Agency and Interaction in Formal Epistemology

Vincent F. Hendricks
Department of Philosophy / MEF
University of Copenhagen
Denmark

Department of Philosophy
Columbia University
New York / USA

CPH / August 2010

1   Formal Epistemology

- Formal epistemology is a fairly recent field of study in philosophy dating back only a decade or so.

- This is not to say that formal epistemological studies have not been conducted prior to the late 1990's, but rather that the term introduced to cover the philosophical enterprise was coined around this time. Predecessors to the discipline include Carnap, Hintikka, Levi, Lewis, Putnam, Quine and other high-ranking officials in formal philosophy.

- Formal epistemology denotes the formal study of crucial concepts in general or mainstream epistemology including knowledge, belief (-change), certainty, rationality, reasoning, decision, justification, learning, agent interaction and information processing.

2 Agency and Interaction

- The point of departure is rooted in two philosophically fundamental and interrelated notions central to formal epistemology [Helzner & Hendricks 10, 12];

    - **agency** – what agents are, and

    - **interaction** – what agents do.

- Agents may be individuals, or they may be groups of individuals working together.

- In formal epistemology across the board various assumptions may be made concerning the relevant features of the agents at issue.

- Relevant features may include the agent's beliefs about its environment, its desires concerning various possibilities, the methods it employs in learning about its environment, and the strategies it adopts in its interactions with other agents in its environment.

- Fixing these features serves to bound investigations concerning interactions between the agent and its environment.

  - The agent's beliefs and desires are assumed to inform its decisions.

  - Methods employed by the agent for the purposes of learning are assumed to track or approximate or converge upon the facts of the agent's environment.

  - Strategies adopted by the agent are assumed to be effective in some sense.

3 AI Methodologies

- Epistemic Logic ⟵

- Interactive Epistemology and Game Theory

- Probability Theory

- Bayesian Epistemology

- Belief Revision Theory

- Decision Theory

- Computational Epistemology (Formal learning theory) ⟵

4 Active Agency

1. 'Agent' comes from the Latin term *agere* meaning 'to set in motion, to do, to conduct, to act'.

2. 'Agency' means 'the acting of an agent' in particular in presence of other agents.

3. An agent may interact or negotiate with its *environment* and/or with *other agents*.

4. An agent may make decisions, follow strategies or methodological recommendations, have preferences, learn, revise beliefs ... call these *agent agendas*.

5. Active Agency = Agents + Agendas

5 Modal Operator Epistemology

Modal operator epistemology is the the cocktail obtained by mixing formal learning theory and epistemic logic in order to study the formal properties of limiting convergence knowledge.

- *The Convergence of Scientific Knowledge*. Dordrecht: Springer, 2001

- *Mainstream and Formal Epistemology*. New York: Cambridge University Press, 2007.

- *Agency and Interaction* [with Jeff Helzner]. New York: Cambridge University Press, 2012.

- + papers [Hendricks 2002—2010].

5.1 Worlds

- An evidence stream ε is an ω-sequence of natural numbers, $i.\ e.,\ \varepsilon \in \omega^\omega$.

- A possible world has the form (ε, n) such that $\varepsilon \in \omega^\omega$ and $n \in \omega$.

- The set of all possible worlds $\mathcal{W} = \{(\varepsilon, n) \mid \varepsilon \in \omega^\omega, n \in \omega\}$.

- $\varepsilon \mid n$ denotes the finite initial segment of evidence stream ε of length n.

- Define $\omega^{<\omega}$ to be the set of all finite initial segments of elements in ω.

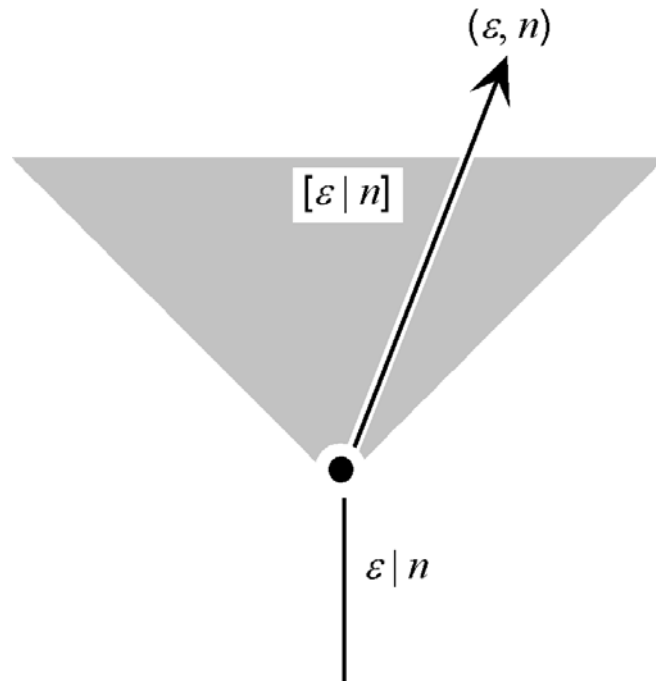- Let $(\varepsilon \mid n)$ denote the set of all infinite evidence streams that extends $\varepsilon \mid n$.

