Active Agents

VINCENT F. HENDRICKS
Department of Philosophy and Science Studies, Roskilde University, P.O. Box 260,
DK-4000 Roskilde, Denmark
E-mail: vincent@ruc.dk

Abstract. The purpose of this survey is twofold: (1) to place some central themes of epistemic logic in a general epistemological context, and (2) to outline a new framework for epistemic logic developed jointly with S. Andur Pedersen unifying some key "mainstream" epistemological concerns with the "formal" epistemological apparatus.

Key words: Active agenthood, epistemic logic, epistemology, learning

Epistemic logic can be tracked back to the Middle Ages. The subject was really boosted to the attention of logicians and philosophers by Von Wright's first systematic surveys in the 1950s. Roughly 10 years later Hintikka published his seminal work on the logic of knowledge and belief. Epistemic and doxastic logics have since then grown into powerful enterprizes enjoying many important applications.

Despite all the important applications classical epistemic logic suffers from certain shortcomings. These shortcomings place the field in a rather odd and awkward position: On the one hand it is a discipline of indispensable importance to for example computer science, linguistics and economics and yet on the other hand it is a discipline devoted to the logic of knowledge and belief but alien to epistemologists and philosophers interested in the theory of knowledge.

Recent results and approaches designed to overcome these shortcomings have brought the logics of knowledge quite close to the theories of knowledge. In fact one may even identify a set of new *logical epistemologies* having essential features in common with important mainstream epistemological proposals.

1. Modelling with Multi-Modalities

In 1970, when epistemic logic was still partly in its infancy, Scott noted:

> Here is what I consider one of the biggest mistakes of all in modal logic: concentration on a system with just one modal operator. The only way to have any philosophically significant results in deontic logic or epistemic logic is to combine these operators with: Tense operators (otherwise how can you formulate

principles of change?); the logical operators (otherwise how can you compare the relative with the absolute?); the operators like historical or physical necessity (otherwise how can you relate the agent to his environment?); and so on and so on. (Scott, 1970: 143)

The focus on mono-modal systems in the 1970s were needed to clarify the mathematical foundations of modal logic in general and to get a firm grip on the basic tools needed for the model theory and proof theory of modal logic. Then in the 1980s multi-modal systems began to appear under the influence of computer science and AI with combinations of dynamic/temporal/epistemic logic. Around 1990, based on this experience the first significant mathematical theory started appearing about either "transfer" of properties of component modal systems to the whole, or "emergent phenomena" of the combined systems. Today, in branching tense logic researchers are mixing alethic and tense logical operators and in the epistemic logic of Fagin, Halpern, Moses and Vardi, epistemic operators are combined with tense logical operators while modelling the knowledge dynamics of an entire system of agents. Sowa's theory of nested graph models based on Dunn's semantics for laws and facts combined with a theory of contexts is a way of simplifying the reasoning tasks in multi-modal reasoning (Sowa, 2000), and can be adapted to Kripke models, situation semantics, temporal models and variants of these.⋆

Epistemic variants of modal logics usually begin with the slippery and much debated notion of "possible worlds" as Hintikka explains:

In order to speak of what a certain person *a* knows and does not know, we have to assume a class ("space") of possibilities. These possibilities will be called scenarios. Philosophers typically call them possible worlds. This usage is a symptom of intellectual megalomania. (Hintikka, 2003: 19)

A difference between a philosophical logician and a philosopher is that while the logician often remains rather agnostic about the ontological significance of the possible worlds and may just refer to them as scenarios, situations, states, contexts or conceptual constructions, the philosopher is usually quite concerned with the metaphysical baggage that comes along with the notion.

Possible worlds are often viewed as largely unanalyzed or even unanalyzable entities complete in their spatio-temporal history whatever they are. Another way to obtain multi-modalities is to decompose or "deconstruct" possible worlds, provide them with some structure that allows one to model and study their temporal extensions, their spatial extensions and the agent to which these extensions are epistemically pertinent. This is in a way what Sowa is doing when levelling contexts, and another way of decomposing worlds and providing them with explicit structure is accomplished by *modal operator theory*.

Modal operator theory (a term coined in Hendricks, 2001; Hendricks and Pedersen, 2000a, 2000b) is the cocktail obtained by mixing epistemic, tense and

---

⋆ I am indebted to Johan van Benthem for explaining to me the history of multi-modalities.

alethic logic with some concepts drawn from formal learning theory[*] and has so far been used to study the validity of limiting convergent knowledge.

The initial setup is adopted from Kelly's computational epistemology in Kelly (1996). First, an evidence stream is what supplies a scientific inquiry method or agent with evidence.[**] An evidence stream, ε, is an ω-*sequence* of natural numbers, i.e., $\varepsilon \in \omega^\omega$. According to the definition it is assumed that the method of inquiry is studying some system with discrete states that may be encoded by natural numbers, so that in the limit the method receives an infinite stream ε of numbers. Hence, an evidence stream $\varepsilon = (a_0, a_1, a_2, \ldots, a_n, \ldots)$ consists of *code* numbers of evidence, i.e., at each state i of inquiry a_i is the code number of the evidence acquired at this state. Let the lower case Greek letters τ, θ, ζ, μ, ς, υ, $\ldots$ denote evidence streams. It is now possible to provide possible worlds with some explicit analyzable structure leaving the metaphysics behind:

> A possible world is a pair consisting of an evidence stream ε and a state coordinate n, i.e., (ε, n), where $\varepsilon \in \omega^\omega$ and $n \in \omega$.

The set of all possible worlds $\mathcal{W} = \{(\varepsilon, n) \mid \varepsilon \in \omega^\omega, n \in \omega\}$. The following notational conventions are introduced with respect to worlds:

- Let all pairs (ε, n), (τ, n'), (θ, m), (ς, m''), $\ldots$, denote possible worlds.
- Let $\varepsilon \mid n$ denote the finite initial segment of an evidence stream also called the *handle*.
- Let $\omega^{<\omega}$ denote the set of all finite initial segments of evidence streams.
- Let $(\varepsilon \mid n)$ denote the set of all infinite evidence streams that extend $\varepsilon \mid n$, also called the *fan*.

The handle $\varepsilon \mid n$ represents the evidence seen up to state n, i.e., $\varepsilon \mid n = a_0, a_1, \ldots, a_{n-1}$. The rest $\varepsilon \backslash \varepsilon \mid n = a_n, a_{n+1}, a_{n+2}, \ldots$ is the evidence that one would observe if the world develops according to (ε, n). The set of possible worlds in the fan is defined by $[\varepsilon \mid n] = (\varepsilon \mid n) \times \omega$. In other words $[\varepsilon \mid n] = \{(\tau, k) \mid k \in \omega$ and $\tau \mid n = \varepsilon \mid n\}$ (Figure 1).

The world fan $[\varepsilon \mid n]$ represents the *minimal background knowledge* of possible empirical possibilities or values that the world may take according to the method.

This simple characterization of evidence streams, state coordinates and possible worlds impose a branching time structure. For any finite time the method has observed some handle of the world, but from that point onwards, the world may "branch" in any way it pleases for all the method knows. So points or moments in time can be defined as finite initial segments of evidence. This indicates a branching time structure and a branching tense logic rather than a linear model

---

[*] Recently redubbed *computational epistemology* by Kelly.
[**] The terms "method" and "agent" are used interchangeably since it is immaterial from the point of view of modal operator theory whether the agent applies a (possibly internal) method or the method operates as an independent cognitive device or disposition.
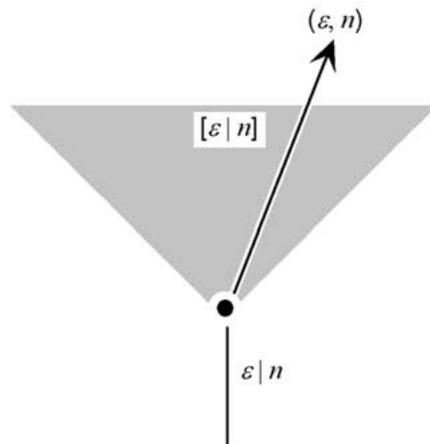
Figure 1. Handle,fan and possible world.

of time. The current structure is closely related to standard Ockhamistic semantics. An Ockhamistic tense structure is typically given by:

1. O is a non-empty set of moments in time,
2. $\prec$ is an irreflexive and transitive earlier-later relation which is backwards linear:

$$\forall n, n', n'' \in O : (n \prec n' \wedge n'' \prec n') \rightarrow$$
$$(n \prec n'' \vee n'' \prec n \vee n = n'').$$

An Ockhamistic model is a triple $(O, \prec, V)$ where V is a valuation function which assigns truth-values $V(n, p)$ to all pairs where the first argument is an element of O and the second argument p is a propositional variable. It is possible to define an Ochamistic tense structure in modal operator theory as a pair $(\mathbb{T}, \prec)$ such that:

1. $\mathbb{T} = \{\varepsilon \mid n \mid \varepsilon \in \omega^\omega, n \in \omega\}$,
2. $t_1 \prec t_2$ iff $\varepsilon \mid n = \tau \mid n$ and $n < m$,

where a *moment* in time t is the branching moment of times such that $t_1 = \varepsilon \mid n$ and $t_2 = \tau \mid m$. In other words, a moment in time is a finite initial segment $t \in (\mathbb{T}, \prec)$. Hence a moment in time is defined as the course of events up until "now." Time is based on events and thus the branching time structure is given by events. We are working on the entire tree of finite sequences of natural numbers where finite initial segments (of events) define moments in time.

