The Logic of Justified Belief Change, Soft Evidence and Defeasible Knowledge

Alexandru Baltag Bryan Renne* Sonja Smets†

University of Amsterdam
Institute for Logic, Language and Computation

Abstract

We present a logic for reasoning about the evidence-based knowledge and beliefs and the evidential dynamics of non-logically-omniscient agents. We do this by adapting key tools and techniques from Dynamic Epistemic Logic, Justification Logic, and Belief Revision so as to provide a lightweight, yet fine-grained approach that characterizes well-known epistemic and doxastic attitudes in terms of the evidential reasoning that justifies these attitudes. We then add the dynamic operations of evidence introduction, evidence-based inference, strong acceptance of new evidence (evidential "upgrade"), and irrevocable acceptance of additional evidence (evidential "update"). We exemplify our theory by providing a formal dynamic account of Lehrer's well-known Gettier-type scenario involving the famous Ferrari and the infamous Messrs. Nogot and Havit.

1 Introduction

As shown by the famous Gettier counterexamples [8], "knowledge" cannot simply be equated with "justified true belief." But what is the missing ingredient in this old Platonic equation? While epistemologists have proposed different answers to fill the gap, all would agree that not just any justification will do in order to turn an item of true belief into knowledge. It is essential that "knowledge" comes equipped with a *correct*, or "good," justification. Taking this insight as our starting point, we offer in this paper a new formalization for a plethora of notions ranging from justified belief to defeasible knowledge, each of which comes with its own justification based on how well an agent's evidence supports her epistemic attitude.

The so-called Defeasibility Theory defines "knowledge" as *true justified belief that is stable under belief revision with any new evidence*: "if a person has knowledge, then that person's justification must be sufficiently strong that it is not capable of being defeated by evidence that he does not possess" (Pappas and Swain [13]). One of the problems is interpreting what "evidence" means in this context. One possible interpretation, considered by at least one author [15], takes "evidence" to mean "*any* proposition," meaning we include possible *misinformation*: "real knowledge" should be robust even in the face of false evidence. This interpretation corresponds to our "infallible knowledge" modality K, which could be called "absolutely unrevisable belief." This is a fully introspective type of knowledge, satisfying all the laws of the modal system S5.

However, the most common interpretation of Defeasibility Theory is to take it as requiring persistence of belief only in the face of "any *true* information." The resulting notion of "knowledge" was formalized by

 Citation: Alexandru Baltag, Bryan Renne, and Sonja Smets. The Logic of Justified Belief Change, Soft Evidence and Defeasible Knowledge. In L. Ong and R. de Queiroz, editors, *Proceedings of the 19th Workshop on Logic, Language, Information and Computation (WoLLIC 2012)*, volume 7456 of *Lecture Notes in Computer Science*, pages 168–190, Buenos Aires, Argentina. Springer-Verlag Berlin Heidelberg, 2012.

*Funded by an Innovational Research Incentives Scheme Veni grant from the Netherlands Organisation for Scientific Research (NWO).

†Funded in part by an Innovational Research Incentives Scheme Vidi grant from the Netherlands Organisation for Scientific Research (NWO) and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement no. 283963.

Stalnaker in [17], and defined there as follows: "an agent knows that φ if and only if φ is true, she believes that φ, and she continues to believe φ if any *true* information is received". This interpretation corresponds to our "defeasible knowledge" modality $\Box$, which is positively (but not necessarily negatively) introspective, satisfying all the axioms of the modal system S4.

In [6], two of the authors of this paper studied these two notions in detail, using a dynamic logic of belief change to make precise the sense in which these modal operators match the above-mentioned characterizations in terms of their potential (in)defeasibility. However, both the above notions of "knowledge" suffer from the problem of logical omniscience. So at best they can be taken to capture some kind of *implicit, or potential, knowledge*. Moreover, Lehrer's conception [10, 12] of defeasible knowledge is more sophisticated: he requires, not only that the belief itself be stable in the face of any true evidence, but also that the *justification* supporting this belief be similarly stable.

In this paper, we formalize the *explicit defeasible knowledge* that can be actually possessed by a (non-logically omniscient) agent. For this, we develop a version of Justification Logic (JL), in the tradition of [2], with the new feature that it borrows concepts from Belief Revision theory to deal with "soft" (fallible) evidence. Furthermore, we combine this approach with ideas and techniques from Dynamic Epistemic Logic (DEL) [4, 5, 6, 18], including important ideas from the temporal DEL literature [14, 16, 22], obtaining a Dynamic Justification Logic that can deal with justified belief change and soft evidential dynamics.

Thus, in essence we bring together the work of two traditions in Logic (DEL and JL), while using models coming from a third tradition (Belief Revision theory). The added value comes from the interplay of these settings, which in particular allows us to capture several of the subtle distinctions made in [10, 11], pointing to scenarios in which an agent has a justified true belief but no good evidence to turn his belief into knowledge. We formalize various types of epistemic-evidential actions, and we use them to give dynamic characterizations of explicit "knowledge" (in both its defeasible and its infallible versions). We provide complete, decidable proof systems for these logics, and apply them to the analysis of one of the well-known Gettier-type counterexamples in the literature.

The approaches in the available literature that are closest to our work are Artemov's paper [1] on the Gettier problem, and the work of van Benthem and Velázquez-Quesada [20, 21] on the dynamics of evidence. In our last section we make a more detailed comparision between these papers and ours. For now, it suffices to say that our solution is closely related to these approaches and was in fact inspired by them, but it is nevertheless original and avoids some of the problems encountered in these works.

2 Belief, Justification, Awareness, and Knowledge

2.1 Syntax

Definition 2.1 (Language)**.** Given a set Φ of atomic sentences, the language $\mathscr{L} := (\mathscr{T}, \mathscr{F})$ consists of the set $\mathscr{T}$ of *evidence terms* t and the set $\mathscr{F}$ of *propositional formulas* (sentences) φ defined by the following double recursion:

$$\begin{array}{lcl} \varphi & ::= & \bot \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid Et \mid t \gg \varphi \mid \Box\varphi \mid K\varphi \mid Y\varphi \quad \text{with } p \in \Phi \\ t & ::= & c_\varphi \mid t \cdot t \mid t + t \end{array}$$

Notation: let $\Diamond$ denote $\neg\Box\neg$, let $\hat{K}$ denote $\neg K\neg$, and let $\hat{Y}$ denote $\neg Y\neg$. The set $\mathsf{sub}(t)$ of *subterms* of a term t is defined by induction on the construction of t as follows: $\mathsf{sub}(c_\varphi) = \{c_\varphi\}$, $\mathsf{sub}(s \cdot u) = \{s \cdot u\} \cup \mathsf{sub}(s) \cup \mathsf{sub}(u)$, $\mathsf{sub}(s + u) = \{s + u\} \cup \mathsf{sub}(s) \cup \mathsf{sub}(u)$. The set $\mathsf{sub}(\varphi)$ of *subformulas* of a formula φ is defined by induction on the construction of φ as follows: $\mathsf{sub}(\bot) := \{\bot\}$, $\mathsf{sub}(p) := \{p\}$, $\mathsf{sub}(\neg\theta) := \{\neg\theta\} \cup \mathsf{sub}(\theta)$, $\mathsf{sub}(\theta \wedge \theta') := \{\theta \wedge \theta'\} \cup \mathsf{sub}(\theta) \cup \mathsf{sub}(\theta')$, $\mathsf{sub}(Et) := \{Et\}$, $\mathsf{sub}(t \gg \theta) := \{t \gg \theta\}$, $\mathsf{sub}(\Box\theta) := \{\Box\theta\} \cup \mathsf{sub}(\theta)$, $\mathsf{sub}(K\theta) := \{K\theta\} \cup \mathsf{sub}(\theta)$, and $\mathsf{sub}(Y\theta) := \{Y\theta\} \cup \mathsf{sub}(\theta)$. We define an operation $(\cdot)^Y : \mathscr{T} \cup \mathscr{F} \to \mathscr{T} \cup \mathscr{F}$ by setting: $(c_\varphi)^Y := c_{(\varphi^Y)}$, $(t \cdot s)^Y := t^Y \cdot s^Y$, and $(t + s)^Y := t^Y + s^Y$ for terms; and $\bot^Y := \bot$, $p^Y := p$, $(\neg\varphi)^Y := \neg\varphi^Y$, $(\varphi \wedge \psi)^Y := \varphi^Y \wedge \psi^Y$, $(Et)^Y := Et^Y$, $(t \gg \varphi)^Y := t^Y \gg \varphi^Y$, $(K\varphi)^Y := YK\varphi$, $(\Box\varphi)^Y := Y\Box\varphi$, $(Y\varphi)^Y := YY\varphi$.

Et says that *evidence t is available* to the agent (though not necessarily accepted by her). $t \gg \varphi$ says that *t is admissible evidence* for φ: if accepted, this evidence supports φ. $\Box\varphi$ says that *the agent (implicitly) defeasibly knows φ*. $K\varphi$ says that *the agent (implicitly) infallibly knows φ*. $Y\varphi$ says that *"yesterday" (i.e.,*

before the last epistemic action) φ *was true.* c_φ is an *evidential certificate*: a "canonical" piece of evidence in support of sentence φ. $t \cdot s$ is a compound evidence, obtained by combining (using Modus Ponens) the two pieces of evidence t and s. Finally, $t + s$ is a body of evidence that aggregates (without performing logical inference) all the evidence provided by t and s; $t \cdot s$ therefore supports both the statements supported by t and those supported by s.

By "defeasible knowledge" $\Box$ we mean here knowledge in the sense of the Defeasibility Theory: justified true belief that cannot be defeated by any new *true* information that the agent might receive. By "infallible" knowledge K we mean "absolutely certain," absolutely unrevisable, fully introspective knowledge: belief that cannot fail to be true, and so it cannot be defeated by any new information (including false testimony). We will later show that our formal operators match these characterizations. Note that both these notions are forms of *implicit* knowledge. We will later introduce the corresponding types of *explicit* knowledge.

