Perfect Recall of Imperfect Knowledge

Andreas Witzel
Courant Institute
of Mathematical Sciences
New York University
New York, USA
awitzel@nyu.edu

ABSTRACT

Perfect recall, intuitively the ability to remember all past mental states, has been predominantly studied in the context of interpreted systems and game theory, which mostly consider S5 systems (of "correct" knowledge). More recently, the notion has become of interest to the epistemic logic community, where weaker systems are not unusual. Building upon recent work where we studied different definitions of perfect recall in Epistemic Temporal Logic (ETL), we argue that the intuitive motivations given there are still valid in such sub-S5 settings. However, definitions that were equivalent in S5 cease to be so without S5, so that these less restrictive settings allow for a more fine-grained comparison of the different definitions and their underlying intuitions.

Categories and Subject Descriptors

F.4.1 [**Mathematical Logic**]: Modal Logic; F.4.1 [**Mathematical Logic**]: Temporal Logic; I.2.3 [**Artificial Intelligence**]: Deduction and Theorem Proving

General Terms

Theory

Keywords

perfect recall, Epistemic Temporal Logic, KD45

1. INTRODUCTION

Perfect recall is an epistemic-temporal notion concerning an agent's ability to remember the past. It does not entail that all knowledge an agent has at some point is preserved forever—in fact, certain (negative) knowledge must be lost as the agent learns. For example "I know that I don't know p" is lost when I learn p. Rather, it means that an agent remembers all the information he once *had*, and can use it to reason about the present.

The notion of perfect recall is well-studied in game theory (see, e.g., [11] and [9, Section 11.1.3]) and in the distributed computing/interpreted systems/temporal logic literature [3, 15, 16, 7, 10] (see [4] for a discussion from the viewpoint of the intersection of both). In recent years, it has also been discussed in the epistemic logic community [12, 13, 14, 8], specifically in the context of Epistemic Temporal Logic (ETL) (see, e.g., [13]). ETL is an epistemic logic (or rather, family of logics) with added event modalities, interpreted on tree models, and it is intended to model agents over time.

In [17], we reviewed and compared several definitions of perfect recall, including one new one, and discussed how they may be seen to capture the intuitive meaning of perfect recall. Here, we argue that these intuitive motivations do not depend on S5 and make sense with weaker forms of knowledge (or beliefs)[1] as well. While the definitions that we discuss here were equivalent in S5, without S5 differences emerge and it becomes clear that they indeed capture different parts of the intuition. We therefore go on to examine exactly how they relate.

The current paper builds upon our previous study of perfect recall in S5 ETL [17], from where we repeat the basic framework (Section 2), definitions (Section 3) and some results for the sake of self-containment. Then, in the main Section 4, we explore sub-S5 settings. Section 5 offers some concluding discussion.

2. EPISTEMIC-TEMPORAL LOGIC (ETL)

We focus on the single-agent case since perfect recall is a property inherent to one agent; all our considerations carry over to the multi-agent case. We consider models over some finite set E of **events**. A **history** $h \in E^*$ is a finite sequence of events, and we denote the empty history (or **root**) by ϵ. We denote sequences simply by listing their elements, possibly preceded by a prefix sequence. For two histories h, h' we write $h \rightsquigarrow_e h'$ if $h' = he$, that is, if h' **extends** h by one event e. We write $h \rightsquigarrow h'$ if $h \rightsquigarrow_e h'$ for some event e. We denote the transitive and reflexive closure of $\rightsquigarrow$ by $\rightsquigarrow^*$, so $h \rightsquigarrow^* h'$ says that h is a **prefix** of h' (possibly h' itself), or vice versa, h' is an **extension** of h. For $h \rightsquigarrow^* h'$, we sometimes also write $h \preceq h'$, and $h \prec h'$ for $h \preceq h'$ with $h \neq h'$. A **protocol** $H \subseteq E^*$ is a set of histories closed under taking prefixes, intuitively representing the allowed evolutions of the system.

An **(epistemic) accessibility relation** is a binary relation $\sim \subseteq H \times H$ on histories.[2] It specifies, for any given

TARK 2011, July 12-14, 2011, Groningen, The Netherlands.
 ISBN 978-1-4503-0707-9, $10.00.

[1] We sometimes use the term *knowledge* somewhat sloppily to also include incorrect "knowledge", which is usually called "belief".

[2] We stick with the customary equivalence-like symbol $\sim$,

history h, the histories that the agent **considers possible** at h, which may or may not include h itself. Various conditions can be imposed on $\sim$, making it capture various notions of knowledge or belief (for details see, e.g., [2]). Most commonly, the relation is assumed to be an *equivalence relation*, making it capture a notion of "correct knowledge". This case is also referred to as **S5**. We focus here on less restrictive cases and refer to them as **sub-S5**. Relevant properties include *reflexivity*, reflecting **truth** (or correctness), *seriality*, reflecting **consistency**, *transitivity*, reflecting **positive introspection**, and *Euclideanness*, reflecting **negative introspection**. A common sub-S5 class is **KD45**, the class of frames with serial, transitive, and Euclidean accessibility relations.