The evidence stream ε is the actual evidence stream. The state coordinate n, for some specified n, is to be thought of as the "age" of the world (ε, n), i.e., n is the
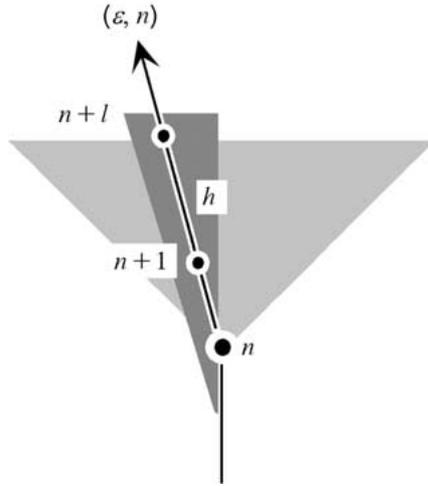
Figure 2. Truth of an empirical hypothesis in a possible world.

time in the branch (or the universal clock). The earlier-later relation is defined with respect to finite initial segments.

The structure $(\mathbb{T}, \prec)$ satisfies the requirements above; the relation is irreflexive, transitive and backwards linearly ordered. There exist, of course, many other Ockhamistic tense structures. One example would be a structure that extends into the indefinite past. The current structure, however, has a starting point.

The agents will eventually need some hypotheses to have knowledge of. Hypotheses will be identified with sets of possible worlds. Define the set of all simple empirical hypotheses

$$\mathcal{H} = P(\omega^\omega \times \omega).$$

So $h \in \mathcal{H}$ iff $h = \{(\varepsilon, n) \mid \varepsilon \in A, n \in N\}$ where $A \subseteq \omega^\omega$ and $N \subseteq \omega$. An empirical hypothesis h is said to be *true* in world (ε, n) iff

$$(\varepsilon, n) \in h \text{ and } \forall l \in \omega : (\varepsilon, n + l) \in h.$$

Truth requires identification and inclusion of the actual world (ε, n) in the hypothesis for all possible future states of inquiry. Attention is restricted to hypotheses which are stably true like laws of nature (Figure 2).

Modal operator theory now allows for the introduction of tense and alethic operators in the following intuitive way where the temporal operators F, G, P and H take their usual meaning.

1. (ε, n) validates Fh *iff* $\exists k > n : (\varepsilon, k)$ validates h.
2. (ε, n) validates Gh *iff* $\forall k > n : (\varepsilon, k)$ validates h.
3. (ε, n) validates Ph *iff* $\exists k < n : (\varepsilon, k)$ validates h.
4. (ε, n) validates Hh *iff* $\forall k < n : (\varepsilon, k)$ validates h.

5. (ε, n) validates $\square h$ *iff* $\forall (\tau, m) \in \mathcal{W} : (\tau, m)$ validates h.

6. (ε, n) validates $\boxdot h$ *iff* $\forall k \in \omega : (\varepsilon, k)$ validates h.

7. (ε, n) validates $\boxminus h$ *iff* $\forall (\tau, m) \in [\varepsilon \mid n] : (\tau, m)$ validates h.

The operator "$\square$" is called universal necessity since it quantifies indifferently over the set of all possible worlds simpliciter. Operator "$\boxdot$" is called temporal necessity because it quantifies indifferently over all times. Finally "$\boxminus$" is called empirical necessity because it ranges over all the empirically possible alternatives relative to the world fan.

The tense and alethic operators defined so far follow relatively standard branching temporal logic. It remains to introduce the epistemic operators in this new framework.

2. From Indices to Functions

2.1. "First Generation Epistemic Logic"

Having defined the space of possibilities, the early epistemic logic – what is referred to as the "first generation epistemic logic" by Hintikka (2003) – proceeded axiomatically. "An agent δ knows that p" (where p is some proposition) is formalized as a modal operator $K_\delta p$ in a formal language which is interpreted using the standard apparatus of modal (alethic) logic. One of the hopes was that cataloguing the possible complete systems of such logics would allow for a picking of the most appropriate or intuitive ones often ranging from **S4** over the intermediate systems **S4.2**–**S4.4** to **S5**.[*] By way of example, Hintikka (1962) settled for **S4** while Lenzen (1978) argued for **S4.2**. The significant difference between alethic logic and epistemic logic at the time was the addition of the agent δ to the syntax. The interesting epistemological question is what roles are assigned the agents in the first generation epistemic logic since they are the ones who apparently have knowledge which is, say, **S4.3** valid? That agents hold the knowledge is also the natural understanding of the notation $K_\delta p$ which Hintikka notes:

> Epistemic logic begins as a study of the logical behavior of the expression of the form "b knows that." One of the main aims of this study is to be able to analyze other constructions in terms of "knows" by means of "b knows that." The basic notation will be expressed in the notation used here by "K_b." This symbolization is slightly misleading in that a formula of the form $K_b S$ the term b for the agent (knower) is intended to be outside the scope of K, not inside as our notation might suggest. (Hintikka and Halonen, 1998)

The problem is that the only roles of the agents in the "first generation epistemic logic" are to serve as indices on the accessibility relation between possible worlds.

---

[*] Refer to Gochet and Gribomont's (2003) forthcoming paper for an excellent survey of epistemic logic and its key issues.

Figure 3. An inactive agent.

Now, epistemic-logical principles or axioms building up modal systems are relative to an agent who may or may not validate these principles or axioms like the *K K*-thesis. Indices on accessibility relations will not suffice for epistemological and cognitive pertinence simply because there is nothing particularly epistemic about being indices. The agents are *inactive* in the first generation epistemic logic (Figure 3).

This pans out in the more principal discussion of whether there is, or should be, a relation between the formal results of epistemic logic and more general epistemological concerns like justification, methodology, reliability and rationality? Some epistemic logicians were of the opinion that there should be no such relation like Lenzen argues:

> The search for the correct analysis of knowledge, while certainly of extreme importance and interest to epistemology, seems not significantly to affect the object of epistemic logic, the question of validity of certain epistemic-logical principles. (Lenzen, 1978)

In contrast, Hintikka has all along pursued the idea that epistemic logic should aim at hooking up with broader epistemological issues. Early on, Hintikka (1962) argued that the axioms or principles of epistemic logic are conditions describing a certain kind of general (strong) *rationality*. Whatever statements can be proved false by application of the epistemic axioms or principles are not inconsistent in the sense that their truth is logically impossible but rather "indefensible." Indefensibility is defined by the agent's sloppiness or incapacity in the past, present or future to follow the implications of what he knows. Defensibility is in turn defined as not falling victim of what Chisholm called "epistemic negligence." The idea of defensibility provides a hint as to the cognitive status of the epistemic axioms. An epistemic statement for which its negation is simply indefensible is referred to as "self-sustaining." Self-sustenance corresponds to the meta-logical concept of validity. In other words, a self-sustaining statement corresponds to a logically valid statement. Thus, a statement which is rationally indefensible to deny. But then the epistemic principles are descriptions of rationality.