Figure 1: Handle of evidence and fan of worlds

- The set of possible worlds in the fan, i.e. background knowledge, is defined as

$$[\varepsilon \mid n] = (\varepsilon \mid n) \times \omega.$$

5.2 Hypotheses

Hypotheses will be identified with sets of possible worlds. Define the set of all simple empirical hypotheses

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

A 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.
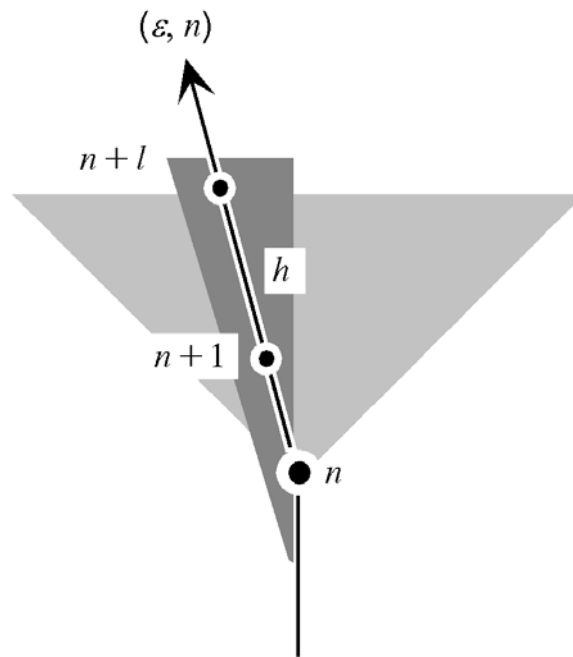
Figure 2: Truth of a hypothesis h in a possible world (ε, n).

5.3 Agents and Inquiry Methods

An inquiry method (or agent) may be either one of discovery or assessment:

A discovery method δ is a function from finite initial segments of evidence to hypotheses, i.e.

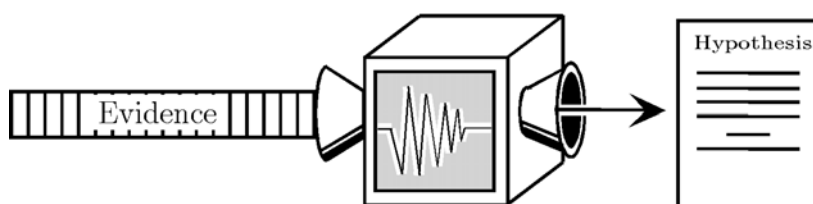$$\delta : \omega^{<\omega} \longrightarrow \mathcal{H}. \tag{1}$$

Figure 3: Discovery method.

The convergence modulus for a *discovery* method (abbreviated cm) accordingly:

Definition 1 $cm(\delta, h, [\varepsilon \mid n]) = \mu k \forall n' \geq k \forall (\tau, n') \in [\varepsilon \mid n] : \delta(\tau \mid n') \subseteq h.$

An assessment method α is a function from finite initial segments of evidence and hypotheses to true/false, i.e.

$$\alpha : \omega^{<\omega} \times \mathcal{H} \longrightarrow \{0, 1\} \tag{2}$$
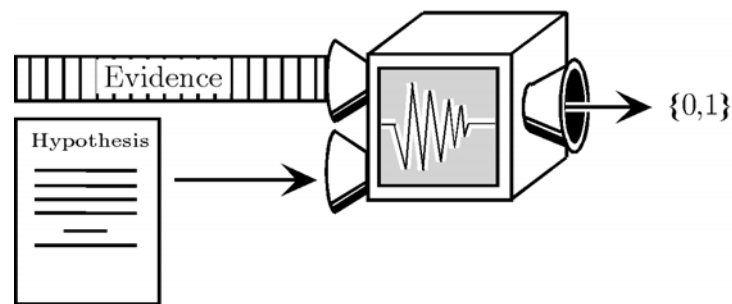

Figure 4: Assessment method.

