A Hyper-intensional Learning Semantics for Inductive Empirical Knowledge

Kevin T. Kelly

September 2, 2013

Abstract

This paper presents a new semantics for inductive empirical knowledge. The epistemic agent is represented concretely as a learner who processes new inputs through time and who forms new beliefs from those inputs by means of a concrete, computable learning program. The agent's belief state is represented hyper-intensionally as a set of time-indexed sentences. Knowledge is interpreted as avoidance of error in the limit and as having converged to true belief from the present time onward. Familiar topics are re-examined within the semantics, such as inductive skepticism, the logic of discovery, Duhem's problem and the articulation of theories by auxiliary hypotheses, the role of serendipity in scientific knowledge, Fitch's paradox and deductive closure of knowability, whether one can know inductively that one knows inductively, whether one can know inductively that one does not know inductively, and whether expert instruction can spread common inductive knowledge through a community, rather than merely exhibit it.

1 Introduction

Science formulates general theories. Can such theories count as knowledge, or are they doomed perpetually to the status of *mere* theories, as the anti-scientific fringe perennially urges? The ancient argument for inductive skepticism urges the latter view: no finite sequence of observations can rule out the (relevant) possibility of future surprises, so universal laws and theories are unknowable.

A popular strategy for responding to skeptical arguments appeals to possible world semantics for subjunctive conditionals. The idea works well in the case of "ultimate" brain-in-a-vat skepticism: if you are normally looking at a cat on a mat, you wouldn't hallucinate the cat if it weren't there, so your belief is be *sensitive* to the truth (Nozick 1981, Roush 2007). Or if you were to believe that there is a cat on the mat, most worlds in which you do so are remote worlds involving systematic hallucinations, so your belief is *safe*

(Sosa 1999, Pritchard 2007, Williamson 2000). But does either idea extend to inductive skepticism regarding laws and theories? If the true law were not $Y = bX+a$, would science have noticed *already*? Are all worlds in which the truth is $Y = cX^2+bX+a$ safely bounded away from $Y = bX + a$ worlds in terms of similarity—regardless how small the coefficient c is?[1] One may fiddle with similarity metrics to obtain almost any desired result, but similarity metrics are ultimately supposed to explain usage concerning counterfactual conditionals, and it remains awkward to deny that science *might* be subject in the future to more of the sorts of revolutionary surprises it has encountered in the past, even with respect to its best-tested theories (Laudan 1981). The best one can expect of even ideally diligent, ongoing scientific inquiry, it seems, is that it roots out error *eventually*. Perhaps allowance for a time lag between the onset of knowledge and error-detection is essential for knowledge of universal laws and theories.

There is a venerable tradition, expounded by Peirce (1878), James (1898), Reichenbach (1949), Carnap (1945), Putnam (1963), and Gold (1967) and subsequently developed by computer scientists and cognitive scientists into a body of work known as formal learning theory (Jain et al. 1999), which models the epistemic agent as a *learner* who processes information through time and who stabilizes, eventually, to true, inductive beliefs. Inductive learning is a matter of finding the truth eventually and of avoiding error eventually. It is natural to think of inductive knowledge that ϕ as having learned that ϕ. Having learned that ϕ implies that one has actually stabilized to true belief that ϕ and that one would have *converged to true belief* whether ϕ otherwise. The proposed semantics is more lenient—one has knowledge that ϕ if and only if one has actually converged to true belief that ϕ (as in having learned) and one would have *avoided error* whether ϕ otherwise—one might simply suspend belief forever if the data are so unexpected that one no longer knows what is going on. Allowance for suspension of belief is more plausible sociologically, and it also turns out to be essential if the consequences of known theories are to be themselves knowable by the same standard.[2]

The proposed semantics is not presented as a definitive analysis of inductive knowledge in the traditional, exacting sense. It is intended merely to be the best available logical framework for representing interactions between inquiry and inductive knowledge and to cast a unifying light on some familiar issues in epistemology and the philosophy of science. Such issues include: how learning produces inductive knowledge, how deductive inference produces new inductive knowledge from old, how inductive knowledge can thrive in a morass of inconsistency, why scientific knowledge should allow for a certain kind of luck or "serendipity", how one can know that one knows, why one can't know that one doesn't

[1]Nozick (1981) and Roush (2007) defend the idea that we would have noticed the failure of known laws already because, if a given uniformity weren't true, some distinct uniformity would have been. But in the polynomial example, all the regularities are law-like. Vogel (1987) presents additional objections to tracking as an adequate account of inductive knowledge.

[2]Alternatively, one could simply stipulate that the deductive consequences of inductive knowledge are known, but then one would have no explanation why or how they are known, aside from the stipulation.

know, the invalidity of Fitch's (1963) paradox of unknowability for inductive knowledge, how students can acquire knowledge without being able to second-guess their teachers, and how common scientific knowledge can emerge in a population of mostly non-experts. One common thread running through the following results is *epistemic parasitism.* Deduction can generate new inductive knowledge from old, a given agent can know inductively that she knows, and education can transfer inductive knowledge from an expert to her pupils but only if the new knowledge is mindlessly parasitic on the old—independent reasons or insistence upon consistency between alternative lines of research can as easily destroy knowledge as further it. Another recurrent theme is how hazardous and misleading it can be to rely upon over-simplified logical languages and models that license unbounded iteration of plausible but vaguely understood axioms.

Inclusion of the entire learning process within models of epistemic logic is consonant with the current trend in epistemic logic (van Benthem 2011) toward more dynamic and potentially explanatory modeling of the agent. Recently, there have been explicit studies of truth tracking and safety analyses of knowledge (Holliday 2013) and of inductive learning within a modal logical framework (Gierasimczuk 2010). Hendricks (2001) proposed to develop learning models for inductive knowledge, itself. This paper carries that proposal to fruition.[3] Johan van Benthem has encouraged the development of connections between learning theory and epistemic logic for decades, so it is a particular pleasure to contribute this study to his festschrift.

2 Syntax

Let $G = \{1, \ldots, N\}$ be indices for a group of N individuals. Let $\mathbf{L}_{\mathsf{atom}} = \{\mathsf{p}_i : i \in \mathbb{N}\}$ be a countable collection of atomic sentences. Define the modal language $\mathbf{L}_{\mathsf{BIT}}$ in the usual way with classical connective $\rightarrow$ and the following modal operators, where Δ is a finite

[3] Some of the underlying ideas were presented informally in (Kelly 2001).

set of sentences.[4]

$$
\begin{array}{rcl}
\bot & : & \textit{contradiction}; \\
[\dot{\mathsf{F}}]\,\phi & : & \text{it is } \textit{henceforth} \text{ the case that } \phi; \\
[\mathsf{F}]\,\phi & : & \text{it is } \textit{henceforth} \text{ the case that } \phi \text{ from now}; \\
\mathsf{N}\,\phi & : & \phi \text{ is the case } \textit{now}; \\
@_t\,\phi & : & \phi \text{ is the case } \textit{at} \text{ time } t; \\
[\,\mathsf{I}\,]_i\,\phi & : & \textit{information} \text{ is available to } i \text{ that } \phi \text{ is now the case}; \\
[\mathsf{D}]_i\,\phi & : & \textit{information} \text{ is available to } i \text{ that what } \phi \text{ says now is} \\
 & & \quad \text{determined to be the case by the learning program } i \text{ has now}; \\
[\mathsf{B}]_i\,\phi & : & \text{the learning method of } i \textit{ believes} \text{ that what } \phi \text{ says now is true}; \\
\mathsf{S}_i\Delta & : & \Delta \text{ is } \textit{doxastically stable} \text{ for } i; \\
[\mathsf{M}]_i\,\phi & : & \text{it is } \textit{methodologically necessary} \text{ for } i \text{ that } \phi \text{ is the case now}; \\
\psi\ \langle\mathsf{MD}]\!\!\rightarrow_{i,\Delta}\ \phi & : & \text{it is } \textit{methodologically feasible} \text{ for } i \text{ to determine that } \phi \text{ is the case,} \\
 & & \quad \text{given that } \psi \text{ is the case, and to do so in a way that holds} \\
 & & \quad i\text{'s belief whether } \Delta \text{ fixed.}
\end{array}
$$

Let $\mathbf{L}_{@\mathsf{BIT}}$ denote the set of all $\mathbf{L}_{\mathsf{BIT}}$ sentences that are prefixed by an operator $@_t$ for some $t \in \mathbb{N}$, so $\mathbf{L}_{@\mathsf{BIT}} \subseteq \mathbf{L}_{\mathsf{BIT}}$.

Extend $\mathbf{L}_{\mathsf{BIT}}$ with definitions as follows. The classical connectives $\neg, \wedge, \vee$, and $\top$ are definable in the usual way. For each box operator $[\mathsf{X}]_i$ listed above or defined later, assume that the *dual* operator is defined as follows:

$$\langle\mathsf{X}\rangle_i\,\phi \ := \ \neg[\mathsf{X}]_i\neg\phi;$$

Define the standard notation:

$$
\begin{array}{rcl}
\mathsf{B}\,\phi & := & [\mathsf{B}]_i\,\phi; \\
\mathsf{G}\,\phi & := & [\mathsf{F}]\,\phi; \\
\mathsf{F}\,\phi & := & \langle\mathsf{F}\rangle\ \phi;
\end{array}
$$

and similarly for $\dot{\mathsf{F}}$, $\dot{\mathsf{G}}$. Introduce the abbreviations:

$$
\begin{array}{rcl}
\mathsf{S}_i\,\delta & := & \mathsf{S}_i\{\delta\}; \\
\psi\,\langle\mathsf{MD}]\!\!\rightarrow_{i,\delta}\ \phi & := & \psi\,\langle\mathsf{MD}]\!\!\rightarrow_{i,\{\delta\}}\ \phi; \\
\langle\mathsf{MD}]_i\ \phi & := & \top\,\langle\mathsf{MD}]\!\!\rightarrow_{i,\emptyset}\ \phi.
\end{array}
$$

[4]The rich base language is consonant with D. Scott's (1970) advice to seek epistemic principles in interactions between operators.

When Γ, Δ are finite subsets of $\mathbf{L}_{\mathsf{BIT}}$ and X_i is an arbitrary modal operator, let:

$$\begin{aligned} \mathsf{X}_i\Gamma &:= \bigwedge_{\gamma\in\Gamma} \mathsf{X}_i\,\gamma; \\ \Delta \to \Gamma &:= \bigwedge_{\delta\in\Delta} \delta \to \bigwedge_{\gamma\in\Gamma} \gamma; \end{aligned}$$

3 Computational Learning Models

Let E denote the set of possible *external worlds*. In a Kantian spirit, learning semantics imposes no structure or restrictions whatever on E. Let $T = \mathbb{N}$ be interpreted as discrete *stages of inquiry*. Let $G = \{1, \ldots, N\}$ be interpreted as a finite set of *agents*. Agent $i \in G$ is assumed to have some overall, discrete, physical sensory or receptive state that will be called the agent's current *input* state. Think of $S = \mathbb{N}$ as code numbers for those states. Sensory states are not assumed to have propositional meanings (they are never assigned truth values) but their occurrence makes propositional information *available*. Let S^* be the set of all finite sequences of input states, so each $\sigma \in S^*$ is a possible *input history*.

It is assumed that learning proceeds by a learning *method* that receives successive inputs and that directs i to believe or to suspend belief on each $\phi \in \mathbf{L}_{\mathsf{BIT}}$ by returning 1 or 0 in response to ϕ. One could go to the trouble of modeling changes to the learner's internal memory state as new inputs are received, but it is easier simply to observe that the learning method determines a total function of type $L : S^* \times \mathbf{L}_{@\mathsf{BIT}} \to \{0, 1\}$: just incrementally simulate the given learning method on the successive input states in history $\sigma \in S^*$, provide input $\phi \in \mathbf{L}_{@\mathsf{BIT}}$, and return the result.

It remains to say what it means for L to be computable. Let Φ^k_c denote the k-ary partial recursive function[5] computed by the Turing machine with Gödel code c.[6] Technically, the k arguments to Φ^k_c are supposed to be natural numbers, so let $\langle . \rangle : \mathbb{N}^* \to \mathbb{N}$ be a fixed, effective, bijective assignment of code numbers to input histories and let $\mathsf{g} : \mathbf{L}_{@\mathsf{BIT}} \to \mathbb{N}$ be a fixed, effective, bijective assignment of code numbers to $\mathbf{L}_{@\mathsf{BIT}}$ sentences. Then define, for each $\sigma \in S^*$ and $\phi \in \mathbf{L}_{@\mathsf{BIT}}$:

$$L_c(\sigma, \phi) \quad = \quad \Phi^2_c(\langle\sigma\rangle\,, \mathsf{g}(\phi)), \tag{1}$$

Say that c is a *learning method* if and only if Φ^2_c is total and Boolean-valued. Let $C \subset \mathbb{N}$ denote the set of all learning methods.

Each learning method covers all future contingencies, but i's learning method can change from time to time through maturation, mishap, revelation, disease, and death. A

[5]In recursive function theory, one can write any number of inputs on the tape of a given Turing machine, but the machine produces *interesting* outputs only for some fixed number of inputs k. The arity superscript k is usually obvious from context and will be omitted.

[6]The standard notation is ϕ_i, but ϕ is also standardly employed in logic as a variable over sentences.

joint method trajectory is a function $\mathbf{c} : (G \times T) \to C$ that assigns a learning method $c \in C$ to each agent $i \in G$ at each time $t \in T$. A *possible world* is a pair $w = (e, \mathbf{c})$, such that $e \in E$ and each $\mathbf{c}$ is a joint method trajectory. The set of all possible worlds is therefore definable as: $W = E \times ((G \times T) \to C)$. Let e_w denote the e component of w and let $\mathbf{c}_w$ denote the $\mathbf{c}$ component. Then one can define the *method assignment function* $\mathbf{c}(i, w, t) = \mathbf{c}_w(i, t)$.

A *proto computational learning model* (PCLM) for agents G is a quadruple $\mathfrak{M}_{t^*} = (E, \mathbf{s}, V, t^*)$ such that E is a non-empty set, $t^* \in T$ and:

$$\begin{aligned} \mathbf{s} &: (G \times W \times T) \to \mathbb{N}; \\ V &: (\mathbf{L}_{\mathsf{atom}} \times T) \to \mathsf{Pow}(W). \end{aligned}$$

Think of t^* as the "now" of the epistemic context under discussion (Kamp 1971). Think of $V(\mathsf{p}, t)$ as the proposition expressed by atomic sentence p in world w at arbitrary time t, so V is the model's *valuation function.* Think of $\mathbf{s}(i, w, t)$ as the signal state that w presents to i at t in w, so $\mathbf{s}$ is the *input assignment function.*

Note that the method assignment function $\mathbf{c}$ and the input assignment function $\mathbf{s}$ both have domain $(G \times W \times T)$. Let $\mathbf{g}$ be a generic such function. For fixed i, w, t, let $g_{i,w,t} = \mathbf{g}(i, w, t)$. For fixed t and w let $\mathbf{g}_{t,w}(i) = \mathbf{g}(i, w, t)$, so that $\mathbf{g}_{w,t} = (g_{w,1,t}, \ldots, g_{w,N,t})$ is the $\mathbf{g}$-*profile* of the agents in w at t. For fixed w, let $\mathbf{g}_w(i, t) = \mathbf{g}(i, w, t)$ so that $\mathbf{g}_w = (\mathbf{g}_{w,0}, \ldots, \mathbf{g}_{w,t}, \ldots)$ is the *joint* $\mathbf{g}$-*trajectory* of the agents in w and the restriction $\mathbf{g}_w|t = (\mathbf{g}_{w,0}, \ldots, \mathbf{g}_{w,(t-1)})$ is the *joint* $\mathbf{g}$-*history* of the agents in w up to t. For fixed w and i let $g_{w,i}(t) = \mathbf{g}_{i,w,t}$, so that $g_{i,w} = (g_{i,w}(0), \ldots, g_{i,w}(t), \ldots)$ is the $\mathbf{g}$-*trajectory* of i in w and $g_{i,w}|t$ is the $\mathbf{g}$-*history* of i in w up to t. Thus, one may speak of the *joint method trajectory* $\mathbf{c}_w$ in w, of the *input history* $s_{i,w}$ of i in w, etc.

The input history $s_{i,w}|t'$ of i in w has no truth value—it is a temporal sequence of physical states—but it *makes available* the following *cumulative propositional information* to i in w at t:

$$\mathbf{I}(i, w, t) \quad = \quad \{w' \in W : s_{i,w}|t = s_{i,w'}|t\}.$$

Call $\mathbf{I}$ the *information* assignment function. In Kripke semantics for modal epistemic logic, available information is represented in terms of the binary relation "w' is possible for all that i has been informed in w (at time t)":

$$\mathcal{I}_{i,t}(w, w') \quad \Leftrightarrow \quad w' \in I_{i,w,t}.$$

For fixed i and t, $(W, \mathcal{I}_{i,t}, V)$ is a standard Kripke model. Since $\mathcal{I}_{i,t}$ is an equivalence relation, the corresponding, available information operator is S5, as is often assumed (van Benthem 2010). Making propositional information available via physical signals is not the same thing as inserting that information directly into i's beliefs—it is still up to i's computable learning function $L_{c_{w,i,t}}$ to interpret the signals, to recover the information they afford, and to incorporate it smoothly into i's hyper-intensional belief system—or to fail to.

Possibilities of error that are incompatible with the information currently available will be deemed irrelevant to learning and knowledge. Furthermore, it does not seem that i needs to have been *informed* of her own learning method—the method merely has to *determine* success in light of available information. Accordingly, define the *determination assignment function*:

$$\mathbf{D}(i, w, t) \quad = \quad \{w' \in I_{i,t,w} : c_{i,w,t^*} = c_{i,w',t^*}\}.$$

Then $D_{i,w,t}$ is the strongest proposition determined at t by the learning strategy possessed by i in w at t^*. The binary relation $\mathcal{D}_{i,t}(w, w')$ is again an equivalence relation that refines $\mathcal{I}_{i,t}(w, w')$.

Since tensed statements have no truth value until the time is specified, the objects of belief for i are understood to be true or false at the time of utterance t^*. The hyper-intensional *belief assignment function* is then defined by:

$$\mathbf{B}'(i, w, t) \quad = \quad \{\phi \in \mathbf{L}_{\mathsf{@BIT}} : L_{c_{i,w,t}}(s_{i,w}|t, @_{t^*}\,\phi) = 1\}.$$

The belief state $B_{i,w,t}$ of i in w at t is decided by the learning method $c_{i,w,t}$ that i actually follows in w at t. So belief is sentential and computationally concrete.