If epistemic logics are not to be pertinent to the knower who are they to be pertinent to? An agent may have knowledge which is **S4.3** valid but what one would really like to know – what really bakes the epistemological noodle – is *why* and *how* the agent has to *behave* in order to gain the epistemic strength that he has in terms of validity. We want active agents in order to make epistemic logic pertinent to

Figure 4. An active agent.

epistemology, computer science, artificial intelligence and cognitive psychology, and the original formalization of a knowing agent suggests this. Inquiring agents are agents who read data, change their minds, interact or have common knowledge, act according to strategies and play games, have memory and act upon it, follow various methodological rules, expand, contract or revise their knowledge bases, etc., all in the pursuit of knowledge – inquiring agents are *active agents* (Figure 4).

2.2. "Second Generation Epistemic Logic"

The idea of *active agents* is admittedly an extended interpretation of one of the characterizing features, and great virtues of, what Hintikka calls the "second generation epistemic logic" in Hintikka (2003): The realization that the agents of epistemic logic should play an active role in the knowledge acquisition, validation and maintenance processes. In Hintikka (2003) and elsewhere, Hintikka observes this obligation by emphasizing the strategies for his new game–theoretical semantics, or epistemic logic as the logic of questions and answers and the search for the best questions to ask.

Game-theory is about strategies for winning games – and it is an agent who may or may not have a winning strategy among the other agents. Fagin, Halpern, Moses and Vardi, Auman, Stalnaker and other logicians studying game theory have demonstrated how epistemic logic uncovers important features of *agent rationality* (Fagin et al., 1996). Additionally, Fagin et al. (1995) stipulate (sometimes) for the multi-agent systems that the agents may possess certain *epistemic properties* – in particular *perfect recall*. The idea of perfect recall is that the interacting agents' knowledge in the dynamic system may grow while the agents still keep track of old information. An agent's current local state encodes all that has happened so far in the system run. That is also an aspect of the active agents paradigm in the new trends of epistemic logic. Ditmarch, van der Hoek and Kooi develop a "dynamic epistemic logic" which studies how information changes, model *actions* with epistemic aspects and provide a new language to reason about these actions. Actions are viewed as multi-agent Kripke models where a precondition function replaces the usual denotation function for propositional variables (Ditmarch et al., 2003). Belief revision theorists like Rott model "informational economy" or "conservatism" and consider cognitive economics and the problem of rational choice

for *agents* (Rott, 2003) and Baltag–Moss–Solecki combine epistemic logic with belief revision theory (Baltag et al., 1999).

Modal operator theory joins the second generation epistemic logic obligation. The theory has at its base rather than as a derivative the idea that whatever epistemic axioms and epistemic systems are possible to validate for some epistemic operator is *acutely sensitive to the methodological behavior of the agent involved*. This will be clear when it is realized that agents may be viewed as functions rather than indices as in modal operator theory.

2.2.1. *Methods, Limiting Convergence and Recommendations*

A scientific inquiry method, in particular a *discovery method*, conjectures hypotheses in response to the evidence received. More specifically, a discovery method δ is a function from finite initial segments of evidence to hypotheses, i.e.,

$$\delta : \omega^{<\omega} \longrightarrow \mathcal{H}. \tag{1}$$

The only immediate criterion of rationality imposed on a discovery method is that it is not allowed to conjecture absurdities. Thus, for any discovery method δ and for any world (τ, n'): $\delta(\tau \mid n') \neq \emptyset$. What remains to be defined is a criterion of successful convergence for a discovery method.

The convergence criterion imposed on discovery methods in Hendricks (2001) and Hendricks and Pedersen (2000a, 2000b, 2004) is a limiting one. Limiting convergence is not a novel concept. Peirce held the view that it is impossible to say anything about the direction of science in the short run but science may all the same asymptotically approach the truth in the long run (Peirce, 1958). James also argues for limiting convergence because knowledge of general laws is impossible if one is required to say when science has succeeded (James, 1960). Since James and Peirce, limiting convergence has become a more and more respected convergence criterion both in philosophy (Boyd, 1984), computer science (Gold, 1967) and in methodology where for an example computational epistemology uses limiting convergence for obtaining certain characterization theorems (Kelly, 1996) while Bayesians apply a limiting convergence criterion to facilitate "almost sure" convergence theorems (Earman, 1992). Limiting convergence simply means that the method is allowed to vacillate some number of times, which cannot be specified in advance, before it reaches its point of convergence. The user of the method, or the agent himself, may not know when this state of stabilization has occurred since no determinate sign of convergence may ever be produced.

In Hendricks (2001) a very strong version of limiting convergence is utilized based on a suggestion made in (Kelly, 1996). The notion implies uniform convergence moduli in all relevant worlds:* δ discovers h in (ε, n) iff

---

* A weaker version of limiting convergence may indeed be entertained such that there is no uniform convergence moduli throughout worlds. This is more in the spirit of Peirce and later Gold.

$$\exists k \forall n' \geq k \forall (\tau, n') \in [\varepsilon \mid n] : \delta(\tau \mid n') \subseteq h \tag{2}$$

for which the convergence modulus

$$cm(\delta, h, (\varepsilon, n)) = \mu k \forall n' \geq k \forall (\tau, n') \in [\varepsilon \mid n] : \delta(\tau \mid n') \subseteq h. \tag{3}$$

The method may be subject to various methodological recommendations, program commands or behavioral patterns. For a well-known example, say that a discovery method δ is *consistent* iff

$$\forall (\tau, n') : \big[[\tau \mid n'] \cap \delta(\tau \mid n') \neq \emptyset\big]. \tag{4}$$

Consistency may be strengthened to another recommendation for which the discovery method δ has *perfect memory* iff

$$\textit{if } (\mu, k) \in \delta(\varepsilon \mid n) \textit{ then}$$

$$(\mu \mid n = \varepsilon \mid n) \textit{ and } \forall l \leq k : (\mu, l) \in \delta(\varepsilon \mid n). \tag{5}$$

Furthermore a discovery method δ is said to be *consistently expectant* iff

$$\textit{if } (\mu, k) \in \delta(\varepsilon \mid n) \textit{ then } [k \geq n \textit{ and } (\mu \mid n = \varepsilon \mid n)]. \tag{6}$$

Perfect memory and consistent expectation are inconsistent: Suppose $(\mu, k) \in \delta(\varepsilon \mid n)$. If the discovery method has perfect memory, then $\forall l \leq k : (\mu, l) \in \delta(\varepsilon \mid n)$. Now if the method simultaneously is consistently expectant, then $(\mu, k) \in \delta(\varepsilon \mid n)$ implies $k \geq n$, but by perfect memory $(\mu, n - 1) \in \delta(\varepsilon \mid n)$. Contradiction! Thus, a discovery method cannot obey perfect memory and consistent expectation at the same time. Methodological recommendations may conflict as Nozick (1993) emphasizes and Kelly (1996) proves on numerous occasions.

2.2.2. *Limiting Convergent Knowledge*

It is possible to define a variety of limiting concepts of knowledge in modal operator theory. It suffices for the current purposes to restrict attention to the following concept of knowledge which may be informally portrayed accordingly (for a thorough epistemological motivation of the concept please refer to Hendricks (2001)):

δ *may* scientifically know *h in the limit iff there exists a possible world which validates δ's knowledge of h. In other words:*

1. *h is true,*
2. *δ conjectures h after some finite evidence sequence has been read and continues to conjecture h in all future.*
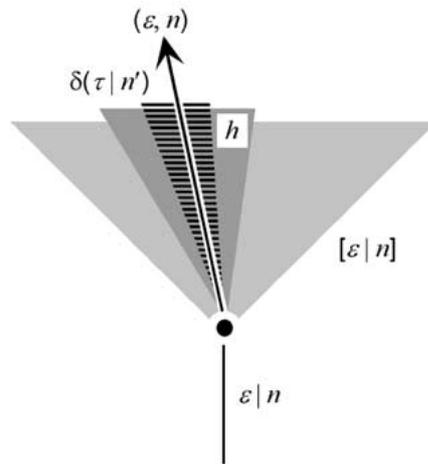
Figure 5. Limiting convergent knowledge.

Thus, there is a time n, such that for each later time n' : δ conjectures h at n' in all possible worlds admitted by the background knowledge in which h is true. Now if we modify this definition by telling the method only to conjecture something entailed by the evidence and the background knowledge then we end up with:[*]

(ε, n) validates $K_\delta h$ *iff* $\forall n' \geq n, \forall (\tau, n') \in [\varepsilon \mid n]$:

1. $\delta(\tau \mid n') \subseteq h$,
2. $(\tau, n') \in \delta(\tau \mid n')$.