An **ETL tree frame** is a tuple $\mathcal{F} = \langle E, H, \sim \rangle$ consisting of a set of events E, a protocol H and an epistemic accessibility relation $\sim$. We will often omit E and H and implicitly assume that any events or histories we talk about belong to E or H, respectively. Such frames can be viewed as temporal trees (induced by the $\rightsquigarrow$ relation) with epistemic accessibilities between nodes. We also consider **forests**, which are (finite) sets of trees with distinct roots (i.e., distinct empty histories $\epsilon_1, \ldots, \epsilon_k$)[3] and possibly interrelating epistemic accessibilities.[4] Although we usually have trees in mind, all our considerations apply to both trees and forests, and we make the distinction explicit where necessary. While we usually omit the roots when writing down histories, we include them where necessary to avoid confusion. We use properties of the epistemic accessibility relation to specify frames with a corresponding relation; for example, by an *S5 frame* we mean a frame with an S5 accessibility relation.

The **language** of ETL that we consider consists of a finite set At of propositional atoms and of all formulas built from those according to the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi \mid \langle e \rangle \varphi,$$

where $p \in \mathsf{At}$ and φ, ψ are formulas. Intuitively, $K\varphi$ means that the agent knows φ, and the event modality $\langle e \rangle \varphi$ means that event e can occur and afterwards φ will hold. The remaining propositional connectives are defined as abbreviations as usual, and the duals of the modalities are denoted by L (dual of K) and $[e]$ (dual of $\langle e \rangle$). We write $\Diamond\varphi$ to abbreviate $\bigvee_{e \in E} \langle e \rangle \varphi$.

A **valuation** $V : \mathsf{At} \to 2^H$ assigns to each atom the set of histories where it is true. We write $\mathcal{F}, V, h \models \varphi$ for a frame $\mathcal{F}$, a valuation V and a history h of the protocol of $\mathcal{F}$ to say that φ is **satisfied** by $\mathcal{F}, V, h$. Satisfaction of formulas is defined inductively as usual, starting with the truth values of atoms as given by V, and with the following semantics:

$$\begin{array}{ll}
\mathcal{F}, V, h \models p & \text{iff } h \in V(p) \\
\mathcal{F}, V, h \models \neg\varphi & \text{iff } \mathcal{F}, V, h \not\models \varphi \\
\mathcal{F}, V, h \models \varphi \wedge \psi & \text{iff } \mathcal{F}, V, h \models \varphi \text{ and } \mathcal{F}, V, h \models \psi \\
\mathcal{F}, V, h \models K\varphi & \text{iff for each } h' \in H \text{ with } h \sim h': \\
 & \mathcal{F}, V, h' \models \varphi \\
\mathcal{F}, V, h \models \langle e \rangle \varphi & \text{iff } he \in H \text{ and } \mathcal{F}, V, he \models \varphi
\end{array}$$

For a binary relation R on histories we write

$$[h]_R = \{h' \mid hRh'\}$$

to denote the **image** of h under R. If R is an equivalence relation, then $[h]_R$ is the equivalence class of h with respect to R. If H is a set of histories, we write

$$[H]_R = \bigcup_{h \in H} [h]_R$$

for the union of the images of all $h \in H$.

3. PERFECT RECALL

In this section, we summarize the discussion of perfect recall notions, repeating the necessary parts from [17].

In game theory, turn-taking typically makes successive situations distinguishable and perfect recall can often be formulated as "remembers all his actions". ETL has no notion of turns and no notion of agency associated with events. An event is just an event and comes with no specification as to who performs it,[5] and so we have to use other ways to express the notion.

3.1 Basic definitions

We start by giving some intuition about the upcoming definitions. As mentioned, the basic idea is that an agent with perfect recall can at any point remember all the information that he had at any previous point in time; he is then able to exclude any possibilities for the current state of the world which are inconsistent with that information. Precisely of what nature is the information that the agent remembers?

Two different intuitions for perfect recall.

There are two related kinds of information and ways in which an agent might detect inconsistencies, the first on the level of epistemic states, and the second on the level of the semantic structures that model them.[6] Consider a perfect-recall agent in some state of the world, in ETL terms a history, h, and some other history h'.

Firstly, if in state h' the agent would have gone through a different sequence of epistemic states than he actually has in h, then he can exclude the possibility of h', since he can recall all his epistemic states.

Secondly, if in the state before h the agent was certain that the world was *not* in a state along history h', then at h the agent can exclude the possibility that the world is in state h', since he can recall his previous assessments. Put differently,

even though our focus here is on accessibility relations that are not equivalence relations.

[3]An alternative formulation is to require H to be closed only under *non-empty* prefixes; if then the (unique) empty history *is* contained in H, we have a tree, and otherwise we have a forest of histories whose roots are their respective first events. We will stick with the slight conceptual abuse of allowing several distinct empty histories and hope that no confusion will arise from it.

[4]The distinction between trees and forests corresponds to the *unique initial state* condition in the interpreted systems literature.

[5]If we do want to attribute certain events to certain agents, then corresponding observability conditions for the agent performing a particular action can be specified separately, and our definitions of perfect recall will not interfere.

[6]See Section 5 for some discussion related to the question of which aspects of the model an agent can access.

if h' is *not* an extension of some history considered possible before, then the agent can exclude the possibility of h' since there would have been no way for the world to evolve to h'.

We give formalizations of these intuitions in the following. While the definitions we give are equivalent in the context of S5 [17], as we will see, without S5 they no longer coincide. However, the intuitions that they formalize are just as valid without S5 as with; they only talk about what the agent considers possible, without relying on, e.g., correctness of the agent's knowledge.

Formalizing the intuitions.