In the long run, we are all dead. So if inductive knowledge depends upon i's actual convergence to true belief, then inductive knowledge is impossible and learning semantics fails to deliver on its promise to sidestep inductive skepticism. One potential such story is that i would converge to true belief if i were to continue to use her current method forever. However, that would make it impossible for i to know inductively that all humans are mortal, since i would be immortal if she were to follow her current method forever. What matters is not what i would have believed had *she* really followed her learning method forever, but simply what her current method would have *directed* her to believe. Then i can know inductively that all humans are mortal because her own learning method would simply chalk herself up as another confirming instance when her obituary is published. She can also know inductively that it is dangerous to drive when intoxicated even though she always believes the contrary when intoxicated, and so forth. That can be accomplished semantically by basing i's future, virtual beliefs on what i's current learning method c_{i,w,t^*} at t^* would direct i to believe in response the inputs $s_{i,w,t}$ that are really available to i in w at t. Accordingly, define the *virtual* belief assignment function as follows:

$$\mathbf{B}(i, w, t) \quad = \quad \{\phi \in \mathbf{L}_{\mathsf{@BIT}} : L_{c_{i,w,t^*}}(s_{i,w}|t, @_{t^*}\,\phi) = 1\}.$$

Virtual belief coincides with belief at t^* but may differ markedly thereafter, and that is a good thing.

The aim is to interpret learnability, knowability, and the feasibility of knowing some things given that you know other things. In order to interpret those modalities, one must entertain counterfactual changes in i's method. However, it is not intended to model counterfactual dependencies between i's method and the external world (e.g., if i were not a Bayesian then the history of statistics would be different) or between the methods of distinct agents (e.g., i is an identical twin of j so their methods would be the same). Instead, it is assumed that a simple substitution operation accomplishes the requisite metaphysical voyage to the nearest possible world in which i's method changes. The joint method assignment $\mathbf{c}[d/i,t]$ that results from substituting learning program d for the learning program c_{i,t^*} of i in $\mathbf{c}$ at t^* as follows:

$$(\mathbf{c}[d/i,t])(i',t') \quad = \quad \begin{cases} d & \text{if } i' = i \ \wedge \ t' = t; \\ \mathbf{c}(i',t') & \text{otherwise.} \end{cases}$$

Let $w = (e, \mathbf{c}) \in W$. Then the world in which i uses program d at t^* in w is:

$$w[d/i,t] \quad = \quad (e, \mathbf{c}[d/i,t]).$$

Counterfactual shifts of method open the door to the medieval problem of information concerning future contingents, for since $\mathbf{s}(w,i,t)$ depends on w, which specifies i's method trajectory $c_{i,w}$, we have the specter that $s_{i,w}|t \neq s_{i,w[d/i,t]}|t$—i.e., that the information informing i's choice of learning method at stage t could be yanked back by nature if that choice were carried out—as when a member of the Calvinist elect decides to sin because she is already saved. Learning semantics simply assumes that the information available to i cannot be affected by future shifts in method. A *computational learning model (CLM)* is, accordingly, a PCLM that satisfies the following, for all $i \in G$, $d \in C$, $w \in W$, and $t \in T$:

$$s_{i,w}|t \quad = \quad s_{i,w[d/i,t]}|t. \tag{2}$$

4 Learning Semantics

Let $\mathfrak{M}_{t^*} = (E, \mathbf{s}, V, t^*)$ be a CLM. Define the proposition $\|\phi\|^t_{\mathfrak{M}_{t^*}}$ expressed by ϕ in $\mathfrak{M}_{t^*}$ as follows. In the base case:

$$\|\mathsf{p}\|^t_{\mathfrak{M}_{t^*}} \quad = \quad V(\mathsf{p}, t).$$

Start with classical, propositional logic:

$$\begin{aligned} \|\bot\|^t_{\mathfrak{M}_{t^*}} &= \emptyset; \\ \|\phi \to \psi\|^t_{\mathfrak{M}_{t^*}} &= (W \setminus \|\phi\|^t_{\mathfrak{M}_{t^*}}) \cup \|\psi\|^t_{\mathfrak{M}_{t^*}}. \end{aligned}$$

For the temporal operators, define:

$$
\begin{aligned}
\|[\dot{\mathsf{F}}]\phi\|^{t}_{\mathfrak{M}_{t^*}} &= \bigcup_{t' \geq t^*} \|\phi\|^{t'}_{\mathfrak{M}_{t'}};\\
\|[\mathsf{F}]\phi\|^{t}_{\mathfrak{M}_{t^*}} &= \bigcup_{t' \geq t^*} \|\phi\|^{t'}_{\mathfrak{M}_{t^*}};\\
\|\mathsf{N}\,\phi\|^{t}_{\mathfrak{M}_{t^*}} &= \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}};\\
\|@_{t'}\,\phi\|^{t}_{\mathfrak{M}_{t^*}} &= \|\phi\|^{t'}_{\mathfrak{M}_{t^*}}.
\end{aligned}
$$

Think of t^* as "now", or the time at which the truth of some statement is to be assessed, and of t as a time that is considered in the evaluation of a future tense operator. Of principal interest in the sequel is the contextual future tense operator $[\mathsf{F}]$, which quantifies over all times from t^* onward without resetting the contextual now. Alternatively, $[\dot{\mathsf{F}}]$ quantifies over all times from t onward and also *resets* the contextual now to the future time visited. That allows one to speak of convergence to the truth at some time in the future.

Information and determination are defined propositionally in the standard way and both are S5 operators for reasons discussed above.

$$
\begin{aligned}
\|[\mathsf{I}]_i\,\phi\|^{t}_{\mathfrak{M}_{t^*}} &= \{w \in W : I_{i,w,t} \subseteq \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}}\};\\
\|[\mathsf{D}]_i\,\phi\|^{t}_{\mathfrak{M}_{t^*}} &= \{w \in W : D_{i,w,t} \subseteq \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}}\}.
\end{aligned}
$$

Virtual belief, on the other hand, is thoroughly and unabashedly hyper-intensional.

$$
\|[\mathsf{B}]_i\,\phi\|^{t}_{\mathfrak{M}_{t^*}} = \{w \in W : \phi \in B_{i,t,w}\}.
$$

Methodological necessity is straightforward—it is a universal quantifier over possible learning strategies for i at t^*:

$$
\|[\mathsf{M}]_i\,\phi\|^{t}_{\mathfrak{M}_{t^*}} = \{w \in W : (\forall c \in C)\ w[c/i, t^*] \in \|\phi\|^{t}_{\mathfrak{M}_{t^*}}\}. \tag{3}
$$

The dual operator $\langle \mathsf{M} \rangle_i$ is crucial for interpreting theses concerning learnability and knowability.

To motivate conditional feasibility, consider the familiar modal logical thesis that i knows that she knows what she knows: $\mathsf{K}_i\,\phi \to \mathsf{K}_i\mathsf{K}_i\,\phi$. The thesis isn't very plausible as it stands—maybe i has no belief whatever concerning $\mathsf{K}_i\,\phi$. It becomes more plausible if it is understood to say that i is in a position to guarantee that she knows that $\mathsf{K}_i\,\phi$ via a computable inference from her current beliefs, *given* that she knows that ϕ. Since i may not be aware of what her learning program is in w, the inferential procedure should work for any learning program compatible with her current information. The idea is not to modify *how* i knows that ϕ, so the operation should also hold i's beliefs whether ϕ fixed. Define $c \equiv_\Delta d$ to hold if and only if the following conditions:

$$\begin{array}{rcl} L_c(\sigma, @_{t^*}\, \delta) & = & L_d(\sigma, @_{t^*}\, \delta); \\ L_c(\sigma, @_{t^*} \neg\delta) & = & L_d(\sigma, @_{t^*} \neg\delta); \end{array}$$

hold for all $\sigma \in I^*$ and $\delta \in \Delta$. Say that total recursive function h *holds* Δ *fixed* if and only if $h(c) \equiv_\Delta c$, for all $c \in C$. Then let $w \in \|\psi \,\langle\mathsf{MD}]\!\!\rightarrow_{i,\delta}\, \phi\|^t_{\mathfrak{M}_{t^*}}$ if and only if there exists total recursive function h that holds Δ fixed such that, for all $u \in I_{i,w,t^*}$:

$$u \in \|\psi\|^t_{\mathfrak{M}_{t^*}} \quad \Rightarrow \quad u[h(c_{i,u,t})/i, t] \in \|\phi\|^t_{\mathfrak{M}_{t^*}}.$$

Conditional feasibility is witnessed by a total recursive (t.r.) transformation h of one learning method into another. The following, standard lemma from recursive function theory facilitates the construction of such functions.

Proposition 1. $(\forall\ t.r.\ f)(\exists\ t.r.\ h)(\forall c, x, y \in \mathbb{N})\ \Phi_{h(c)}(x, y) = f(c, x, y)$.

Inference—even deductive inference—can be subtly treacherous in learning semantics. Suppose that i contemplates changing her learning strategy $c = c_{i,w,t^*}$ to d, which generates exactly the same belief state concerning premise δ that c does, in every possible input situation. Assumption (2) guarantees that d results in the same belief whether δ that c does given the same inputs, but the change from c to d could modify or even shut off the flow of *future* inputs to i because other agents may detect the change in i (e.g., they may be subjects in a poorly blinded social psychology experiment and may desert in protest when they discover the morally unflattering conclusions i intends to publish about them if the experiment vindicates δ). Furthermore, the change from c to d could make δ false if the truth of δ depends on what some or all of the agents believe (e.g., i is a major player in the market). Either way, i's election to adopt inferential strategy d could be empirically or semantically *self-defeating*, in the sense that premise δ of the intended inference becomes untestable or false as a consequence of the inference being performed.

Fortunately, it is part of good scientific practice to choose premises and experimental designs that prevent one's valid inferences from being self defeating, so it is useful to have vocabulary expressing that such preventive measures have successfully been carried out for some intended premise δ. It is too strong to say that the inputs to i would be exactly the same whether she uses c or d because i would presumably receive some signals dependent upon her own beliefs. It suffices that c and d do not base their beliefs concerning δ on any inputs that would change if i were to replace c with d. Define $w \in \|\mathsf{S}_i\, \Delta\|^t_{\mathfrak{M}_{t^*}}$ to hold if and only if for all $d \in C$ such that $c_{i,w,t^*} \equiv_\Delta d$ and for all $u \in D_{i,w,t^*}$, $t \geq t^*$, and $\delta \in \Delta$:

$$u \in \|\delta\|^{t^*}_{\mathfrak{M}_{t^*}} \quad \Leftrightarrow \quad u[d/i, t^*] \in \|\delta\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{4}$$

$$L_{c_{i,u,t^*}}(s_{i,u}|t, @_{t^*}\, \delta) \quad = \quad L_d(s_{i,u[d/i,t^*]}|t, @_{t^*}\, \delta); \tag{5}$$

$$L_{c_{i,u,t^*}}(s_{i,u}|t, @_{t^*} \neg\delta) \quad = \quad L_d(s_{i,u[d/i,t^*]}|t, @_{t^*} \neg\delta). \tag{6}$$

That concludes the inductive definition of $\|\phi\|^{t}_{\mathfrak{M}_{t^*}}$. Define validity in a model and logical validity as follows:

$$\begin{aligned} \mathfrak{M}_{t^*} \models \phi \;\; &\Leftrightarrow \;\; W = \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \\ \models \phi \;\; &\Leftrightarrow \;\; \mathfrak{M}_{t^*} \models \phi, \text{ for each CLM } \mathfrak{M}_{t^*}. \end{aligned}$$

Note that validity in a model initializes time to the model's current epistemic context time t^*. Finally, logical entailment and equivalence are defined as follows:[7]

$$\begin{aligned} \phi \models \psi \;\; &\Leftrightarrow \;\; \models (\phi \rightarrow \psi); \\ \phi \equiv \psi \;\; &\Leftrightarrow \;\; \models (\phi \leftrightarrow \psi). \end{aligned}$$

5 Example: Outcomes of a Repeated Experiment

CLMs accommodate a boggling range of learning situations, but a simple collection of models suffices to illustrate many of the results that follow. Assume that each agent i observes the successive values of a repeated experiment whose outcomes are effectively coded as natural numbers. Identify possible external worlds with infinite outcome sequences $\varepsilon : \mathbb{N} \rightarrow \mathbb{N}$. Let E_0 denote the set of all such sequences. Define, for $k \in \mathbb{N}$:

$$\begin{aligned} \mathbf{s}_0(i, (\varepsilon, \mathbf{c}), t) \;\; &= \;\; \varepsilon(t); \\ V_0(\mathsf{p}_k, t) \;\; &= \;\; \{\epsilon \in E_0 : \varepsilon(t) = k\} \times \mathsf{C}^N; \\ \mathfrak{N}_{t^*} \;\; &= \;\; (E_0, \mathbf{s}_0, V_0, t^*). \end{aligned}$$

Temporal operators allow for compact expression of a range of increasingly complex statements:

$$\begin{aligned} \mathsf{p}_k \;\; &: \;\; \text{the current outcome is } k; \\ \mathsf{G}\mathsf{p}_k \;\; &: \;\; \text{the outcome will be } k; \\ \mathsf{F}\mathsf{p}_k \;\; &: \;\; \text{the outcome is } k \text{ from now on}; \\ \mathsf{F}\mathsf{G}\,\mathsf{p}_k \;\; &: \;\; \text{the outcome will stabilize to value } k; \\ \mathsf{G}\mathsf{F}\,\mathsf{p}_k \;\; &: \;\; \text{the outcome is } k \text{ infinitely often.} \end{aligned}$$

A hypothesis is currently *objective* for i just in case i cannot alter its truth value by changing her learning method:

$$\mathsf{O}_i\,\phi \;\; := \;\; [\,\mathsf{I}\,]_i(\phi \leftrightarrow [\mathsf{M}]_i\,\phi).$$

[7] N.b. substitution of equivalents for equivalents under temporal operators does not preserve validity (Kamp 1971). For example, $\models \mathsf{G}(\phi \leftrightarrow \phi)$ and $\phi \equiv \mathsf{N}\,\phi$, but $\not\models \mathsf{G}(\phi \leftrightarrow \mathsf{N}\,\phi)$.

One special feature of model $\mathfrak{N}_{t^*}$ is that inputs do not depend on methods, so:

$$\mathfrak{N}_{t^*} \quad \models \quad \mathsf{O}_i\,\phi \to \mathsf{S}_i\,\phi. \tag{7}$$

Another special property of $\mathfrak{N}_{t^*}$ is that it is *empirical* in the sense that the truth of objective statements supervenes on the input stream:

$$(s_{i,w} = s_{i,u} \ \wedge \ w \in \|\mathsf{O}_i\,\phi \ \wedge \ \phi\|^t_{\mathfrak{N}_{t^*}}) \quad \Rightarrow \quad u\|\phi\|^t_{\mathfrak{N}_{t^*}}. \tag{8}$$

6 Example: Agency, Games, and Experimentation

The agents in model $\mathfrak{N}_{t^*}$ are isolated, passive scientists who receive inputs from a fixed, non-reactive nature and change their beliefs accordingly. But even an isolated scientist can manipulate nature. Moreover, in a social system, the actions of the agents are observable by other agents, resulting in potential cascades of interactive effects. Although $\mathbf{L}_{\mathsf{BIT}}$ has no vocabulary describing acts other than belief, CLMs can represent arbitrarily complex social interactions involving such acts. The trick is to locate agents' diachronic strategies for non-doxastic actions within the "external world" $e \in E$. Therefore, all of the valid theses of learning semantics are valid for game-theoretic applications.

Here is one way to do it. Let $X \subseteq \mathbb{N}$ be a set of potential actions. Assuming that the actions are observable by all of the agents, let $S = X^N$. Then S^* contains all possible finite play histories. Let A denote the set of all $a \in \mathbb{N}$ such that Φ_a is total with range included in X. The *disposition to act* computed by a looks at the current input history and chooses how to act:

$$A_a(\sigma) \quad = \quad \Phi_a(\langle\sigma\rangle).$$

Dispositions to act can change through time just as dispositions to believe can. A *joint disposition trajectory* $\mathbf{a} : (G \times T) \to A$ assigns a profile of dispositions to the agents at each time. In purely social applications, the "external world" e can be identified with $\mathbf{a}$, so possible worlds are pairs $w = (\mathbf{a}, \mathbf{c})$. In experimental science, one agent can represent nature and the rest of the agents can be used to model socially distributed scientific inquiry. The real inputs available to the agents at a given stage are generated by the action dispositions of the agents at earlier stages. Let $\sigma * s$ denote the concatenation of signal $s \in S$ to finite sequence $\sigma \in S^*$.

$$\begin{aligned}
\dot{s}_{i,(\mathbf{a},\mathbf{c})}|0 \quad &= \quad ();\\
\dot{s}_{i,(\mathbf{a},\mathbf{c})}|(t+1) \quad &= \quad \dot{s}_{i,(\mathbf{a},\mathbf{c})}|t * (A_{a_{1,t+1}}(\dot{s}_{i,(\mathbf{a},\mathbf{c})}|t), \ldots, A_{a_{N,t+1}}(\dot{s}_{i,(\mathbf{a},\mathbf{c})}|t)).
\end{aligned}$$

In the long run, all the players of an infinite game are dead, as are the dispositional properties of societies, economies, and terrestrial organisms. Hence, it is often more natural to think of the agents as studying one another's and nature's *current* reactive dispositions, just as was done for belief. Information gathered by means of earlier dispositions remains available.

$$\begin{aligned} s_{i,(\mathbf{a},\mathbf{c})}|t &= \dot{s}_{i,(\mathbf{a},\mathbf{c})}|t; \\ s_{i,(\mathbf{a},\mathbf{c})}|(t+1) &= s_{i,(\mathbf{a},\mathbf{c})}|t * (A_{a_{1,t+1}}(s_{i,(\mathbf{a},\mathbf{c})}|t), \ldots, A_{a_{N,t+1}}(s_{i,(\mathbf{a},\mathbf{c})}|t)). \end{aligned}$$

Either way, assumption (2) is satisfied, so a CLM results when a valuation function V is specified.

In game theory, each agent receives some utility in each world at each time, as a result of what all the agents do. The utilities may also shift through time if we interpret the agents as playing different games from time to time. All of that can be absorbed into the definition of V—for example, some atomic sentence could be interpreted to say that the players are currently in a (virtual) Nash equilibrium with respect to the game that V tacitly assumes them to be playing at the time.

7 Correctness and Error

Define "i is in *error* that ϕ" as follows:

$$\mathsf{E}_i\,\phi \;:=\; \mathsf{B}_i\,\phi \,\wedge\, \mathsf{N}\neg\phi.$$

Define "i is in error *whether* ϕ" similarly, where the tilde is a mnemonic that the intended reading is "whether":

$$\tilde{\mathsf{E}}_i\,\phi \;:=\; \mathsf{E}_i\,\phi \,\vee\, \mathsf{E}_i\neg\phi.$$

According to that definition, one cannot be in error whether ϕ unless one believes that ϕ or believes that $\neg\phi$. That is straightforward, if belief is deductively closed, but in the present, hyper-intensional framework it is very weak—e.g.:

$$\tilde{\mathsf{E}}_i\,\phi \;\not\equiv\; \tilde{\mathsf{E}}_i\neg\phi; \tag{9}$$

and belief that ϕ does not count as an error whether $\neg\phi$. However, in order to interpret successful learning whether ϕ, all that is required is some unambiguous convention for i "getting ϕ wrong", and the proposed convention suffices in a minimal, maximally paraconsistent way. Other computationally feasible conditions could plausibly be added (cf. section 16.3 below) and, as a matter of fact, the results that follow are all provable under the much more stringent condition that avoidance of error whether ϕ implies avoidance of all error whatever. However, it is a considerable advantage of the present approach that inductive knowledge is feasible *before* science is entirely error-free.