The definition of knowledge is depicted in Figure 5. It is a strong notion of knowledge, especially given the latter condition since it implies entailment of the truth by the evidence and the background knowledge. The method is *infallible*.

Observe that since the agent is a function from finite initial segments of evidence to hypotheses there is no knowledge without the method conjecturing a hypothesis. Thus, the agent is *actively within the scope of the operator* so $K_\delta h$ in modal operator theory is genuinely faithful to the intended meaning of the standard formalization of the knowledge operator.

2.3. MODAL PROPOSITIONAL LOGIC

All the set-theoretical intuitions of the previous sections are completely formalizable in a modal propositional logic $\mathcal{L}$ with some modifications to the standard setup. In short, the syntax includes an infinite supply of propositional variables, brackets, the Boolean operators, the unary modal operators F, G , H, P, $\square$, $\boxdot$, $\boxminus$, K_δ. The construction of well-formed formulas follows the usual recursive recipe.

---

[*] This methodological recommendation is called *infallibility* in Hendricks (2001).

A model $\mathbb{M}$ is a triple $\langle \mathcal{W}, \varphi, \delta \rangle$ where $\mathcal{W}$ is a non-empty set of possible worlds, φ is a denotation function from propositional variables into $P(\mathcal{W})$ and δ a discovery function: $\delta : \omega^{<\omega} \longrightarrow P(\mathcal{W})$. A few more modifications to the standard setup are required but the reader is referred to Hendricks (2001) for the details.[*]

Once the syntax and semantics are fixed one may then ask the following pair of questions:

1. *Which epistemic axioms can be validated by an epistemic operator based on the definition of limiting convergent knowledge for discovery methods?*
2. *Does the validity of the various epistemic axioms relative to the method depend upon enforcing methodological recommendations?*

The first question is close to the first generation issue of epistemic logic, the second question a variety of the second generation issue. It is possible to prove the following theorem:

*If knowledge is defined as limiting convergence, then knowledge validates **S4** iff the discovery method has consistent expectations.* Hendricks (2001: 203).

Neither axiom (T): $K_\delta h \rightarrow h$ nor axiom (K): $K_\delta(h \rightarrow h') \rightarrow K_\delta h \rightarrow K_\delta h'$ require methodological recommendations for their validity; they are valid given the definition of knowledge as it stands. But axiom (4): $K_\delta h \rightarrow K_\delta K_\delta h$, or the KK-thesis (and also sometimes referred to as positive introspection), does require you to entertain a methodological recommendation even if the discovery method is infallible![**] Some more detail pertaining to the KK-thesis and the stated theorem will be provided in the last section. It can additionally be shown for instance that perfect memory is an impediment to validating axiom (4). Thus, methodological recommendations may be classified as to whether they are *boosting* in the sense that a methodological recommendation is *conducive* to validating epistemic axioms and systems, *debilitative* in the sense that the methodological recommendation is an *impediment* to validating epistemic axioms and systems, or *neutral* if it is neither boosting nor debilitative.[‡]

It should also be noted that any method defined for limiting convergence fails to validate **S5** no matter which methodological recommendations are in play. Intuitively the reason is that the axiom of wisdom (also known as negative introspection), axiom (5): $\neg K_\delta h \rightarrow K_\delta \neg K_\delta h$, requires the method to turn non-convergence into convergence but when knowledge is defined by convergence this is impossible. This is an interesting philosophical result for the following reason:

---

[*] Some formal sloppiness in this presentation will occur for ease of readability. By way of example, in the modal formalization, A denotes the proposition defining a hypothesis h. However not to introduce too much new notation we continue to write h except where explicitly stated otherwise.

[**] Consistency and perfect memory do not help either in validating **S4**.

[‡] These notions are similar to Kelly's distinction between permissive and restrictive architectures.

According to the Academic skepticism of Arcesilaus and Carneades there is no irrefutable and guaranteed standard by which one can unequivocally ascertain the truth/falsity status of the judgements made. Any knowledge claim put forth introduces some element which transcends immediate experience. But then one would have to know something, namely the specific element introduced, which possibly could not be false. Then again, the evidence for this element would have to rest on perception and reasoning which in the first place were found to be unreliable. Hence the reason for the classical Socratic dictum of *epistemic modesty*: "All I know is that I know nothing." Suppose epistemic modesty is propositionally formalized as an axiom schema $K_\delta \neg K_\delta h$ for arbitrary h. Under the assumption that the Academic skeptic is willing to take the logical consequences of epistemic modesty he also commits himself to the axiom of wisdom because $K_\delta \neg K_\delta h$ trivially implies $\neg K_\delta h \rightarrow K_\delta \neg K_\delta h$ given the introduction rule for $\rightarrow$. The fact that one cannot converge to axiom (5) is accordingly a strong result against Academic skepticism because even if the Academics are allowed the use of the infallible method that never makes mistakes they cannot themselves converge to the axiom which they take to witness the impossibility of knowledge!*

The general lesson to be learned from the theorem is that whatever epistemic axioms or systems the method is capable of validating is immensely sensitive to the very behavior of the method. Various methodological recommendations may have to be imposed on the method in order to obtain the desired epistemic strength – active agents again.

Modal operator theory provides a rich framework for other investigations of which a few highlights will be mentioned. Some axioms including tense modalities turn out to be valid like

$$K_\delta h \rightarrow \begin{cases} 1. & F K_\delta h \\ 2. & G K_\delta h \end{cases} \tag{7}$$

which by the way is independently discussed by Fagin, Halpern, Moses and Vardi in Fagin et al. (1995). Axioms including tense and alethic modalities may also be investigated for validity – see further Hendricks (2001: chap. 13).

So far attention has been restricted to discovery methods. One may equally well define an assessment method α as a function from finite initial segments of evidence and hypotheses to $\{0, 1\}$ where 0 denotes false and 1 denotes truth

$$\alpha : \mathcal{H} \times \omega^{<\omega} \longrightarrow \{0, 1\}, \tag{8}$$

which is a standard method of justification assessing hypotheses in the light of incoming evidence. Successful limiting convergence can be defined for assessment such that α decides h in the limit in (ε, n) iff

1. if h is true in (ε, n), then

$$\exists k \geq n, \forall n' \geq k, \forall (\tau, n') \in [\varepsilon \mid n] : \alpha(h, \tau \mid n') = 1,$$

---

* Refer to Hendricks (2001: 210) for the detailed proof.

2. if h is false in (ε, n), then

$$\exists k \geq n, \forall n' \geq k, \forall(\tau, n') \in [\varepsilon \mid n] : \alpha(h, \tau \mid n') = 0$$

with the following limiting convergence modulus

$$cm(\alpha, h, (\varepsilon, n))$$
$$= \mu k \geq n, \forall n' \geq k, \forall(\tau, n') \in [\varepsilon \mid n] : \alpha(h, \varepsilon \mid n) = \alpha(h, \tau \mid n').$$

The assessment α verifies h in the limit when h is true and refutes h in the limit when h is false.

It turns out that discovery methods can *induce* assessment methods in the following way assuming the truth of the hypothesis in question. This is a natural assumption as the inquiry methods are used in knowledge definitions with truth as a precondition:

If a discovery method δ discovers h in (ε, n) in the limit, then there exists a limiting assessment method α which verifies h in (ε, n) in the limit.

Assume that δ discovers h in (ε, n) in the limit and let $cm(\delta, h, (\varepsilon, n))$ be its convergence modulus. Define α in the following way:

$$\alpha(h, \varepsilon \mid n) = 1 \text{ iff } \delta(\varepsilon \mid n) \subseteq h.$$

It is clear that if $n' \geq cm(\delta, h, (\varepsilon, n))$ then for all $(\tau, n') \in [\varepsilon \mid n] : \delta(\tau \mid n') \subseteq h$. Consequently $\alpha(h, \tau \mid n') = 1$ and therefore

$$cm(\alpha, h, (\varepsilon, n)) = cm(\delta, h, (\varepsilon, n)).$$

Assessment methods can also induce discovery methods along similar lines. This is interesting because a limiting notion of knowledge defined using an assessment method rather than a discovery method also validates **S4**. This information may in turn be used when *knowledge transmissibility* is studied in what has been labelled *multiple method systems* (Hendricks, 2001). Knowledge transmissibility was first studied by Hintikka (1962). Hintikka investigated whether

$$K_a K_b p \rightarrow K_a p \tag{9}$$

held for his definition of knowledge where p is some arbitrary proposition and a, b are agents. In a certain sense knowledge transmissibility is rather trivial here because it is essentially the iterated version of axiom (T) with different agents. As long as the agents index the same possible worlds knowledge transmissibility holds in the sense of (9). But in the current scheme of things knowledge transmissibility is far from trivial because there are agents or methods of different natures based on either discovery or assessment. Thus we have to consider in all generality whether

$$K_\Theta K_\Xi A \rightarrow K_\Theta h \tag{10}$$

is valid for arbitrary inquiry methods Θ, $\Xi \in \{\alpha, \beta, \gamma, \delta\}$ where γ, δ are discovery methods while and α, β are assessment methods. This means a classification of the transmissibility instances paraphrased as:

Uniform Transmissibility. Is it possible that a discovery method δ having knowledge of the fact that another discovery method γ has knowledge of some hypothesis h, may obtain knowledge of this hypothesis h and similarly for assessment?

