

The models described so far generate only few logical principles. The operator K behaves as in every normal modal logic, i.e. it satisfies a rule of necessitation and a distribution scheme:

RN $\qquad A \,/\, \mathsf{K}A,$

K. $\qquad \mathsf{K}(A \to B) \to (\mathsf{K}A \to \mathsf{K}B).$

But the $\langle\mathsf{K}\rangle$-clause by itself generates only a rule of monotonicity,

RMp. $\qquad A \to B \,/\, \langle\mathsf{K}\rangle A \to \langle\mathsf{K}\rangle B.$

(Suppose that (1) $\forall x,\ x \models A \to B$ and that (2) $a \models \langle\mathsf{K}\rangle A$. We need to show that $a \models \langle\mathsf{K}\rangle B$, i.e. that $\exists S'$ with $S \leq S'$ and $S'_a \models B$. From (2) we get that $\exists S'$ with $S \leq S'$ and $S'a \models A$. So, by (1), we also have $S'_a \models B$.)

For more logic, we need to consider richer structures. Our aim now is to build up a logical basis of plausible strength on which to test the effects of adding the knowability-thesis.

We start by strengthening the K-operator so as to furnish the resources tapped in the Fitch-argument.[15] This is done by requiring that the relation S be *reflexive* and *transitive* so as validate the principles

T $\qquad \mathsf{K}A \to A,$

4. $\qquad \mathsf{K}A \to \mathsf{K}\mathsf{K}A.$

Next we make $\langle\mathsf{K}\rangle$ a normal modal operator, as K already is. For this, we need two conditions:

Continuation. $\qquad \exists S' : S \leq S',$

Combination. $\qquad$ If $S \leq S'$ & $S'_a \subseteq X$ and $S \leq S''$ & $S''_a \subseteq Y$, then $\exists R : S \leq R$ & $R_a \subseteq X \cap Y.$

The two conditions correspond[16] respectively to the schemas

RNp $\qquad A \,/\, \langle\mathsf{K}\rangle A,$

Mp. $\qquad \langle\mathsf{K}\rangle A \wedge \langle\mathsf{K}\rangle B \to \langle\mathsf{K}\rangle(A \wedge B).$

[15] In fact, as pointed out above, a little less suffices for the Fitch-argument.

[16] Correspondence is here meant in the sense that a schema is valid in a class of structures just in case the structures satisfy the condition in question. Thus, as to RNp, clearly the condition validates the rule. For the converse (the necessity of the condition for the rule) assume that if $[\![A]\!] = W$, then $\exists S' : S \leq S'$ and $(\forall a)\ S'_a \models \langle\mathsf{K}\rangle A$. Let $A = \top$. Then $\exists S' : S \leq S'$; so the condition holds. — Likewise, the validity of Mp follows immediately from the condition. For the converse consider a model with relations S, S' and S'' such that $S \leq S'$ and $S \leq S''$ and choose atoms P and Q such that $S'_a \subseteq [\![P]\!]$ and $S''_a \subseteq [\![Q]\!]$. Then $a \models \langle\mathsf{K}\rangle P$ and $a \models \langle\mathsf{K}\rangle Q$. Now the condition fails if and only if there is no R with $S \leq R$ and $R_a \subseteq [\![P \wedge Q]\!]$. But this is just the condition under which $a \not\models \langle\mathsf{K}\rangle(P \wedge Q)$. So, if the condition fails, so does the schema.

The further correspondences claimed below follow similarly. Here is, to illustrate once more, the argument for the necessity of Knowability (see the next footnote) for PotK: Consider a model with (1) $a \models P$, for some P, and (2) for all S', if $S \leq S'$, then $S'_a \not\models P$. So the condition fails. But from (2) we have $a \not\models \langle\mathsf{K}\rangle P$, whence, given (1), $a \not\models P \to \langle\mathsf{K}\rangle P$.

Given RMp and Mp, we can quickly derive

Kp. $$\langle\mathsf{K}\rangle(A \to B) \to (\langle\mathsf{K}\rangle A \to \langle\mathsf{K}\rangle B).$$

(Combination looks complicated. What can we say in favour of it, apart from that it corresponds to Mp and thus delivers the goods? Suppose that both S' and S'' are available from S. If we follow these paths from a so as to develop our knowledge state S_a, we shall know on the one path that we inhabit the X-worlds and on the other path that we are somewhere in the Y-worlds. In such a situation there should be a way to combine the evidence so as to cut down on the number of open possibilities. Combination allows just this: There should be a way of coming to know that we are in the meet of X and Y.)

Further, let us minimally connect knowledge and potential knowledge. Knowledge had is a limiting case of knowledge potentially had, which we formulate thus:

C. $$\mathsf{K}A \to \langle\mathsf{K}\rangle A.$$

The schema is matched by the following condition:

Preservation. $$S_a \subseteq X \;\Rightarrow\; \exists S' : S \leq S' \;\&\; S' \subseteq X.$$

(Given C, RNp follows from RN. Likewise, Preservation implies Continuation.)