Correctness that ϕ is absence of error whether ϕ together with belief that ϕ. Correctness *whether* ϕ might plausibly be defined as correctness that ϕ or correctness that $\neg\phi$, but to maintain the minimal focus on ϕ and $\neg\phi$ so far as learning whether ϕ is concerned, correctness whether ϕ is defined as belief whether ϕ together with absence of error whether ϕ.[8]

$$\begin{aligned}\tilde{\mathsf{B}}_i\,\phi &:= \quad \mathsf{B}_i\,\phi \;\wedge\; \mathsf{B}_i\neg\phi;\\ \mathsf{C}_i\,\phi &:= \neg\tilde{\mathsf{E}}_i\,\phi \;\wedge\; \mathsf{B}_i\,\phi;\\ \tilde{\mathsf{C}}_i\,\phi &:= \neg\tilde{\mathsf{E}}_i\,\phi \;\wedge\; \tilde{\mathsf{B}}_i\,\phi.\end{aligned}$$

8 Inductive Learning

In formal learning theory, *inductive learning* whether ϕ is understood as guaranteed convergence of i's current learning method to correct belief whether ϕ:

$$\tilde{\mathsf{L}}_i\phi \;:=\; [\mathsf{D}]_i\mathsf{F}\mathsf{G}\tilde{\mathsf{C}}_i\,\phi.$$

The truth conditions for $\tilde{\mathsf{L}}_i\,\phi$ can be expressed entirely in terms of the semantics of ϕ along with i's concrete learning program c and the concrete inputs it receives. Recalling that $c_{i,u,t^*} = c_{i,w,t^*}$, for each $u \in D_{i,w,t}$, we have that $w \in \|\mathsf{L}_i\,\phi\|^t_{\mathfrak{M}_{t^*}}$ if and only if for all $u \in D_{i,w,t}$:

$$\begin{aligned} u \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;&\Rightarrow\; (\lim_{t\to\infty} L_{c_{i,w,t^*}}(s_{i,u}|t, @_t\,\phi) \;\;= 1 \;\wedge \qquad (10)\\ &\qquad \lim_{t\to\infty} L_{c_{i,w,t^*}}(s_{i,u}|t, @_t\,\neg\phi) = 0);\\ u \notin \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;&\Rightarrow\; (\lim_{t\to\infty} L_{c_{i,w,t^*}}(s_{i,u}|t, @_t\,\neg\phi) = 1 \;\wedge \qquad (11)\\ &\qquad \lim_{t\to\infty} L_{c_{i,w,t^*}}(s_{i,u}|t, @_t\,\phi) \;\;= 0).\end{aligned}$$

That is essentially equivalent to saying, in formal learning theory, that i's *current* method c_{i,w,t^*} *decides* ϕ in the limit (Kelly 1996), except that learning semantics allows that the data depend on the learning method, whereas most learning theoretic analyses do not.

Regarding death in the long run, note that:

$$\models \;\; ([\mathsf{D}]_i\mathsf{F}\mathsf{G})\,\phi \to ([\mathsf{D}]_i\mathsf{G})([\mathsf{D}]_i\mathsf{F}\mathsf{G})\,\phi;$$

so learning does imply learning forever, if the epistemic context is held fixed:

$$\models \;\; \tilde{\mathsf{L}}_i\,\phi \to [\mathsf{D}]_i\mathsf{G}\tilde{\mathsf{L}}_i\,\phi.$$

At the same time, i may have been diagnosed with a terminal illness, so there exist models $\mathfrak{M}_{t^*}$ in which:

$$\mathfrak{M}_{t^*} \;\;\models\;\; \tilde{\mathsf{L}}_i\,\phi \;\wedge\; [\,\mathsf{I}\,]_i\dot{\mathsf{F}}\dot{\mathsf{G}}\neg\tilde{\mathsf{L}}_i\,\phi.$$

[8] I am indebted to Ted Shear for pointing out a flaw in an earlier version of this definition.

9 Inductive Learnability

Just as the theory of computability concerns what can be computed, rather than how we actually compute, formal learning theory focuses on learnability—the feasibility of learning—rather than on learning, itself. Learning semantics affords at least four distinct grades of feasibility, whose entailments are immediate from their truth conditions:

$$\langle \mathsf{M} \rangle_i [\mathsf{D}]_i \, \phi \quad \models \quad \langle \mathsf{MD}]_i \, \phi \quad \models \quad \langle \mathsf{M} \rangle_i \, \phi \quad \models \quad \langle \mathsf{M} \rangle_i \langle \mathsf{D} \rangle_i \, \phi. \tag{12}$$

In the case of learnability, all of these concepts collapse to $\langle \mathsf{M} \rangle_i \tilde{\mathsf{L}}_i \, \phi$, because the last entails the first, since $[\mathsf{D}]_i$ is an S5 operator:

$$\langle \mathsf{M} \rangle_i [\mathsf{D}]_i \tilde{\mathsf{L}}_i \, \phi \quad \equiv \quad \langle \mathsf{MD}]_i \, \tilde{\mathsf{L}}_i \, \phi \quad \equiv \quad \langle \mathsf{M} \rangle_i \tilde{\mathsf{L}}_i \, \phi \quad \equiv \quad \langle \mathsf{M} \rangle_i \langle \mathsf{D} \rangle_i \tilde{\mathsf{L}}_i \, \phi. \tag{13}$$

Concretely, $w \in \| \langle \mathsf{M} \rangle_i \tilde{\mathsf{L}}_i \, \phi \|^t_{\mathfrak{M}_{t^*}}$ if and only if there exists $d \in C$ such that (10) and (11) hold with d substituted for c_{i,w,t^*}, for all $u \in I_{i,w[d/i,t^*],t}$. If ϕ satisfies $\mathsf{O}_i \, \phi$ in $\mathfrak{N}_{t^*}$, one can also substitute $I_{i,w,t}$ for $I_{i,w[d/i,t^*],t}$, in which case the truth conditions for learnability are essentially the same as the conditions for decidability in the limit in (Kelly 1996).[9]

Universal truths and existential truths about the future are inductively learnable in the empirical model $\mathfrak{N}_{t^*}$—just believe the universal hypothesis until it is refuted and believe its negation thereafter, and do the dual thing in the existential case:

$$\mathfrak{N}_{t^*} \quad \models \quad \langle \mathsf{M} \rangle_i \tilde{\mathsf{L}}_i \mathsf{G} \, \mathsf{p}_k; \tag{14}$$

$$\mathfrak{N}_{t^*} \quad \models \quad \langle \mathsf{M} \rangle_i \tilde{\mathsf{L}}_i \mathsf{F} \, \mathsf{p}_k. \tag{15}$$

But not every empirical hypothesis is inductively learnable. In his "Paralogisms of Pure Reason" Immanuel Kant (1782/1787) observes that hypotheses like the finite or infinite divisibity of matter or the existence of a first moment in time "outpace all possible experience". Suppose that the laboratory returns a 1 whenever a currently fundamental particle is split and returns a 0 when an attempted split fails. Then the finite divisibility of matter can be expressed as $\mathsf{FG} \, \mathsf{p}_0$ and the finite divisibility of matter can be expressed as $\mathsf{GF} \, \mathsf{p}_1$. One intuitive sign that these hypotheses might pose empirical difficulties is that neither is refuted by any finite sequence of inputs. In fact, the following are non-learnable in $\mathfrak{N}_{t^*}$, for all k:

$$\mathfrak{N}_{t^*} \quad \models \quad \neg \tilde{\mathsf{L}}_i \mathsf{FG} \, \mathsf{p}_k; \tag{16}$$

$$\mathfrak{N}_{t^*} \quad \models \quad \neg \tilde{\mathsf{L}}_i \mathsf{GF} \, \mathsf{p}_k. \tag{17}$$

It suffices to show, via a standard, learning theoretic diagonal argument that no c satisfies convergence conditions (10) and (11).

[9] The differences concern mere conventions for coding the acceptance, rejection, or suspension of belief of i with respect to ϕ.

Learning semantics is a flexible framework for inductive learning and learnability that allows one, for the first time, to rigorously iterate the learning operator, in order to analyze precisely such statements as that it is learnable whether someone else is learning whether ϕ. But the focus of this paper is on inductive knowledge, to which we now turn.

10 Inductive Knowledge

Agent i *has learned* whether (that) ϕ if and only if i is learning whether ϕ and, henceforth, i virtually and correctly believes whether ϕ:

$$\begin{aligned} \tilde{\mathsf{Led}}_i\,\phi &:= \mathsf{G}\tilde{\mathsf{C}}_i\,\phi \;\wedge\; \tilde{\mathsf{L}}_i\,\phi; \\ \mathsf{Led}_i\,\phi &:= \mathsf{G}\mathsf{C}_i\,\phi \;\wedge\; \tilde{\mathsf{L}}_i\,\phi \;\equiv\; \tilde{\mathsf{Led}}_i\,\phi \;\wedge\; \phi. \end{aligned}$$

Having learned inductively whether ϕ may sound odd, since the culmination of inductive inquiry depends on what i's current learning method would do in the future. But such locutions are actually quite common: e.g., "I have quit smoking for good".

It is natural from a learning perspective to expect that inductive knowledge is just having learned. But there is a weighty consideration to the contrary: learnability is not preserved under logical consequence—recall (14), (16), and (17) and that in temporal logic:

$$\models\; \mathsf{G}\,\phi \;\rightarrow\; \mathsf{GF}\,\phi; \tag{18}$$

$$\models\; \mathsf{G}\,\phi \;\rightarrow\; \mathsf{FG}\,\phi. \tag{19}$$

Since having learned entails learnability, it follows that knowledge as having learned cannot be closed under logical consequence. And the examples sound bad: we would know that the laws of quantum mechanics apply invariably, but not that they apply infinitely often or all but finitely often. It sounds better to say that we know that a predicate holds infinitely often *because* we know that it holds invariably.

Pursuing that idea, suppose that i's only reason for believing that $\mathsf{GF}\,\phi$ is that she believes $\mathsf{G}\,\phi$ and suppose that her reason for believing $\mathsf{G}\,\phi$ is that it has stood up to severe testing so far (it might have been refuted by the data at each stage of inquiry). It is a traditional theme in the philosophy of science that general theories are not testable until they are *articulated* with auxiliary assumptions (Duhem 1914). Semantically speaking, "articulation" amounts to the substitution of a logically stronger, testable hypothesis for the untestable hypothesis, itself. Thus, one may think of $\mathsf{G}\,\phi$ as a testable articulation of $\mathsf{GF}\,\phi$, since it posits a particularly simple *way* in which $\mathsf{GF}\,\phi$ might be true. Then i stabilizes to true belief that $\mathsf{GF}\,\phi$ as soon as she stabilizes to true belief that $\mathsf{G}\,\phi$, so the actual convergence requirement is also met for $\mathsf{GF}\,\phi$. But what if $\mathsf{G}\,\phi$ were to be refuted, say at time t? Maybe i has plausible ideas about how to re-articulate $\mathsf{GF}\,\phi$ (e.g., as

$@_{t+1}\mathsf{G}\,\phi$). In order to learn by such a strategy, i would require a contingency plan for re-articulating $\mathsf{GF}\,\phi$ that somehow hits upon a true articulation eventually in *every* possible world in which $\mathsf{GF}\,\phi$ is true. But it has already been shown that no such contingency plan exists for $\mathsf{GF}\,\phi$, since it is not learnable—intuitively, there are *uncountably many* potential such articulations, most of which posit uncomputable input streams that computable i cannot even conceive of, much less hit upon by luck.

Another venerable theme in the philosophy of science is that there is "no logic of discovery" (Hempel 1945, Popper 1935), which means, roughly, that science does not have, and need not have, for purposes of empirical justification, an explicit contingency plan for what to propose when old hypotheses are refuted. A standard argument for that conclusion is historical rather than learning theoretic.[10] The chemist Kekulé famously claimed to come up with the idea that benzene molecules are cyclic by dreaming of a snake biting its tail (Hempel 1945, Benfey 1958). Kekulé's benzene hypothesis was a testable articulation of the atomic theory of matter that stood up to test. It does not seem to count against our knowledge of atomic theory, in light of subsequent testing, that Kekulé possessed no systematic contingency plan for dreaming up alternative structures had the ring structure failed. Scientists refer to luck that does not undermine scientific knowledge as *serendipity*. Kekule's dream was serendipitous in that sense, as is all luck in hitting upon a true articulation of a hypothesis. Since untestable hypotheses like $\mathsf{GF}\,\phi$ cannot be learned, they can be known only via serendipity. So serendipity, the practice of testing testable surrogates for untestable hypotheses, and the slogan that there is "no logic of discovery" are all tightly bound to a *logical* consideration—the closure of knowability under logical consequence.

Suppose that i is commanded by her thesis advisor to investigate $\mathsf{GF}\,\phi$ by severely testing $\mathsf{G}\,\phi$. We know that i lacks a full logic of discovery for $\mathsf{GF}\,\phi$, since $\mathsf{GF}\,\phi$ is not learnable. Suppose, plausibly, that she has far less—if $\mathsf{G}\,\phi$ is ever refuted, she has no idea what is going on, suspends belief forever whether $\mathsf{GF}\,\phi$, and switches to a more rewarding career in business. *If* her advisor was right (serendipity), then she has already converged to true belief that $\mathsf{GF}\,\phi$ and, since her belief that $\mathsf{GF}\,\phi$ is based *solely* on her belief that $\mathsf{G}\,\phi$, she is also guaranteed to root out *error* with respect to $\mathsf{GF}\,\phi$ eventually. Her (actual) convergence to true belief that the untestable hypothesis is true is serendipitous, but her eventual avoidance of error is not lucky at all—it is guaranteed by her strategy to suspend belief forever if $\mathsf{GF}\,\phi$ is refuted. So in terms of rooting out error, i is no worse off for her dearth of new ideas than she would have been had she possessed a full learning strategy for $\mathsf{GF}\,\phi$.

In light of the preceding considerations, it is proposed that inductive knowledge that ϕ is actual convergence to true belief that ϕ along with guaranteed, eventual avoidance

[10] A notable exception is (Putnam 1963), which argues against the logic of discovery based on a proof that it is impossible to exactly identify the input sequence even assuming that the sequence is computable.

of error whether ϕ:[11]

$$\begin{aligned}\tilde{\mathsf{K}}_i\,\phi &:= \mathsf{G}\tilde{\mathsf{C}}_i\,\phi \;\wedge\; [\mathsf{D}]_i\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi;\\ \mathsf{K}_i\,\phi &:= \mathsf{G}\mathsf{C}_i\,\phi \;\wedge\; [\mathsf{D}]_i\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi \;\equiv\; \tilde{\mathsf{K}}_i\,\phi \;\wedge\; \phi.\end{aligned}$$

As an immediate consequence, having learned implies knowing and learning implies that you will know, but not conversely, which secures the relevance of learning to epistemology without making it necessary for knowledge:

$$\models \tilde{\mathsf{L}}_i\,\phi \rightarrow \tilde{\mathsf{K}}_i\,\phi; \tag{20}$$

$$\models \mathsf{L}_i\,\phi \rightarrow \mathsf{K}_i\,\phi. \tag{21}$$

In terms of concrete learning methods, the first conjunct of $\tilde{\mathsf{K}}\,\phi$ is true in w at t if and only if:

$$\begin{aligned} w \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;\;\Rightarrow\;\; &((\forall t \geq t^*)\; L_{c_{i,w,t^*}}(s_{i,w}|t, @_t\,\phi) = 1 \;\wedge \\ &(\forall t \geq t^*)\; L_{c_{i,w,t^*}}(s_{i,w}|t, @_t\,\neg\phi) = 0);\end{aligned} \tag{22}$$

$$\begin{aligned} w \notin \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;\;\Rightarrow\;\; &((\forall t \geq t^*)\; L_{c_{i,w,t^*}}(s_{i,w}|t, @_t\,\phi) = 0 \;\wedge \\ &(\forall t \geq t^*)\; L_{c_{i,w,t^*}}(s_{i,w}|t, @_t\neg\phi) = 1);\end{aligned} \tag{23}$$

and the second conjunct is true in w at t if and only if for all $u \in I_{w,i,t}$:

$$u \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;\;\Rightarrow\;\; \lim_{t\to\infty} L_{c_{i,w,t^*}}(s_{i,u}|t, @_t\,\neg\phi) = 0; \tag{24}$$

$$u \notin \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;\;\Rightarrow\;\; \lim_{t\to\infty} L_{c_{i,w,t^*}}(s_{i,u}|t, @_t\,\phi) = 0. \tag{25}$$

Note that (24) and (25) weaken the corresponding conditions (10) and (11) for having learned.

[11] Hendricks (2001) presents several concepts of empirical knowledge, the closest of which to the following proposal is "realistic reliable true belief" or *RRT knowledge*. Hendricks' informal gloss of RRT knowledge (p. 181) amounts to the following proposal in the present notation: $\mathsf{Krrt}_i\,\phi := \mathsf{G}\,\phi \;\wedge\; \tilde{\mathsf{L}}_i\,\phi$ (the operator $[\mathsf{D}]_i$ is dropped from the $\tilde{\mathsf{L}}_i\,\phi$ condition in the accompanying formal statement—presumably unintentionally). RRT knowledge is very different from inductive knowledge as defined below. First of all, RRT knowledge requires that $\mathsf{G}\,\phi$, which would make it impossible for i to know, for example, that she believes that ϕ, if that belief state is transient. Learning semantics sidesteps that difficulty by evaluating the proposition believed at the "now" of utterance. Second, RRT knowledge does not require $\mathsf{GB}_i\,\phi$, so RRT knowledge does not even imply belief that ϕ, much less stable belief that ϕ—it may be years until the learning process succeeds. Finally, RRT knowledge does imply learning whether ϕ, which implies that RRT knowability cannot be closed under deductive consequence, as has just been explained. Hendricks' claim that RRT knowledge validates the axioms of modal system S4 (proposition 12.3, p. 208) is therefore false. The discrepancy is explained by the fact that, just prior to the proof of proposition 12.3, Hendricks inadvertently modifies the concept of RRT knowledge a second time (p. 194) to $\mathsf{G}\,\phi$ conjoined with the existence of a future time t' such that it is determined now that i believes that ϕ forever after t'—whether or not ϕ is true.