Non-Uniform Transmissibility. Is it possible that a discovery method δ having knowledge of the fact that another assessment method α has knowledge of some hypothesis h, may obtain knowledge of this hypothesis h and similarly starting with an assessment method?

Transmissibility in modal operator theory constitutes an example of how communication effects the cognitive states and capacities of the methods involved in the knowledge exchange. It turns out, given inducement, that the answers to both of the questions stated immediately above are affirmative. The latter result is interesting in that it partially breaks the back of the classical, and much celebrated, dichotomy between assessment and discovery in the philosophy of science and methodology: There is as much a logic of discovery as there is a logic of justification as far as knowledge acquisition is concerned and the different methods may even reliably transmit knowledge to one another.

Successful transmissibility is also sensitive to the goal of inquiry. Some philosophers argue that truth is a way too ambitious aim and that scientific inquiry may plausibly carry on settling for less like consistency, simplicity, empirical adequacy, unification, etc. This making do with less rather than more is also partly responsible for the great divide between scientific realism and scientific anti-realism. It is possible to define other cognitive goals besides truth in the current paradigm. In (Hendricks, 2001) a goal is defined which is congruent with Van Fraassen's notion of empirical adequacy (van Fraassen, 1980). According to Van Fraassen, truth implies empirical adequacy but empirical adequacy does not necessarily entail truth. Switching the goal of inquiry from truth to empirical adequacy also changes the epistemic attitude from one of knowledge to one of *acceptance*. Suppose then that an arbitrary method of either assessment or discovery *accepts* some hypothesis and consider again the possible knowledge transmissibility instances between two methods in (10). Inducement is preserved in both directions since a discovery method is able to output the hypothesis that the assessment method is currently pointing to, and the opposite direction also passes muster because the assessment method is able to verify whatever hypothesis the discovery method is conjecturing. While inducement is preserved, knowledge transmissibility is *not* always preserved. In the uniform case where the inquiry methods are of a fixed inquiry type a discovery method δ may *accept* that another discovery method γ

has *knowledge* of h. This does not imply however that δ may swop epistemic attitudes from acceptance to knowledge of h by watching over what γ does exactly because empirical adequacy does not imply truth. In the non-uniform case with mixed cognitive goals the situation is sometimes similar in that transmissibility also can fail from time to time. An assessment method α may *accept* that another discovery method γ has *knowledge* of h but again this does not entail a swop of attitudes exchanging acceptance for knowledge and empirical adequacy for truth – the "cognitive exchange rate" is wrong. While a method may use inducement to communicate convergence, observing stabilization does not necessarily add up to communicating the goal of the inquiry process.

Transmissibility may be construed as a conversation taking place between two agents in which one of the agents is either fully or only partly aware of the information of the other agent. Now whether the inquiring agent may obtain "first-hand" information by studying the answering agent's stabilization is highly sensitive to the two agents sharing the same cognitive goal. Even if the agent answers *adequately*, the information passed along may not suffice for the inquiring agent to pick up the *true* message.

3. Epistemic Logic and Epistemology

Contemporary epistemological studies are roughly either carried out in

1. a *mainstream* and traditionally rather informal way using largely common-sense considerations and concentrating on sometimes folksy and sometimes exorbitantly speculative examples/counter-examples, or
2. a *formal* way by applying a variety of tools and methods from logic, computability theory or probability theory to the theory of knowledge.

The two traditions have unfortunately proceeded largely in isolation from one another. It may sometimes be hard to see how the formal results hook-up with more traditional epistemological themes. But the two approaches have much in common epistemologically and methodologically and may significantly benefit from one another.

3.1. FORCING

Epistemology is a reply to skepticism. Skeptics have since the days of old cited *prima facie* possibilities of error as the most substantial arguments against knowledge claims. A contemporary response to skepticism is to invoke *forcing* – a term coined in Hendricks (2001). Forcing is a kind of "logical sufficiency." Perhaps it is more of an heuristic principle than really a thesis:

Whenever knowledge claims are challenged by alleged possibilities of error, the strategy is to show that the possibilities of error fail to be genuine in the relevant sense.

By way of examples, the influential *epistemic reliabilism* of Goldman, Nozick's elegant formulation of the *counterfactual reliabilism* and Lewis' new *modal epistemology* are informal epistemological proposals observing the forcing relation.

Epistemic reliabilism Goldman (1979, 1986, 1992) acknowledges the agent's limited cognitive abilities and accordingly deflates the agent's epistemic responsibilities. The idea is to substitute the excessive requirements often proposed by skepticism for knowledge acquisition with more moderate conditions. Hence a justified belief may itself very well be false, but its mode of acquisition must in general yield truth. Besides the requirement that the target belief must be true in order to gain knowledge, its mode of acquisition must rule out all relevant possibilities of error. Thus, the forcing is given by the mode of acquisition: The mode of acquisition may not be able to exclude the possibility that a Cartesian demon is feeding systematically misleading sensations, or that they are vat-generated impressions. Then again these are not realistic possibilities of error.⋆ To combat the skeptic it suffices to note that infallible or certainty-producing methods are not required for knowledge.

Nozick's (1981) counterfactual reliabilistic knowledge definition (a similar one was advocated earlier by Dretske (1970)) is also a forcing proposal. The inherent decision procedure and the counterfactual semantics require the agent to succeed in all possible worlds sufficiently close to the actual world in order to acquire knowledge. The agent may not know that he is not a brain in a vat – but then again, that possibility of error is so remote. Asking a physicist whether it is viable possibility of error that his voltmeter is calibrated incorrectly while measuring the voltage drop over some LRC circuit is reasonable, asking him then whether it is a realistic possibility of error that he is a brain in a vat is just silly and far out.

Lewis' modal epistemology assumes knowledge of a great many things in a variety of different (conversational) contexts (Lewis, 1996). Considering brains in vats and Cartesian demons is to "epistemologize" which may make the knowledge we thought we had evaporate into thin air. Hence, all we need are rules to slice off possible worlds and then describe how we avoid error and gain truth in the ones left dictated by the current context (as humans actually do).

It turns out that a host of formal epistemological proposals share the forcing heuristic as well. Knowledge claims may be restricted by the constraints imposed on the accessibility relation between possible worlds, which may be viewed as the forcing foundation for epistemic logic. Also, probability theory, in particular Bayesian epistemology, requires the agent to succeed in circumstances with high

---

⋆ ... even though the method may not be able to determine this which is the entire point of the brains and the demons. Kelly has shown that global underdetermination exactly like demons and brains equal the impossibility of reliable inquiry (Kelly, 1996).

probability and dispose of sets of possibilities carrying low probability (Howson and Urbach, 1989; Earman, 1992). Thus, if the truth lies in worlds ascribed 0 or infinitesimal priors the truth will never be found. That is forcing. Computational epistemology requires success in all worlds in accordance with the background knowledge of empirical possibilities which likewise is to pay homage to the forcing thesis. Finally, modal operator theory requires success in only possible worlds consistent with what has been observed until now which satisfies the forcing relation.

From the forcing perspective what separates the different epistemological proposals, whether mainstream or formal, are the criteria imposed for circumscribing the set of relevant worlds over which to succeed for knowledge to come about. Additionally, if one also observes the difference between what may be called a *first person perspective on inquiry* following Lewis (1996) in which

— it is considered what an agent can solve, can do or defend considering the available means for an end given the epistemic environment he is sunk into

and a *third person perspective on inquiry* in which

— it is considered what an agent could solve, could do or defend considering the best means for an end independently of the epistemic environment he is sunk into

the mainstream and formal approaches have even more to say to each other as will become apparent below. A detailed treatment of forcing epistemology is to be found in Hendricks (2003).