The first notion uses the idea of local-state sequences, or "epistemic experiences", meaning sequences of epistemic states that the agent has gone through. Repetitions of identical states are ignored, since the agent has no way of discriminating between two states in which he has the same epistemic state. Similar as in game theory (cf. [9, Section 11.1.3]), we identify an agent's epistemic state with the set of worlds that he considers possible (the difference being that without S5, this does not yield "information sets" in the game theoretic sense).

Definition 3.1. Given an ETL frame and a history $e_1 \dots e_\ell$ (with root ϵ), the agent's **epistemic experience** is the sequence

$$\mathrm{EE}(e_1 \dots e_\ell) := [\epsilon]_\sim \ [e_1]_\sim \ [e_1 e_2]_\sim \ \dots \ [e_1 \dots e_\ell]_\sim$$

of epistemic states he has gone through.

We say that the epistemic experiences in two histories h, h' are **equivalent modulo stutterings**, in symbols $\mathrm{EE}(h) \approx \mathrm{EE}(h')$, iff the sequences with all repetitions of subsequent identical sets removed are equivalent.

An ETL frame has **perfect recall with respect to epistemic experience** ($\mathsf{PR_{ee}}$) iff, whenever $h \sim h'$, we have $\mathrm{EE}(h) \approx \mathrm{EE}(h')$.

The second definition is a slight (but equivalent) variant of a technical notion most commonly used in the interpreted systems literature.[7] By rephrasing, in [17] we provided it with an independent motivation.

Definition 3.2. An ETL frame has **perfect recall with respect to history consistency** ($\mathsf{PR_{hc}}$) iff for any histories h, h' and event e with $he \sim h'$, there is some history h'' with $h \sim h'' \rightsquigarrow^* h'$ (i.e., some prefix of h' is epistemically accessible from h).

Rephrasing this condition, we get that for each history h and event e, we have

$$[he]_\sim \subseteq [[h]_\sim]_{\rightsquigarrow^*} \ .$$

This formulation suggests the intuitive reading discussed above: A frame has $\mathsf{PR_{hc}}$ if all histories considered possible after some event are extensions of histories considered possible before the event. That is, if the agent was certain that a certain history was *im*possible before some event, then after that event he will remember that assessment and still consider (any extension of) that history impossible.

In [17] we went on to propose a refinement of $\mathsf{PR_{hc}}$, which also gave rise to a simple axiomatization of perfect recall. It is "local" in the sense that it avoids jumping to arbitrary prefixes of histories.

[7]See [6, p. 204] (who call perfect recall "no forgetting") and [15, Proposition 2.1(a)].

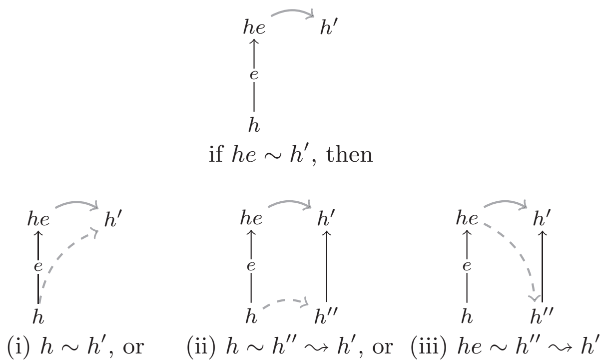


Figure 1: Illustrating $\mathsf{PR^\ell_{hc}}$. Gray, bent arrows indicate accessibilities.

Definition 3.3. An ETL frame has $\mathsf{PR_{hc}}$**, local version** ($\mathsf{PR^\ell_{hc}}$) iff for each history h and event e, we have

$$[he]_\sim \subseteq [h]_\sim \cup [[h]_\sim]_\rightsquigarrow \cup [[he]_\sim]_\rightsquigarrow$$

Put differently, for each h, h', e with $he \sim h'$, either of the following holds:

(i) $h \sim h'$

(ii) $h \sim h'' \rightsquigarrow h'$ for some h''

(iii) $he \sim h'' \rightsquigarrow h'$ for some h''.

(Note that the word "some" is somewhat misleading, since h'' is the uniquely determined direct predecessor of h'.)

See Figure 1 for an illustration of this definition.

The intuitive interpretation of $\mathsf{PR^\ell_{hc}}$ gives a more fine-grained account of the agent's possibilities for reasoning than $\mathsf{PR_{hc}}$ does. As we will see in Section 4, the fine-grainedness of the version we proposed indeed makes a difference.

Let us walk through it and compare it with $\mathsf{PR_{hc}}$. We consider the history he and refer to h as "before e" and to he as "after e". The definition says that any history considered possible after e either

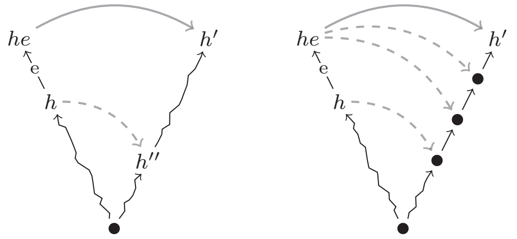
(i) was considered possible already before e, i.e., the agent didn't notice e, nor time passing; or

(ii) is an extension by one event of a history considered possible before e, i.e., the agent correctly thinks one event occurred, though he may not be certain which one; or

(iii) is an extension by one event of another history considered possible after e.

As illustrated in Figure 2, this last condition inductively bridges the gap to $\mathsf{PR_{hc}}$, and allows the agent to consider possible that several events occur while really just e is happening. The difference, as compared to $\mathsf{PR_{hc}}$, is that now there is a stricter consistency requirement. If the agent considers possible that several events have happened, he must obviously be unable to detect some of them, since really just one event happened. Given that, he must also consider possible the intermediate histories along these several events—either from the history after e or from before e. Exactly this consistency requirement is inductively captured by the last condition (in interplay with the second condition).