11 Inductive Knowability

Learning semantics again affords the following notions of inductive knowability, in descending strength:

$$\langle\mathsf{M}\rangle_i[\mathsf{D}]_i\mathsf{K}_i\,\phi \quad\models\quad \langle\mathsf{MD}]_i\,\mathsf{K}_i\,\phi \quad\models\quad \langle\mathsf{M}\rangle_i\mathsf{K}_i\,\phi \quad\models\quad \langle\mathsf{M}\rangle_i\langle\mathsf{D}\rangle_i\mathsf{K}_i\,\phi. \tag{26}$$

This time, the distinctions matter, due to the actual convergence requirement $\mathsf{G}\tilde{\mathsf{C}}_i\,\phi$ for knowledge. But the first three can be neglected, since they imply a version of inductive skepticism, namely, that if it is logically valid that ϕ is knowable by i, then i has the information that she has the power to have correct belief that ϕ now, which implies that she has the information that she has the power to make ϕ true now.[12]

$$\models (\langle\mathsf{M}\rangle_i[\mathsf{D}]_i\mathsf{K}_i\,\phi \;\vee\; \langle\mathsf{MD}]_i\,\mathsf{K}_i\,\phi \;\vee\; \langle\mathsf{M}\rangle_i\mathsf{K}_i) \quad\Rightarrow\quad \models [\mathsf{I}]_i\langle\mathsf{M}\rangle_i\,\phi. \tag{27}$$

That leaves the fourth, *weak* option, which requires only that it be feasible for i to make it *possible* that she knows now—an idea consonant with allowance for serendipity in inductive knowledge:

$$\begin{aligned}
\langle\mathsf{MD}\rangle_i\,\phi \;&:=\; \langle\mathsf{M}\rangle_i\langle\mathsf{D}\rangle_i\tilde{\mathsf{K}}_i\,\phi &(28)\\
&\equiv\; \langle\mathsf{M}\rangle_i\langle\mathsf{D}\rangle_i(\mathsf{G}\tilde{\mathsf{C}}_i\,\phi \;\wedge\; [\mathsf{D}]_i\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi) &(29)\\
&\equiv\; \langle\mathsf{M}\rangle_i(\langle\mathsf{D}\rangle_i\mathsf{G}\tilde{\mathsf{C}}_i\,\phi \;\wedge\; [\mathsf{D}]_i\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi); &(30)
\end{aligned}$$

where the last equivalence is due to $[\mathsf{D}]_i$ being S5. Condition (30) expands to the existence of $d \in C$ such that for some $u \in I_{w,i,t}$:

$$\begin{aligned}
u[d/i,t] \in \|\phi\|^t_{\mathfrak{M}_{t^*}} \;&\Rightarrow\; ((\forall t \geq t^*)\; L_d(s_{u,i}|t, @_t\,\phi) \;\;= 1 \;\wedge &(31)\\
&\qquad (\forall t \geq t^*)\; L_d(s_{u,i}|t, @_t\,\neg\phi) = 0);\\
u[d/i,t] \notin \|\phi\|^t_{\mathfrak{M}_{t^*}} \;&\Rightarrow\; ((\forall t \geq t^*)\; L_d(s_{u,i}|t, @_t\,\phi) \;\;= 0 \;\wedge &(32)\\
&\qquad (\forall t \geq t^*)\; L_d(s_{u,i}|t, @_t\,\neg\phi) = 1);
\end{aligned}$$

and for all $u \in I_{w,i,t}$:

$$\begin{aligned}
u[d/i,t] \in \|\phi\|^t_{\mathfrak{M}_{t^*}} \;&\Rightarrow\; \lim_{t\to\infty} L_d(s_{i,u}|t, @_t\,\neg\phi) = 0; &(33)\\
u[d/i,t] \notin \|\phi\|^t_{\mathfrak{M}_{t^*}} \;&\Rightarrow\; \lim_{t\to\infty} L_d(s_{i,u}|t, @_t\,\phi) \;\;= 0. &(34)
\end{aligned}$$

Conditions (31) and (32) are trivially satisfiable by dogmatically believing that ϕ and conditions (33) and (34) are trivially satisfiable by skeptically suspending belief whether ϕ. But the conditions are not jointly trivial—the possibility of having converged to the truth risks the possibility of error infinitely often, unless one has an appropriate strategy

[12] The third option does yield a non-trivial interpretation of knowability *whether*.

for when to suspend judgment, as Popper (1935) recommended. For example, weak knowability can fail when even the total input stream does not determine the truth of ϕ in any world. In that case, say that ϕ is *globally underdetermined*—venerable examples include "the Absolute is lazy" and Poincare's (1904) perfect trade-off between shrinking forces and geometry.

The logical positivists attempted to rule out globally underdetermined hypotheses by deeming them meaningless, on empiricist grounds, but they lacked an explicit story connecting global determination with knowability. Here is one. Recall the strategy, discussed above, of guessing a testable articulation ψ of ϕ, believing ϕ until ψ is refuted, and suspending judgment thereafter. It witnesses the following, liberal knowability condition for objective hypotheses in $\mathfrak{N}_{t^*}$:

Proposition 2. *Suppose that* $w \in \|\mathsf{O}_i\, \phi\|^{t^*}_{\mathfrak{N}_{t^*}}$ *and there exists* $u \in I_{i,w,t^*}$ *such that* $u \in \|\phi\|^{t^*}_{\mathfrak{N}_{t^*}}$ *and* $s_{i,u}$ *is computable. Then* $w \in \|\langle \mathsf{MD} \rangle_i\ \mathsf{K}_i\, \phi\, \|^{t^*}_{\mathfrak{N}_{t^*}}$.

As a corollary, we have the following, knowability result, in contrast to the non-learnability results (16) and (17) above. Just let u satisfy $s_{i,u,t} = s_{i,w,t}$ for $t < t^*$ and $s_{i,u,t} = k$ for $t \geq t^*$:

$$\mathfrak{N}_{t^*} \models \langle \mathsf{MD} \rangle_i (\mathsf{K}_i \mathsf{G}\, \mathsf{p}_k \wedge \mathsf{K}_i \mathsf{F}\, \mathsf{p}_k \wedge \mathsf{K}_i \mathsf{FG}\, \mathsf{p}_k \wedge \mathsf{K}_i \mathsf{GF}\, \mathsf{p}_k). \tag{35}$$

The restriction to $\mathfrak{N}_{t^*}$ is crucial because there are models in which the Institutional Review Board shuts down the research project as soon as i starts to draw useful conclusions from it. It is also crucial that ϕ is true in some computable world. For example, take the setting to be $\mathfrak{N}_{t^*}$ restricted to worlds that present binary data. Add a new atomic sentence q with the valuation $V(\mathsf{q}) = \{w \in W : s_{i,u} = g\}$, where g is a fixed, total, non-computable, binary-valued function. Call the resulting model $\mathfrak{B}_{t^*}$. Then we have:

$$w \notin \|\langle \mathsf{MD} \rangle_i \mathsf{K}_i\, \mathsf{q}\|^{t^*}_{\mathfrak{B}_{t^*}}. \tag{36}$$

The restriction to binary sequences in the preceding, negative result is crucial. If the range of inputs at each stage might be infinite, then one can add an atomic sentence to $\mathfrak{N}_{t^*}$ that is knowable but true only in worlds that are empirically *infinitely* uncomputable (cf. Kelly 1996, 7.19).

12 Fitch's Paradox Redux

Proposition 2 provides a plausibly broad, sufficient condition for weak inductive knowability. It was also shown that the condition can fail—concrete, methodological issues like underdetermination, uncomputability, or ethical considerations can stand in the way. The knowability literature in traditional epistemic logic has focused on the more arcane possibility of unknowability due to epistemic self-reference. Consider the *Moore sentence* for ϕ, defined as follows:

$$\mathsf{Mo}_i\,\phi \;:=\; \phi \;\wedge\; \neg\mathsf{K}_i\,\phi.$$

The Moore sentence is not knowable in standard epistemic logic, for suppose that i knows it. Then since knowledge is true, $\mathsf{Mo}_i\,\phi$ is also be true, so $\neg\mathsf{K}_i\,\phi$ is true. But since Mo_i is known, so is conjunct ϕ of $\mathsf{Mo}_i\,\phi$, so $\mathsf{K}_i\,\phi$ is true. Contradiction. The proof requires only (i) that the conjuncts of a known conjunction are known and (ii) that knowledge is true.

That is hardly surprising in itself, but it leads in one step[13] to *Fitch's paradox*, the statement that any agent for whom every truth is knowable is already omniscient:

$$(\forall\phi)\ (\phi \rightarrow \Diamond_i\mathsf{K}_i\,\phi) \rightarrow (\forall\phi)\ (\phi \rightarrow \mathsf{K}_i\,\phi). \tag{37}$$

For contraposition, suppose that the consequent of (37) is false. Then unknowable $\mathsf{Mo}_{ii}\,\phi$ is true, which implies the denial of the antecedent.

Fitch's paradox is not all that paradoxical after the "gotcha" moment when one realizes that the denial of the consequent is the self-referential Moore sentence. Nevertheless, there is a specialist literature devoted to refuting Fitch's paradox, some authors going so far as to blame proof by contraposition (Williamson 1993). Therefore, it may be of interest to check whether the proof of (37) is valid in learning semantics, when the hazy modality $\Diamond_i$ is sharpened to weak methodological feasibility $\langle\mathsf{MD}\rangle_i$:

$$(\forall\phi)\ (\phi \rightarrow \langle\mathsf{MD}\rangle_i\mathsf{K}_i\,\phi) \rightarrow (\forall\phi)\ (\phi \rightarrow \mathsf{K}_i\,\phi). \tag{38}$$

The step from i knowing that $\mathsf{Mo}_i\,\phi$ to i knowing that ϕ evidently fails due to hyper-intensionality—i need not even believe that ϕ. That makes it weakly possible to know the Moore sentence of an arbitrary, objective, knowable statement ϕ in $\mathfrak{N}_{t^*}$:

$$\mathfrak{N}_{t^*} \;\models\; (\mathsf{O}_i\,\phi \;\wedge\; \langle\mathsf{MD}\rangle_i\mathsf{K}_i\,\phi) \rightarrow \langle\mathsf{MD}\rangle_i\mathsf{K}_i\,\mathsf{Mo}_i\,\phi. \tag{39}$$

However, the inference from $\phi \;\wedge\; \psi$ to ϕ is about as easy and plausible as deduction gets, so one can reasonably expect i to be *conjunctively cogent* with respect to the Moore sentence in the following sense:

$$\mathsf{Cocomo}_i\,\phi \;:=\; [\,\mathsf{I}\,]_i(\mathsf{B}_i\mathsf{Mo}_i\,\phi \leftrightarrow (\mathsf{B}_i\,\phi \;\wedge\; \mathsf{B}_i\neg\mathsf{K}_i\,\phi)).$$

Can $\mathsf{Mo}_i\,\phi$ be known even by a conjunctively cogent agent? Surprisingly, yes—at least in the typical case in which i has not been informed outright that knowable, objective ϕ is true:

$$\mathfrak{N}_{t^*} \;\models\; (\mathsf{O}_i\,\phi \;\wedge\; \neg[\mathsf{D}]_i\,\phi \;\wedge\; \langle\mathsf{MD}\rangle_i\mathsf{K}_i\,\phi) \;\rightarrow\; \langle\mathsf{MD}\rangle_i(\mathsf{K}_i\mathsf{Mo}_i\,\phi \;\wedge\; \mathsf{Cocomo}_i\,\phi). \tag{40}$$

[13]The ingenious step was taken by Alonzo Church (2009) in an anonymous referee report on Fitch's manuscript.

Of course *some* sort of aphasia is required for that dubious feat, but it is now more plausibly located in learning, rather than in failure to perform a trivial deductive inference.[14] Think of i as taking ϕ as an object of blind faith, but whenever she is asked whether she knows that ϕ, she is awakened from her dogmatic slumber to consider the evidence whether ϕ. Since her dogmatism precludes her from knowing that ϕ, the knowability of $\mathsf{Mo}_i\,\phi$ reduces to that of ϕ, so she knows the former based on the evidence for the latter. Thus, i *trades* knowledge that ϕ for knowledge that $\mathsf{Mo}_i\,\phi$. Some such trade is inevitable, since it is impossible to know both that ϕ and that $\mathsf{Mo}_i\,\phi$ because knowledge is true:

$$\mathfrak{N}_{t^*} \models \neg\langle\mathsf{MD}\rangle_i(\mathsf{K}_i\,\phi \;\wedge\; \mathsf{K}_i\mathsf{Mo}_i\,\phi). \tag{41}$$

Nonetheless, one *can* know the conjunction of $\phi \;\wedge\; \mathsf{Mo}_{ii}\,\phi$ of the two jointly unknowable statements, and so forth, by the same trick—as long as one does not know that ϕ.

$$\mathfrak{N}_{t^*} \models (\neg[\mathsf{D}]_i\,\phi \;\wedge\; \langle\mathsf{MD}\rangle_i\mathsf{K}_i\,\phi) \to \tag{42}$$

$$\to \langle\mathsf{MD}\rangle_i(\mathsf{K}_i(\mathsf{Mo}_i\,\phi \;\wedge\; \phi) \;\wedge\; \mathsf{Cocomo}_i(\mathsf{Mo}_i\,\phi \;\wedge\; \phi)). \tag{43}$$

None of that adds up to a counterexample to (38), whose validity in learning semantics remains an open question. But the interest of that question is merely technical, because every CLM can be augmented with an atomic sentence diag with the semantics "i believes neither me nor my negation":

$$\|\mathsf{diag}_i\|^t_{\mathfrak{M}_{t^*}} = \{w \in W : L_{c_{i,w,t^*}}(s_{i,w}|t, @_{t^*}\,\mathsf{diag}) \neq 1 \;\wedge\; L_{c_{i,w,t^*}}(s_{i,w}|t, @_{t^*}\neg\mathsf{diag}) \neq 1\}.$$

Then neither diag nor $\neg\mathsf{diag}$ is knowable, but one side or the other must be true, so (38) is valid in each such model. There is nothing empirically or cognitively ineffable about diag—its truth supervenes in a concrete, computational way on what i's learning program does in response to inputs. The restriction of proposition 2 to objective statements fends off such examples.

More relevantly for the topics that follow, learning semantics allows for the construction of monsters like diag that invalidate just about any standard-looking thesis of epistemic logic: e.g., "i does not believe that she knows me". The purpose of the doxastic stability operator S_i is to protect general theses of epistemic logic from the self-referential onslaught. Under the hypothesis that $\mathsf{S}_i\,\phi$ obtains, knowledge, learning and having learned are preserved under counterfactual changes of method that do not modify the agent's current learning disposition with respect to ϕ.

Proposition 3. *Suppose that* $u \in \|\mathsf{S}_i\,\Delta\|^{t^*}_{\mathfrak{M}_{t^*}}$ *and* $d \equiv_\Delta c_{i,u,t^*}$ *and* $\phi \in \Delta$*. Then:*

$$u \in \|\mathsf{K}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;\Rightarrow\; u[d/i, t^*] \in \|\mathsf{K}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}};. \tag{44}$$

and similarly for $\tilde{\mathsf{K}}_i$*,* $\tilde{\mathsf{L}}_i$*,* $\tilde{\mathsf{Led}}_i$ *and* Led_i*.*

[14] Alternative learning strategies within the same agent are a familiar theme in the epistemology literature—e.g., (Nozick 1981).

13 Epistemic Logic Redux

Here is a standard menu of potential axioms of modal epistemic logic:

$$\begin{array}{lll}
\text{N} & : & \mathsf{K}_i\,\phi,\ \text{if}\ \models \phi; \\
\text{K} & : & \mathsf{K}_i(\phi \to \psi) \to (\mathsf{K}_i\,\phi \to \mathsf{K}_i\,\psi); \\
\text{T} & : & \mathsf{K}_i\,\phi \to \phi; \\
\text{B} & : & \phi \to \mathsf{K}_i\neg\mathsf{K}_i\neg\phi; \\
4 & : & \mathsf{K}_i\,\phi \to \mathsf{K}_i\mathsf{K}_i\,\phi; \\
0.2 & : & \neg\mathsf{K}_i\neg\mathsf{K}_i\,\phi \to \mathsf{K}_i\neg\mathsf{K}_i\neg\phi; \\
0.3 & : & \mathsf{K}_i(\mathsf{K}_i\,\phi \to \mathsf{K}_i\,\psi) \vee \mathsf{K}_i(\mathsf{K}_i\,\psi \to \mathsf{K}_i\,\phi); \\
0.4 & : & \phi \to (\neg\mathsf{K}_i\neg\mathsf{K}_i\phi \to \neg\mathsf{K}_i\,\phi); \\
5 & : & \neg\mathsf{K}_i\,\phi \to \mathsf{K}_i\neg\mathsf{K}_i\,\phi.
\end{array}$$

In conventional possible worlds models for epistemic logic, world accessibility is just a formal trick for assigning a propositional *knowledge state* to agent i in world w as follows:

$$K_{w,i,t} \quad = \quad \{u \in W : R_{i,t}(w,u)\}.$$

The usual idea is to select plausible principles from the menu and then to *impose* them on the assigned knowledge states. Modal semantics then serves as a silent bookkeeper that faithfully manages the iteration of operators subject to those assumptions. For example, principle T says, plausibly, that knowledge is true. In standard possible worlds models, that corresponds to the imposition of reflexivity on the model's accessibility relation. Learning semantics also validates T in its standard form:

$$\text{T:} \qquad \models \mathsf{K}_i\,\phi \to \phi. \tag{45}$$

The rest of the principles on the list are plainly false, unless they are re-interpreted vaguely in terms of abilities, obligations, or ideals. But then it would be better to distinguish knowledge, itself, from the additional *je ne sais quoit* in order to shine a cold, logical light on both and on their interactions. Conditional feasibility $\langle\mathsf{MD}]\!\!\to_{i,\phi}$ provides one clear and plausible interpretation. It allows one to say that there exists some effective way for i to modify her learning method that is guaranteed to achieve the situation described in the consequent given that the antecedent is true. The question addressed in the following section is which, if any, of the traditional candidate axioms is valid under that interpretation, and under what restrictions.