Finally, and in accordance with condition K1, we make $\langle\mathsf{K}\rangle$ a factive operator,

Tp. $$\langle\mathsf{K}\rangle A \to A,$$

by requiring that

p-Reflexivity. $$S \leq S' \;\Rightarrow\; S'aa.$$

Now K and $\langle\mathsf{K}\rangle$ look very much alike as to logical properties. Only the unidirectional C introduces an asymmetry, and the introspection schema

4p. $$\langle\mathsf{K}\rangle A \to \langle\mathsf{K}\rangle\langle\mathsf{K}\rangle A$$

is conspicuously absent. The schema is an instance of what will be our version of the knowability-thesis, namely

PotK. $$A \to \langle\mathsf{K}\rangle A.$$

It would therefore be imprudent at this point to include 4p in the process of working up to a logical basis on which to try out the effects of adding the knowability-thesis in the form of PotK. Indeed, if we weakened PossK to its instance

$$\Diamond\mathsf{K}A \to \Diamond\mathsf{K}\Diamond\mathsf{K}A$$

—which is the $\Diamond\mathsf{K}$-analogue to 4p—, then, on plugging in for A the Fitch-sentence $\Diamond\mathsf{K}P \wedge \neg\mathsf{K}\Diamond\mathsf{K}P$ (a knowable proposition whose knowability is unknown), we could still derive by the familiar Fitch-reasoning that

$$\Diamond\mathsf{K}P \to \mathsf{K}\Diamond\mathsf{K}P.$$

This may not be as obviously problematic as $P \to \mathsf{K}P$ but its passing from possibility to actuality (in the consequent) is at least dubious. So rather than adding 4p separately, we shall consider it implicitly when we add in a moment the full force of PotK to the conditions joined so far.

We call structures of the type $(W, S, \leq)$ that satisfy the conditions above (S-Reflexivity and -Transitivity, Combination, Preservation and p-Reflexivity) *potential knowledge structures* and let $\mathbb{P}$ denote the class of all such structures. These structures are designed to model how subjects attempt to sharpen their knowledge under the influence of evidence. As this is not in general a foolproof procedure, some potential knowledge states will plead ignorance where the original state has knowledge. If things go well enough, loss of knowledge on one issue can be compensated for by progress on some other issue. In any case $\leq$ is not such that it induces a cumulative ordering on potential knowledge states.

Let us now consider the knowability-thesis Ver with potential knowledge filling the rôle of knowability:

PotK. $$A \to \langle\mathsf{K}\rangle A.$$

The schema is valid in potential knowledge structures satisfying, for example, the condition that knowledge in a world can be *perfected* to an omniscient state. This is to assume that among the potential knowledge states accessible from what is known about a world a is one in which *everything* is known about a:[17]

Pefectibility. $$\exists T : S \leq T \;\&\; (Tab \;\Rightarrow\; a = b).$$

Call structures in $\mathbb{P}$ that satisfy Perfectibility *perfect potential knowledge structures*. The point to be made is that while PotK is valid in such structures, a trivialising schema such as

(13) $$\langle\mathsf{K}\rangle A \to \mathsf{K}A,$$

is not. This may be demonstrated directly by small models.

[17] The condition of perfectibility is stronger than what is needed for the validity of PotK. We adopt it here for the sake of simplicity. For a sufficient *and* necessary condition we need to switch the order of the quantifiers: Instead of there being a single state in which every truth is known, it suffices (and is also necessary for PotK) that for every truth there is a state in which it is known, i.e.

Knowability. $$a \in X \;\Rightarrow\; \exists S' : S \leq S' \;\&\; S'_a \subseteq X.$$

Alternatively, we may reason as follows. If PotK entailed (13), then the embarrassing Fitch-conclusion

$$A \to \mathsf{K}A \tag{14}$$

would follow. Now (14) corresponds to the condition

Omniscience. $\qquad Sab \;\Rightarrow\; a = b$

and evidently Perfectibility does not imply Omniscience. Thus, perfect potential knowledge models cannot force the validity of (14) and, hence, they cannot validate (13) either. So the verifiability thesis as expressed by PotK, i.e. in terms of potential knowledge, cannot be fitched.

We have thus shown that PotK carries no commitment to the conclusion of the Fitch-argument. Still, it is instructive to observe at which step the argument is blocked.