Definition 2.2 (Admissibility)**.** *Admissibility* is the smallest binary relation $\gg \subseteq \mathscr{T} \times \mathscr{F}$ satisfying the following conditions: (1) $c_\varphi \gg \varphi$; (2) if $t \gg (\psi \Rightarrow \varphi)$ and $s \gg \psi$, then $(t \cdot s) \gg \varphi$; (3) if $t \gg \varphi$ or $s \gg \varphi$, then $(t + s) \gg \varphi$. Note that admissibility is both a syntactic meta-relation and a symbol in the language. It will be clear from context which is which.

Lemma 2.3 (Temporal Admissibility)**.** $t \gg \varphi$ implies $t^Y \gg \varphi^Y$.

Proof. By induction on the construction of t. $\Box$

Lemma 2.4 (Computability of Admissibility)**.** The map $t \mapsto \{\varphi \mid t \gg \varphi\}$ of type $\mathscr{T} \to \wp(\mathscr{F})$ is computable, and for every t the set $\{\varphi \mid t \gg \varphi\}$ is finite.

Proof. The map is recursive, and the finiteness of $\{\varphi \mid t \gg \varphi\}$ can be proved by induction on the complexity of terms. $\Box$

Definition 2.5. $\mathscr{T}^e := \{t \in \mathscr{T} \mid \exists \varphi : t \gg \varphi\}$ is the *set of admissible terms.*

Definition 2.6 (Propositional Content)**.** For every term $t \in \mathscr{T}$, we define the *propositional content* con_t *of* t as the conjunction of all the formulas for which t is admissible evidence: $\mathsf{con}_t := \bigwedge\{\theta \mid t \gg \theta\}$. For $t \notin \mathscr{T}^e$, this is the conjunction of an empty set of formulas, so in this case (if we intrepret $\bigwedge$ as infimum in the complete Boolean algebra of propositions) we get *tautologically* true content: $\mathsf{con}_t = \top := \neg\bot$.

Definition 2.7 (Implicit Belief, Implicit Acceptance, Implicit Evidence)**.** We introduce the following abbreviations for the language $\mathscr{L}$:

$B\varphi$	:=	$\Diamond\Box\varphi$	says that *the agent (implicitly) believes* φ,
$A(t)$	:=	$\bigwedge_{c_\varphi \in \mathsf{sub}(t)} B\varphi$	says that *the agent (implicitly) accepts evidence* t,
$G(t)$	:=	$\bigwedge_{c_\varphi \in \mathsf{sub}(t)} \Box\varphi$	says that t *is good (implicit) evidence*,
$I(t)$	:=	$\bigwedge_{c_\varphi \in \mathsf{sub}(t)} K\varphi$	says that t *is infallible (implicit) evidence*, and
$t : \varphi$	:=	$A(t) \wedge t \gg \varphi$	says that t *is (implicit) evidence for belief of* φ.

Like the implicit knowledge notions K and $\Box$, implicit belief suffers from logical omniscience. We now introduce the corresponding explicit notions, which reflect the beliefs, knowledge and justifications that are actually possessed by a (non-logically-omniscient) agent.

Definition 2.8 (Explicit Belief and Knowledge)**.** We introduce the following additional abbreviations for the language $\mathscr{L}$:

$B^e\varphi$	:=	$B\varphi \wedge Ec_\varphi$	says that *the agent explicitly believes* φ,
$\Box^e\varphi$	:=	$\Box\varphi \wedge Ec_\varphi$	says that *the agent explicitly defeasibly knows* φ,
$K^e\varphi$	:=	$K\varphi \wedge Ec_\varphi$	says that *the agent explicitly infallibly knows* φ, and
$t :^e \varphi$	:=	$t : \varphi \wedge Et$	says that t *is explicit evidence for belief of* φ.

2.2 Semantics

Definition 2.9 (Model)**.** A *model* $M = (W, [\![\cdot]\!], \sim, \geq, \rightsquigarrow, E)$ is a structure consisting of a nonempty set W of *possible worlds*; a *valuation map* $[\![\cdot]\!] : \Phi \to \wp(W)$; binary relations $\sim$, $\geq$, and $\rightsquigarrow$ on W, with $\sim$ ("epistemically indistinguishable from") representing epistemic possibility/indistinguishability, $\geq$ ("no more plausible than") representing relative plausibility, and $\rightsquigarrow$ ("is the temporal predecessor of") representing immediate temporal precedence (going forward in time from a moment to the next moment); as well as an *evidence map* $E : W \to \wp(\mathscr{T})$; all satisfying the following conditions:

- $\sim$ is an equivalence relation and $\geq$ is a preorder.[1]
- *Indefeasibility*: $w \geq v \Rightarrow w \sim v$.
- *Local Connectedness*: $w \sim v \Rightarrow (w \geq v \vee v \geq w)$.
- *Propositional Perfect Recall*: $(w \rightsquigarrow v \sim v') \;\Rightarrow\; \exists w'(w \sim w' \rightsquigarrow v')$.
- *Evidential Perfect Recall*: $w \rightsquigarrow w' \Rightarrow \{t^Y \mid t \in E(w)\} \subseteq E(w')$.
- *Uniqueness of Past*: $(w' \rightsquigarrow w \wedge w'' \rightsquigarrow w) \Rightarrow w' = w''$.
- *Persistence of Facts*: $w \rightsquigarrow w' \Rightarrow (w \in [\![p]\!] \Leftrightarrow w' \in [\![p]\!])$.
- *(Implicit) Evidential Introspection*: $w \sim v \Rightarrow E(w) = E(v)$.
- *Subterm Closure*: If $t \cdot t' \in E(w)$ or $t + t' \in E(w)$, then $t \in E(w)$ and $t' \in E(w)$.

  This says that a compound evidence is actually available to the agent only if its component pieces of evidence are available.
- *Certification of Evidence*: If $t \in E(w)$ and $t \gg \varphi$, then $c_\varphi \in E(w)$.

  This says that *every actual evidence in support of a sentence φ can be converted into a certificate of correctness*: a canonical piece of evidence c_φ that certifies it. All explicit knowledge can be certified.

A *pointed model* is a pair (M, w) consisting of a model M and a designated world w in M called the "actual world."

Many authors in the Belief Revision literature require their models to satisfy some version of the following requirement:

Definition 2.10. *The Best Worlds Assumption* applies to a model iff for every non-empty set $P \subseteq W$ of indistinguishable worlds (i.e., such that $w \sim w'$ for all $w, w' \in P$), the set

$$\min P := \{w \in P \mid w' \geq w \text{ for all } w' \in P\}$$

(consisting of the most plausible worlds in P) is also non-empty.

The Best Worlds Assumption is useful, since it allows for a very natural and intuitive definition of (conditional) belief $B(\varphi|P)$. Some authors (e.g., Grove [9]) weaken this condition to cover only the sets P that are *definable* by some sentence ψ in their language: this is indeed enough to define syntactical conditional belief operators $B(\varphi|\psi)$. However, in this paper, we will consider an even stronger condition, called *standardness*:

Definition 2.11 (Standard Model)**.** A model $M = (W, [\![\cdot]\!], \sim, \geq, \rightsquigarrow, E)$ is said to be *standard* if both the strict converse-plausibility relation $<$ and the immediate temporal predecessor relation $\rightsquigarrow$ are well-founded. This means that there are no infinite chains $w_0 > w_1 > w_2 > \cdots$ of more and more plausible worlds, and there are no infinite chains $w_0 \leftsquigarrow w_1 \leftsquigarrow w_2 \leftsquigarrow \cdots$ going back in time. Observe that well-foundedness implies acyclicity, so in a standard model there are no temporal loops. Note also that the well-foundedness of $\rightsquigarrow$ together with the Propositional Perfect Recall condition imply "temporal perfect recall": $w \rightsquigarrow v$ implies $w \not\sim v$. Similarly, note that *every standard model satisfies the Best Worlds Assumption.*[2]

[1] A *preorder* is a reflexive and transitive binary relation. For a preorder $\geq$, we denote by $>$ the strict version given by $t > s := (s \geq t) \wedge (t \not\geq s)$. We denote by $\leq$ and $<$ the converse relations.

[2] Indeed, it is easy to see that this condition follows from the well-foundedness of $<$ together with the above Local Connectedness assumption.

Definition 2.12 (Truth)**.** We now define a *satisfaction relation* $(M, w) \models \varphi$ between pointed models (M, w) and formulas $\varphi \in \mathscr{F}$. We also denote $(M, w) \models \varphi$ in the more familiar way by $w \models_M \varphi$, omitting the subscript M when M is fixed.

$$
\begin{array}{lll}
w \not\models \bot & & \\
w \models p & \text{iff} & w \in [\![p]\!] \\
w \models \neg\varphi & \text{iff} & w \not\models \varphi \\
w \models \varphi \wedge \psi & \text{iff} & w \models \varphi \text{ and } w \models \psi \\
w \models Et & \text{iff} & t \in E(w) \\
w \models t \gg \varphi & \text{iff} & t \gg \varphi \\
w \models \Box\varphi & \text{iff} & v \models \varphi \text{ for every } v \leq w \\
w \models K\varphi & \text{iff} & v \models \varphi \text{ for every } v \sim w \\
w \models Y\varphi & \text{iff} & v \models \varphi \text{ for every } v \rightsquigarrow w
\end{array}
$$

Given a model $M = (W, [\![\cdot]\!], \sim, \geq, \rightsquigarrow, E)$, we can extend the valuation map $[\![\cdot]\!]$ to *all* sentences, by putting $[\![\varphi]\!] = \{w \in W \mid w \models \varphi\}$. *Validity* $\models \varphi$ means that $(M, w) \models \varphi$ for every *standard* pointed model (M, w).

The following result shows that belief, as defined above, fits with its most widely accepted definition in standard models:

Lemma 2.13. In a *standard* model $M = (W, [\![\cdot]\!], \sim, \geq, \rightsquigarrow, E)$, *"belief" is the same as "truth in the most plausible worlds"*:

$$w \models_M B\varphi \quad \text{iff} \quad w' \models_M \varphi \text{ for all } w' \in \min\{w' \in W \mid w \sim w'\} \ .$$

2.3 Example of the Gettier Problem

The following example of a Gettier problem is adapted from [11]. Our (unnamed) agent (Lehrer's "Claimant," who we assume to be a woman) is the professor of a class consisting of two students, Mr. Nogot and Mr. Havit. Let us denote by p the sentence "Mr. Nogot owns a Ferrari" and by q the sentence "Mr. Havit owns a Ferrari."