13.1 Deductive Cogency

Let Δ be a finite set of premises and let Γ be a set of conclusions. Suppose that Δ implies Γ, in light of i's information. Maybe she knows neither. But is there any concrete, inferential disposition i could set up in herself to guarantee that if she knows the premises in Δ then she knows the conclusions in Γ as well? Yes, if the premises are inferentially stable, for learning semantics validates the following principle for finite, disjoint Δ, $\Gamma \subseteq \mathbf{L}_{\mathsf{BIT}}$ and for arbitrary, finite superset Δ' of Δ that is disjoint from Γ:

$$\text{FD:} \qquad \models \ (\mathsf{S}_i\Delta' \ \wedge \ [\,\mathsf{I}\,]_i\,(\Delta \to \Gamma) \ \wedge \ \mathsf{K}_i\Delta) \ \ \langle \mathsf{MD}] \!\!\to_{i,\Delta'} \ \mathsf{K}_i\Gamma; \tag{46}$$

When $\Delta = \emptyset$ and $\Gamma = \{\phi\}$, thesis (46) collapses to a feasible version of the rule N of necessitation:

$$\text{FN:} \qquad \models \ [\,\mathsf{I}\,]_i\,\phi \ \langle \mathsf{MD}] \!\!\to_{i,\Delta'} \ \mathsf{K}_i\Gamma. \tag{47}$$

When $\Delta = \{\psi, \ \psi \to \phi\}$ and $\Gamma = \{\phi\}$, thesis (46) collapses to a feasible version of the standard axiom K:

$$\text{FK:} \qquad \models \ (\mathsf{S}_i\Delta' \ \wedge \ \mathsf{K}_i\,\psi \ \wedge \ \mathsf{K}_i(\psi \to \phi)) \quad \langle \mathsf{MD}] \!\!\to_{i,\Delta'} \ \mathsf{K}_i\phi. \tag{48}$$

One may not infer rashly from FN and FK, as one may from the corresponding, traditional axioms N and K, that the knowledge of i is closed under logical consequence or even that it might someday be. The extension of knowledge by deduction must proceed, as it does in the real world, by dint of feasible, cognitive exertion. The local inferential modification that witnesses thesis (46) is *pure deductive inference*—believing the conclusions in Γ if and *only* if one believes the premises in Δ and never believing the negation of any conclusion in Γ. Then convergence to correct belief that Δ in the actual world results in convergence to true belief in Γ in the actual world and guaranteed, eventual avoidance of error regarding the premises in Δ results in guaranteed, eventual avoidance of error regarding conclusions in Γ. In that sense, pure deductive inference makes knowledge of Γ *epistemically parasitic* on knowledge that Δ. If the parasitic relationship is disrupted because i has independent reasons for believing some conclusion $\gamma \in \Gamma$, i might be disposed to fall into error with respect to γ infinitely often in some possible worlds compatible with current information, so it is crucial that i's *only* reason for believing Γ is its deducibility from Δ.

The validity of (46) is closely bound to allowance for serendipity. It has already been shown in terms of $\mathsf{G}\,\mathsf{p}_k$ and $\mathsf{GF}\,\mathsf{p}_k$ that (46) fails for learning:

$$\text{Thesis (46) is invalid with } \tilde{\mathsf{L}}_i, \tilde{\mathsf{Led}}_i, \mathsf{Led}_i \text{ in place of } \mathsf{K}_i. \tag{49}$$

Serendipity raises a cautionary moral about the role of deduction in natural science. The world of science is a “dappled” pastiche of mutually incompatible models and theories and missed connections (Cartwright 1999). Heisenberg and Schrödinger even battled over logically equivalent hypotheses, each of which was rigorously tested over distinct domains of phenomena.[15] When contradictions are found, scientists steer around them until some

[15]For a version of the history, cf. (van der Werden 1973). Learning semantics allows for the possibility that each scientist knew his own formulation of quantum mechanics at the same time he disputed the competing formulation. Even neighborhood semantics (Scott 1970), which models belief as a set of propositions, cannot model that situation.

other experts resolve them. When new logical connections are found between formerly disparate research programs, caution is exercised regarding the drawing of inferences from one program to the other until they are cross-checked by new data. Learning semantics explains that logical conservatism. For suppose that there are two independent research programs studying hypotheses ϕ and ψ, respectively, on the basis of entirely disparate phenomena. Suppose that the current articulation of ϕ is standing up well, but ψ looks bad—its last five articulations were refuted and the current one is in trouble. What to do? Inferring ψ from ϕ would generate new knowledge that ψ from knowledge that ϕ if inquiry whether ϕ has culminated. But if inquiry whether ψ has culminated in knowledge that $\neg\psi$, then inferring ψ from ϕ would *destroy* knowledge that $\neg\psi$. The contrapositive inference from $\neg\psi$ to $\neg\phi$ is fraught with a similar risk of destroying knowledge that ϕ. Hyper-intensional refusal to fire either inference is guaranteed to preserve knowledge of whichever hypothesis is known and leaves the door open to empirical evidence to resolve the conflict—hardly a slam-dunk argument for the ideal of deductive closure in empirical science.

13.2 KK

Suppose that i knows that ϕ. Evidently, she may fail to know that she knows that ϕ—she may not even conceive of the question whether she knows that ϕ unless she is challenged. Or ϕ may say "i does not believe that she knows me". But inattention and Moorean tricks aside, is i even *capable* of knowing that she knows, even though no bell rings (James 1896) when inductive inquiry succeeds? The answer may appear to be negative:

> ...[Learning in the limit] does not entail that [the learner] knows he knows the answer, since [the learner] may lack any reason to believe that his hypotheses have begun to converge. (Martin and Osherson 1998).

Learning semantics delivers a strong, positive verdict: i cannot know *infallibly* that she knows inductively, but there is an easy and natural inferential strategy i can adopt to know *inductively* that she knows inductively that ϕ, and so on, to arbitrary iterations. Define iterated knowledge as follows:

$$\begin{aligned} \mathsf{K}_i{}^0\phi &:= \phi; \\ \mathsf{K}_i{}^{k+1}\phi &:= \mathsf{K}_i\mathsf{K}_i{}^k\phi. \end{aligned}$$

Define the following sets of sentences:

$$\begin{aligned} K_i^k(\phi) &= \{\mathsf{K}_i{}^{k'}\phi : k' \leq k\}; \\ K_i^\omega(\phi) &= \bigcup_{k\in\mathbb{N}} K_i^k(\phi). \end{aligned}$$

Then for each finite Δ containing ϕ and disjoint from K_i^ω, we have:[16]

$$\text{F4}^*: \quad \models \ (\mathsf{S}_i\Delta \ \wedge \ \mathsf{K}_i\,\phi) \ \langle\mathsf{MD}]\!\!\rightarrow_{i,\Delta} \ K_i^\omega(\phi). \tag{50}$$

As a consequence, we have the following, feasible version of the standard (infeasible) reflection principle 4, for each k:

$$\text{F4}: \quad \models \ (\mathsf{S}_i\Delta \ \wedge \ \mathsf{K}_i\,\phi) \ \langle\mathsf{MD}]\!\!\rightarrow_{i,\Delta} \ \mathsf{K}_i^{\,k}\,\phi. \tag{51}$$

A simple inferential strategy that witnesses (51) is for i to believe at t that she knew that ϕ at t^* if she never stopped believing that ϕ from t^* until t and to believe that did not know that ϕ if the alternative case obtains. That inference is intuitive: if i remembers that she retracted ϕ between t^* and the current time t, then the retraction shakes her confidence that she knew that ϕ already at t^*. Otherwise, from i's viewpoint, she had persuasive evidence for ϕ at t^* and nothing in particular has dissuaded her since then, so of course she thinks she knew that ϕ at t^*.

In contrast to the situation for deductive closure, learning that one is learning is easy—learning implies that it is determined that one is learning and whatever is determined can be learned by believing it no matter what and never believing its negation. Having learned whether one has learned whether and having learned that one has learned that are both valid by the same inferential strategy invoked to validate (51). So we have:

$$\text{Thesis (51) remains valid with } \tilde{\mathsf{K}}_i, \tilde{\mathsf{L}}_i, \tilde{\mathsf{Led}}_i, \mathsf{Led}_i \text{ in place of } \mathsf{K}_i. \tag{52}$$

13.3 The Insidious Unknown Unknown

For Plato (1949), the least flattering epistemic condition is *hubris*: failure even to know that one does not know. The next step is to know, at least, that one does not know. That motivates one to enter the path of inquiry, or seeking to know, which Plato optimistically assumed would lead both to knowledge and to knowledge that one knows at the moment of infallible "recollection". It has been shown that learning is guaranteed to culminate in inductive knowledge and that serendipity allows even for acquisition of knowledge of unlearnable truths. Furthermore, it has just been shown that one can come to know that one knows at the very same time, just as Plato proposed, even without Plato's commitment to infallible recollection. But what about Plato's fateful first step—coming to know that one does not know? In that case, learning semantics delivers a negative verdict:

$$\text{F5}': \quad \mathfrak{N}_{t^*} \not\models (\mathsf{S}_i\,\phi \ \wedge \ \neg\mathsf{K}_i\,\phi) \ \langle\mathsf{MD}]\!\!\rightarrow_{i,\phi} \ \mathsf{K}_i\neg\mathsf{K}_i\,\phi. \tag{53}$$

[16]Strictly speaking, one must restrict $K_i^\omega(\phi)$ to a finite set for the statement to be well-formed, but the proof of validity works for the unrestricted version.

The convergence required for knowing that one knows parasitically tracks the convergence of knowledge itself. But failure to know inductively may be witnessed only by failure to converge in the distant future, and the requirement to have converged already to true belief that one will not converge in the future due to unforseen surprises occasions the problem of induction, with which we began. For example, suppose that i has seen enough evidence to convince her that $\mathsf{G}\,\mathsf{p}_k$ until such time as some non-k input is received, at which time she drops her belief that $\mathsf{G}\,\mathsf{p}_k$. Call i's learning method c. Method c yields inductive knowledge that $\mathsf{G}\,\mathsf{p}_k$ in the constantly k world w in which $\mathsf{G}\,\mathsf{p}_k$ is true. Now, suppose that i possesses some magical inferential technique h that guarantees i knowledge now that she does not know that $\mathsf{G}\,\mathsf{p}_k$ if she does not know that $\mathsf{G}\,\mathsf{p}_k$ and that the inferential technique manages to avoid altering i's beliefs whether $\mathsf{G}\,\mathsf{p}_k$. In particular, learning method $h(c)$ must be guaranteed to yield knowledge immediately that c does not produce knowledge that $\mathsf{G}\,\mathsf{p}_k$. Let w_m be the "grue-like" world in which i receives input k until stage m and $k+1$ thereafter. Statement $\mathsf{G}\,\mathsf{p}_k$ is false in w_m, so $h(c)$ stabilizes to belief that $\neg\mathsf{K}_i\mathsf{G}\,\mathsf{p}_k$ immediately in w_m, for *every* m. So $h(c)$ converges to $\neg\mathsf{K}_i\mathsf{G}\,\mathsf{p}_k$ in world w, since w_m agrees empirically with w until m. But, ironically, i *knows* that $\mathsf{G}\,\mathsf{p}_k$ in w because $\mathsf{G}\,\mathsf{p}_k$ is objective in $\mathfrak{N}_{t^*}$ and h holds i's beliefs whether ϕ fixed. So $h(c)$ fails to avoid error in the limit whether $\neg\mathsf{K}_i\mathsf{G}\,\mathsf{p}_k$.

In fact, slight variants of the reflective problem of induction just described suffice to invalidate feasible versions of all of the proposed axioms between .4 and 5, so among the standard axioms, only T, FD, and F4 are valid in learning semantics:

$$\text{FB:}\quad \mathfrak{N}_{t^*} \not\models (\mathsf{S}_i\,\phi \wedge \neg\phi)\ \langle\mathsf{MD}]\!\!\rightarrow_{i,\phi}\ \mathsf{K}_i\neg\mathsf{K}_i\,\phi; \tag{54}$$

$$\text{F.2:}\quad \mathfrak{N}_{t^*} \not\models (\mathsf{S}_i\,\phi \wedge \neg\mathsf{K}_i\neg\mathsf{K}_i\neg\phi)\ \langle\mathsf{MD}]\!\!\rightarrow_{i,\phi}\ \mathsf{K}_i\neg\mathsf{K}_i\,\phi; \tag{55}$$

$$\begin{aligned}\text{F.3:}\quad \mathfrak{N}_{t^*} \not\models\ & ((\mathsf{S}_i\,\phi \wedge \mathsf{S}_i\,\phi \wedge \mathsf{K}_i\neg\mathsf{K}_i\,\phi)\ \langle\mathsf{MD}]\!\!\rightarrow_{i,\phi,\psi}\ \mathsf{K}_i\neg\mathsf{K}_i\,\psi)\ \vee \\ & \vee ((\mathsf{S}_i\,\phi \wedge \mathsf{S}_i\,\phi \wedge \mathsf{K}_i\neg\mathsf{K}_i\,\psi)\ \langle\mathsf{MD}]\!\!\rightarrow_{i,\phi,\psi}\ \mathsf{K}_i\neg\mathsf{K}_i\,\phi);\end{aligned} \tag{56}$$

$$\text{F.4:}\quad \mathfrak{N}_{t^*} \not\models (\mathsf{S}_i\,\phi \wedge \neg\phi \wedge \neg\mathsf{K}_i\neg\phi)\ \langle\mathsf{MD}]\!\!\rightarrow_{i,\phi}\ \mathsf{K}_i\neg\mathsf{K}_i\,\phi. \tag{57}$$

It suffices to let $\phi = \mathsf{G}\,\mathsf{p}_k$ and $\psi = \mathsf{G}\,\mathsf{p}_{k'}$, for distinct k, k'.

The same examples refute the corresponding versions of (53-57) for knowing whether, having learned whether, and having learned that:

$$\text{Theses (53-57) remain invalid with } \tilde{\mathsf{K}}_i, \tilde{\mathsf{Led}}_i, \mathsf{Led}_i \text{ in place of } \mathsf{K}_i. \tag{58}$$

However, it is trivially feasible for i to be learning whether i is not learning whether ϕ when i is not learning whether ϕ—it suffices for i to believe that she is not learning whether ϕ no matter what, since learning begins with operator $[\mathsf{D}]_i$:

$$\text{F5L:}\quad \models (\mathsf{S}_i\,\phi \wedge \neg\tilde{\mathsf{L}}_i\,\phi)\ \langle\mathsf{MD}]\!\!\rightarrow_{i,\phi}\ \tilde{\mathsf{L}}_i\neg\tilde{\mathsf{L}}_i\,\phi. \tag{59}$$

14 Joint Inductive Knowledge

Plato's original question in the *Meno* (1949) was not what knowledge *is*, but whether virtue can be *taught*. He merely assumed that knowledge can be taught, but when knowledge is inductive, that answer is not so obvious. Of course, a knowledgable expert can *exhibit* her inductive knowledge to her pupils, but can she *transfer* it to her pupils, so that they *know* inductively what she knows inductively, rather than merely what she believes? The transmission of inductive knowledge from expert to pupil throws a skeptical curve of its own: it is too strict to require that the pupil have access to information concerning the expert's learning method—Physics 101 does not presuppose Psychology 601. But the proposed, individualistic semantics for inductive knowledge requires something like that, since the pupil must avoid error in the limit no matter what learning method the mentoring expert might be using. That strict standard makes sense, if the pupil is a cognitive scientist studying her mentor's cognitive architecture or if the pupil is playing a competitive game against her mentor, but when the pupil is merely learning from her mentor, it is more natural to allow her knowledge to supervene *jointly* on her own learning strategy and on her mentor's. In that spirit, this section presents an alternative, *joint* version of learning semantics that is friendlier to joint epistemic efforts like education. In the following section, it is shown how it is jointly feasible for the expert and a room full of pupils to acquire common knowledge of the expert's inductive knowledge.

Let $w \in W$, $\mathbf{c}_{w,t} = (c_{w,1,t}, \ldots c_{w,N,t})$ and $\mathbf{d} \in C^N$. Then let $u[\mathbf{d}/t]$ denote the result of substituting $\mathbf{d}$ for $\mathbf{c}_{w,t}$ in w at t. A *joint* CLM satisfies the following, joint invariance postulate, for each $i \in G$, $w \in W$, $\mathbf{d} \in C^N$, and $t \in T$:

$$s_{i,w}|t \quad = \quad s_{i,w[\mathbf{d}/t']}|t. \tag{60}$$

Joint information and determination are defined as follows:

$$\begin{aligned} I_{G,w,t} &= \bigcup_{i \in G} I_{i,w,t}; \\ D_{i,w,t} &= \{u \in I_{G,w,t} : \mathbf{c}_{u,t} = \mathbf{c}_{w,t}\}; \end{aligned}$$

with corresponding operators:

$$\begin{aligned} \|[\mathsf{I}]_G\,\phi\|^t_{\mathfrak{M}_{t^*}} &= \{w \in W : I_{G,w,t} \subseteq \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}}\}; \\ \|[\mathsf{D}]_G\,\phi\|^t_{\mathfrak{M}_{t^*}} &= \{w \in W : D_{G,w,t} \subseteq \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}}\}. \end{aligned}$$

Joint information is weaker than individual information, but joint determination compensates, somewhat, by holding everyone's method fixed, as though they were a team achieving a single goal. Joint information and determination are no longer S5 operators. The transmission of common knowledge assumes that property, so it is useful to have a concise notation for expressing it in the object language:

$$\|\mathsf{IS5}_G\|^{t^*}_{\mathfrak{M}_{t^*}} = \{w \in W : (\forall u \in I_{G,w,t^*})\ I_{G,u,t^*} = I_{G,u,t^*}\}.$$

Define joint inductive knowledge for i as before, but with joint determination in place of personal determination:

$$\mathsf{K}_{G,i}\,\phi \quad := \quad \mathsf{GC}_i\,\phi \;\wedge\; [\mathsf{D}]_G\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi.$$

Joint methodological possibility expresses the existence of a methodological coordination among the agents that brings about ϕ:

$$\|\langle\mathsf{M}\rangle_G\,\phi\|^t_{\mathfrak{M}_{t^*}} = \{w \in W : (\exists\mathbf{d} \in C^N)\; w[\mathbf{d}/t^*] \in \|\phi\|^t_{\mathfrak{M}_{t^*}}\}.$$

To define joint conditional feasibility, let $\mathbf{h} = (h_1, \ldots, h_N)$ be an N-sequence of total recursive functions, let $\mathbf{h}(\mathbf{c}) = (h_1(c_1), \ldots, h_N(c_N))$, and let $\boldsymbol{\Delta}$ be an N-sequence of finite subsets of $\mathbf{L}_{\mathsf{BIT}}$. Define:

$$\mathbf{c} \equiv_{\boldsymbol{\Delta}} \mathbf{d} \quad\Leftrightarrow\quad (\forall i \in G)\; c_i \equiv_{\Delta_i} d_i.$$

Say that $\mathbf{h}$ *holds* $\boldsymbol{\Delta}$ *fixed* if and only if $\mathbf{c} \equiv_{\boldsymbol{\Delta}} \mathbf{h}(\mathbf{c})$, for all $\mathbf{c} \in C^N$. Then let $\|\psi\;\langle\mathsf{MD}]\!\!\rightarrow_{G,\boldsymbol{\Delta}}\;\phi\|^t_{\mathfrak{M}_{t^*}}$ denote the set of all $w \in W$ for which there exists N-sequence $\mathbf{h}$ of total recursive functions such that $\mathbf{h}$ holds $\boldsymbol{\Delta}$ fixed and for all $u \in I_{G,w,t}$:

$$u \in \|\psi\|^t_{\mathfrak{M}_{t^*}} \quad\Rightarrow\quad u[\mathbf{h}/t^*] \in \|\phi\|^t_{\mathfrak{M}_{t^*}}.$$