The convergence modulus for an assessment is defined in the following way:

Definition 2 $cm(\alpha, h, [\varepsilon \mid n]) = \mu k \geq n, \forall n' \geq k, \forall (\tau, n') \in [\varepsilon \mid n] : \alpha(h, \varepsilon \mid n) = \alpha(h, \tau \mid n')$.

5.4 Knowledge Based on Discovery

(ε, n) validates $K_\delta h$ iff

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

2. $\forall n' \geq n, \forall (\tau, n') \in [\varepsilon \mid n] : \delta(\tau \mid n') \subseteq h$.

The discovery method may additionally be subject to certain agendas (methodological recommendations) like

- perfect memory

- consistency

- infallibilility etc.

5.5 Knowledge Based on Assessment

(ε, n) validates $K_\alpha h$ iff

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

2. α *decides* h *in the limit in* $[\varepsilon \mid n]$:

   (a) if $(\varepsilon, n) \in h$ and $\forall l \in \omega : (\varepsilon, n + l) \in h$ *then* $\exists k \geq n, \forall n' \geq k, \forall (\tau, n') \in [\varepsilon \mid n] : \alpha(h, \tau \mid n') = 1,$

   (b) if $(\varepsilon, n) \notin h$ or $\exists l \in \omega : (\varepsilon, n + l) \notin h$ *then* $\exists k \geq n, \forall n' \geq k, \forall (\tau, n') \in [\varepsilon \mid n] : \alpha(h, \tau \mid n') = 0.$

6 Multi-Modal Systems

The above set-theoretical characterization of inquiry lends itself to a multi-modal logic. The modal language $\mathcal{L}$ is defined accordingly:

$$A ::=$$

$$\mid a \mid A \wedge B \mid \neg A \mid K_\delta A \mid K_\alpha A \mid [A!]B \mid I_\delta A \mid I_\alpha A$$

Operators for alethic as well as tense may also me added to the language.

Definition 3 *Model*

A model $\mathbb{M} = <\mathcal{W}, \varphi, \delta, \alpha>$ *consists of:*

1. *A non-empty set of* possible worlds $\mathcal{W}$,

2. *A* denotation function φ :*Proposition Letters* $\longrightarrow$ $P(\mathcal{W})$,*i. e.,* $\varphi(a) \subseteq \mathcal{W}$.

3. Inquiry methods

   (a) $\delta : \omega^{<\omega} \longrightarrow P(\mathcal{W})$

   (b) $\alpha : \omega^{<\omega} \times \mathcal{H} \longrightarrow \{0, 1\}$

Definition 4 *Truth Conditions*

Let $\varphi_{\mathbb{M},(\varepsilon,n)}(A)$ denote the truth value in (ε,n) of a modal formula A given $\mathbb{M}$, defined by recursion through the following clauses:

1. $\varphi_{\mathbb{M},(\varepsilon,n)}(a) = 1 \ iff \ (\varepsilon,n) \in \varphi(a) \ and \ \forall l \in \omega :$
   $(\varepsilon, n+l) \in \varphi(a)$
   $for \ all \ propositional \ variables \ a, b, c, \dots ,\ .$

2. $\varphi_{\mathbb{M},(\varepsilon,n)}(\neg A) = 1 \ iff \ \varphi_{\mathbb{M},(\varepsilon,n)}(A) = 0,$

3. $\varphi_{\mathbb{M},(\varepsilon,n)}(A \wedge B) = 1 \ iff \ both \ \varphi_{\mathbb{M},(\varepsilon,n)}(A) = 1 \ and \ \varphi_{\mathbb{M},(\varepsilon,n)}(B) = 1;$
   $otherwise \ \varphi_{\mathbb{M},(\varepsilon,n)}(A \wedge B) = 0.$

4. $\varphi_{\mathbb{M},(\varepsilon,n)}(K_\delta A) = 1$ iff

   (a) $(\varepsilon, n) \in [A]_{\mathbb{M}}$ and $\forall l \in \omega : (\varepsilon, n + l) \in [A]_{\mathbb{M}}$,