Mr. Nogot tells our agent that he owns a Ferrari and shows her the title papers and a picture of him driving a Ferrari. This testimonial evidence supports sentence p, and so it is admissible for p; hence, we will denote this evidence by c_p. The evidence term c_p is thus *available* to our agent (since it was made available to her by Mr. Nogot). Let us further assume that this evidence is *accepted* by her: on the basis of c_p, she believes p (i.e., that Mr. Nogot owns a Ferrari). Moreover, let us assume that this belief is actually false: in fact, Mr. Nogot does not own a Ferrari, he just lied to our agent, forged the car title and faked the picture (using Photoshop to edit himself into the driver's seat). Furthermore, let us assume that, unknown to our agent, Mr. Havit actually does own a Ferrari.

Based on her available accepted evidence c_p and using propositional inference, our agent concludes that *some student in her class owns a Ferrari* $(p \vee q)$. This belief is *true* (since in fact Mr. Havit owns one), it is *justified* (given Mr. Nogot's testimony and the rules of logic), but it is *not "knowledge"* in Lehrer's sense. Of course, $p \vee q$ is not infallible knowledge (in the absolutely certain sense captured by the operator K above), since the agent posseses no "hard" evidence for $p \vee q$. Indeed, testimonial evidence is "soft": the fact that Mr. Nogot claims that he owns a Ferrari is still consistent with the possibility that nobody in the class owns any car. Moreover, our agent's justified belief in $p \vee q$ is easily *defeasible, even by true evidence*: if in the future she would (correctly) learn that Mr. Nogot does not in fact own a Ferrari $(\neg p)$, then she would be forced to drop her (correct) belief in $p \vee q$. Hence, this true justified belief is not "knowledge," even in the fallible, defeasible sense, captured in our formalism by the operator $\Box$.

Here is a simple model of the epistemic situation described in this story:

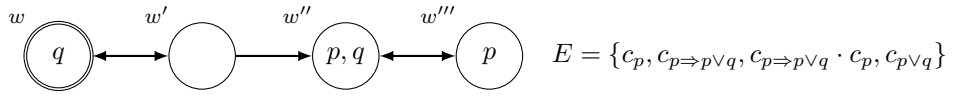

Figure 1: The Nogot-Havit scenario

The set of possible worlds is $W = \{w, w', w'', w'''\}$ and the valuation is $[\![p]\!] = \{w'', w'''\}, [\![q]\!] = \{w, w''\}$. In Figure 1, we represent each possible world by a circle (labeled with the name of the world and encompassing the atomic sentences true at that world). The double-circled world indicates the real world or current state of affairs (in which q is true and p is false; i.e., Mr. Havit has a Ferrari and Mr. Nogot does not). We represent the plausibility relations $\geq$ by horizontal arrows (pointing from a world w to all the worlds $v \leq w$ that are at least as plausible as w), but we omit the arrows that can be obtained by reflexivity (looping) and transitivity (arrow composition). The one-way arrow from w' to w'' (and the one-way arrows, obtained by transitivity, from w to both w'' and w''' and from w' to w''') show the p-worlds are more plausible than the $\neg p$-worlds. As a consequence, the agent implicitly believes p (since p is true in all the most plausible worlds w'' and w'''). The epistemic indistinguishability relation is not directly represented but can be recovered by closing the horizontal arrows under transitivity, reflexivity and *symmetry*. So here all the four worlds are epistemically indistinguishable, which expresses the fact that the agent has no hard evidence concerning p and q, and thus she has no infallible knowledge (K) concerning their truth values: all four Boolean combinations are epistemically possible (in the sense of K). The available evidence is the same at all four worlds (in agreement with the condition of Implicit Evidential Introspection) and consists of Nogot's testimonial evidence c_p, logical evidence $c_{p \Rightarrow p \vee q}$ supporting the axiom $p \Rightarrow p \vee q$, inferential evidence $c_{p \Rightarrow p \vee q} \cdot c_p$ (obtained by combining the previous two in accordance with Modus Ponens) supporting $p \vee q$, and a certificate $c_{p \vee q}$ confirming that she derived $p \vee q$. According to our definitions, the agent has (both implicit and) explicit true justified belief in $p \vee q$: the sentence $(p \vee q) \wedge B^e(p \vee q) \wedge (c_{p \Rightarrow p \vee q} \cdot c_p) : (p \vee q)$ holds at the real world w. However, she does not have (either implicit or explicit) knowledge of $p \vee q$ (either in the infallible sense of K or in the defeasible sense of $\Box$).

In this model, all evidence is accepted: $c_{p \Rightarrow p \vee q}$ is infallibly so (since the agent has implicit infallible knowledge of axioms) and c_p is incorrectly accepted (since the agent implicitly believes p but p is in fact false). But this is *not* a general requirement in our setting: availability does not imply acceptance. Indeed, mutually inconsistent evidence terms might be available (in the sense that the agent is aware of them, can compute them or is considering them), while in our models, belief is always consistent. For instance, our agent may be aware of some very weak evidence *against* p, say the fact (denoted by $c_{\neg p}$) that she never actually saw Mr. Nogot in a Ferrari, but she might choose to reject such evidence. In this case, she still keeps the same (implicit and explicit) beliefs as in the above story, as illustrated by the following model:

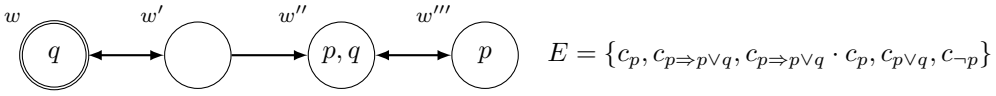

Figure 2: The Nogot-Havit scenario with additional evidence $c_{\neg p}$

Note that in both the above models, the agent has *no explicit introspection* about her beliefs or about her justifications! She simply does not consider such issues. If we want to model a situation in which the agent uses introspection to become aware of her explicit belief in p, then we obtain the model in Figure 3.

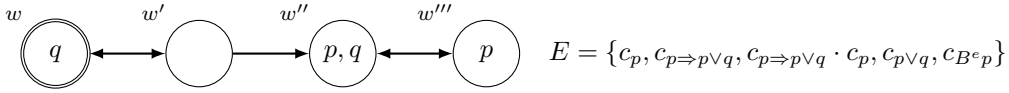

Figure 3: The Nogot-Havit scenario with introspection of belief

The fact that there are no $\rightsquigarrow$-arrows in any of these diagrams simply expresses the fact that we chose the current moment as the starting point (moment 0) in our story. Of course, a more accurate representation would include the *history* of how our agent came to believe $p \vee q$. A possible such history could be given by the model in Figure 4.

The real world at the current moment (previously denoted by w) is now denoted by w_3. The vertical arrows represent the immediate temporal precedence relation, going from one moment to the next moment. So the time flows downward in this diagram. According to this history, originally (at the "true" initial

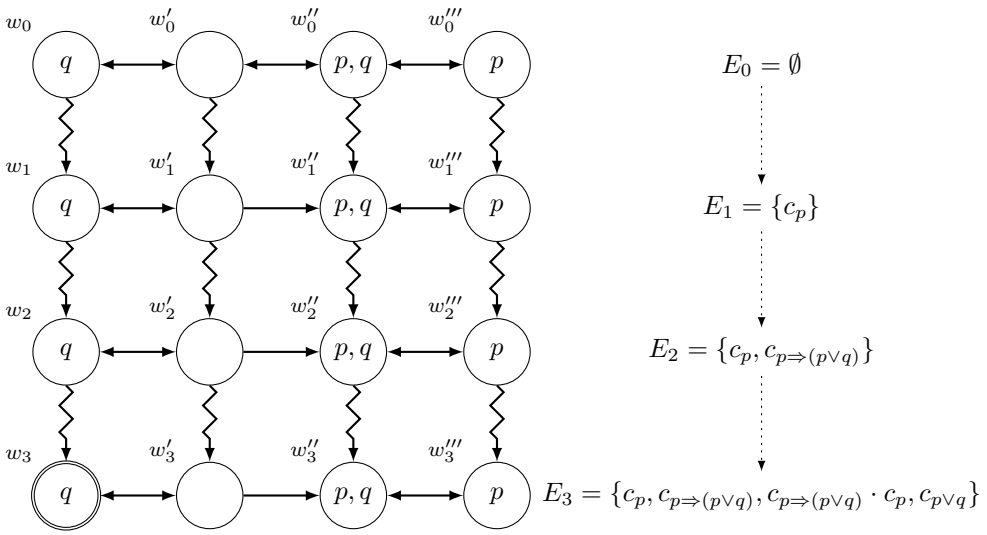

Figure 4: Temporal development leading to the Nogot-Havit scenario

moment w_0) the agent had no evidence ($E = \emptyset$), and no non-trivial beliefs about p and q, so she considered all four Boolean combinations to be equally plausible. After this, she received and accepted Mr. Nogot's testimonial evidence c_p, leading her to explicitly believe p (at moment w_1), and thus implicitly (but not yet explicitly!) believe $p \vee q$. At the next moment w_2, she thought about the logical axiom $p \Rightarrow p \vee q$, became aware of its applicability to this particular instance, and so the infallible evidence $c_{p\Rightarrow p\vee q}$ became available to her. She then used Modus Ponens, thereby computing the evidence term $c_{p\Rightarrow p\vee q} \cdot c_p$ that supports the conclusion $p\vee q$, certified this derivation by adding $c_{p\vee q}$ to her evidence set, and therefore acquired an explicit belief in $p \vee q$ (at the current moment w_3).

2.4 Proof System

Definition 2.14 (Theory)**.** JB, the *theory of justified belief*, is defined in Table 1.