It remains only to define a joint version of inferential stability. Let $w \in \|\mathsf{S}_{G,i}\boldsymbol{\Delta}\|^t_{\mathfrak{M}_{t^*}}$ hold if and only if for all $\mathbf{d} \in C^N$ such that $\mathbf{c}_{w,t^*} \equiv_{\boldsymbol{\Delta}} \mathbf{d}$ and for all $u \in D_{G,w,t^*}$, $t \geq t^*$, and $\delta \in \Delta_i$:

$$\begin{aligned} u \in \|\delta\|^{t^*}_{\mathfrak{M}_{t^*}} &\quad\Leftrightarrow\quad u[\mathbf{d}/t^*] \in \|\delta\|^{t^*}_{\mathfrak{M}_{t^*}}; && (61)\\ L_{c_{i,u,t^*}}(s_{i,u}|t, @_{t^*}\,\delta) &\quad=\quad L_{d_i}(s_{i,u[\mathbf{d}/t^*]}|t, @_{t^*}\,\delta); && (62)\\ L_{c_{i,u,t^*}}(s_{i,u}|t, @_{t^*}\neg\delta) &\quad=\quad L_{d_i}(s_{i,u[\mathbf{d}/t^*]}|t, @_{t^*}\neg\delta). && (63)\end{aligned}$$

A joint version of proposition 3 holds:

Proposition 4. *Suppose that $\phi \in \Delta_i$ and $u \in \|\mathsf{S}_{G,i}\boldsymbol{\Delta}\|^{t^*}_{\mathfrak{M}_{t^*}}$ and let $\mathbf{d} \in C^N$ satisfy $\mathbf{d} \equiv_{\boldsymbol{\Delta}} \mathbf{c}_{u,t^*}$. Then:*

$$u \in \|\mathsf{K}_{G,i}\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \quad\Rightarrow\quad u[\mathbf{d}/t^*] \in \|\mathsf{K}_{G,i}\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{64}$$

15 Common Inductive Knowledge

Given the joint perspective outlined in the preceding section and some basic assumptions about how the expert and pupils interact, it is jointly feasible for the expert and her pupils to acquire the expert's inductive, theoretical knowledge that ϕ. It suffices that the pupil believe that ϕ if the expert does and suspend belief that ϕ otherwise—just as a scientist is entitled to suspend belief when the current articulation of her hypothesis is refuted. The pupil thereby becomes an epistemic parasite of the expert, just as the expert becomes an epistemic parasite of herself when she infers deductive consequences of what she knows. Educated pupils can serve, in turn, as experts, resulting in a cascade of joint scientific knowledge through the population—as long as the pupils have knowledgable instructors.

It is a further question whether the pupils and the expert jointly know that they know, know that they know that they know, etc, all the way to joint, *common inductive knowledge* that ϕ. Define joint, *mutual knowledge* to degree n as follows:

$$\begin{aligned} \mathsf{K}_G{}^0\, \phi &:= \phi; \\ \mathsf{K}_G{}^{k+1}\, \phi &:= \bigwedge_{i \in G} \mathsf{K}_{G,i} \mathsf{K}_G{}^k\, \phi. \end{aligned}$$

Define *common knowledge* that ϕ as the set of sentences:

$$K^{\omega}_G(\phi) \;=\; \{\mathsf{K}_G{}^k\, \phi : k \in \mathbb{N}\}.$$

It is plausible that a completely trusted, infallible, public announcement that ϕ can generate common knowledge that ϕ. It is less obvious that fallible, common *inductive* knowledge is feasible in a room full of pupils who place blind trust in their instructor. Learning semantics yields a positive result based on epistemic parasitism and serendipity, in close analogy to the preceding proof of positive introspection.

The expert must communicate with the pupils in some way in order to instruct them. It is assumed that the pupil somehow receives sufficient information to correctly believe whether the expert believes that ϕ—it is not assumed that the expert actually causes the pupil to believe that ϕ, although that is always nice. Let $e \in G$ be the designated expert and let $G_- = G \setminus \{e\}$ be the set of pupils. Define the operator "*e teaches* the pupils in G_- whether ϕ" as follows:

$$\mathsf{T}_{G,e}\, \phi \;:=\; \bigwedge_{j \in G_-} [\mathsf{I}]_G \mathsf{G}\tilde{\mathsf{C}}_j \mathsf{B}_e\, \phi.$$

Now it is possible to state and prove the *joint feasibility of common inductive knowledge* thesis, which is valid if Δ_e contains ϕ and Δ_i is disjoint from $K^{\omega}_G(\phi)$, for all $i \in G$:

$$\text{FC:} \quad \models (\mathsf{IS5}_G \;\wedge\; \mathsf{T}_{G,e}\, \phi \;\wedge\; \mathsf{S}_{G,e}\boldsymbol{\Delta} \;\wedge\; \mathsf{K}_{G,e}\, \phi) \;\; \langle \mathsf{MD}] \!\!\rightarrow_{G,\boldsymbol{\Delta}} \; K^{\omega}_G(\phi). \tag{65}$$

Although the FC principle concerns common inductive knowledge generated and promulgated by a single expert, it sets the stage for a series of similar results that involve common inductive knowledge generated through the cooperation of a team of experts—a topic of current interest in social epistemology (e.g., Mayo-Wilson 2011).

In dynamic epistemic logic, there are models in which public announcements generate common knowledge of what has been announced (van Benthem 2010). But how do public announcements result in anything more than common knowledge of the fact that the announcement was made? Plausibly, common knowledge of what has been announced is common inductive knowledge grounded in the community's joint strategy to disbelieve sources caught in inconsistencies or lies. One potential extension of FC is to validate the possibility of common inductive knowledge of what is reported in a public announcement in such a model.

A familiar assumption in game theory is that the agents have common knowledge of rationality (Aumann 1995). But how is such knowledge possible? Recall the game-theoretic model described in section 6 above. Violation of the kth level of mutual rationality are detectable by horizontal play in a centipede game of corresponding length. If all of the agents have the disposition to continue playing down at the first move in ever longer centipede games, learning semantics provides a determinate, explanatory, account of how common knowledge of rationality is jointly feasible in such a group. And if the agents are all disposed to cooperate a bit by playing sideways, the group can just as easily develop inductive common knowledge of partial cooperation![17]

16 Extensions

The proposed semantics of inductive knowledge is schematic and lenient by design, which opens the door to potentially fruitful refinements and restrictions, a few of which are sketched in a preliminary way below.

16.1 Inductive Knowledge of Stochastic Theories

Most real scientific applications are statistical, so probabilities should be added to learning semantics. Think of $s_{1,w,t}, \ldots, s_{N,w,t}$ as random variables jointly distributed in w at t according to $\mathbf{P}_{w,t}$. Think of $\mathbf{P}_{w,t}$ as the underlying, joint probability distribution in world w at time t. Of course, chance dispositions can change—even coins wear out if they are flipped too many times. Inductive learning from stochastic inputs is standardly defined in terms of convergence in probability rather than actual convergence. Since the past is settled, the inputs presented prior to t' in w are assigned probability 1 by $\mathbf{P}_{w,t}$. Therefore, the chance of error at t^* is tacitly referred to some prior *reference time* t^{**}—e.g., the time at which the experimental design for the current epistemic context (w, t^*) was set up. Let $\mathsf{K}_i\, \phi$ be true in w at t^* relative to reference time t^{**} if and only if (i) the learning method c_{i,w,t^*} of i correctly believes that ϕ in w at t^* in light of the actual sample inputs received in w; (ii) according to $\mathbf{P}_{w,t^{**}}$, for every $t \geq t^*$, method c_{i,w,t^*} is probably correct that ϕ at t^* and its probability of correctness rises to 1 as t increases and (iii) according to $\mathbf{P}_{w,t^{**}}$, the chance that i adopts a method at t^* that commits an error whether ϕ at t

[17]This application was suggested via personal communication by Jennifer Juhn.

drops (non-monotonically) to 0.[18]

It is anticipated that stochastic learning semantics validates principles very similar to those validated above, but there is an important difference. In the non-probabilistic learning semantics developed above, the outcomes of future random events are inductively knowable (Hendricks 2001). In the stochastic semantics just sketched, they are unknowable, since convergence in probability is grounded in chances operative at reference time t^{**} prior to the predicted events and a high chance in w at t^{**} of believing what will happen in a future coin flip implies failure to drive the chance of error toward 0 in alternative worlds. That does not rule out inductive knowledge of stochastic laws and theories, since they predict only chances of random outcomes, not the outcomes themselves.

16.2 Empirical Justification and Ockham's Razor

The proposed learning semantics also says nothing about confirmation by evidence, which is traditionally identified with the "justification" condition for inductive knowledge. One expedient response is simply to tack on some ad hoc condition of "sufficient confirmation", but then the requirement is not explained in terms of truth-conduciveness. Of course, information is required to converge to the truth, but no amount of information is "sufficient" to do so—semantically, the question whether $\mathsf{G}\,\mathsf{p}_k$ remains isomorphic to itself as one restricts it to any compatible, finite data set. Conditional probabilities may converge to 1 as evidence accumulates, but those probabilities are just opinions, and i is already has her opinions. As for the Bayesian ideal of coherence, learning semantics' allowance for scientific knowledge in the face of outright inconsistency is an intended improvement.

But scientists do have systematic, short-run preferences for simpler or more cross-testable theories, a disposition known as *Ockham's razor*, and it would be nice to have a unified, learning semantic explanation of that preference. Of course, a systematic bias toward simple theories can be encoded in prior probabilities imposed upon the semantics, but that does not explain why one should have such a bias (Glymour 1980). It would be better to *explain* Ockham's razor in terms of truth-conduciveness.

Here is a way to do so, following (Kelly 2010). It is plausible, from the viewpoint of virtue epistemology (Sosa 1980), that "knowledge" is a socially sanctioned encomium that motivates improved truth-conduciveness in the community. Then one would expect the standards for knowledge to shift in order to provide achievable aims in alternative epistemic contexts of varying intrinsic difficulty. Truth-conduciveness is just learning ability. Recall that learning was dropped as a condition for knowing because $\mathsf{GF}\,\mathsf{p}_k$ is not learnable but follows deductively from learnable $\mathsf{G}\,\mathsf{p}_k$. From a virtue epistemological

[18]Method c_{i,w,t^*} may have been chosen after peeking at the sample taken from t^{**} to t^*. That is the actual method generating i's actual belief, so it is the one that matters for knowledge at t^*. But the peeking may have affected i's probability of error in other worlds. Condition (iii) therefore considers the possibility of dependence of the i's method on the sample.

viewpoint, that is not a reason to drop the learning requirement entirely—learning should be required whenever it is feasible. Furthermore, learning, itself, is not a one-size-fits-all concept. Virtuous pursuit of truth, like virtuous pursuit of anything, should be direct and efficient—the best designed anti-aircraft missile may need to swerve to intercept an evasive target, but a missile that chases its own tail for no reason is surely defective. The epistemological equivalent of swerving is retraction of one's prior belief whether ϕ in light of new information. So the virtuous pursuit of truth whether ϕ, and, hence, knowledge whether ϕ, should require that i learn with *minimal retractions*, if learning is possible at all. Otherwise, serendipitous knowledge in the sense defined above suffices. Retraction minimization makes sense for agents in CLMs, but $\mathbf{L}_{\mathsf{BIT}}$ must be enriched in order to express it.

The method of believing that $\mathsf{G}\,\mathsf{p}_k$ until it is refuted and disbelieving it thereafter results in knowledge that $\mathsf{G}\,\mathsf{p}_k$ if $\mathsf{G}\,\mathsf{p}_k$ is true, because the method learns with one retraction (from $\mathsf{G}\,\mathsf{p}_k$ to $\neg\mathsf{G}\,\mathsf{p}_k$) and no learning method learns whether $\mathsf{G}\,\mathsf{p}_k$ with fewer retractions. It remains true that pure deduction of $\mathsf{GF}\,\mathsf{p}_k$ from knowledge that $\mathsf{G}\,\mathsf{p}_k$ results in knowledge that $\mathsf{GF}\,\mathsf{p}_k$ because $\mathsf{GF}\,\mathsf{p}_k$ is not learnable. For a more interesting example, consider the statement that exactly one of inputs 0, 1 will occur in the future:

$$\delta \;:=\; (\mathsf{F}\,\mathsf{p}_0 \;\vee\; \mathsf{F}\,\mathsf{p}_1) \;\wedge\; (\mathsf{G}\neg\mathsf{p}_0 \;\vee\; \mathsf{G}\neg\mathsf{p}_1). \tag{66}$$

Statement δ is learnable by saying no until either 0 or 1 is observed, by saying yes until the alternative digit is observed, and by converging to no thereafter, which adds up to two retractions ending with no. No better learning performance is possible. Suppose that i believes δ *prior* to seeing either 0 or 1. Then nature can withhold both 0 and 1 until, on pain of not eliminating error in the limit, i is forced to say no. Now nature can show digit 0 and withhold 1 as long as it takes to force i to say yes, and so on, resulting in three retractions ending with no, so i is not optimally truth-conducive (i.e., virtuous), and, hence, does not know. But after seeing the first digit, i is as efficient as anyone who shares her doxastic history, so i knows that δ from that moment onward.

Think of waiting for 0 or for 1 as waiting for an *empirical effect* predicted by δ. To relate the logical idea to real scientific inquiry, consider the following story. Scientist i arbitrarily believes hypothesis ϕ that the standard particle theory is correct, except that one must add one new kind of particle—the platon. The reactivity of the platon is an unknown parameter that must be estimated from the data and that may be arbitrarily small. Suppose that i's theory happens to be true. Suppose, further, that i's laboratory is diligently seeking some predicted effect of the platon by means of a huge detector that is guaranteed to notice some effect of the platon, eventually, with a delay depending on the platon's unknown reactivity. As it happens, i's confidence wavers and she stops believing in the platon if no tell-tale effects are detected within the next four years (think of the "null" results of experimental attempts to detect the ether drift). However, if ϕ is true, the detector does capture a noticeable effect eventually, after which i again believes that ϕ.

Then nature can reveal effects of yet another particle, the sophiston, causing i to retract ϕ and admit the new particle as well. So i says yes, no, yes, no, which adds up to two 3 retractions. Had she refrained from believing in the platon until its tell-tale signs were observed, she would at worst have learned the truth with 2 retractions—no, yes, no. So optimal truth-conduciveness—i.e., epistemic virtue—implies that i should not multiply entities without empirical necessity.

Recall the example of polynomial laws with free parameters, with which we began. Each such law has the same learning complexity as δ—it is learnable with two retractions ending with no. Therefore, it cannot be known until empirical effects bounding each of the law's coefficients away from 0 have been observed. Thus, one can know only the *simplest* such form compatible with experience at any given time. Ockham's razor is not an ad hoc, additional condition for knowledge—it has just been derived from optimal truth-conduciveness.

16.3 Coherence

The thoroughgoing hyper-intensionality of learning semantics is a refreshing change from habitual over-rationalization in epistemic logic, but a bit more coherence should be imposed. It is easy to get a perfect score on a multiple choice test if you get to choose every answer. Something like that can happen in learning semantics. Recall that scientist i can know that the input sequence is ε by guessing that it is ε until ε is refuted. Suppose that scientist i simultaneously believes *every* hypothesis of the form "the input stream is exactly primitive recursive sequence ε", and is disposed to drop each such hypothesis when it disagrees with the data. Suppose, by serendipity, that the true input stream ε is primitive recursive, so the hypothesis corresponding to ε is true. Then i knows that the future will conform to ε, even though i believes every possible primitive recursive input stream compatible with current information. That sounds too easy. Furthermore, for someone as aphasic as i, the very concept of belief is called into question. What would i predict to happen at the next stage? Certainly not what she "knows" will happen, since she cannot pick her known theory out of the raft of her alternative, incompatible beliefs. Science may be incoherent overall, but each of its insular paradigms is coherent enough to generate consensus concerning determinate predictions. So normal science within a paradigm is not trivial the way i's knowledge is, even though science is globally incoherent. In a more sophisticated version of learning semantics, paradigms would be individuated by *questions* and success in a question would require that the agent never believe more than one answer to the question. Other logical foibles would still be permitted.

16.4 Prediction

The predictions of a known law are knowable with serendipity via thesis KD—plausibly, believe a given prediction as long as you believe the law that entails it and drop belief in both if the prediction is refuted. But one can also serendipitously know a given prediction in isolation—make a lucky guess about what will happen at $t \geq t^*$ and retract that guess if it fails to be vindicated at t. That sounds too lenient–it is nothing more than lucky true belief.

Alternatively, predictions are not knowable in advance in the retraction-minimal sense, since one can achieve 0 retractions if one waits until the the predicted event occurs—ironically, they are unknowable because they are *easier* to learn than laws are and, hence, are subject to a higher standard of virtue (a general reason to doubt closure of knowability under logical consequence in virtue epistemology). Hence, thesis FD fails for retraction minimal knowledge, since the predictions of known laws are unknowable. That sounds awkward—some think that only the predictions matter in science and they turn out not to be knowable until it is too late to act upon them!

One response is that predictive knowledge of observable events is different from inductive knowledge of laws and theories. Virtues are dispositions. From that viewpoint, predictive knowledge may imply that one has a disposition to *continue* to correctly predict events of the same type in the future and that one would eventually avoid mistakes regarding predictions of that type in alternative worlds. Then knowing a law implies predictive knowledge of its predictions, but not conversely—a reliable predictive method would suffice for predictive knowledge.

It would be better to have a univocal account of inductive knowledge for theories and for decidable statements, but there may not be one. Prediction is where normative talk adapted to direct perception collides with normative talk adapted to universal statements and scientific method, so it is to be expected that the concept of predictive knowledge is subject to a cross-current of epistemic standards and intuitions.

17 Acknowledgments

This work was supported generously by the John Templeton Foundation, award number 24145. The author has had the basic ideas in mind since the early 1990s and has enjoyed discussions on the topic (in approximately temporal order, and keeping in mind the author's fading memory) with Oliver Schulte, Vincent Hendricks, Stig Andur Pedersen, Nuel Belnap, Johan van Benthem, Horacio Arlo-Costa, Hans Kamp, Fred Dretske, Peter van der Schraaf, Jonathan Vogel, Sherrilyn Roush, Ivan Verano, Hanti Lin, Timothy Williamson, Gregory Wheeler, Alexandru Baltag, Sonja Smets, Gerhard Schurz, Liam Bright, Wesley Holliday, Thomas Icard, Alexandru Radelescu, Clark Glymour, Kevin Zollman, Eric Martin, Konstantin Genin, and Ted Shear, none of whom should be blamed for the outcome. The author is also indebted to Ted Shear for providing detailed corrections on a late draft. Preliminary versions of the system have been presented in talks at Carnegie Mellon University (2006), the Logic and Methodology Workshop at Stanford (2011), and the workshop on Games Interactive Rationality, and Learning in Lund (2012,

2013).