Intuitively it is thus clear that $\mathsf{PR^\ell_{hc}}$ is at least as strong a condition as $\mathsf{PR_{hc}}$.



(a) $\mathsf{PR}_{\mathsf{hc}}$ (b) $\mathsf{PR}^{\ell}_{\mathsf{hc}}$

Figure 2: $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ inductively bridges the gap to $\mathsf{PR}_{\mathsf{hc}}$, but it imposes additional conditions on the intermediate states. (As shown in [17], in S5 the two conditions are equivalent.)

Lemma 3.4 ([17]). *Any ETL frame that has $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ also has $\mathsf{PR}_{\mathsf{hc}}$.*

Proof. A simple induction on the length of h' in the definition of $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ proves the claim. □

Furthermore, $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ is axiomatizable in ETL.

Theorem 3.5 ([17]). *An ETL frame has $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ iff it validates the following formula for each event e (recall that $\Diamond\varphi$ abbreviates $\bigvee_{e'}\langle e'\rangle\varphi$):*

$$\langle e\rangle Lp \rightarrow Lp \vee L\Diamond p \vee \langle e\rangle L\Diamond p \ .$$

Together with the normal modal logic axioms and deduction rules, it is sound and complete with respect to the class of ETL frames with $\mathsf{PR}^{\ell}_{\mathsf{hc}}$.

This ends the review of the basic notions and necessary prerequisites from [17], and we now turn to examining the various definitions in sub-S5 settings. While we saw in [17] that all of the presented notions are equivalent in S5 (see also [15]), we will now see that the picture is less clear-cut without S5.

4. SUB-S5 SETTINGS

To our knowledge, perfect recall has only been considered in the context of S5 in the literature, a likely reason being that both communities that have studied the notion most (interpreted systems and game theory) virtually exclusively consider S5 settings. However, it can make perfect sense, for example, to say of a misinformed agent that he correctly remembers all information he has ever had, even if that information itself is not correct. We here explore this idea.

It is important to see that the motivation and justifications for the definitions of perfect recall we gave did not assume S5 knowledge. Each of the notions captured a particular way of not losing information, and they make sense even without S5. We therefore simply use the same basic definitions as in [17], but take a new look at how they relate.

Note that although $[h]_\sim$ does not necessarily contain h itself, we do have the following fact.

Fact 4.1. *If $\sim$ is a transitive and Euclidean relation, then it is an equivalence relation on $[h]_\sim$ for any h (cf. [5, Theorem 3.3]). Consequently, for any h, h' with $[h]_\sim \cap [h']_\sim \neq \emptyset$, we have $[h]_\sim = [h']_\sim$.*

4.1 Contrasting the notions

First note that $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ implies $\mathsf{PR}_{\mathsf{hc}}$, since Lemma 3.4 did not assume S5. However, without any assumptions about the frames, none of the other mutual implications among $\mathsf{PR}_{\mathsf{hc}}$, $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ and $\mathsf{PR}_{\mathsf{ee}}$ remain. This is witnessed by Figure 3, illustrating that epistemic experience and history consistency reflect two different ways of remembering past information.

The following list gives intuitive interpretations of the various situations depicted.

(a) An agent that only has perfect recall with respect to epistemic experience may at some point be certain that a particular history can be excluded, but later on "forget" this piece of information. In particular, at e_1e_3 the agent considers e_2e_3 possible, even though he never considered e_2 possible.

(b), (c) An agent that only has perfect recall with respect to history consistency may at some state consider another state possible, although in that other state his epistemic experience would have been different. For example, an agent at e_1 in (c) is (mistakenly) certain that he is at e_2e_3, even though in that state his previous epistemic state would have been $\{e_2\}$, which contradicts his actual epistemic experience.

(d) The intuition is similar to the previous case, but here we see that $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ is more fine-grained than $\mathsf{PR}_{\mathsf{hc}}$. An agent at e_1 thinks he is at e_2e_3, even though he never considered e_2 possible. He thus thinks himself at the endpoint of a history whose unfolding he deemed impossible, and in that sense, he loses information. $\mathsf{PR}_{\mathsf{hc}}$ grants this agent the label of perfect recall, while $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ denies it.

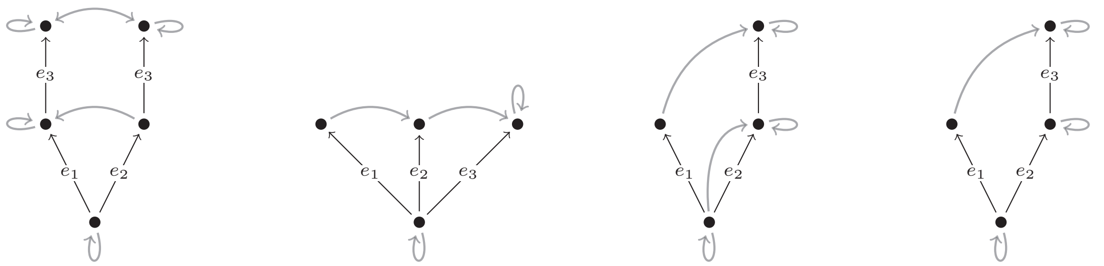
Note that, while these phenomena reflect some kind "*forgetting*", they do not at first glance constitute a coherent, rational method of *belief revision*. A full-fledged doxastic logic is needed in order to really model agents that reconsider their previous assessments and deal with "unwanted" memories properly.