3.2. Justification and Methodology

Since Plato's *Meno* and *Theatetus* epistemology has sought to identify the necessary and sufficient conditions for the possession of knowledge. The result has been the much celebrated tripartite definition of knowledge as justified true belief.

The justification condition of this definition has probably received the greatest attention of the three components from Plato to contemporary epistemology because it is apparently difficult to define it properly as Lenzen emphasizes:

> Though there is basic agreement that *something* must be added to true belief to obtain knowledge, what precisely this "something" is, remains far from being evident. Because of the vagueness of such notions as "having sufficient reasons for believing", "being justified in believing", it is difficult to make a decision concerning the adequacy of (5), i.e., that knowledge implies justification. Lenzen (1978: 28)

In the standard definition, any claim to knowledge requires that the satisfaction of the belief condition is "adequately" connected to the satisfaction of the truth

condition. The two conditions alone are thought to be jointly insufficient to secure knowledge since some true beliefs may be the fortunate result of lucky conjectures, various accidental inferences, evidence collected under obscured perceptual circumstances etc. Now, such beliefs should on the standard analysis obviously not count as knowledge since the belief and truth conditions are inadequately connected to each other due to the questionable *means* by which the *de facto* true beliefs have been derived. According to the justification condition, if some argument or other justificational structure can be provided which describe why the first two conditions are properly connected, then the agent may be said to have secure indication that a known proposition is true.

On top of Lenzen's point of vagueness pertaining to the justificational issue, Gettier's *succès de scandale* counterexamples to the thesis that knowledge is justified true belief did nothing much but to make matters even worse (Gettier, 1963). Already Russell anticipated the counterexamples in the late 1940s but in the impressively short paper from 1963 Gettier explicitly demonstrated that the celebrated definition of knowledge as true justified belief is flawed – by so doing Gettier fixed the agenda for the epistemological community to this day.* Suppose Milton has strong evidence for, and is accordingly justified in believing, the proposition "Rachél owns a Ford car." Now, Milton has another friend, Angelo, of whom he does not know his current location but randomly picks San Francisco as Angelo's location just now. Given the introduction rule for $\vee$, Milton is accordingly justified in believing the proposition "Rachél owns a Ford car $\vee$ Angelo is in San Francisco." As the story goes, Rachél does not own a Ford car but unknown to Milton Angelo is by accident but in fact in San Francisco so the disjunctive proposition remains true. In other words, if Milton claims to have knowledge in terms of justified true belief he has knowledge for the wrong reasons: He claims to have knowledge because he apparently has reasons to believe that Rachél owns a Ford, which she does not, but all the same, truth is preserved because unknown to Milton, Angelo is in San Francisco. Thus Milton's original reasons for believing and being justified are undercut but on the standard definition of knowledge he may still be said to have knowledge! That's a Gettier-case.** Note that the Gettier-cases are automatically handled by the proposed definition of limiting convergent knowledge in Section 2.2.2 since there is no way the evidence for (h or h') can be undercut if the evidence entails (h or h').

To avoid the Gettier-cases some mainstream epistemologists have appealed to *reliability*. For an example, Goldman argues that the belief in some hypothesis is justified if and only if the method by which the belief is formed is reliable in

---

* I am indebted to Robert van Rooy for pointing out to me Russell's anticipation of what has later become known as the Gettier-paradoxes.

** Another counterexample originally provided by Gettier depends on the introduction rule for $\exists$ and a substitution assumption but will not be dealt with here.

the sense of producing more true convictions than false ones in the actual world.[*] Nozick insists on a "heavier" strategy by appealing to a strongly reliable recursive procedure with a halting condition imposed in all nearby worlds, while Lewis in his modal epistemology speaks of the rule of reliability and refers in a footnote to a type of nomic sufficiency account à la Armstrong. Observe that independently of what reliability is supposed to mean, *it is a criterion imposed on the inquiry method or agent* generating the belief:

The justification condition of epistemology ends up in methodology as the study of how science arrives at its posited truths, i.e., how beliefs are justified by the canons, norms or recommendations and internal workings of the method applied or agent in question.

Whatever is vague about justification has obviously to be resolved and can be resolved methodologically as Sankey recently noted:

> These are questions about the truth-conduciveness of the method. While they relate directly to the epistemic status of the method, they bear *indirectly on the nature of rational justification*. For if use of method conduces to truth, then, given the relation between method and justification, the warrant provided by the method is warrant with respect to truth. Sankey (1999: 1)

Also a mainstream epistemologist like Bonjour stresses a similar point:

> An adequate epistemological theory must establish a connection between its account of justification and its account of truth: i.e., it must be shown that justification, as viewed by that theory, is truth-conducive, that one who seeks justified beliefs is at least likely to find true ones. Bonjour (1976: 75)

Methodological recommendations, truth-conduciveness, reliability, convergence, strategies for winning games, changing your beliefs economically and reliably[**] and the like are at the very *core* of many formal epistemological proposals. Computational epistemology scrutinizes the feasibility of recommendations for getting to the truth reliably for both ideal and computationally bounded agents; game-theory models rationality among agents possibly through the use of epistemic logic; belief revision concentrates on informational economy and the agent's rational change of beliefs etc. In general what the mainstream epistemologists are looking for seems to be what the formal epistemologists have to offer: A good example of where formal epistemology meets mainstream is when Stalnaker suggests using belief revision to get out of the Gettier-cases (Stalnaker, 1996): Milton knows that *p* iff Milton believes *p* and learning no further true information would

---

[*] Sometimes Goldman does advocate a many world view but is to this day indecisive on the issue. The (possibly limiting) ratio of true beliefs over false beliefs remains at the core of his reliabilistic theory unfortunately reinviting the Gettier paradoxes which reliability was introduced to block.

[**] See Kelly et al. (1996) for a logical reliabilistic analysis of belief revision. Also Martin and Osherson discuss belief revision in the light of computational epistemology (Martin and Osherson, 1998).

lead him to change his mind. Given the standard analysis of belief revision, this analysis gives rise to system **S4.2**.⋆

This article is concluded with yet another example of the demand and supply situation taken from modal operator theory hopefully revealing the fruitful epistemological and methodological interactions between mainstream and formal theories of knowledge.

4. KK'ing Diachronically

Also before, but especially after Hintikka's publication of *Knowledge and Belief* in 1962, and to this day, epistemologists have been concerned with two philosophical themes as they relate to *Knowledge and Belief*: (1) The problem of logical omniscience, and (2) the plausibility of the KK thesis both of which are consequences of the logical epistemology advocated in the book.

The logical omniscience problem will not be dealt with here but an elegant treatment of the problem together with a new logic of knowledge and belief for agents with limited reasoning powers and a way of modelling first vs. third person perspectives on inquiry is to be found in Segerberg's recent work (Segerberg, 2003). Wójcicki's new approach may also be extended to cover similar issues (Wójcicki, 2003).

The KK plausibility is still unsettled. A mainstream epistemologist like Nozick abandons it because the agent may not be tracking the fact that he is tracking. Not having self-awareness also supports James' distinction between absolutist's philosophy and pragmatism. One may not infallibly know when one has converged to the fact that one has converged to the correct answer. Formal learning theorists like Martin and Osherson are of the same opinion:

> This does not entail that Ψ knows he knows the answer, since (as observed above) Ψ may lack any reason to believe that his hypotheses have begun to converge. Martin and Osherson (1998: 13)

In the previous section a concept of knowledge was introduced based on the idea of limiting convergence, and yet limiting convergence is often cited as one of the primary reasons for not validating the KK-thesis! But if one wants to validate the KK-thesis and simultaneously entertain a limiting concept of knowledge how is it possible to have the cake and eat it too?

Given that both tense and alethic modalities can be treated in modal operator theory together with epistemic modalities it makes sense to distinguish between two interpretations of epistemic axioms:

⋆ I am indebted to Robert van Rooy for suggesting Stalnaker's solution as an example. Hintikka embraces a similar understanding of knowledge when he explains "If somebody says I know that *p* in the strong sense, he implicitly denies that any further information would have lead him to alter his view" (Hintikka, 1962: 21). Thus changing one's mind about *p* implies that one did not know *p* to begin with.