Lemma 2.15 (Derivable Principles)**.** We have the following.

1. Application for Admissibility: $\vdash (s \gg \varphi) \Rightarrow (t \gg (\varphi \Rightarrow \psi) \Leftrightarrow (t \cdot s) \gg \psi)$.
2. Application: $\vdash (s \gg \varphi) \Rightarrow \big(t : (\varphi \Rightarrow \psi) \wedge s : \varphi \Leftrightarrow (t \cdot s) : \psi\big)$.
3. Weakening of Justified Belief: $\vdash t : \varphi \Rightarrow B\varphi$.
4. Certification of Implicit Belief: $\vdash B\varphi \Rightarrow c_\varphi : \varphi$.
5. Weakening of Justified Defeasible Knowledge: $\vdash (t \gg \varphi) \wedge G(t) \Rightarrow \Box\varphi$.
6. Certification of Defeasible Knowledge: $\vdash \Box\varphi \Rightarrow (c_\varphi \gg \varphi) \wedge G(c_\varphi)$.
7. Weakening of Justified Infallible Knowledge: $\vdash (t \gg \varphi) \wedge I(t) \Rightarrow K\varphi$.
8. Certification of Infallible Knowledge: $\vdash K\varphi \Rightarrow (c_\varphi \gg \varphi) \wedge I(c_\varphi)$.

Proof. (1) follows by the definition of admissibility. (2) follows by the definition of $u : \theta$, (1), and classical reasoning. (4) follows by the definition of $A(c_\varphi)$. (6) follows by the definition of $G(c_\varphi)$. (8) follows by the definition of $I(c_\varphi)$.

Recalling that $u : \theta = A(u) \wedge u \gg \theta$ and $A(u) = \bigwedge_{c_\theta \in \mathsf{sub}(u)} B\theta$, the proof of (3) is by induction on the construction of t. Base case: $\vdash c_\varphi : \varphi \Rightarrow B\varphi$ follows by the definition of $A(c_\varphi)$ and classical reasoning.

AXIOM SCHEMES

Classical Logic:	Axioms of Classical Propositional Logic
Knowledge of Available Evidence:	$\vdash Et \Rightarrow KEt$
Subterm Closure:	$\vdash E(t \cdot s) \Rightarrow Et \wedge Es$
	$\vdash E(t + s) \Rightarrow Et \wedge Es$
Certification of Evidence:	$\vdash t \gg \varphi \wedge Et \Rightarrow Ec_\varphi$
Admissibility:	$\vdash t \gg \varphi$ whenever $t \gg \varphi$
	$\vdash \neg(t \gg \varphi)$ whenever $t \not\gg \varphi$
Infallible Knowledge:	S5 axioms for K
Defeasible Knowledge:	S4 axioms for $\Box$
Indefeasibility:	$\vdash K\varphi \Rightarrow \Box\varphi$
Local Connectedness:	$\vdash K(\varphi \vee \Box\psi) \wedge K(\psi \vee \Box\varphi) \Rightarrow (K\varphi \vee K\psi)$
Normality of Y:	$\vdash Y(\varphi \Rightarrow \psi) \Rightarrow (Y\varphi \Rightarrow Y\psi)$
Propositional Perfect Recall:	$\vdash YK\varphi \Rightarrow KY\varphi$
Evidential Perfect Recall	$\vdash YEt \wedge \neg Y\bot \Rightarrow Et^Y$
Uniqueness of Past:	$\vdash \neg Y\varphi \Rightarrow Y\neg\varphi$
Persistence of Facts:	$\vdash Yp \Leftrightarrow (\neg Y\bot \Rightarrow p)$

RULES

$$\frac{\varphi \Rightarrow \psi \quad \varphi}{\psi}\ (\text{MP}) \qquad \frac{\varphi}{\Box\varphi}\ (\Box\text{N}) \qquad \frac{\varphi}{K\varphi}\ (K\text{N}) \qquad \frac{\varphi}{Y\varphi}\ (Y\text{N})$$

Table 1: Table 1. The theory JB

Induction step: assuming $\vdash s_i : \theta \Rightarrow B\theta$ and for each $i \in \{1, 2\}$ and $\theta \in \mathscr{F}$, we wish to show that $\vdash (s_1 \cdot s_2) : \varphi \Rightarrow B\varphi$. Let $S := \{\psi \mid s_1 \gg (\psi \Rightarrow \varphi) \wedge s_2 \gg \psi\}$. If $S = \emptyset$, then $\vdash \neg(s_1 \cdot s_2) \gg \varphi$ and so we have $\vdash (s_1 \cdot s_2) : \varphi \Leftrightarrow \bot$ and hence $\vdash (s_1 \cdot s_2) : \varphi \Rightarrow B\varphi$. So let us assume that $S \neq \emptyset$. We then have by classical reasoning that $\vdash (s_1 \cdot s_2) : \varphi \Leftrightarrow s_1 : (\psi \Rightarrow \varphi) \wedge s_2 : \psi$ for an arbitrarily selected $\psi \in S \neq \emptyset$. By the induction hypothesis and classical reasoning, we then have $\vdash (s_1 \cdot s_2) : \varphi \Rightarrow B(\psi \Rightarrow \varphi) \wedge B\psi$. Applying modal reasoning, it follows that $\vdash (s_1 \cdot s_2) : \varphi \Rightarrow B\varphi$.

Recalling that $G(u) = \bigwedge_{c_\theta \in \mathsf{sub}(u)} G\theta$, the proof of (5) is by induction on the construction of t. Base case: $t = c_\varphi$ and the result follows by the definition of $G(c_\varphi)$ and classical reasoning. Induction step: assuming $\vdash (s_i \gg \theta) \wedge G(s_i) \Rightarrow \Box\varphi$ for each $i \in \{1, 2\}$ and $\theta \in \mathscr{F}$, we wish to show that $\vdash (s_1 \cdot s_2) \gg \varphi \wedge G(s_1 \cdot s_2) \Rightarrow \Box\varphi$. We define $S := \{\psi \mid s_1 \gg (\psi \Rightarrow \varphi) \wedge s_2 \gg \psi\}$. If $S = \emptyset$, then $\vdash \neg(s_1 \cdot s_2) \gg \varphi$ and hence $\vdash (s_1 \cdot s_2) \gg \varphi \wedge G(s_1 \cdot s_2) \Leftrightarrow \bot$, from which the result follows by classical propositional reasoning. So let us assume that $S \neq \emptyset$. Choosing an arbitrary $\psi \in S$, we have $\vdash (s_1 \cdot s_2) \gg \varphi \wedge G(s_1 \cdot s_2) \Leftrightarrow s_1 \gg (\psi \Rightarrow \varphi) \wedge s_2 \gg \psi \wedge G(s_1) \wedge G(s_2)$ by classical propositional reasoning, the definition of admissibility, and the definition of $G(u)$. But we then have $\vdash (s_1 \cdot s_2) \gg \varphi \wedge G(s_1 \cdot s_2) \Rightarrow \Box(\psi \Rightarrow \varphi) \wedge \Box\psi$ by the induction hypothesis and classical reasoning and therefore that $\vdash (s_1 \cdot s_2) \gg \varphi \wedge G(s_1 \cdot s_2) \Rightarrow \Box\varphi$ by modal reasoning.

The argument for (7) is similar to that for (5). □

A common criticism of epistemic modal logic is that it suffers from *logical omniscience*: the agent believes all logical consequences of her beliefs, including in particular all valid formulas. But in JB, *only implicit belief $B\varphi$ and implicit knowledge—either infallible $K\varphi$ or defeasible $\Box\varphi$—satisfies logical omniscience.* That is, $K\varphi$ (or $\Box\varphi$) says only that the agent can come to infallibly (or defeasibly) know φ *only in principle.* Implicit knowledge may be thought of as "potential knowledge" of φ that the agent might in principle obtain, though perhaps she will never have this knowledge in actuality.

Explicit knowledge $K^e\varphi$ (or $\Box^e\varphi$) is very different. This represents the agent's *actual knowledge*, in that $K^e\varphi = K\varphi \wedge Ec_\varphi$ and $\Box^e\varphi = \Box\varphi \wedge Ec_\varphi$ say that the agent not only has the potential to realize her implicit knowledge of φ but also that *she has in fact gone through the trouble of obtaining and correctly validating the certificate of correctness c_φ for φ* (i.e., Ec_φ). Therefore, *explicit knowledge does not satisfy logical omniscience.*

Definition 2.16 (Iterated Axioms and Logical Terms)**.** An *iterated axiom* is a formula of the form

$$\underbrace{X_1X_2X_3\cdots X_n}_{\text{zero or more } X_i\text{'s}}\varphi \ ,$$

where each $X_i \in \{\Box, K, Y\}$ and φ is an axiom. The set of *logical terms* is the smallest set that contains certificates c_φ for each iterated axiom φ and is closed under the evidence-combining operator $t \cdot s$ (for Modus Ponens).

The logical terms are those that are built by applying the inference operator $\cdot$ to certificates of knowledge c_φ for iterated axioms φ. We may think of logical terms as the logical arguments we use to justify iterated axioms and their logical consequences. The forthcoming Theorem 2.18 shows that the agent can in principle always find purely logical justification to support infallible knowledge of logical truths.

Lemma 2.17 (Necessitation Elimination)**.** For each $\varphi \in \mathscr{F}$, we have $\vdash \varphi$ iff φ is provable from iterated axioms without the use of necessitation rules (i.e., KN, $\Box$N, or YN).

Proof. By induction on the number of necessitations. □

Theorem 2.18 (Theorem Internalization)**.** For each $\varphi \in \mathscr{F}$, we have $\vdash \varphi$ iff there exists a logical term t such that $\vdash I(t) \wedge t \gg \varphi$.

Proof. The right-to-left direction follows by Lemma 2.15(7), so we focus on the left-to-right direction. We write $\vdash^* \varphi$ to mean that φ is provable from iterated axioms without the use of necessitation rules. By Lemma 2.17, it suffices for us to prove by induction on the proof length that $\vdash^* \varphi$ implies there is a term t such that $\vdash^* (t \gg \varphi) \wedge I(t)$. Proceeding, we recall that $I(s) = \bigwedge_{c_\psi \in \mathsf{sub}(s)} K\psi$.