The following straightforward result enables us to identify the settings in which the different notions of perfect recall can be meaningfully compared.

Proposition 4.2. *Any ETL frame that has $\mathsf{PR}_{\mathsf{ee}}$ is transitive and Euclidean.*

Proof. This is obvious from Definition 3.1. For example, for any three histories h, h', h'', if $h \sim h'$ and $h \sim h''$, then $\mathsf{PR}_{\mathsf{ee}}$ implies that $[h]_\sim = [h']_\sim = [h'']_\sim$. Since $h', h'' \in [h]_\sim$, we also get $h' \in [h'']_\sim$ and $h'' \in [h']_\sim$. □

This result is not very surprising, given that $\mathsf{PR}_{\mathsf{ee}}$ requires of an agent to be able to assess his own epistemic experience—including at the current state. It implies that any reflexive ETL frame with $\mathsf{PR}_{\mathsf{ee}}$ is already an S5 ETL frame. On the level of agents, an agent who has correct beliefs and $\mathsf{PR}_{\mathsf{ee}}$ has in fact already correct (S5) knowledge. However, perfect recall does not *require* the agent to have correct beliefs (in fact, perfect recall by itself is compatible with believing falsum). For example, KD45 is a common sub-S5 setting in which perfect recall is a meaningful notion.



(a) $\mathsf{PR}_{\mathsf{ee}}$, but not $\mathsf{PR}_{\mathsf{hc}}$ nor $\mathsf{PR}^{\ell}_{\mathsf{hc}}$

(b) $\mathsf{PR}_{\mathsf{hc}}$ and $\mathsf{PR}^{\ell}_{\mathsf{hc}}$, but not $\mathsf{PR}_{\mathsf{ee}}$

(c) $\mathsf{PR}_{\mathsf{hc}}$ and $\mathsf{PR}^{\ell}_{\mathsf{hc}}$, but not $\mathsf{PR}_{\mathsf{ee}}$

(d) $\mathsf{PR}_{\mathsf{hc}}$, but not $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ nor $\mathsf{PR}_{\mathsf{ee}}$

Figure 3: $\mathsf{PR}_{\mathsf{hc}}$, $\mathsf{PR}^{\ell}_{\mathsf{hc}}$, and $\mathsf{PR}_{\mathsf{ee}}$ compared on ETL frames, gray arrows depicting the accessibilities. $\mathsf{PR}_{\mathsf{hc}}$ is violated in (a) since $e_1e_3 \sim e_2e_3$ but there is no prefix of e_2e_3 that is accessible from e_1. $\mathsf{PR}_{\mathsf{ee}}$ is violated in (b) since $\mathrm{EE}(e_1) \not\approx \mathrm{EE}(e_2)$, and in (c) and (d) since $\mathrm{EE}(e_1) \not\approx \mathrm{EE}(e_2e_3)$. $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ is violated in (d) since $e_1 \sim e_2e_3$ but $\epsilon \not\sim e_2e_3$ and $\epsilon \not\sim e_2$ and $e_1 \not\sim e_2$.

Given that transitivity and Euclideanness are inherent to $\mathsf{PR}_{\mathsf{ee}}$, we continue our comparison within the corresponding class of frames. In the following, we use **introspective** to mean "transitive and Euclidean".

4.2 Characterizing $\mathsf{PR}_{\mathsf{ee}}$ locally

Given the fact that $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ no longer characterizes $\mathsf{PR}_{\mathsf{ee}}$ and the original definition of $\mathsf{PR}_{\mathsf{ee}}$ is somewhat unwieldy, it can be useful to have a local condition on histories and accessibilities that corresponds to it. It turns out that we can re-use $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ by slightly modifying the frame in question.

For a frame $\mathcal{F}$ with accessibility relation $\sim$, we will use $\dot{\mathcal{F}}$ and $\dot{\sim}$ to denote the S5 closure. We need the following small technical condition: We say that a frame satisfies **persistent insanity** if, whenever $[h]_\sim = \emptyset$ and $h \preceq h'$, then $[h']_\sim = \emptyset$. Intuitively, once a corresponding agent has inconsistent beliefs, he will remain in that pitiful condition forever.

Proposition 4.3. *An introspective ETL frame satisfying persistent insanity has* $\mathsf{PR}_{\mathsf{ee}}$ *iff its S5 closure has* $\mathsf{PR}^{\ell}_{\mathsf{hc}}$.[8]

Proof. As shown in [17], $\mathsf{PR}_{\mathsf{ee}}$ and $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ are equivalent in S5, so $\mathsf{PR}_{\mathsf{ee}}$ and $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ are equivalent on the S5 closure of the frame. We can thus prove the claim by showing that $\mathsf{PR}_{\mathsf{ee}}$ is invariant under taking this closure.

To see that this is indeed the case, take any pair h, h' of histories, the accessibility relation $\sim$ of an introspective frame satisfying persistent insanity, and its S5 closure $\dot{\sim}$. With Fact 4.1, it is easy to see that, as long as $[h]_\sim \neq \emptyset$, we have $[h]_\sim = [h']_\sim$ iff $[h]_{\dot{\sim}} = [h']_{\dot{\sim}}$. Inductively it follows that the equivalence of epistemic experiences is invariant under taking the S5 closure as long as $[h]_\sim \neq \emptyset$; persistent insanity ensures that $\mathsf{PR}_{\mathsf{ee}}$ is also satisfied for any h' extending an h with $[h]_\sim = \emptyset$. ☐

To see that persistent insanity is indeed needed for this result, consider Figure 3(a) with the accessibilities $e_2 \sim e_1$ and $e_1 \sim e_1$ removed. The resulting frame still has $\mathsf{PR}_{\mathsf{ee}}$, but its S5 closure does not have $\mathsf{PR}^{\ell}_{\mathsf{hc}}$.