Synchronicity: An epistemic (or doxastic or combined) axiom is *synchronic* if the consequent obtains by the very same time the antecedent obtains.

Diachronicity: An epistemic (or doxastic or combined) axiom is *diachronic* if the consequent either obtains later or would have obtained later than the antecedent even if things had been otherwise.

Malcolm's autoepistemology defends the KK-thesis synchronically from a first person perspective (Malcolm, 1952). There are other models of first person knowledge operators validating KK synchronically in particular Moore (1995), Fitting (1983) and Arló-Costa (1998). Actually these models yield a stronger logic than **S4** – they validate **S5**. On the other hand, Nozick denies KK at least synchronically from the 1st person point of view. Then there is Levi's epistemological program which essentially is a garden-variety of a first person view in which the main issue in the semantics for Levi is Ramsey's distinction between *the logic of truth* and *the logic of consistency* rather than first and third person distinctions (Levi, 1997). These two sets of distinctions are obviously related but not exactly identical. In the paper Levi argues against the validity of the KK-principle as an axiom of an epistemic logic of truth which, somewhat simplified, is tantamount to denying that KK is an axiom for a third person knowledge operator. What Levi really argues is that the KK-principle is valid as a principle regulating the consistency of a rational epistemic agent while the logic of truth for epistemic agents need not be regulated by such a principle. Lewis (1996) seems to follow suit and their underlying suggestion must in the end be that if there is a universal third person logic of knowledge, such a logic is probably rather weak.

With these two interpretations of the epistemic axioms in hand return to the question of how to have the cake and eat it too and the current line of defense for KK. To have knowledge of a hypothesis is to have reached a modulus of convergence after which the method continues to project the conjecture over all later times and possible worlds. Now, knowledge of a hypothesis h is a subset of the hypothesis h. To have knowledge of knowledge of a hypothesis h must be to reach a modulus of convergence only *after* convergence to knowledge of h has arisen. This is because knowledge of knowledge of a hypothesis is a subset of knowledge of a hypothesis, so knowledge of knowledge can only happen once knowledge of the hypothesis has obtained. Hence

$$K_\delta K_\delta h \subseteq K_\delta h \subseteq h. \tag{11}$$

One has to *force strategically* to validate KK.

Suppose the method has perfect memory and hence remembers the past evidence. Then KK becomes impossible to validate. Perfect memory demands that the method starts to force for KK at a time l in worlds prior to the modulus of convergence has been reached for mere knowledge of h which only happens at a later stage $n > l$. Forcing for KK prior to knowledge of h is impossible because
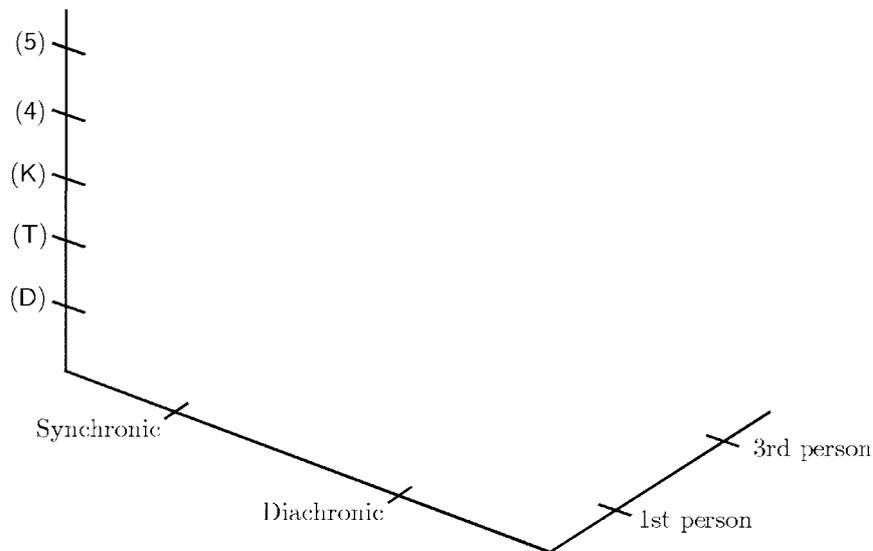
Figure 6. 3D epistemic strength.

there may exist a world λ required for KK veering off the actual world ε before the knowledge of h has arisen. Since the method has perfect memory it will attempt to "crawl" below n to get this world λ in the KK-conjecture. If the method crawls below n and captures λ, then λ will be in $K_\delta K_\delta h$ but not necessarily in $K_\delta h$ and thus violating (11) above.

Suppose on the other hand, that the method entertains *consistent expectation.* Consistent expectation implies that any additional convergence and forcing in terms of knowing that one knows h takes place at a point in time $n' > n > l$ where the method *has* converged to h and hence forces already. Adding knowledge to knowledge requires forcing and then later some more forcing.*

This is fairly easy to see if one stands outside looking in, i.e., if one adopts a third person perspective on inquiry as modal operator theory does. It is less obvious how one could stand in the epistemic environment and know that if one has consistent expectations, then one is eventually going to know that one knows when one does. The validity of KK is up to the method rather than to the world, and hinges on the diachronic interpretation of the KK thesis, but given consistent expectation, *not* unwarranted confidence in the status of one's own earlier beliefs – *the method already knows* if you look at it from the third person point of view, but not necessarily from the first.

In sum, it seems that when determining strength of epistemic operators in a way pertinent to both epistemic logic and epistemology alike – and thus construct a logical epistemology – we have to consider at least 3 dimensions besides the agent involved (Figure 6).

---

* For more detail, refer to Hendricks (2001) and Hendricks and Pedersen (2000a).

5. Conclusion

The point of introducing the two interpretations of epistemic axioms; the point of distinguishing between first and third persons perspectives on inquiry and the point of forcing are to square away confusion and misunderstandings between mainstream and formal epistemologies. To open up for profitable interactions and exchanges between the two approaches. For example, in criticizing some position, whether formal or informal, without noticing that the criticism is based on a third person perspective and the position advocated is first person may turn out to be criticizing an apple for not being an orange. Similarly if an epistemic axiom is advocated diachronically, then it is not advocated synchronically or indifferently. A mainstream epistemological paradigm may be a forcing strategy, and a formal one may be too – the question is then how they do force respectively putting them on par for comparison. Admittedly, epistemology becomes significantly more complex with these additional parameters added in and we are going to need all the help we can get from formal and mainstream epistemology alike. That is just the way it goes with active agents.

Modal operator theory adopts themes from logic, linguistics and computer science. The paradigm is about adding an active long-run perspective to epistemic logic in which agents are learning mechanisms over time merging branching alethic-temporal logic with formal learning theory. In particular, on top of the model of all branching evidence streams a formal language is introduced including epistemic modalities whose indicies are learning mechanisms, and then the idea is to look for reflections between principles in the logic and the structural features of the learning mechanisms. Modal logics are important to linguistics and computer science but so is learning obviously. Formal learning theory has its roots in computability theory, not in philosophy or epistemology, and was first really put to use in linguistics as a tool for modelling language acquisition and a child's abilities to converge to a grammar for its natural language. Modal operator theory is a setting for combining the axiomatic approach of modal logic with the learning approach of linguistics and computer science.

References

Arló-Costa, H., 1998, "Qualitative and probabilistic models of full belief," pp. 25–43 in *Proceedings of Logic Colloquim'98*, S. Buss, P. Hajék, and R. Pudlák, eds., Lecture Notes on Logic, Vol. 13, A.K. Peters, ASL.

Baltag, A., Moss, L.S., and Solecki, S., 1999, "The logic of public annoucement and common knowledge," pp. 43–56 in *Proceedings of TARK 1998*, I. Gilboa, ed., San Mateo, CA: Morgan Kaufmann.

Bonjour, L., 1976, "The coherence theory of empirical knowledge," *Philosophical Studies* **30**, 281–312. Reprinted as pp. 70–89 in *Contemporary Readings in Epistemology*, M.F. Goodman and R.A. Snyder, eds., Englewood Cliffs, NJ: Prentice Hall.

Boyd, R., 1984, "The current state of scientific realism," pp. 41–82 in *Scientific Realism*, J. Leplin, ed., Berkeley, CA: University of California Press.

Braüner, T., Hasle, P., and Øhrstrøm, P., 1998, "Determinism and the origins of temporal logic," pp. 116–133 in *Proceedings of the Second International Conference on Temporal Logic*, H. Barringer, M. Fisher, D. Gabbay, and G. Gough, eds., Applied Logic Series. Dordrecht: Kluwer Academic Publishers.