- Case: φ is an iterated axiom.

  Since φ is an iterated axiom, c_φ is a logical term and $\vdash^* K\varphi$. Hence $\vdash^* I(c_\varphi)$. Further, we have $c_\varphi \gg \varphi$ by the definition of admissibility and hence $\vdash c_\varphi \gg \varphi$. Conclusion: $\vdash^* (c_\varphi \gg \varphi) \wedge I(c_\varphi)$.

- Case: φ follows by MP from $\psi \Rightarrow \varphi$ and ψ.

  By the inner induction hypothesis, there exist logical terms t and s such that $\vdash^* t \gg (\psi \Rightarrow \varphi) \wedge I(t)$ and $\vdash^* (s \gg \psi) \wedge I(s)$. Hence $\vdash^* (t \cdot s) \gg \psi$ and $\vdash^* I(t \cdot s)$. Conclusion: $\vdash^* ((t \cdot s) \gg \varphi) \wedge I(t \cdot s)$. □

Theorem 2.19 (Soundness and Completeness for Non-Standard Models)**.** JB is sound and strongly complete with respect to the class of all models.

Proof. Soundness is by induction on the length of derivation. We omit the details. Completeness is by way of a canonical model construction. We define the *canonical model* $\Omega := (W^\Omega, [\![\cdot]\!], \sim, \geq, \rightsquigarrow, E)$ by setting

$$\begin{array}{rcl}
W^\Omega & := & \{\Gamma \subseteq \mathscr{F} \mid \Gamma \text{ is maximal consistent}\} \ , \\
[\![p]\!] & := & \{\Gamma \in W \mid p \in \Gamma\} \ , \\
\Gamma \sim \Delta & \text{iff} & \{\theta \mid K\theta \in \Gamma\} \subseteq \Delta \ , \\
\Gamma \geq \Delta & \text{iff} & \{\theta \mid \Box\theta \in \Gamma\} \subseteq \Delta \ , \\
\Gamma \rightsquigarrow \Delta & \text{iff} & \{\theta \mid Y\theta \in \Delta\} \subseteq \Gamma \ , \text{ and} \\
E(\Gamma) & := & \{t \in \mathscr{T} \mid Et \in \Gamma\} \ .
\end{array}$$

It is easy to see that Ω is a model: most properties follow by standard correspondence theory [7], while the properties of Evidential Perfect Recall, Knowledge of Available Evidence, Subterm Closure, and Certification of Evidence follow by modal reasoning using axioms of the same name.

What remains is for us to prove the *Truth Lemma*: for each $\Gamma \in W^\Omega$ and each $\theta \in \mathscr{F}$, we have $\theta \in \Gamma$ iff $\Gamma \models_\Omega \theta$. The proof is by induction on the construction of θ. All steps of this induction are standard [7], except the ones refering to formulas of the form Et or $t \gg \varphi$. For formulas Et, to have $Et \in \Gamma$ is what it means to have $t \in E(\Gamma)$, which is itself equivalent to $\Gamma \models Et$ by the definition of truth. For formulas $t \gg \varphi$, it follows immediately from the Admissibility axioms (Table 1), maximal consistency, and the definition of truth that we have $(t \gg \varphi) \in \Gamma$ iff $\Gamma \models t \gg \varphi$. This completes the proof of the Truth Lemma. Strong completeness follows immediately in the usual way [7]. □

Theorem 2.20 (Completeness for Standard Models, Finite Model Property)**.** JB is sound and weakly complete with respect to the class of *standard* models. Moreover, it is also weakly complete with respect to the class of *finite* standard models.

Proof Sketch. First, we unravel the canonical model to Ω obtain another model $\Omega \times \mathbb{Z}$ in which $\rightsquigarrow$ is acyclic. For this, we take copies (w,k) of each world w in Ω and each integer $k \in \mathbb{Z}$. Accordingly, the set of worlds of our new model $\Omega \times \mathbb{Z}$ will be $W^{\Omega} \times \mathbb{Z}$ with $(w,k) \sim (w',k')$ iff $k = k'$ and $w \sim w'$, $(w,k) \geq (w',k')$ iff $k = k'$ and $w \geq w'$, $(w,k) \rightsquigarrow (w',k')$ iff $k' = k-1$ and $w \rightsquigarrow w'$, $E(w,k) = E(w)$, and $[\![p]\!] = \{(w,k) \mid w \in [\![p]\!]\}$. We obtain a (non-standard) model $\Omega \times \mathbb{Z}$, in which the relation $\rightsquigarrow$ is acyclic. One can easily check (by induction on formulas) that $(w,k) \models_{\Omega\times\mathbb{Z}} \varphi$ iff $w \models_{\Omega} \varphi$ for each $k \in \mathbb{Z}$ and $\varphi \in \mathscr{F}$.

Fix a consistent formula ψ and a world v in Ω satisfying $v \models_{\Omega} \psi$ and hence $(v,0) \models_{\Omega\times\mathbb{Z}} \psi$. Take the submodel $M := (\Omega \times \mathbb{Z})_{(v,0)}$ generated by $(v,0)$ via the relations $\sim$, $\geq$, and $\leftsquigarrow$; i.e., M is the restriction of (the relations, functions, and valuation of) the model $\Omega \times \mathbb{Z}$ to the set

$$W^M := \{(w,k) \mid (w,k) \rightsquigarrow^* (v',0) \text{ for some } v' \sim v\} \ .$$

Here we use the iterated temporal arrows $\rightsquigarrow^n$ and $\rightsquigarrow^*$, which are defined inductively by setting $w \rightsquigarrow^0 w'$ iff $w = w'$, $w \rightsquigarrow^{n+1} w'$ iff there exists w'' such that $w \rightsquigarrow^n w'' \rightsquigarrow w'$, and $w \rightsquigarrow^* w'$ iff there exists some $n \in \mathbb{N}$ such that $w \rightsquigarrow^n w'$. It is easy to see that the set W^M is closed (as a submodel of $\Omega \times \mathbb{Z}$) under the relations $\sim$, $\geq$, and $\leftsquigarrow$. (The proof for $\sim$ uses Propositional Perfect Recall, and the proof for $\geq$ uses Indefeasibility and the result for $\sim$; see Definition 2.11.) So M is indeed a generated submodel, and hence by standard results in modal logic about generated submodels [7], it follows that for every $(w,n) \in W^M$ and every formula φ, we have $(w,n) \models_M \varphi$ iff $w \models_{\Omega} \varphi$. Hence $(v,0) \models_M \psi$. It is also easy to see that each temporal layer of this model is connected: $(w,n) \sim (w',n')$ holds in M iff $n = n'$.

Let now m be the modal Y-depth of formula ψ; that is, m is the maximum number of nested Y-modalities occuring in ψ. For each $0 \leq n \leq m$, let $\Psi_n := \mathsf{sub}_n(\psi)$ be the set of all subformulas of ψ of modal Y-depth less than or equal to n. We construct a new model M' by "cutting" M to depth m (i.e., deleting all worlds (w,n) having $n > m$) and applying to the n^{th} temporal layer of the resulting submodel (for each nonnegative $n \leq m$) the transitive filtration with respect to the set Ψ_{m-n}. More precisely, we define an equivalence relation $\equiv$ on W^M by

$$(w,n) \equiv (w',n') \text{ iff } (n = n') \wedge \forall \varphi \in \Psi_{m-n}\big((w,n) \models_M \varphi \Leftrightarrow (w',n') \models_M \varphi\big) \ .$$

The set W' of possible worlds in our new model M' will consist of all the $\equiv$-equivalence classes of worlds of depth at most m:

$$W' = \{\overline{(w,n)} \mid (w,n) \in W^M \text{ and } 0 \leq n \leq m\} \ ,$$

where $\overline{(w,n)} = \{(w',n') \in W^M \mid (w,n) \equiv (w',n')\}$. For $R \in \{\sim, \rightsquigarrow\}$, we take the induced relations on classes (the "smallest filtration"):

$$\overline{(w,n)} R \overline{(w',n')} \text{ iff } \exists (v,k) \in \overline{(w,n)}, \exists (v',k') \in \overline{(w',n')} : (v,k) R (v',k') \ .$$

In the case of $\sim$, this amounts to $\overline{(w,n)} \sim \overline{(w',n')}$ iff $n = n'$, and in the case of $\rightsquigarrow$, this boils down (with the aid of Propositional Perfect Recall) to

$$\begin{aligned} \overline{(w,n)} \rightsquigarrow \overline{(w',n')} \text{ iff } \ & (n' = n-1) \wedge \\ & \exists w''\big(w'' \rightsquigarrow w' \wedge \forall \varphi \in \Psi_{m-n}(w \models_{\Omega} \varphi \Leftrightarrow w'' \models_{\Omega} \varphi)\big) \ . \end{aligned}$$

For $\geq$, we take the transitive filtration:

$$\overline{(w,n)} \geq \overline{(w',n')} \text{ iff } (n' = n) \wedge \forall \varphi \in \Psi_{m-n} \, (w \models_{\Omega} \Box\varphi \Rightarrow w' \models_{\Omega} \Box\varphi) \ .$$

The valuation is defined as in any filtration. Formulas of the form Et and $t \gg \varphi$ are treated in the same way that filtration treats atomic formulas. (For formulas Et, this works because all worlds in M belonging to the same temporal layer n are $\sim$-indistinguishable and so agree on the truth values of formulas Et by Evidential Perfect Recall. For formulas $t \gg \varphi$, it works because all worlds in Ω agree on the truth values of formulas $t \gg \varphi$.)

The resulting model M' is finite and inherits all the properties of the previous models, in particular $\rightsquigarrow$ is acyclic and $>$ is transitive and irreflexive (and hence also acyclic). Since every acyclic relation on a finite set is well-founded, M' is a standard model. Finally, it is trivial to check (by induction on n) that, for every $0 \leq n \leq m$, every world $\overline{(w, m-n)} \in W'$ and every formula $\varphi \in \Psi_n$, we have $(w,n) \models_{M'} \varphi \Leftrightarrow (w,n) \models_M \varphi$. In particular, we obtain $(v,0) \models_{M'} \psi$. □

Corollary 2.21 (Decidability)**.** The logic JB is decidable.