[8] Note that this is indeed a local condition: On introspective frames the S5 closure is the symmetric and reflexive closure, without any need of iterating through the accessibility relation (cf. Fact 4.1).

Corollary 4.4. *A KD45 ETL frame has* $\mathsf{PR}_{\mathsf{ee}}$ *iff its S5 closure has* $\mathsf{PR}^{\ell}_{\mathsf{hc}}$.

Proof. Immediate, since KD45 frames are introspective and vacuously satisfy persistent insanity. ☐

Remark 4.5. Note that $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ implies persistent insanity, so Proposition 4.3 applies to all introspective frames with $\mathsf{PR}^{\ell}_{\mathsf{hc}}$.

As witnessed by Figure 3 and by Figure 4 later on, the notions $\mathsf{PR}_{\mathsf{ee}}$ and $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ are incomparable in the sense that each one is stronger than the other one under certain circumstances. Proposition 4.3 gives an insight as to why this is so: By applying the $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ condition to the S5 closure of a frame, on the one hand the antecedent in this condition becomes more permissive, but on the other hand so does the consequent. Thus, the condition gets both strengthened and weakened.

Now that we have contrasted our basic notions of perfect recall and provided and discussed separate local characterizations, we proceed to characterize the combination of the notions. We use PR to denote the combination of perfect recall notions, $\mathsf{PR}_{\mathsf{ee}}$ plus $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ (and thus $\mathsf{PR}_{\mathsf{hc}}$), describing perfect-recall agents that can reason both about their epistemic experience and about history consistency.

Our considerations so far hold both for ETL trees and ETL forests. Now, however, the distinction becomes important. We start by focusing on trees.

4.3 Characterizing PR on trees

It turns out that on introspective trees, $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ captures both aspects of perfect recall, much like it (and $\mathsf{PR}_{\mathsf{hc}}$) did on S5 frames. This allows us to define and axiomatize PR on introspective trees, reusing the results we obtained in [17].

Theorem 4.6. *An introspective ETL tree has* PR *iff it has* $\mathsf{PR}^{\ell}_{\mathsf{hc}}$.

Note that this result is not in contradiction with the examples in Figure 3, since (b) is not transitive and (c) is not Euclidean. For the proof, we need the following auxiliary results.

Observation 4.7. *For any introspective ETL frame with* $\mathsf{PR}^{\ell}_{\mathsf{hc}}$ *and histories* h_1, h_2 *with* $h_1 \preceq h_2$ *and* $h_1 \sim h_2$, *for each* $h_1' \preceq h_1$ *there is* $h_2' \preceq h_2$ *such that* $h_1' \sim h_2'$.

Proof. The claim can be shown with a simple induction on h_1, using Lemma 3.4 and transitivity of $\preceq$. ☐

Lemma 4.8. *For any introspective ETL frame with* $\mathsf{PR}^\ell_{\mathsf{hc}}$ *and histories* h_1 *and* $h_2' \preceq h \preceq h_2$*, if* $h_1 \sim h_2$ *and* $h_1 \sim h_2'$ *then* $h_1 \dot{\sim} h$.

Proof. With Euclideanness, we obtain $h_2 \sim h_2'$. Let h_2'' be the *shortest* prefix of h_2 such that $h_2 \sim h_2''$ (note that $h_2'' \preceq h_2' \preceq h$). Observation 4.7 implies that there must be $h' \preceq h_2''$ with $h \sim h'$. Another application of Observation 4.7 then yields that there is $h'' \preceq h'$ such that $h_2'' \sim h''$, and by transitivity we get $h_2 \sim h''$. Now if $h' \prec h_2''$ then $h'' \prec h_2''$, contradicting that h_2'' is the shortest prefix accessible from h_2. So $h' = h_2''$, thus $h \sim h_2''$. Since $h_1 \sim h_2 \sim h_2''$, we obtain the claim. ☐

Lemma 4.9. *If an introspective ETL tree has* $\mathsf{PR}^\ell_{\mathsf{hc}}$*, then so does its S5 closure.*

Proof. Take any introspective tree $\mathcal{F}$ with $\mathsf{PR}^\ell_{\mathsf{hc}}$. To see that its S5 closure $\dot{\mathcal{F}}$ also has $\mathsf{PR}^\ell_{\mathsf{hc}}$, let h, e, h' be such that $he \dot{\sim} h'$. We need to show that $\dot{\sim}$ satisfies one of the three conditions in the definition of $\mathsf{PR}^\ell_{\mathsf{hc}}$.