18 Bibliography

Aumann, R. (1995). "Backward Induction and Common Knowledge of Rationality". *Games and Economic Behavior* 8: 619.

Benfey, O. T. (1958). "August Kekulé and the Birth of the Structural Theory of Organic Chemistry in 1848". *Journal of Chemical Education* 35: 21-23.

Carnap, R. (1945). "On Inductive Logic". *Philosophy of Science* 12: 72-97.

Duhem, P. (1914) *La théorie physique son objet et sa structure*, 2nd ed., Paris: Chevalier et Rivire.

Cartwright, N. (1999) "The Dappled World: A Study of the Boundaries of Science". Cambridge: Cambridge University Press.

Church, A. (2009) "Referee Reports on Fitch's 'A Definition of Value'," *New Essays on the Knowability Paradox.* J. Salerno, ed. Oxford: Oxford University Press: 1320.

Fitch, F. (1963). "A Logical Analysis of some Value Concepts". *The Journal of Symbolic Logic* 28:135-142.

Gierasimczuk, N. (2010) *Knowing One's Limits: Logical Analysis of Inductive Inference.* ILLC Dissertation Series DS-2010-11, Institute for Logic, Language, and Computation.

Glymour, C. (1980) *Theory and Evidence.* Princeton: Princeton University Press.

Gold, E. M. (1967). "Language Identification in the limit". *Information and Control* 10:447-474.

Hempel, C. (1945) "Studies in the Logic of Confirmation". *Mind* 213: 1-26.

Hendricks, V. (2001). *The Convergence of Scientific Knowledge.* Dordrecht: Springer.

Holliday, W. (2012). "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism". *Journal of Philosophical Logic*, forthcoming.

Jain, S., D. Osherson, J. Royer, A. Sharma. (1999) *Systems that Learn: An Introduction to Learning Theory*, 2nd ed. New York: Bradford.

James, W. (1896). "The Will to Believe". *The New World*:

Kamp, H. (1971). "Formal Properties of 'Now"'. *Theoria* 27:227-274.

Kant, I. (1782/1787). *Critique of Pure Reason.* P. Guyer and A. Wood, trans. Cambridge: Cambridge University Press. 1997.

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

Kelly, K. T. (2001). "The Logic of Success". *British Journal for the Philosophy of Science, special millennium issue* 51: 639-666.

Kelly, K. T. (2010)"Simplicity, Truth, and Probability", *Handbook for the Philosophy of Statistics.* Prasanta S. Bandyopadhyay and Malcolm Forster eds., Dordrecht: Elsevier.

Laudan, L. (1981). "A Confutation of Convergent Realism". *Philosophy of Science* 48: 19-49.

Martin, E. and D. Osherson (1998) "Elements of Scientific Inquiry". Cambridge: M.I.T. Press.

Mayo-Wilson, C. (2011). "The Problem of Piece-meal Induction". *Philosophy of Science* 78: 864-874.

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

Peirce, C. S. (1878). "How to Make our Ideas Clear". *Popular Science Monthly.* 12: 286-302.

Poincare, H. (1904). *Science and Hypothesis.* New York: The Walter Scott Publishing Co.

Plato (1949). *Meno.* B. Jowett, trans. Indianapolis: Bobbs-Merrill.

Popper, K. R. (1935). *Logik der Forschung.* Vienna: Springer.

Putnam, H. (1963). "'Degree of Confirmation' and Inductive Logic". *The Philosophy of Rudolf Carnap*, ed. A. Schilpp. LaSalle: Open Court.

Pritchard, D. (2007). "Anti-Luck Epistemology". *Synthese* 158: 277-98.

Reichenbach, H. (1949). *The Theory of Probability.* London: Cambridge University Press.

Roush, S. (2007). *Tracking Truth: Knowledge, Evidence, and Science.* Oxford: Oxford University Press.

Roush, S. (2010). "Closure on Skepticism". *Journal of Philosophy* 107: 243-256.

Scott, D. (1970). "Advice in Modal Logic". *Philosophicawl Problems in Logic*, ed. K. Lambert: 143-173.

Sosa, E. (1999). "How to Defeat Opposition to Moore". *Philosophical Perspectives* 13: 141-54.

Sosa, E., (1980). "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge," *Midwest Studies in Philosophy* 5: 325.

Vogel, J. (1987). "Tracking, Closure, and Inductive Knowledge". *The Possibility of Knowledge: Nozick and His Critics.* S. Luper-Foy, ed. Totowa: Rowman and Littlefield.

van Benthem, J. (2011) *Logical Dynamics of Information and Interaction*, Cambridge: Cambridge University Press.

van Benthem, J. (2010) *Modal Logic for Open Minds*, Standord: CSLI Lecture Notes.

van der Werden, (1973) "From Matrix Mechanics and Wave Mechanics to Unified Quantum Mechanics". *The Physicist's Conception of Nature*, J. Mehra, ed. Dordrecht: Reidel. pp. 276-293.

1993. "Verificationism and Non-Distributive Knowledge". *Australasian Journal of Philosophy* 71: 7886.

Williamson, T. (2000). *Knowledge and its Limits.* Oxford: Oxford University Press.

19 Proofs of Propositions

Proof of proposition 2. Just let $L_d(\sigma, \psi)$ return 1 if $\psi = @_{t^*}\,\phi$ and σ is an initial segment of $s_{i,u,t}$ and return 0 otherwise. □

Proof of proposition 3. Abbreviate:

$$\begin{aligned} c &= c_{i,u,t^*}; \\ x &= u[d/i, t^*]. \end{aligned}$$

Assume that $\phi \in \Delta$ and that:

$$d \equiv_\Delta c; \tag{67}$$

$$u \in \|\mathsf{S}_i \Delta\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{68}$$

$$u \in \|\mathsf{K}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{69}$$

From (69) we have:

$$u \in \|\mathsf{GC}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{70}$$

$$y \in \|\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}, \text{ for all } y \in D_{i,u,t^*}. \tag{71}$$

It suffices to show that:

$$
\begin{aligned}
x &\in \|\mathsf{GC}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (72)\\
y &\in \|\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}},\ \text{for all } y \in D_{i,x,t^*}. && (73)
\end{aligned}
$$

From (67-68), we have that:

$$
\begin{aligned}
u \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} &\Leftrightarrow x \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (74)\\
u \in \|\mathsf{G}[\mathsf{B}]_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} &\Leftrightarrow x \in \|\mathsf{G}[\mathsf{B}]_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (75)\\
u \in \|\mathsf{G}\langle\mathsf{B}\rangle_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} &\Leftrightarrow x \in \|\mathsf{G}\langle\mathsf{B}\rangle_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. && (76)
\end{aligned}
$$

So requirement (72) follows from (70).

For requirement (73), let $y \in D_{i,x,t^*}$. Then $s_{i,y}|t^* = s_{i,x}|t^* = s_{i,u[d/i,t^*]}|t^*$. So $s_{i,y}|t^* = s_{i,u}|t^*$, by (2). Let $z = y[c/i,t^*]$. So $s_{i,z}|t^* = s_{i,u}|t^*$, again by (2) and, hence, $z \in D_{i,u,t^*}$. So it follows from (71) that:

$$
z \in \|\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \qquad (77)
$$

and from (67-68) that:

$$
\begin{aligned}
y \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} &\Leftrightarrow z \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (78)\\
y \in \|\mathsf{FG}[\mathsf{B}]_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} &\Leftrightarrow z \in \|\mathsf{FG}[\mathsf{B}]_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (79)\\
y \in \|\mathsf{FG}\langle\mathsf{B}\rangle_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} &\Leftrightarrow z \in \|\mathsf{FG}\langle\mathsf{B}\rangle_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. && (80)
\end{aligned}
$$

Requirement (73) follows directly from (77-80). □

Proof of proposition 4. Let $\mathbf{d} \in C^N$ and let $u \in W$. Abbreviate:

$$
\begin{aligned}
\mathbf{c} &= \mathbf{c}_{i,u,t^*};\\
x &= u[\mathbf{d}/t^*].
\end{aligned}
$$

Assume that $\phi \in \Delta_i$ and that:

$$
\begin{aligned}
&d_i \equiv_\phi c_i; && (81)\\
&u \in \|\mathsf{S}_{G,i}\boldsymbol{\Delta}\|^{t^*}_{\mathfrak{M}_{t^*}}; && (82)\\
&u \in \|\mathsf{K}_{G,i}\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. && (83)
\end{aligned}
$$

Proceed as in the preceding proof, with D_{G,u,t^*}, D_{G,x,t^*} in place of D_{i,u,t^*}, D_{i,x,t^*}. The argument for requirement (72) is the same as before. For requirement (73), let $y \in D_{G,x,t^*}$. So $y \in D_{i,x,t^*}$, for some $i \in G$. Then $s_{i,y}|t^* = s_{i,x}|t^* = s_{i,u[\mathbf{d}/t^*]}|t^*$. So $s_{i,y}|t^* = s_{i,u}|t^*$, by (60). Let $z = y[\mathbf{c}/t^*]$. So $s_{i,z}|t^* = s_{i,u}|t^*$, again by (60) and, hence, $z \in D_{i,u,t^*} \subseteq D_{G,u,t^*}$. Continue as in the preceding proof. □

20 Proofs of Selected Statements

Proof of (14) *and* (15). Let $w \in W$ be given. To witness the first claim, define learning method c so that:

$$L_c(\sigma, \phi) \quad = \quad \begin{cases} 1 & \text{if } \phi = @_{t^*}\mathsf{G}\,\mathsf{p}_k \text{ and } (\forall t : t^* \leq t \leq \mathsf{lh}(\sigma))\ \sigma(t) = k; \\ 1 & \text{if } \phi = @_{t^*}\neg\mathsf{G}\,\mathsf{p}_k \text{ and } (\exists t : t^* \leq t \leq \mathsf{lh}(\sigma))\ \sigma(t) \neq k; \\ 0 & \text{otherwise.} \end{cases}$$

The method that witnesses the second claim is similar, except that $\neg$ and $\neq$ are moved from the second clause to the first. □

Proof of (16) *and* (17). The proof of the second statement is similar to that of the first. For the first statement, suppose for contradiction that c satisfies (10) and(11). It suffices to construct $\varepsilon \in E_0$ such that (10) and (11) are both false in arbitrary world w such that $e_w = \varepsilon$. A purely learning theoretic argument suffices. Construct ε by adding chunks in successive stages as follows, where $c = h(c_{w',i,t^*})$. At stage 0, present σ. Let $n > 0$. At stage $2n$, present k until L_c returns 1 for $@_{t^*}\mathsf{FG}\,\mathsf{p}_k$. Learning function L_c must return 1 for $@_{t^*}\mathsf{FG}\,\mathsf{p}_k$ eventually, because if L_c never takes the bait, you continue to present k and L_c fails to converge to belief that $@_{t^*}\mathsf{FG}\,\mathsf{p}_k$ even though it is true, contradicting the hypothesis. At that point, proceed to stage $2n + 1$. At stage $2n + 1$, the demon presents $k + 1$ until L_c returns 0 for $@_{t^*}\mathsf{FG}\,\mathsf{p}_k$. Learning function L_c must return 0 for $@_{t^*}\mathsf{FG}\,\mathsf{p}_k$ eventually, because if L_c never takes the bait, you continue to present $k + 1$ and L_c fails to converge to belief that $@_{t^*}\neg\mathsf{G}\,\mathsf{p}_k$ even though it is true, contradicting the hypothesis. At that point, proceed to stage $2n + 2$. You pass through each stage, producing ε that satisfies (*). □

Proof of (36). The proof follows (Kelly 1996, proposition 7.15). Suppose the contrary. Then we can use the witnessing L_d and $u \in I_{i,w,t^*}$ to compute $g(t)$, for $t \geq t^*$ (for $t < t^*$ use a lookup table). Say that finite input sequence σ of length t is t'-dead if and only if $L_d(\sigma', @_{t^*}\,\phi) = 0$, for each extension σ' of σ of length t'. By (31), $g|(t+1)$ is never t'-dead, but by König's lemma and (34), there exists $t' \geq t + 1$ such that every σ of length $t + 1$ that is distinct from $g|t$ is t'-dead. Then $g|(t + 1)$ is the unique sequence σ that is not t'-dead. Return the last entry of that sequence. □

Proof of (39). By hypothesis, ϕ is knowable in w at t^*. Since ϕ is knowable, let L_c and world $u \in I_{i,w,t^*}$ witness that fact. Let L_d suspend belief whether ϕ in all circumstancces (to avoid knowing that ϕ) and then believe, deny, or suspend belief for both $\neg\mathsf{K}_i\,\phi$ and $\mathsf{Mo}_i\,\phi$ whenever L_c does the same for ϕ. Recall that in $\mathfrak{N}_{t^*}$, (i) the inputs to i do not depend on i's learning method and (ii) the truth value of ϕ does not depend on i's learning method. Due to L_d's dogmatic belief that ϕ, the case hypothesis, and (i) and (ii), there is no world in I_{i,w,t^*} in which $\mathsf{K}_i\phi$ is true, so we have that $[\,\mathsf{I}\,]_i(\mathsf{Mo}_i\,\phi \leftrightarrow \phi)$ is true in w. So by (i) and (ii), agent i knows that $\mathsf{Mo}_i\,\phi$. But, by construction, i is conjunctively cogent with respect to $\mathsf{Mo}_i\,\phi$. □

Proof of (40). Let L_d be as in the proof of (39), except that this time L_d believes that ϕ no matter what (to ensure that i does not know that ϕ in any world in I_{i,w,t^*}, since ϕ is false in some such world) and believes both that $\mathsf{Mo}_i\,\phi$ and that $\neg\mathsf{K}_i\,\phi$ whenever L_c believes that ϕ. □

Proof of (41). Carry out the construction given in the proof of (40) for $\phi \wedge \mathsf{Mo}_i\,\phi$ rather than for $\mathsf{Mo}_i\,\phi$. □

Proof of (46). Let Δ, Γ be finite and mutually disjoint subsets of $\mathbf{L}_{\mathsf{BIT}}$. Let $\Delta \subseteq \Delta'$ and $\Delta' \cap \Gamma = \emptyset$. Define total recursive g such that:

$$g(c, \langle\sigma\rangle, \mathsf{g}(\phi)) = \begin{cases} 1 & \text{if } \phi = @_{t^*}\gamma \;\;\wedge\;\; \gamma \in \Gamma \;\wedge\; (\forall\delta \in \Delta)\; L_c(\sigma, @_{t^*}\delta) = 1; \\ 0 & \text{if } \phi = @_{t^*}\neg\gamma \;\wedge\; \gamma \in \Gamma; \\ L_c(\sigma, \phi) & \text{otherwise.} \end{cases}$$

Apply proposition 1 to obtain total recursive h such that $L_{h(c)}(\sigma, \phi) = g(c, \langle\sigma\rangle, \mathsf{g}(\phi))$. By the definition of h and the fact that Δ' is disjoint from Γ, we have that:

$$c \;\; \equiv_{\Delta'} \;\; h(c); \tag{84}$$

for each $c \in C$ and that for all $z \in W$, $t \in T$ and $\gamma \in \Gamma$:

$$L_{h(c)}(s_{i,z}|t, @_{t^*}\neg\mathsf{K}_i{}^k\,\gamma) = 0; \tag{85}$$

$$L_{h(c)}(s_{i,z}|t, \;\; @_{t^*}\mathsf{K}_i{}^k\,\gamma) = 1 \;\Leftrightarrow\; (\forall\delta \in \Delta)\, L_{h(c)}(s_{i,z}|t', @_{t^*}\,\delta) = 1. \tag{86}$$

Suppose that $u \in I_{w,i,t^*}$ satisfies:

$$u \;\; \in \;\; \|\mathsf{S}_i\Delta'\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{87}$$

$$u \;\; \in \;\; \|[\,\mathsf{I}\,]_i(\Delta \rightarrow \Gamma)\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{88}$$

$$u \;\; \in \;\; \|\mathsf{K}_i\Delta\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{89}$$

Abbreviate:

$$c \;\; = \;\; c_{i,u,t^*};$$
$$x \;\; = \;\; u[h(c)/i, t^*].$$

So from (84),(87) and (89), obtain via proposition 3 that for each $\delta \in \Delta$:

$$x \;\; \in \;\; \|\mathsf{K}_i\,\delta\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{90}$$

So for each $\delta \in \Delta$:

$$x \;\; \in \;\; \|\mathsf{GC}_i\,\delta\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{91}$$

$$y \;\; \in \;\; \|\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\delta\|^{t^*}_{\mathfrak{M}_{t^*}}, \text{ for all } y \in D_{i,x,t^*}. \tag{92}$$

It suffices to show the following requirements, for each $\gamma \in \Gamma$:

$$x \in \|\mathsf{GC}_i\,\gamma\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{93}$$

$$y \in \|\mathsf{FG}\neg\tilde{\mathsf{E}}_i\,\gamma\|^{t^*}_{\mathfrak{M}_{t^*}}, \text{ for all } y \in D_{i,x,t^*}. \tag{94}$$

Let $\gamma \in \Gamma$. For requirement (93), we have by (2) that $x \in I_{i,u,t^*}$, so (88) and (91) yield that:

$$x \in \|\gamma\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{95}$$

So (91) and (95), together with properties (85-86), yield requirement (93). For requirement (94), suppose that $y \in D_{i,x,t^*}$. So by (2), $y \in I_{i,u,t^*}$. So (*) together with (88) and (92) yield requirement (94). □

Proof of statement (51). Define total recursive f as follows:

$$f(c, \langle\sigma\rangle, \mathsf{g}(\psi)) = \begin{cases} 1 & \text{if } (\exists k)\ \psi = @_{t^*}\mathsf{K}_i{}^k\phi\ \wedge \\ & \quad (\forall t' : t^* \leq t' \leq t)(\psi = @_{t^*}\mathsf{K}_i{}^k\phi\ \wedge\ L_c(\sigma|t', \phi) = 1); \\ 0 & \text{if } (\exists k)\ \psi = @_{t^*}\mathsf{K}_i{}^k\phi\ \wedge \\ & \quad (\exists t' : t^* \leq t' \leq t)(\psi = @_{t^*}\mathsf{K}_i{}^k\phi\ \wedge\ L_c(\sigma|t', \phi) = 0); \\ 0 & \text{if } (\exists k)\ \psi = @_{t^*}\neg\mathsf{K}_i{}^k\phi\ \wedge \\ & \quad (\forall t' : t^* \leq t' \leq t)(\psi = @_{t^*}\mathsf{K}_i{}^k\phi\ \wedge\ L_c(\sigma|t', \phi) = 1); \\ 1 & \text{if } (\exists k)\ \psi = @_{t^*}\neg\mathsf{K}_i{}^k\phi\ \wedge \\ & \quad (\exists t' : t^* \leq t' \leq t)(\psi = @_{t^*}\mathsf{K}_i{}^k\phi\ \wedge\ L_c(\sigma|t', \phi) = 0); \\ L_c(\sigma, \phi) & \text{otherwise.} \end{cases}$$

Apply proposition 1 to obtain h such that $L_{h(c)}(\sigma, \psi) = f(c, \langle\sigma\rangle, \mathsf{g}(\psi))$, for all $c \in \mathbb{N}$. Suppose that Δ includes ϕ and is disjoint from $K_i^\omega(\phi)$. By the definition of h, we have that for all $c \in C$:

$$c \ \equiv_\Delta \ h(c); \tag{96}$$

so h holds Δ fixed, and that for all $z \in W$, $t \in T$, and $k \in \mathbb{N}$:

$$L_{h(c)}(s_{i,z}|t, @_{t^*}\mathsf{K}_i{}^k\,\phi) = 1 \ \Leftrightarrow\ (\forall t' : t^* \leq t' \leq t)\, L_{h(c)}(s_{i,z}|t', @_{t^*}\,\phi) = 1; \tag{97}$$

$$L_{h(c)}(s_{i,z}|t, @_{t^*}\neg\mathsf{K}_i{}^k\,\phi) = 1 \ \Leftrightarrow\ (\exists t' : t^* \leq t' \leq t)\, L_{h(c)}(s_{i,z}|t', @_{t^*}\,\phi) = 0. \tag{98}$$

Suppose that $u \in I_{w,i,t^*}$ satisfies:

$$u \in \|\mathsf{S}_i\Delta\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{99}$$

$$u \in \|\mathsf{K}_i\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{100}$$

Abbreviate:

$$\begin{aligned} c &= c_{i,u,t^*}; \\ x &= u[h(c)/i, t^*]. \end{aligned}$$

From (96), (99) and (100), obtain via proposition 3 that $x \in \|\mathsf{K}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} = \|\mathsf{K}_i{}^1\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. Therefore, $x \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} = \|\mathsf{K}_i{}^0\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. So we have the base case $x \in \|K^1(\phi)\|^{t^*}_{\mathfrak{M}_{t^*}}$.