Ditmarch, H.v., van der Hoek, W., and Kooi, B., 2003, "On epistemic action languages," in *Knowledge Contributors*, V.F. Hendricks, S.A. Pedersen, and K.F. Jørgensen, eds., Dordrecht: Kluwer Academic Publishers, forthcoming.

Dretske, F., 1970, "Epistemic operators," *The Journal of Philosophy* **67**, 1007–1022.

Earman, J., 1992, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, Cambridge, MA: MIT Press.

Fagin, R., Halpern, J.Y., Moses, Y., and Vardi, M.Y., 1995, *Reasoning about Knowledge*, Cambridge, MA: MIT Press.

Fagin, R., Halpern, J.Y., Moses, Y. and Vardi, M.Y., 1996, "Common knowledge revisited," *Annals of Pure and Applied Logic* **96**, 89–105.

Fitting, M., 1983, "Moore's non-monotonic logic and **S5**," CUNY.

Gettier, E., 1963, "Is knowledge justified true relief?," *Analysis* **23**, 121–123.

Gochet, P. and Gribomont, P., 2003, "Epistemic logic," in *Handbook of the History and Philosophy of Logic*, D.M. Gabbay and J. Woods, eds., Amsterdam: Elsevier, to appear.

Gold, E.M., 1967, "Language identification in the limit," *Information and Control* **10**, 447–474.

Goldman, A., 1979, "What is justified true belief," pp. 1–23 in *Justification and Knowledge*, G.S. Pappas, ed., Dordrecht: D. Reidel.

Goldman, A., 1986, *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.

Goldman, A., 1992, "Epistemic folkways and scientific epistemology," pp. 155–175 in *Liaisons: Philosophy Meets Cognitive and Social Sciences*, A. Goldman, ed., Cambridge, MA: MIT Press.

Hendricks, V.F., 2001, *The Convergenc of Scientific Knowledge – A View from the Limit*, Studia Logica Library Series: Trends in Logic, Vol. 9, Dordrecht: Kluwer Academic Publishers.

Hendricks, V.F., 2002, "Active agents," ΦNEWS **2**, 5–39.

Hendricks, V.F., 2003, *FORCING Epistemology*, to appear.

Hendricks, V.F. and Pedersen, S.A., 1998, "Assessment and discovery in the limit scientific inquiry," in *Philosophical Dimensions of the Unity of Science – Proceedings of the XI International Congress on Logic, Methodology and the Philosophy of Science*, J. Cachro, S. Hanuszewicz, G. Kurczewski, and A. Rojszczak, eds., Dordrecht: Kluwer Academic Publishers, to appear.

Hendricks, V.F. and Pedersen, S.A., 2000a, "KK-ing diachronically," presented at *LC2000 – European Congress of the Association of Symbolic Logic*, Paris, France, July 23–31, 2000.

Hendricks, V.F. and Pedersen, S.A., 2000b, "The companion to modal operator theory – A program in philosophy online," Department of Philosophy and Science Studies, Roskilde University, Denmark, http://www.mot.ruc.dk

Hendricks, V.F. and Pedersen, S.A., 2004, *Operators in Philosophy of Science*, in preparation.

Hendricks, V.F., Pedersen, S.A. and Jørgensen, K.F., eds., 2003, *Knowledge Contributors*, Synthese Library Series (special volume), Dordrecht: Kluwer Academic Publishers, forthcoming.

Hintikka, J., 1962, *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell: Cornell University Press.

Hintikka, J., 2003, "A second generation epistemic logic and its general significance," in *Knowledge Contributors*, V.F. Hendricks, S.A. Pedersen, and K.F. Jørgensen, eds., Dordrecht: Kluwer Academic Publishers, forthcoming.

Hintikka, J. and Halonen, I., 1998, "Epistemic logic," *Routledge Encyclopedia of Philosophy*, Volume 1, London: Routledge.

Hintikka, J. and Hintikka, M., 1988, *The Logic of Epistemology and the Epistemology of Logic: Selected Essays*, Dordrecht: Kluwer Academic Publishers.

Howson, C. and Urbach, P., 1989, *Scientific Reasoning: The Bayesian Approach*, La Salle, IL: Open Court.

James, W., 1960, "The will to believe," *Essays in Pragmatism*, Hafner Publishing Company.

Kelly, K., 1996, *The Logic of Reliable Inquiry*, New York: Oxford University Press.

Kelly, K., Schulte, O., and Hendricks, V.F., 1996, "Reliable belief revision," pp. 157–175 in *Logic, Methodology, and Philosophy of Science X*, M.L.D. Chiara, ed., Dordrecht: Kluwer Academic Publishers.

Lenzen, W., 1978, "Recent work in epistemic logic," *Acta Philosophica Fennica* **30**, 1–219.

Lenzen, W., 2003, "Knowledge, belief, and subjective probability – Outlines of a unified system of epistemic/doxastic logic," in *Knowledge Contributors*, V.F. Hendricks, S.A. Pedersen, and K.F. Jørgensen, eds., Dordrecht: Kluwer Academic Publishers, forthcoming.

Levi, I., 1997, "The logic of full belief," pp. 40–69 in *The Covenant of Reason: Rationality and the Commitments of Thought*, A.K. Peters, ed., Cambridge: Cambridge University Press.

Lewis, D., 1996, "Elusive knowledge," *Australasian Journal of Philosophy* **7**, 549–567.

Malcolm, N., 1952, "Knowledge and belief," *Mind* **LXI**, 242. Reprinted as pp. 272–279 in *Contemporary Readings in Epistemology*, M.F. Goodman and R.A. Snyder, eds., Englewood Cliffs, NJ: Prentice Hall (1993).

Martin, E. and Osherson, D., 1998, *Elements of Scientific Inquiry*, Cambridge, MA: MIT Press.

Meyer, J.-J.Ch. and van der Hoek, W., 1995, *Epistemic Logic for AI and Computer Science*, Cambridge Tracts in Theoretical Computer Science, Cambridge: Cambridge University Press.

Moore, R.C., 1995, *Logic and Represenation*, CSLI Lecture Notes, Vol. 39, Stanford, CA: CSLI.

Nozick, R., 1981, *Philosophical Explanations*, Cambridge, MA: Harvard University Press.

Nozick, R., 1993, *The Nature of Rationality*, Cambridge, MA: Harvard University Press.

Peirce, C.S., 1958, *Charles S. Peirce: Selected Writings*, P. Wiener, ed., New York: Dover Publications.

Rott, H., 2003, "Interpreting belief revision: What is 'economic' in the theory of belief change," in *Knowledge Contributors*, V.F. Hendricks, S.A. Pedersen, and K.F. Jørgensen, eds., Dordrecht: Kluwer Academic Publishers, forthcoming.

Sankey, H., 1999, "A realist approach to method and truth," Lecture delivered to the Department of Philosophy and Science Studies, Roskilde University, Denmark, May 14.

Scott, D., 1970, "Advice on modal logic," pp. 143–173 in *Philosophical Problems in Logic*, K. Lambert, ed., Dordrecht: D. Reidel.

Segerberg, K., 2003, "Two attempts at deconstructing epistemic logic," in *Knowledge Contributors*, V.F. Hendricks, S.A. Pedersen, and K.F. Jørgensen, eds., Dordrecht: Kluwer Academic Publishers, forthcoming.

Sowa, J.F., 2000, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Pacific Grove, CA: Brooks/Cole.

Sowa, J.F., 2003, "Laws, facts, and contexts: Foundations for multimodal reasoning," in *Knowledge Contributors*, V.F. Hendricks, S.A. Pedersen, and K.F. Jørgensen, eds., Dordrecht: Kluwer Academic Publishers, forthcoming.

Stalnaker, R., 1996, "Knowledge, belief and counterfactual reasoning in games," *Economics and Philosophy* **12**, 133–163.

van Fraassen, B., 1980, *The Scientific Image*, New York: Oxford University Press.

Wójcicki, R., 2003, "Referential semantics," in *Knowledge Contributors*, V.F. Hendricks, S.A. Pedersen, and K.F. Jørgensen, eds., Dordrecht: Kluwer Academic Publishers, forthcoming.

Zanardo, A., 1996, "Branching-time logic with quantification over branches: The point of view of modal logic," *Journal of Symbolic Logic* **61**, 1–39.