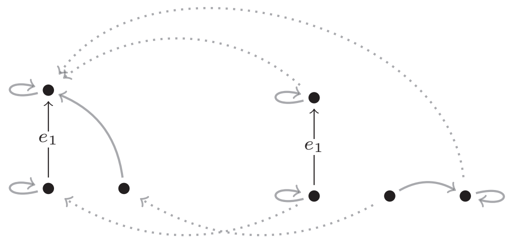
Since $\dot{\sim}$ is the symmetric and reflexive closure of $\sim$ (cf. footnote 8), we have either of these cases:

- $he = h'$. Since $h \dot{\sim} h$, condition (ii) in the definition of $\mathsf{PR}^\ell_{\mathsf{hc}}$ obtains.
- $he \sim h'$. Since $\mathcal{F}$ has $\mathsf{PR}^\ell_{\mathsf{hc}}$, $\sim$ satisfies one of the three conditions of $\mathsf{PR}^\ell_{\mathsf{hc}}$, thus so does $\dot{\sim}$.
- $h' \sim he$. If $h' \neq \epsilon$ then the same argument as in the previous case applies. Otherwise, $h' = \epsilon \sim he$. Euclideanness of $\sim$ yields $he \sim he$, and since $\sim$ satisfies $\mathsf{PR}^\ell_{\mathsf{hc}}$, we have either of these three cases:
  - (i) $h \sim he$. Since $h' \sim he$, we get $h \dot{\sim} h'$, so $\dot{\sim}$ satisfies condition (i) of $\mathsf{PR}^\ell_{\mathsf{hc}}$.
  - (ii) $h \sim h$. Since $h' = \epsilon \preceq h$, Observation 4.7 yields that there is $h'' \preceq h$ such that $h' \sim h''$. Now we have $h' \sim he$, $h' \sim h''$ and $h'' \preceq h \preceq he$, so Lemma 4.8 applies and yields $h' \dot{\sim} h$. Symmetry of $\dot{\sim}$ yields $h \dot{\sim} h'$, so $\dot{\sim}$ satisfies condition (i) of $\mathsf{PR}^\ell_{\mathsf{hc}}$.
  - (iii) $he \sim h$. Together with $h' \sim he$ we get $h' \dot{\sim} h$. Symmetry of $\dot{\sim}$ again yields $h \dot{\sim} h'$, condition (i) of $\mathsf{PR}^\ell_{\mathsf{hc}}$. ☐

We can now straightforwardly prove the stated result.

Proof of Theorem 4.6. PR implies $\mathsf{PR}^\ell_{\mathsf{hc}}$ by definition. To see that the reverse direction holds, take any introspective ETL tree $\mathcal{F}$ that has $\mathsf{PR}^\ell_{\mathsf{hc}}$. Due to Lemma 4.9, its S5 closure also has $\mathsf{PR}^\ell_{\mathsf{hc}}$, and with Proposition 4.3 and Remark 4.5 it follows that $\mathcal{F}$ also has $\mathsf{PR}_{\mathsf{ee}}$. ☐

Since the proof of Theorem 3.5 (as presented in [17]) did not use S5, we immediately obtain an axiomatization of perfect recall in the class of introspective ETL trees.



(a) Forest $\mathcal{F}$ (two trees) (b) Forest $\mathcal{F}'$ (three trees)

Figure 4: (a) $\mathcal{F}$ is an introspective forest that has $\mathsf{PR}^\ell_{\mathsf{hc}}$, but not $\mathsf{PR}_{\mathsf{ee}}$. (b) $\mathcal{F}'$ has $\mathsf{PR}^\ell_{\mathsf{hc}}$ and $\mathsf{PR}_{\mathsf{ee}}$, and $\mathcal{F}$ is its bounded morphic image via the bounded morphism depicted with dotted arrows. So PR is not modally definable on forests.

4.4 Characterizing PR on forests

Theorem 4.6 does not apply to forests, as witnessed by Figure 4(a). Before we look at how to characterize PR here, we note that, unlike on S5 forests, defining PR on introspective forests generally is impossible in ETL (the same holds for $\mathsf{PR}_{\mathsf{ee}}$).

Proposition 4.10. PR *is not modally definable on introspective ETL forests (and thus not on general ETL forests either).*

Proof. $\mathcal{F}'$ in Figure 4 has PR, while its bounded morphic image $\mathcal{F}$ does not. Since modally definable properties are closed under bounded morphic images (cf. [1]), the claim follows. ☐

From Proposition 4.3 and Remark 4.5 it is clear that any introspective frame has PR iff both it and its S5 closure have $\mathsf{PR}^\ell_{\mathsf{hc}}$. However, with an additional slight restriction, we can continue to use $\mathsf{PR}^\ell_{\mathsf{hc}}$ to characterize PR. From Figure 4, it is intuitively clear that accessibilities from some root to a later state in some (different) tree are problematic: In such cases, $\mathsf{PR}^\ell_{\mathsf{hc}}$ is vacuously satisfied, while $\mathsf{PR}_{\mathsf{ee}}$ may not hold.

To fix this, call an ETL forest $\mathcal{F}$ **initially synchronous** if, for any two roots ϵ, ϵ' and history h with $\epsilon \preceq h$ and $\epsilon' \sim h$, we also have $\epsilon' \sim \epsilon$. That is, the agent at least considers it possible that indeed no time has passed initially, although he may immediately lose synchronicity and also consider later states possible. We then get the following.

Lemma 4.11. *If an introspective and initially synchronous ETL forest has* $\mathsf{PR}^\ell_{\mathsf{hc}}$*, then so does its S5 closure.*

Proof. The proof is analogous to that of Lemma 4.9, with one additional observation: If $\epsilon' \sim h$ for some history h with root ϵ, then initial synchronicity yields $\epsilon' \sim \epsilon$. Euclideanness then yields $\epsilon \sim h$, and with Theorem 4.6 it follows that $\mathrm{EE}(\epsilon) \approx \mathrm{EE}(h)$. Due to Fact 4.1, we also have $\mathrm{EE}(\epsilon') \approx \mathrm{EE}(\epsilon)$, so $\mathrm{EE}(\epsilon') \approx \mathrm{EE}(h)$ by transitivity of $\approx$. ☐

Theorem 4.12. *An introspective and initially synchronous ETL forest has* PR *iff it has* $\mathsf{PR}^\ell_{\mathsf{hc}}$.