Next, assume for induction that $x \in \|K^{k+1}(\phi)\|^{t^*}_{\mathfrak{M}_{t^*}}$. So:

$$x \in \|\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{101}$$

and, therefore:

$$x \in \|\mathsf{G}\mathsf{C}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{102}$$

$$y \in \|\mathsf{F}\mathsf{G}\neg\tilde{\mathsf{E}}\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}, \text{ for all } y \in D_{i,x,t^*}\,. \tag{103}$$

For $x \in \|K^{k+2}(\phi)\|^{t^*}_{\mathfrak{M}_{t^*}}$, it suffices to show that: $x \in \|\mathsf{K}_i\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. For that, it suffices, in turn, to show:

$$x \in \|\mathsf{G}\mathsf{C}_i\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{104}$$

$$x \in \|\mathsf{F}\mathsf{G}\neg\tilde{\mathsf{E}}\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}, \text{ for all } y \in D_{i,x,t^*}\,. \tag{105}$$

Requirement (104) expands to the requirements:

$$x \in \|\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{106}$$

$$x \in \|\mathsf{G}[\mathsf{B}]_i\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{107}$$

$$x \in \|\mathsf{G}\langle\mathsf{B}\rangle_i\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{108}$$

Requirement (106) is just (101). Hence, (102) yields:

$$x \in \|\mathsf{G}[\mathsf{B}]_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{109}$$

$$x \in \|\mathsf{G}\langle\mathsf{B}\rangle_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{110}$$

Requirements (107-108) follow from (109-110) and properties (97-98) of h.

For requirement (105), suppose that $y \in D_{i,x,t^*}$. It suffices to show that for all $y \in D_{i,x,t^*}$:

$$y \in \|\mathsf{G}\mathsf{F}[\mathsf{B}]_i\neg\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \Rightarrow y \notin \|\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{111}$$

$$y \in \|\mathsf{G}\mathsf{F}[\mathsf{B}]_i\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \Rightarrow y \in \|\mathsf{K}_i\mathsf{K}_i{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{112}$$

(113)

For requirement (111), suppose that:

$$y \in \|\mathsf{GF}[\mathsf{B}]_i \neg \mathsf{K}_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}} \tag{114}$$

Then by property (98) of h, there exists $t \geq t^*$ such that $y \notin \|\mathsf{B}_i\, \phi\|^{t}_{\mathfrak{M}_{t^*}}$, so by property (97), we have that $y \notin \|\mathsf{B}_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. So $y \notin \|\mathsf{K}_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}$.

For requirement (112), suppose that:

$$y \ \in \ \|\mathsf{GF}[\mathsf{B}]_i \mathsf{K}_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}} \tag{115}$$

For the consequent $y \in \|\mathsf{K}_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, it suffices, as usual, to show the requirements:

$$\begin{aligned} y &\ \in \ \|\mathsf{GC}_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (116)\\ z &\ \in \ \|\mathsf{FG}\neg\tilde{\mathsf{E}}{\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}, \text{ for all } z \in D_{i,y,t^*}\ . && (117) \end{aligned}$$

Requirement (117) is just (103), since $D_{i,y,t^*} = D_{i,u,t^*}$. Requirement (116) expands to:

$$\begin{aligned} y &\ \in \ \|{\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (118)\\ y &\ \in \ \|\mathsf{G}[\mathsf{B}]_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}; && (119)\\ y &\ \in \ \|\mathsf{G}\langle\mathsf{B}\rangle_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}. && (120) \end{aligned}$$

For requirement (118), we have from (115) and property (97) of h that $y \in \|\mathsf{GF}[\mathsf{B}]_i {\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. So $y \in \|{\mathsf{K}_i}^k\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, by (103). For requirement (119), note that (115), along with property (97) of h implies that $y \in \|\mathsf{G}[\mathsf{B}]_i\, \phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, which implies requirement (119) in light of property (97) and requirement (120) in light of property (98). □

Proof of statement (52). For the Led_i case, follow the proof of (51) with Led_i in place of K_i and $\tilde{\mathsf{C}}_i$ in place of $\tilde{\mathsf{E}}_i$. For the $\tilde{\mathsf{L}}_i$ case, make corresponding substitutions and ignore the actual convergence requirements. For the $\widetilde{\mathsf{Led}}_i$ case, add cases for actual convergence to true belief that $\neg\phi$. For the $\tilde{\mathsf{K}}_i$ case, do the same, but retain $\tilde{\mathsf{C}}_i$ in place of $\tilde{\mathsf{E}}_i$. □

Proof of statement (53). Let $w = (\varepsilon, \mathbf{c})$ be a world in $\mathfrak{N}_{t^*}$. Let total recursive h hold belief whether $\phi = \mathsf{G}\,\mathsf{p}_k$ fixed. Let c^* be as in the proof of statement (14). Let $\mathbf{c} \in C^N$ and let $w_{\varepsilon'} = (\varepsilon', \mathbf{c}[c^*/i]_{t^*})$, for arbitrary $\varepsilon' \in E_0$. Let $\tau(t) = \varepsilon(t)$ for $t < t^*$ and let $\tau(t) = k$ for $t \geq t^*$. Let $\tau_t(t') = \tau(t')$ for $t' \geq t$ and let $\tau_t(t') = k+1$ for $t' \geq t$. It is easy to verify that for all $t \geq t^*$:

$$\begin{aligned} w_\tau &\ \in \ \|\mathsf{K}_i \mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}}; && (121)\\ w_{\tau_t} &\ \in \ \|\neg\mathsf{K}_i \mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}}. && (122) \end{aligned}$$

Since the truth of $\mathsf{G}\,\mathsf{p}_k$ does not depend on methods in $\mathfrak{N}_{t^*}$, we have for all $t \geq t^*$ that:

$$w_{\tau_t} \ \in \ I_{w,i,t^*} \cap \|\mathsf{S}_i\, \mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}} \cap \|\neg\mathsf{K}_i \mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}}. \tag{123}$$

So it suffices to show that $w_{\tau_t}[h(c^*)/i, t^*] \notin \|\mathsf{K}_i\neg\mathsf{K}_i\mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}}$. For that it suffices to show that at least one of the following statements holds:

$$w_{\tau_t}[h(c^*)/i, t^*] \notin \|\mathsf{GC}_i\neg\mathsf{K}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{124}$$

$$w_{\tau_t}[h(c^*)/i, t^*] \notin \|[\mathsf{D}]_i\mathsf{FG}\neg\tilde{\mathsf{E}}_i\neg\mathsf{K}_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{125}$$

Case 1: $w_{\tau_t}[h(c^*)/i, t^*] \notin \|\mathsf{GB}_i\neg\mathsf{K}_i\mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}}$, for some $t \geq t^*$. So (124) holds, in light of (122).

Case 2: $w_{\tau_t}[h(c^*)/i, t^*] \in \|\mathsf{GB}_i\neg\mathsf{K}_i\mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}}$, for all $t \geq t^*$. Then since $\tau|t = \tau_t|t$, for each $t \geq t^*$, we have that $w_\tau[h(c^*)/i]_{\geq t^*} \in \|\mathsf{GB}_i\neg\mathsf{K}_i\mathsf{G}\,\mathsf{p}_k\|^{t^*}_{\mathfrak{N}_{t^*}}$. Note that $w_\tau \in I_{w_{\tau_t},i,t^*}$ by construction and (2). So (125) holds, in light of (121). □

Proof of statements (54-57). One merely has to check that the respective antecedents of the various conditionals are satisfied by each world w_{τ_t} in the proof of (53). For (54), observe that $\neg\phi$ is true in w_{τ_t}, by construction. For (55), observe that c^* suspends belief concerning $\neg\mathsf{K}_i\neg\phi$. For (57), observe both that c^* suspends belief concerning $\neg\phi$ and that $\neg\phi$ is true in w_{τ_t}. For (56), let $w \in W$ and let total recursive h hold both ϕ and ψ fixed. To refute the second disjunct of (56) in w, let c^{**} follow the strategy of c^* with respect to ϕ, except that c^{**} believes that $\neg\mathsf{K}_i\,\psi$ no matter what. Then, due to c^{**}'s suspension of belief whether ψ at t^*, we have that c^{**} witnesses the truth of $\mathsf{K}_i\neg\mathsf{K}_i\,\psi$ in every world, so the argument for (53) establishes the falsehood of the second disjunct of (56) in w. Reversing the roles of ϕ and ψ establishes that the first disjunct of (56) is also false in w. □

Proof of statement (65). Define total recursive f_e just as in the proof of (51), except that ${\mathsf{K}_i}^k\phi$ is replaced with ${\mathsf{K}_G}^k\phi$. For $j \in G_-$, define total recursive f_j just like f_e, but with the condition $L_c(\sigma|t', \mathsf{B}_i\,\phi) = 1$ in place of condition $L_c(\sigma|t', \phi) = 1$. Apply proposition 1 to each f_i to obtain respective, total recursive function h_i. Let $\mathbf{h} = (h_1, \ldots, h_N)$.

Suppose that $\phi \in \Delta_e$ and that $\Delta_i \cap K^\omega_G = \emptyset$, for each $i \in G$. By the definition of $\mathbf{h}$, we have that for all $\mathbf{c} \in C^N$:

$$\mathbf{c} \quad \equiv_{\mathbf{\Delta}} \quad \mathbf{h}(\mathbf{c}); \tag{126}$$

so $\mathbf{h}$ holds $\mathbf{\Delta}$ fixed, and we also have that for all $i \in G$, $z \in W$, $t \in T$, and $k \in \mathbb{N}$:

$$L_{h(c)}(s_{i,z}|t, @_{t^*}\neg{\mathsf{K}_G}^k\,\phi) = 1 \quad \Leftrightarrow \quad L_{h(c)}(s_{i,z}|t, @_{t^*}{\mathsf{K}_G}^k\,\phi) = 0;. \tag{127}$$

Suppose that $u \in I_{w,i,t^*}$ satisfies:

$$u \quad \in \quad \|\mathsf{IS5}_G\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{128}$$

$$u \quad \in \quad \|\mathsf{T}_{G,e}\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{129}$$

$$u \quad \in \quad \|\mathsf{S}_{G,e}\mathbf{\Delta}\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{130}$$

$$u \quad \in \quad \|\mathsf{K}_{G,e}\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{131}$$

Abbreviate:

$$\begin{aligned}\mathbf{c} &= \mathbf{c}_{i,u,t^*};\\ x &= u[\mathbf{h}(\mathbf{c})/i, t^*].\end{aligned}$$

From (126), (130) and (131), obtain via proposition 4 that:

$$x \in \|\mathsf{K}_{G,e}\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{132}$$

Note that for $j \in G_-$ and $z \in W$ we have by the definition of $\mathbf{h}$ that:

$$L_{h_e(c_e)}(s_{i,z}|t, @_{t^*}{\mathsf{K}_G}^k\,\phi) = 1 \;\Leftrightarrow\; (\forall t' : t^* \le t' \le t)\, L_{h_e(c_e)}(s_{i,z}|t', @_{t^*}\,\phi) = 1; \tag{133}$$

$$L_{h_j(c_j)}(s_{i,z}|t, @_{t^*}{\mathsf{K}_G}^k\,\phi) = 1 \;\Leftrightarrow\; (\forall t' : t^* \le t' \le t)\, L_{h_j(c_j)}(s_{i,z}|t', @_{t^*}\mathsf{B}_{G,e}\,\phi) = 1; \tag{134}$$

Let $y \in D_{G,x,t^*} \subseteq I_{G,x,t^*}$. So $y \in I_{G,u,t^*}$ by (60). Then by (129), we have for all $j \in G_-$ that $y \in \|\mathsf{G}\tilde{\mathsf{C}}_j\mathsf{B}_e\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. Hence, by (133-134), we have for all $i \in G$, $y \in D_{G,x,t^*}$, and $k \in \mathbb{N}$:

$$L_{h_i(c_i)}(s_{i,y}|t, @_{t^*}{\mathsf{K}_G}^k\,\phi) = 1 \;\Leftrightarrow\; (\forall t' : t^* \le t' \le t)\, L_{h_e(c_e)}(s_{i,y}|t', @_{t^*}\,\phi) = 1; \tag{135}$$

By (132), (127), and (135), we have that $x \in \|\mathsf{K}_{G,j}\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, for all $j \in G_-$, so again by (132) we have $x \in \|{\mathsf{K}_G}^1\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, and hence, that $x \in \|\phi\|^{t^*}_{\mathfrak{M}_{t^*}} = \|{\mathsf{K}_G}^0\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. Thus, we have the base case $\|K^1_G(\phi)\|^{t^*}_{\mathfrak{M}_{t^*}}$.

Next, assume for induction that $x \in \|K^{k+1}_G(\phi)\|^{t^*}_{\mathfrak{M}_{t^*}}$ and show that $x \in \|K^{k+2}_G(\phi)\|^{t^*}_{\mathfrak{M}_{t^*}}$. By the induction hypothesis, we have, for each $i \in G$ that:

$$x \in \|\mathsf{K}_{G,i}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{136}$$

and, therefore:

$$x \in \|\mathsf{G}\mathsf{C}_i{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{137}$$

$$y \in \|\mathsf{F}\mathsf{G}\neg\tilde{\mathsf{E}}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}},\ \text{for all } y \in D_{G,x,t^*}\ . \tag{138}$$

For $x \in \|K^{k+2}_G(\phi)\|^{t^*}_{\mathfrak{M}_{t^*}}$, it suffices to show, for each $i \in G$, that: $x \in \|\mathsf{K}_{G,i}\mathsf{K}_{G,i}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. For that, it suffices, in turn, to show:

$$x \in \|\mathsf{G}\mathsf{C}_i\mathsf{K}_{G,i}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{139}$$

$$x \in \|\mathsf{F}\mathsf{G}\neg\tilde{\mathsf{E}}\mathsf{K}_{G,i}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}},\ \text{for all } y \in D_{G,x,t^*}\ . \tag{140}$$

Requirement (139) expands to the requirements:

$$x \in \|\mathsf{K}_{G,i}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{141}$$

$$x \in \|\mathsf{G}[\mathsf{B}]_i\mathsf{K}_{G,i}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{142}$$

$$x \in \|\mathsf{G}\langle\mathsf{B}\rangle_i\mathsf{K}_{G,i}{\mathsf{K}_G}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{143}$$

Requirement (141) is just (136). Hence, (137) yields:

$$x \in \|\mathsf{G}[\mathsf{B}]_i \mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{144}$$
$$x \in \|\mathsf{G}\langle\mathsf{B}\rangle_i \mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{145}$$

Requirements (142-143) follow from (144-145) and properties (127) and (135) of **h**.

For reuirement (140), suppose that $y \in D_{G,x,t^*}$. It suffices to show that for all $y \in D_{G,x,t^*}$:

$$y \in \|\mathsf{GF}[\mathsf{B}]_i\neg\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;\Rightarrow\; y \notin \|\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{146}$$
$$y \in \|\mathsf{GF}[\mathsf{B}]_i\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \;\Rightarrow\; y \in \|\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{147}$$

For requirement (146), suppose that $y \in \|\mathsf{GF}[\mathsf{B}]_i\neg\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. Then by properties (135) and (127) of h, we have that $y \notin \|\mathsf{G}[\mathsf{B}]_i\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. So $y \notin \|\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$.

For requirement (147), suppose that:

$$y \in \|\mathsf{GF}[\mathsf{B}]_i\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}} \tag{148}$$

For the consequent $y \in \|\mathsf{K}_{G,i}\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, it suffices, as usual, to show the requirements:

$$y \in \|\mathsf{GC}_i\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{149}$$
$$z \in \|\mathsf{FG}\neg\tilde{\mathsf{E}}_i\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}},\ \text{for all } z \in D_{G,y,t^*}\ . \tag{150}$$

Requirement (150) is just (138), since $D_{G,y,t^*} = D_{G,u,t^*}$ by (128).[19] Requirement (149) expands to:

$$y \in \|\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{151}$$
$$y \in \|\mathsf{G}[\mathsf{B}]_i\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}; \tag{152}$$
$$y \in \|\mathsf{G}\langle\mathsf{B}\rangle_i\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}. \tag{153}$$

For requirement (151), we have from (148) and property (135) of h that $y \in \|\mathsf{GF}[\mathsf{B}]_i\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$. So $y \in \|\mathsf{K}_G{}^k\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, by (138). For requirement (152), note that (148), along with property (135) of **h** implies that $y \in \|\mathsf{G}[\mathsf{B}]_i\,\phi\|^{t^*}_{\mathfrak{M}_{t^*}}$, which again, in light of property (135) implies requirement (152). Requirement (153) is then immediate by property (127) of **h**. □

[19]This is the proof's only appeal to the S5 property for information.