   (b) $\forall n' \geq n, \forall (\tau, n') \in [\varepsilon \mid n] : \delta(\tau \mid n') \subseteq [A]_{\mathbb{M}}$

5. $\varphi_{\mathbb{M},(\varepsilon,n)}([A!]B) = 1$ iff
   if $\varphi_{\mathbb{M},(\varepsilon,n)}(A) = 1,$ $then$ $\varphi_{\mathbb{M},(\varepsilon,n)|A}(B) = 1.$

6. $\varphi_{\mathbb{M},(\varepsilon,n)}(I_\Xi A) = 1$ iff $\exists(\tau, m)\exists(\mu, m') \in [\varepsilon \mid n] : \tau \mid n = \mu \mid n$ and $\varphi_{\mathbb{M},(\tau,m)}(A) = 1$ and $\varphi_{\mathbb{M},(\mu,m')}(\neg A) = 1$ for $\Xi \in \{\delta, \alpha\}$.

6.1 Results

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?*

Theorem 1 *If knowledge is defined as limiting convergence, then knowledge validates* **S4** *iff the discovery method / assessment method is subject to certain methodological constraints.*

Many other results have been obtained pertaining to knowledge acquisition over time, the interplay between knowledge acquisition and agendas etc.

7   Transmissibility and Agendas

Already in *Knowledge and Belief* from Hintikka considered whether

$$K_\beta K_\gamma A \rightarrow K_\beta A \qquad (3)$$

is valid (or self-sustainable in Hintikka's terminology) for arbitrary agents β, γ.

Now 3 is simply an iterated version of Axiom **T** for different agents and as long β, γ index the same accessability relation the claim is straightforward to demonstrate.

From an active agent perspective the claim is less obvious.

The reason is agenda-driven or methodological. Inquiry methods β, γ may − or may not − be of the same type:

1. Again a **discovery method** δ is a function from finite initial segments of evidence to hypotheses, i.e. $\delta : \omega^{<\omega} \longrightarrow \mathcal{H}$

2. Again an **assessment method** α is a function from finite initial segments of evidence and hypotheses to true/false, i.e. $\alpha : \omega^{<\omega} \times \mathcal{H} \longrightarrow \{0, 1\}$

If knowledge is subsequently defined either on discovery or assessment, then 3 is not immediately valid unless discovery and assessment methods can "mimic" or induce eachothers' behavior in the following way:

Theorem 2 *If a discovery method δ discovers h in a possible world (ε, n) in the limit, then there exists a limiting assessment method α which verifies h in (ε, n) in the limit.*

Proof. 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:

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

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

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

∎

Similarly, but conversely:

Theorem 3 *If an assessment method α verifies h in (ε, n) in the limit, then there exists a limiting discovery method δ which discovers h in (ε, n) in the limit.*

Proof. Similar construction as in proof of Theorem 2. ∎

Using inducement it is easily shown that

Theorem 4 $K_\delta A \leftrightarrow K_\alpha A$.

The theorem can be easily proved since theorems 2, 3 and 4 provide the assurance that a discovery method can do whatever an assessment method can do and vice versa:

$$\vdash K_\delta K_\alpha A \to K_\delta A$$

(i)	$K_\alpha A \to A$	Axiom **T**
(ii)	$K_\delta(K_\alpha A \to A)$	$(i), (N)$
(iii)	$K_\delta(K_\alpha A \to A) \to (K_\delta K_\alpha A \to K_\delta A)$	$(ii),$ Axiom **K**
(iv)	$K_\delta K_\alpha A \to K_\delta A$	$(ii), (iii), (MP)$

Let there be given a finite set of discovery agents $\Delta = \{\delta_1, \delta_2, \delta_3, ..., \delta_n\}$, a finite set of assessment agents $\Lambda = \{\alpha_1, \alpha_2, \alpha_3, ..., \alpha_n\}$ and let theorem 1 hold for all agents in Δ, Λ. Now it may be shown that 3 holds for agents of different types:

Theorem 5 $\forall \delta_i \in \Delta : K_{\delta_i} K_\alpha A \rightarrow K_{\delta_i} A$
if theorem 2 holds.

Theorem 6 $\forall \alpha_i \in \Lambda : K_{\alpha_i} K_\delta A \rightarrow K_{\alpha_i} A$
if theorem 3 holds.