The argument starts by choosing an embarrassing instance of PossK, namely

$$P \wedge \neg \mathsf{K}P.$$

Using the (K-hedged) factivity of knowledge, KT, and the distributivity of K over $\to$, K, it proceeds as follows (as recalled in Section 3):

$$\cfrac{\cfrac{\cfrac{P \wedge \neg \mathsf{K}P}{\Diamond \mathsf{K}(P \wedge \neg \mathsf{K}P)}\,\text{PossK}}{\Diamond(\mathsf{K}P \wedge \mathsf{K}\neg\mathsf{K}P)}\,\text{K} \qquad \cfrac{\cfrac{\cfrac{}{\mathsf{K}\neg\mathsf{K}P \to \neg\mathsf{K}P}\,\text{KT}}{\Box(\mathsf{K}\neg\mathsf{K}P \to \neg\mathsf{K}P)}\,\text{RN}}{\neg\Diamond(\mathsf{K}P \wedge \mathsf{K}\neg\mathsf{K}P)}\,\equiv}{\cfrac{\bot}{P \to \mathsf{K}P}}$$

Suppose now that we applied PotK instead of PossK to the Fitch-formula $P \wedge \neg \mathsf{K}P$, thus obtaining

$$\langle \mathsf{K} \rangle (P \wedge \neg \mathsf{K}P),$$

saying that it is potentially known that P is an unknown truth—which sounds perfectly alright to the proponent of Ver. We could further use the converse of Mp (derivable from RMp) to move to

$$\langle \mathsf{K} \rangle P \wedge \langle \mathsf{K} \rangle \neg \mathsf{K}P.$$

But, again, this is unobjectionable. The formula records the unsurprising fact that P can be known (in the potentiality sense) and that it can also be known that P, as a matter of fact, is unknown. Note that using PotK the argument comes to a halt exactly at the same point at which the use of ActK would block it. With ActK we obtain the conclusion that it is possible to know that actually P and that P is actually unknown. This rings true but, as we have seen, it invites the embarrassing question as to what knowledge in a non-actual world of ongoings in the actual world consists in. Possible knowledge of the actual world is, by definition, transworld-knowledge. Potential knowledge in a given world is intraworld-knowledge.

7. Conclusion

The knowability principle in terms of potential knowledge expresses the expectation that for every true proposition there exists a potential state of knowledge in which that proposition is accepted. The expectation may be over-optimistic. The point of the present exercise is not to recommend epistemic optimism but to show that it is not ill-founded on merely logical grounds. Even if rooted in the very strong condition that a perfect state of knowledge in which all truths are known is within potential reach, the knowability principle does not collapse the distinction between truth and knowledge as the Fitch-argument would have it.

Acknowledgements. I am grateful to those who have commented on various occasions on earlier versions of this paper, in particular, Johan van Benthem, Ali Esmi, Dominik Kauß, Manfred Kupffer, Hannes Leitgeb, Isaac Levi, Oliver Petersen, Hans Rott, Sonja Smets, and Roy Sorensen. Two anonymous referees have commented in uncommon detail and have given me excellent advice which, I hope, I have succeeded in turning into an improved version of the paper.

References

[1] Alchourrón, C., Gärdenfors, P., and Makinson, D. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic 50* (1985), 510–530.

[2] Darwiche, A., and Pearl, J. On the logic of iterated belief revision. *Artificial Intelligence 89* (1997), 1–29.

[3] Edgington, D. The paradox of knowability. *Mind 93* (1985), 557–558.

[4] Edgington, D. Possible knowledge of unknown truth. *Synthese 173* (2010), 41–52.

[5] Fara, M. Knowability and the capacity to know. *Synthese 173* (2010), 53–73.

[6] Fitch, F. B. A logical analysis of some value-concepts. *The Journal of Symbolic Logic 28* (1963), 133–142.

[7] Fuhrmann, A. When hyperpropositions meet ... *Journal of Philosophical Logic 28* (1999), 559–74.

[8] Fuhrmann, A. Modal modifiers and hyperrelational frames. *forthcoming* (2013+).

[9] Hansson, S. O. A survey of non-prioritized belief revision. *Erkenntnis 50* (1999), 413–427.

[10] James, W. *The Meaning of Truth* (1909). Dover, New York, 2002.

[11] Levi, I. *The Fixation of Belief and Its Undoing.* Cambridge Univ. Press, Cambridge, 1991.

[12] Lindström, S. Situations, truth and knowability — A situation-theoretic analysis of a Paradox by Fitch. In *Action, Language and Cognition*, Ejerhed and Lindström, Eds. Umea, 1993.

[13] Moore, G. Russel's 'Theory of Descriptons' (1944). In *Philosophical Papers*. George Allen and Unwin, London, 1959, pp. 151–195.

[14] Rabinowicz, W., and Segerberg, K. Actual truth, possible knowledge. *Topoi 13* (1994), 101–115.

[15] Rott, H. *Change, Choice and Inference.* Clarendon Press, Oxford, 2001.

[16] Salerno, J., Ed. *New Essays on the Knowability Paradox.* Oxford University Press, Oxford, 2009.

[17] Spohn, W. *The Laws of Belief: Ranking Theory and Its Philosophical Applications.* Oxford University Press, Oxford, 2012.

[18] Tennant, N. *The Taming of the True.* Clarendon Press, Oxford, 1997.

[19] Williamson, T. On the paradox of knowability. *Mind 96* (1987), 256–261.

[20] Williamson, T. *Knowledge and Its Limits.* Oxford University Press, Oxford, 2000.

ANDRÉ FUHRMANN
INSTITUT FÜR PHILOSOPHIE
GOETHE-UNIVERSITÄT
60629 FRANKFURT a. M.
fuhrmann@em.uni-frankfurt.de