Proof. The size of the finite model M' constructed in the above proof is bounded by $N = m \cdot 2^{|\mathsf{sub}(\psi)|}$, where m is the modal Y-depth of ψ. Hence we can simply investigate one by one all models (up to isomorphism) of size at most N, checking whether ψ is satisfied in any of them. □

It is common in Justification Logic to have "evidence internalization terms" $!t$ and $?t$ that allow the agent to introspectively verify his evidence or lack thereof according to the following schemes:

$$\begin{array}{lrcl} \text{PC.} & t:\varphi & \Rightarrow & !t:(t:\varphi) \\ \text{NC.} & \neg t:\varphi & \Rightarrow & ?t:(\neg t:\varphi) \end{array}$$

PC ("Positive Checker") says that if the agent has potential evidence t for φ, then she can in principle use $!t$ (pronounced "bang t") to check that t is indeed potential evidence for φ. NC ("Negative Checker") says that if t is not potential evidence for φ, then the agent can in principle check this as well using $?t$. PC is typically required in order for the Theorem Internalization result to hold. However, as we saw above, positive checker is not needed to prove this result for JB. The reason is that our certificates c_φ allow us to recover a form of PC. In fact, certificates allow us to recover a form of NC as well. Indeed, the following schemes are derivable in our system

$$\begin{array}{lrcl} \text{PC'.} & t:\varphi & \Rightarrow & (c_{t:\varphi}):(t:\varphi) \\ \text{NC'.} & \neg t:\varphi & \Rightarrow & (c_{\neg t:\varphi}):(\neg t:\varphi) \end{array}$$

The next result is a kind of "doxastic internalization," showing that all the implicit beliefs of a rational agent are in principle justifiable, all her explicit beliefs are explicitly justified, and all her (infallible, or at least, defeasible) knowledge can be given a correct (i.e., infallible, or at least "good") justification.

Theorem 2.22 (Knowledge and Belief Internalization)**.** For each $\varphi \in \mathscr{F}$:

$$\begin{array}{lll} w \models_M B\varphi & \text{iff} & \text{there is a term } t \text{ such that } w \models_M t:\varphi \text{ ;} \\ w \models_M B^e\varphi & \text{iff} & \text{there is a term } t \text{ such that } w \models_M t:{}^e\varphi \text{ ;} \\ w \models_M \Box\varphi & \text{iff} & \text{there is a term } t \text{ such that } w \models_M t:\varphi \wedge G(t) \text{ ;} \\ w \models_M \Box^e\varphi & \text{iff} & \text{there is a term } t \text{ such that } w \models_M t:{}^e\varphi \wedge G(t) \text{ ;} \\ w \models_M K\varphi & \text{iff} & \text{there is a term } t \text{ such that } w \models_M t:\varphi \wedge I(t) \text{ ;} \\ w \models_M K^e\varphi & \text{iff} & \text{there is a term } t \text{ such that } w \models_M t:{}^e\varphi \wedge I(t) \text{ .} \end{array}$$

In words: something is (implicit or explicit) belief iff it is (implicitly or explicitly) justifiable by some (implicitly or explicitly accepted) evidence; something is (implicit or explicit) defensible knowledge iff it is (implicitly or explicitly) justifiable by some good evidence; something is (implicit or explicit) infallible knowledge iff it is (implicitly or explicitly) justifiable by some infallible evidence.

Proof. The right-to-left directions of the statements about implicit knowledge and belief follow by soundness, the validity $\models t:\varphi \Rightarrow t \gg \varphi$, and Lemma 2.15 parts (3), (5), and (7). For the left-to-right directions of the statements about implicit knowledge and belief, take $t := c_\varphi$. We have $\models c_\varphi \gg \varphi$ by the definitions of admissibility and truth. Hence $w \models c_\varphi : \varphi$ by the assumption on the left, which implies $B\varphi = A(c_\varphi)$. In the case of the implicit knowledge statements, the additional conjunction on the right side of the "iff" follows by the assumption on the left and the fact that $G(c_\varphi) = \Box\varphi$ and $I(c_\varphi) = K\varphi$. Results for the explicit knowledge and belief statements follow by taking $t = c_\varphi$ and recalling that $\models c_\varphi \gg \varphi$. □

3 Evidence Dynamics

In this paper we will consider only four types of epistemic actions. (1) The first type is the action $t+$ by which an evidence term t becomes available: the agent can form this term, or becomes aware of the possibility of this evidence, without necessarily accepting it. A special example of this is $c_\varphi+$, for some axiom instance φ: this represents the "logical" action of becoming aware of an axiom instance. Other examples include the actions $c_{B\varphi}+$ and $c_{\neg B\varphi}+$: these represent acts of introspection by which the agent becomes aware of some of her implicit beliefs and non-beliefs. (2) The second type is the action $t\otimes s$ by which, given previously available evidence terms t and s, the agent forms a new term $t\cdot s$ representing the logical action of performing a Modus Ponens step. (3) The third type is the action $t!$ of updating with some "hard" evidence t (coming from an absolutely infallible source). This corresponds to the standard DEL update (all the worlds that do not fit evidence t are deleted), except that its input is a new piece of evidence t rather than a proposition. Morever, this is an "explicit" update: the new evidence becomes *available to* (and accepted with absolute certainty by) the agent, although only in its "past" form (t^Y), since it is evidence about the world as it was *before* the update. Similarly, all the previously available evidence is still available but only its "past" form as evidence about the situation before the update.[3] (4) The fourth type is the action $t\Uparrow$ of upgrading with some "soft" evidence t coming from a strongly trusted (though not infallible) source. The new evidence is (strongly) accepted (although not infallibly known). Modulo the same differences as in the case of update, this is essentially an explicit version of the action called "radical upgrade" in the DEL literature and "lexicographic revision" in the Belief Revision literature. All worlds that fit the new evidence become more plausible than the worlds that do not fit it.

There are of course many other possible epistemic actions. In particular, one can define an explicit version of the action $t\uparrow$ known as "conservative upgrade" in the DEL literature and as "minimal revision" in the Belief Revision literature: only the most plausible worlds fitting the new evidence become the most plausible overall. But for simplicity, in this paper we restrict ourselves to the four types of actions mentioned above.

Definition 3.1 (Language with Updates)**.** $\mathscr{L}^{act} := (\mathscr{T}^{act}, \mathscr{F}^{act})$ is the extension of the basic language $\mathscr{L} = (\mathscr{T}, \mathscr{F})$ obtained by adding modal operators $[\alpha]$ for epistemic actions $\alpha \in \{t+, t\otimes s, t!, t\Uparrow\}$, for every $t, s \in \mathscr{T}$. (Note that this not only extends the set of formulas: due to our terms c_φ, it also extends the set of terms.) The notions of subterm, subformula, admissibility, and model are lifted to $\mathscr{L}^{act}$ in the obvious way. We assign the following informal readings to the new modal formulas:

$[t+]\varphi$	says that *after making evidence t available, φ is true*,
$[t\otimes s]\varphi$	says that *after combining evidence t and s (by Modus Ponens), φ is true*,
$[t!]\varphi$	says that *after updating with hard evidence t, formula φ is true*, and
$[t\Uparrow]\varphi$	says that *after (radically) upgrading with soft evidence t, formula φ is true.*

For every action α we define a sentence pre_α, called the *precondition* of α, and a set of terms $\mathscr{T}(\alpha)$ called the *evidence set* of α:

$$
\begin{array}{lcl}
\mathsf{pre}_{t+} & := & \mathsf{pre}_{t\Uparrow} = \top \\
\mathsf{pre}_{t!} & := & \mathsf{con}_t = \bigwedge\{\theta \mid t \gg \theta\} \\
\mathsf{pre}_{t\otimes s} & := & Et \wedge Es \\
 & & \\
\mathscr{T}(t+) = \mathscr{T}(t!) = \mathscr{T}(t\Uparrow) & := & \mathsf{sub}(t) \cup \{c_\theta \mid s \gg \theta \text{ for some } s \in \mathsf{sub}(t)\} \\
\mathscr{T}(t\otimes s) & := & \{t\cdot s\} \cup \{c_\theta \mid t\cdot s \gg \theta\}
\end{array}
$$

The precondition pre_α captures the condition of possibility of action α: actions $t+$ and $t\Uparrow$ can always happen, $t!$ can only happen if t really is hard information (i.e., if the proposition supported by t is actually true), and $t\otimes s$ can only happen if t and s are already available. The evidence set $\mathscr{T}(\alpha)$ consists of all the evidence terms that become available due to α.

[3] This is needed in order to deal with Moore sentences. However, we want to endow the agent with some basic insight of the principle that epistemic actions do not change ontic facts: she should be instantly aware of this, without having to perform additional inference steps to derive this. This explains our definition of φ^Y, which leaves unchanged all the purely propositional formulas (i.e., Boolean combinations of atoms), so that for such formulas we have $(c_\varphi)^Y = c_\varphi$. In effect, epistemic actions with factual evidence will actually produce evidence about the current world as it is after the update.

To say that a formula is *reduced* means that it does not contain a subformula of the form $[\alpha]\varphi$. Note that the notion of *subformula* is lifted to $\mathscr{L}^{act}$ from that given in Definition 2.1 in the obvious way. So, for example, φ is *not* a subformula of $c_{[c_\varphi!]\varphi} \gg [c_\varphi!]\varphi$.