Proof. Analogously to Theorem 4.6, this follows from Lemma 4.11, Proposition 4.3 and Remark 4.5. ☐

So on introspective and initially synchronous ETL forests, $\mathsf{PR}^\ell_{\mathsf{hc}}$ again captures both aspects of perfect recall, the one based on history consistency as well as the one based on epistemic experience. This shows how closely related, even if subtly different, the two views are.

5. CONCLUSIONS

Following [17], we looked at two different ways of "not losing information", that is, accessing and reasoning with one's memories. The first one has been the fundamental definition in the literature on perfect recall. It assumes that a perfect-recall agent can use differences in past epistemic states in order to distinguish present states. The second one is a consistency condition on the histories considered possible. It has been endowed with its own independent motivation in [17], where also a refinement was proposed.

While the two notions have been well-studied in S5, where they coincide, we here argued that their underlying motivations do not depend on S5. We therefore examined them in sub-S5 settings, where they no longer coincide. Since the notions capture independently motivated ways of reasoning with memories, we examined and characterized them individually as well as jointly.

Given that the two notions use different aspects of ETL frames, some discussion is needed concerning the access that we assume an agent to have.

It is a general issue in modeling agents to what extent the model faithfully represents an agent's internal workings, and to what extent it represents the modeler's external perspective. What we mean if we say that an agent "does not lose information", of course, depends on what information we ascribe to him in the first place. ETL is agnostic as to whether the agent has direct access to the semantic structures constituting a model or whether they are just a representation for the modeler, and whether the logic language is supposed to reflect the agent's "mentalese" or whether it is just a way for the modeler to talk *about* the agent. Depending on the intended interpretation, one may exclude or include certain features in what is considered the agent's information, and one may accept or reject certain methods for the agent to access and reason with his memories.

Since ETL does not specify these issues, we simply examined what can be said *if* the agent has access to certain aspects of the model. Outside of S5, where the notions do not coincide, it depends on the modeled situation which definition of perfect recall is the right one.

An interesting question for further research is whether there are additional aspects of reasoning with memories, which might not play a role in S5, or which might be conflated in S5 with the ones we discussed, but become relevant in other settings.

Acknowledgments

This work came out of discussions with Benedikt Löwe and Cédric Dégremont. Thanks also to Krzysztof Apt, Can Başkent, and Johan van Benthem for comments.

The author was funded by an NSF Expeditions in Computing grant.

References

[1] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, 2001.

[2] Brian F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.

[3] Ronald Fagin, Joseph Y. Halpern, Moshe Y. Vardi, and Yoram Moses. *Reasoning about knowledge*. MIT Press, 1995.

[4] Joseph Y. Halpern. On ambiguities in the interpretation of game trees. *Games and Economic Behavior*, 20(1): 66–96, 1997.

[5] Joseph Y. Halpern. The relationship between knowledge, belief, and certainty. *Annals of Mathematics and Artificial Intelligence*, 4(3):301–322, 1991.

[6] Joseph Y. Halpern and Moshe Y. Vardi. The complexity of reasoning about knowledge and time. I. Lower bounds. *Journal of Computer and System Sciences*, 38(1):195–237, 1989.

[7] Joseph Y. Halpern, Ron van der Meyden, and Moshe Y. Vardi. Complete axiomatizations for reasoning about knowledge and time. *SIAM Journal on Computing*, 33 (3):674–703, 2004.

[8] Alistair Isaac and Tomohiro Hoshi. Synchronizing diachronic uncertainty. *Journal of Logic, Language and Information*, 2010.

[9] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, 1994.

[10] Rohit Parikh and Ramaswamy Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12(4):453–467, 2003.

[11] Michele Piccione and Ariel Rubinstein. The Absent-Minded driver's paradox: Synthesis and responses,. *Games and Economic Behavior*, 20(1):121–130, 1997.

[12] Johan van Benthem. Games in Dynamic-Epistemic logic. *Bulletin of Economic Research*, 53(4):219–248, 2001.

[13] Johan van Benthem and Eric Pacuit. The tree of knowledge in action: Towards a common perspective. In Guido Governatori, Ian Hodkinson, and Yde Venema, editors, *Advances in Modal Logic Volume 6 (Proceedings of the AiML 2006 Conference)*, pages 87–106, London, 2006. College Publications.

[14] Johan van Benthem, Jelle Gerbrandy, Tomohiro Hoshi, and Eric Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38(5):491–526, 2009.

[15] Ron van der Meyden. Axioms for knowledge and time in distributed systems with perfect recall. In *Proceedings of the Ninth IEEE Symposium on Logic in Computer Science*, pages 448–457, Los Alamitos, CA, 1993. IEEE Computer Society.

[16] Ron van der Meyden and Ka-shu Wong. Complete axiomatizations for reasoning about knowledge and branching time. *Studia Logica*, 75(1):93–123, 2003.

[17] Andreas Witzel. Characterizing perfect recall using next-step temporal operators in S5 Epistemic Temporal Logic. *Journal of Logic and Computation*, 2011. Advance Access published May 25, 2011, doi:10.1093/logcom/exr010.