8 Public Annoucement

The next two theorems show that the axiom relating public annoucement to knowledge given the standard axiomatization of public annoucement logic with common knowldge holds for knowledge based on discovery and knowledge based on assessment.

Theorem 7 $\forall \delta_i \in \Delta : [K_\alpha A!]K_{\delta_i}B \leftrightarrow (K_\alpha A \rightarrow (K_{\delta_i}K_\alpha A \rightarrow [K_\alpha A!]B))$ *if theorem 2 holds.*

Theorem 8 $\forall \alpha_i \in \Lambda : [K_\delta A!]K_{\alpha_i}B \leftrightarrow (K_\delta A \rightarrow (K_{\alpha_i}[K_\delta A \rightarrow [K_\delta A!]B]))$ *if theorem 3 holds.*

This is a variation of the original knowledge prediction axiom which states that "some a knows B after an announcement A iff (if A is true, a knows that after the announcement of A, B will be the case)":

9 Pluralistic Ignorance

Q: What is the clock-frequency on the bus?

A: I have no idea!

Q: Well it would be good to know now that you are selling the product, no?

A: Listen, I don't think you can find any of my co-workers either that would know!

And then I got really angry with the guy behind the counter ...

- The phenomenon appears when a group of decision-makers have to act or believe at the same time given a public signal.

- Example: Starting up a new philosophy class.

- Pluralistic ignorance arises when the individual decision-maker in a group lacks the necessary information for solving a problem at hand, and thus observes others hoping for more information.

- When everybody else does the same, everybody observes the lack of reaction and is consequently lead to erroneous beliefs.

- We all remain ignorant.

- But ignorance is fragile – *The Emperor's New Clothes*

9.1 Ingredients of Pluralistic Ignorance

1. A finite set ignorant agents either based on discovery of assessment or both:

   (a) $\Delta = \{\delta_1, \delta_2, \delta_3, ..., \delta_n\}$

   (b) $\Lambda = \{\alpha_1, \alpha_2, \alpha_3, ..., \alpha_n\}$

2. A public annoucement:

   (a) $[A!]B$

3. At least one knowing agent based on either discovery or assessment:

   (a) $K_\alpha A$

   (b) $K_\delta A$

4. Inducement theorems 2 and 3.

9.2 Resolving Pluralistic Ignorance using Knowledge Transmissibility

- **Theorem (A):**
  $$\forall \delta_i \in \Delta : [I_{\delta_i} A \wedge [K_\alpha A!] A] \rightarrow [K_\alpha A!] K_{\delta_i} A$$
  if theorem 2 holds.

- **Theorem (B):**
  $$\forall \alpha_i \in \Lambda : [I_{\alpha_i} A \wedge [K_\delta A!] A] \rightarrow [K_\delta A!] K_{\alpha_i} A$$
  if theorem 3 holds.

- In plain words theorem (A) says that if

- it holds for all agents $\delta_i \in \Delta$ that they are ignorant of A and

- that after it has been publicly annouced that α knows A, then A is the case, then

- after it has been publicly annouced that α knows A,

- α's knowledge of A will be transferred to every $\delta_i \in \Delta$

- provided that every $\delta_i \in \Delta$ can mimic α's epistemic behavior given the public annoucement based on theorem 2.

And similarly for theorem (B)

10 NEW WAYS TO GO

- The former law professor at Harvard, Cass Sunstein, and his collaborators have empirically studied a host of social epistemic phenomena besides pluralistic ignorance:

  - **Informational cascades**: An informational cascade occurs when people observe the actions of others and then make the same choice that the others have made, independently of their own private information signals. This can sometimes lead to error when you override your own correct evidence just to conform to others.

  - **Belief polarization**: Belief polarization is a phenomenon in which a disagreement becomes more extreme as the different parties consider evidence on the issue. It is one of the effects of confirmation bias: the tendency of people to search for and interpret evidence selectively, to reinforce their current beliefs or attitudes.

- **Believing false rumors**: He said that, that she said, that John knows, that ...

- ... for more, see for example Sunstein's book, *Going to Extremes: How Like Minds Unite and Divide*, OUP 2009.

- Between (dynamic) epistemic logic, interactive epistemology, decision theory, belief revision theory, probability theory and credence etc. we have the necessary formal machinery to analyze, model, simulate and resolve a host of these phenenomena and then check the results against extensive empirical material.

- So if you are fishing for a PhD- or research project, here is a pond to try ...