Definition 3.2 (Truth for $\mathscr{L}^{act}$)**.** Given a model $M = (W, [\![\cdot]\!], \sim, \geq, \rightsquigarrow, E)$, a world $w \in W$ and an epistemic action $\alpha \in \{t+, t\otimes s, t!, t\Uparrow\}$, we will use the notation w^α to denote the ordered pair (w, α), and we will use this to formally represent the "updated" world resulting from performing action α in world w. The satisfaction relation $(M, w) \models \varphi$ is defined as an extension of our previous definition of truth (Definition 2.12) obtained by adding the following clauses for dynamic modalities $[\alpha]\varphi$, with $\alpha \in \{t+, t \otimes s, t!, t\Uparrow\}$:

$$x \models_M [\alpha]\varphi \quad \text{iff} \quad x^\alpha \models_{M[\alpha]} \varphi \quad \text{with} \quad M[\alpha] := (W^\alpha, [\![\cdot]\!]^\alpha, \sim^\alpha, \geq^\alpha, \rightsquigarrow^\alpha, E^\alpha)$$

and

$$\begin{array}{lcl}
W^\alpha & := & W \cup \{w^\alpha \mid w \in [\![\mathsf{pre}_\alpha]\!]\} \\
E^\alpha(w) & := & E(w) \text{ for } w \in W \\
E^\alpha(w^\alpha) & := & \{u^Y \mid u \in \mathscr{T}(\alpha) \cup E(w)\} \\
[\![p]\!]^\alpha & := & [\![p]\!] \cup \{w^\alpha \in W^\alpha \mid w \in [\![p]\!]\} \\
\sim^\alpha & := & \sim \cup \{(w^\alpha, v^\alpha) \mid w \sim v\} \\
\rightsquigarrow^\alpha & := & \rightsquigarrow \cup \{(w, w^\alpha) \mid w \in [\![\mathsf{pre}_\alpha]\!]\} \\
\geq^\alpha & := & \geq \cup \{(w^\alpha, v^\alpha) \mid w \geq v\} \text{ for } \alpha \in \{t+, t\otimes s, t!\} \\
\geq^{t\Uparrow} & := & \geq \cup \{(w^{t\Uparrow}, v^{t\Uparrow}) \mid \ (w \notin [\![\mathsf{con}_t]\!] \wedge v \in [\![\mathsf{con}_t]\!]) \vee \\
 & & \qquad\qquad (w \notin [\![\mathsf{con}_t]\!] \wedge w \geq v) \vee (v \in [\![\mathsf{con}_t]\!] \wedge w \geq v)\} \\
\geq^{t\Uparrow} & := & \geq \cup \{(w^{t\Uparrow}, v^{t\Uparrow}) \mid w \geq v\} \text{ for } t \notin \mathscr{T}^e
\end{array}$$

3.1 Example Continued

We now generate the temporal progression of Lehrer's Nogot-Havit scenario using our epistemic actions. First, we begin from a situation of complete ignorance represented by the model in Figure 5.

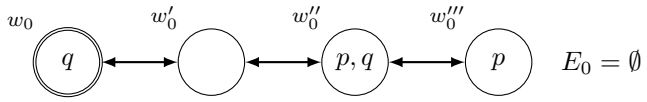

Figure 5: Initial situation with complete ignorance

We successively apply the following actions: first the upgrade $c_p\Uparrow$, by which our agent upgrades with (i.e., accepts as soft evidence) Mr. Nogot's false testimony; then the logical action $c_{p\to p\vee q}+$, by which she becomes aware of evidence for the axiom $p \to p \vee q$; then the action $c_{p\to p\vee q} \otimes c_p$, by which she combines her pieces of evidence using Modus Ponens, acquiring explicit justified belief in $p \vee q$. The result of these transformations is exactly the model in Figure 4.

But now we can go one step further in the future. Suppose that the agent receives hard evidence (from an infallible source such as Lehrer's Critic, who always tells the truth) that Mr. Nogot does not own a Ferrari. We can interpret this as an update $c_{\neg p}!$, which can be applied to the model in Figure 4 to yield the model in Figure 6 below. Our agent has been "Gettierized": some new true evidence defeated her true justified belief!

3.2 An Explicit Version of Moore Sentences

Finally, to explain the need for the Y operator and justify the way we defined evidential dynamics, let us consider the situation after the first step above, as shown in Figure 7. Our agent has just received Mr. Nogot's (lying) testimony c_p and performed an upgrade $c_p\Uparrow$. Notice that now she explicitly believes p, but she is not yet aware of it: she has not yet performed any introspection acts, so she does not explictly know that she explicitly believes p. She simply has not reflected upon her own beliefs. Now suppose the infallible Critic intervenes, giving her hard evidence $c_{\neg p}$ that $\neg p$, and, at the same time, she reflects upon her beliefs, becoming infallibly aware of her explicit belief in p, justified by her "hard" introspective evidence $c_{B^e p}$. Note

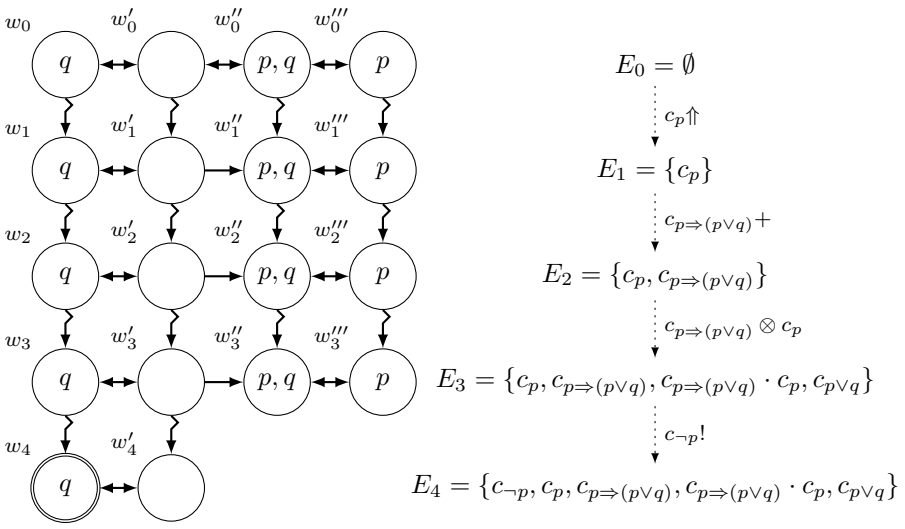

Figure 6: "Gettierization" of the agent from the Nogot-Havit scenario

that, unlike the classical examples of Moore sentences learnt by a fully introspective agent, in this case *both these pieces of hard evidence are non-redundant*: the agent explicitly learns something new from each of them. We can interpret this action as an update $(c_{\neg p} + c_{B^e p})!$, which applied to the model in Figure 7 yields the model in Figure 8. Note the new evidence set: the agent did not simply add the new term (and its subterms $c_{B^e p}$ etc.) to her set. Adding $c_{B^e p}$ would in any case be useless, since by now (in the new model) she already believes $\neg p$, so the evidence $c_{B^e p}$ would simply be rejected. But this action does give her a new, important (and correct!) introspective piece of evidence, namely $c_{Y B^e p}$: she explicitly learns that she used to believe p (before the update). This is indeed new: at the time when she was holding the explicit belief in p, she was not introspective about it. Now that she is aware of this (past) belief, she does not hold it anymore!

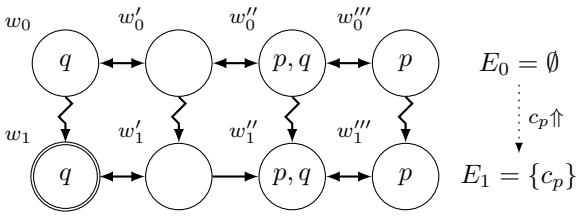

Figure 7: After action $c_p\Uparrow$

3.3 Robustness of Knowledge

Using dynamics, we can now show in what sense our notions of knowledge capture their intended interpretations:

Theorem 3.3. *Something is (explicit) defeasible knowledge iff it is (explicitly) believed to have been true no matter what new* hard *(and hence true) evidence is received:*

$$w \models \Box\varphi \quad \text{iff} \quad w \models [t!]BY\varphi \text{ for every evidence term } t\,;$$
$$w \models \Box^e\varphi \quad \text{iff} \quad w \models [t!]B^eY\varphi \text{ for every evidence term } t\,.$$

Something is (explicit) infallible knowledge iff it is (explicitly) believed to have been true no matter what new

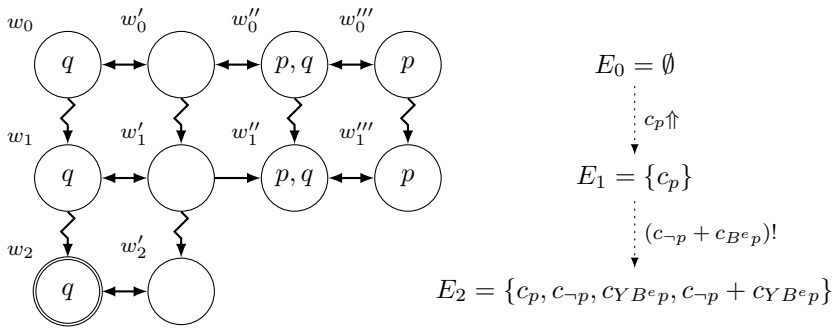

Figure 8: After action $(c_{\neg p} + c_{B^e p})!$

soft *(possibly false) evidence is received:*

$$\begin{array}{lll} w \models K\varphi & \text{iff} & w \models [t\Uparrow]BY\varphi \text{ for every evidence term } t\,; \\ w \models K^e\varphi & \text{iff} & w \models [t\Uparrow]B^eY\varphi \text{ for every evidence term } t\,. \end{array}$$

Definition 3.4 (Theory)**.** JB, the *theory of dynamic justified belief*, is defined in Table 2.

	Axiom Schemes
JB:	Axioms and rules of JB (Table 1)
Persistence of Facts:	$[\alpha]p \iff (\mathsf{pre}_\alpha \Rightarrow p)$
Functionality:	$[\alpha]\neg\varphi \iff (\mathsf{pre}_\alpha \Rightarrow \neg[\alpha]\varphi)$
Conjunction Distributivity:	$[\alpha](\varphi \wedge \psi) \iff [\alpha]\varphi \wedge [\alpha]\psi$
Evidence Dynamics:	$[\alpha]Et^Y$ for $t \in \mathscr{T}(\alpha)$
	$[\alpha]Et^Y \iff (\mathsf{pre}_\alpha \Rightarrow Et)$ for $t \notin \mathscr{T}(\alpha)$
	$[\alpha]Es \iff \neg\mathsf{pre}_\alpha$ for $s \notin \{t^Y \mid t \in \mathscr{T}\}$
Admissibility Dynamics:	$[\alpha](t \gg \varphi) \iff (\mathsf{pre}_\alpha \Rightarrow t \gg \varphi)$
Knowledge Dynamics:	$[\alpha]K\varphi \iff (\mathsf{pre}_\alpha \Rightarrow K[\alpha]\varphi)$
	$[\alpha]\Box\varphi \iff (\mathsf{pre}_\alpha \Rightarrow \Box[\alpha]\varphi)$ for $\alpha \in \{t+, t \otimes s, t!\}$
	$[t\Uparrow]\Box\varphi \iff \Box(\neg\mathsf{con}_t \Rightarrow [t\Uparrow]\varphi) \wedge$
	$(\mathsf{con}_t \Rightarrow \Box[t\Uparrow]\varphi \wedge K(\neg\mathsf{con}_t \Rightarrow [t\Uparrow]\varphi))$
Temporal Dynamics:	$[\alpha]Y\varphi \iff (\mathsf{pre}_\alpha \Rightarrow \varphi)$

Table 2: Table 2. The theory JB

Lemma 3.5 (Reduction)**.** For each $\varphi \in \mathscr{F}^{act}$, there is a reduced $\varphi^\dagger \in \mathscr{F}^{act}$ such that $\vdash \varphi \Leftrightarrow \varphi^\dagger$.

Theorem 3.6 (Soundness and Completeness)**.** $\vdash \varphi$ iff $\models \varphi$ for each $\varphi \in \mathscr{F}^{act}$.

4 Conclusion and Comparison with Related Work

Our framework is a variant of the traditional Justification Logic semantics for justified belief and knowledge. But a key difference lies in our semantics for justified belief $t : \varphi$. According to the traditional Justification Logic semantics (the "Fitting semantics") [1, 2, 3], the agent has justified belief $t : \varphi$ if and only if: (1) t is admissible for φ and (2) the agent believes φ (in the sense that $B\varphi$ holds for an appropriate modal operator B). Here the notion of admissibility is weaker than ours: the traditional semantics does not require that each atomic piece of evidence is admissible for a finite number of formulas (let alone for a unique formula); further, the traditional account does not require that a compound piece of evidence $t \cdot s$ or $t + s$ be admissible for a formula if and only if the constituents t and s are admissible for certain related formulas

(only the "if" direction is required). For example, while both the traditional account and ours agree that $(s \gg \varphi) \Rightarrow (t \gg (\varphi \Rightarrow \psi) \Rightarrow (t \cdot s) \gg \psi)$, only ours also ensures that $(s \gg \varphi) \Rightarrow ((t \cdot s) \gg \psi \Rightarrow t \gg (\varphi \Rightarrow \psi))$. It is therefore possible in the traditional semantics that a compound piece of evidence $t \cdot s$ is admissible for a formula without its constituent pieces of evidence t and s having any relation to this formula on their own. We forbid this in our setup: admissibility always provides an "evidential chain" that links the formulas for which a compound piece of evidence is admissible to the finitely many formulas for which its constituent pieces of evidence are admissible. This evidential chain plays an important role in our semantics of (implicit) justified knowledge and belief: the agent has (implicit) justified knowledge or belief if and only if she has a piece of evidence t whose structure describes a step-by-step evidential reasoning process along a well-formed evidential chain from basic assumptions—all of which are supported by acceptable, good, or infallible certificates—to a final conclusion that is justified by the reasoning process encoded by the evidential chain. (Here the emphasis is on the "only if" direction: she cannot have justified knowledge or belief without having a proper justification!) We therefore have not only that the resulting compound piece of evidence is admissible for the conclusion φ, but that this evidence necessarily encodes a meaningful and evidentially sound argument for t (that begins with premises c_ψ whose certified ψ's are validated according to belief, defeasible knowledge, or infallible knowledge, and that proceeds stepwise via Modus Ponens $s_1 \cdot s_2$ and monotonic combination of evidence $s_1 + s_2$ to the eventual conclusion φ). In contrast, the traditional semantics admits the possibility that a combined piece of evidence $t \cdot s$ is admissible for an assertion φ without any evidential chain linking admissible formulas of t and of s to φ, and, further, that one or more of t or s is individually faulty (in the sense that none of the formulas for which it is admissible is believed or known). In such a circumstance, the agent has a "justified belief" in φ based on a compound piece of evidence $t \cdot s$ whose components t and s are individually unrelated to φ and are not all reliable evidence for those things for which they are admissible. We expressly forbid such a possibility in our semantics: compound pieces of evidence always provide a proper evidential chain from reliable certificates to a justified conclusion.

In our setting we consider different ways in which an agent can change his evidence and update his beliefs and knowledge: the agent can be confronted with new evidence coming from an external source, he can reach new conclusions by bringing pieces of available evidence together or he can become aware of how to perform a specific inference on evidence. Related work in the Dynamic Epistemic Logic literature can be found in [20, 21], where van Benthem and Velázquez-Quesada provide a logical system that can handle information changes that include an agent's acts of inference. These authors start from a particular type of awareness logic and enhance it with the dynamic features that are common in Dynamic Epistemic Logic. Their focus lies on how implicit and explicit knowledge can be related, specifically addressing the question "what do agents have to do to make their implicit knowledge explicit?" [20]. In [21], their work was extended to consider implicit and explicit belief and belief revision. If we compare the "evidence sets" in our models with van Benthem and Velázquez-Quesada's "awareness sets," we see that their setting can be thought of as a special case of ours. Indeed as noted in [21], their setting is the special case in which each formula can have one unique justification, namely the formula itself. Another important difference between the two approaches points to how one can deal with the synchronization between explicit and implicit knowledge under epistemic actions that change the model at hand. Indeed as noted also in [20], an update with epistemic higher-order information (i.e., information that refers to the knowledge of beliefs of the agent) can easily push the explicit and implicit knowledge of the agent "out of sync." An example is our explicit version of a Moore-sentence–type announcement in Figure 8. The evidential updates defined by van Benthem and Velázquez-Quesada are simply adding the new evidence to the awareness set (or, in our setting, to the evidence set). But in the case of a Moore announcement, this would add evidence that is already obsolete! Van Benthem and Velázquez-Quesada attempt to solve this problem by proposing in [20, 21] a different definition of "explicit knowledge," which in our setting would correspond to $K^e\varphi = K(\varphi \wedge Ec_\varphi)$. In our view, this variant definition, although interesting, is not enough to solve the synchronization problem because it loses track of the past. According to us, this problem simply cannot be solved if one does not keep track of the past: our solution is to enhance our models with a temporal operator that allows us to refer back to the previous state before the epistemic action happened. Hence all new evidence that becomes available to the agent comes in its "past" form: it always is evidence about the world as it was before the learning or reasoning action took place. In the case of purely propositional information, this will still be valid evidence about the current state of the world. But in the case of doxastic information, it will simply be (new!) evidence about the agent's past beliefs: a previously non-introspective agent becomes aware of some of her past beliefs, even if in the meantime she

already changed them!

Overall, our explicit treatment of justification, awareness, dynamics and epistemics tackles several of the issues that have been left open in the work of [20]. In future work we aim to combine our approach with the semantic treatment of "evidence" in [19], where van Benthem and Pacuit use neighborhood models for evidence and develop a theory of "evidence management."

References

[1] S. Artemov. The logic of justification. *Review of Symbolic Logic*, 1(4):477–513, 2008.

[2] S. Artemov and M. Fitting. Justification logic. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2011 edition, 2011.

[3] S. Artemov and E. Nogina. Introducing justification into epistemic logic. *Journal of Logic and Computation*, 15(6):1059–1073, 2005.

[4] A. Baltag and L. S. Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.

[5] A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK VII)*, pages 43–56, Evanston, Illinois, USA, 1998.

[6] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Texts in Logic and Games, pages 9–58. Amsterdam University Press, 2008.

[7] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

[8] E. Gettier. Is justified true belief knowledge? *Analysis*, 23:121–123, 1963.

[9] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[10] K. Lehrer. *Theory of Knowledge*. Routledge, London, 1990.

[11] K. Lehrer. *Theory of Knowledge*. Westview Press, United States, 2000.

[12] K. Lehrer and T. J. Paxson. Knowledge: Undefeated justified true belief. *Journal of Philosophy*, 66:225–237, 1969.

[13] G. Pappas and M. Swain, editors. *Essays on Knowledge and Justification*. Cornell Univ. Press, Ithaca, NY, 1978.

[14] B. Renne, J. Sack, and A. Yap. Dynamic epistemic temporal logic. In X. He, J. Horty, and E. Pacuit, editors, *Logic, Rationality, and Interaction, Second International Workshop, LORI 2009, Chongqing, China, October 8–11, 2009, Proceedings*, volume 5834/2009 of *Lecture Notes in Computer Science*, pages 263–277. Springer Berlin/Heidelberg, 2009.

[15] H. Rott. Stability, strength and sensitivity: Converting belief into knowledge. *Erkenntnis*, 61:469–493, 2004.

[16] J. Sack. Temporal languages for epistemic programs. *Journal of Logic, Language, and Information*, 17(2):183–216, 2008.

[17] R. Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128:169–199, 2006.

[18] J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.

[19] J. van Benthem and E. Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99(1):61 – 92, 2011.

[20] J. van Benthem and F. Velazquez Quesada. The dynamics of awareness. *Synthese*, 177:5–27, 2010.

[21] F. R. Velazquez Quesada. *Small steps in dynamics of information*. ILLC dissertation series DS-2011-02, University of Amsterdam ILLC, 2011.

[22] A. Yap. Dynamic epistemic logic and temporal modality. In P. Girard, O. Roy, and M. Marion, editors, *Dynamic Formal Epistemology*, volume 351 of *Synthese Library*, chapter 3, pages 33–50. Springer, 2011.