

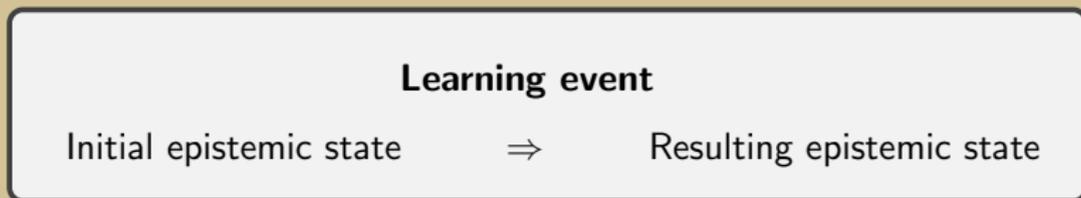
Dynamic introspection CSLI Workshop

Michael Cohen

Stanford University

Introduction

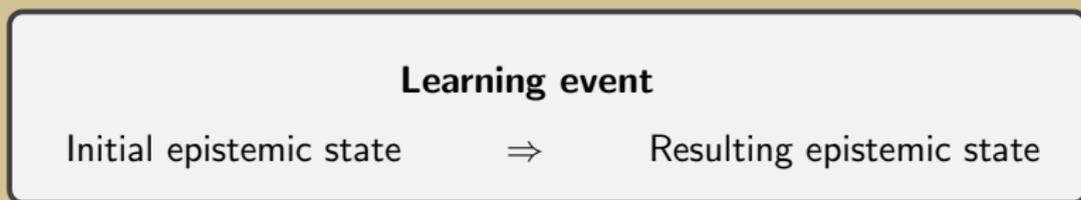
- Here is a very simple, “bare-bones” account of learning:



- Process of elimination of epistemic possibilities

Introduction

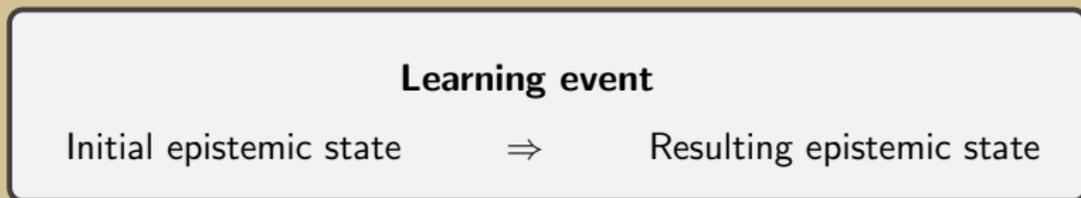
- Here is a very simple, “bare-bones” account of learning:



- Process of elimination of epistemic possibilities
- Appears in: Bayesian updating, many dynamic epistemic logics, Stalnaker’s account of assertion, dynamic semantics.
- Learning as a monotonic function between epistemic states (models).

Introduction

- Here is a very simple, “bare-bones” account of learning:



- Process of elimination of epistemic possibilities
- Appears in: Bayesian updating, many dynamic epistemic logics, Stalnaker’s account of assertion, dynamic semantics.
- Learning as a monotonic function between epistemic states (models).
- My project: study the epistemic *introspection assumptions* that are built into this account.

Toy example

Learning event

Initial epistemic state \Rightarrow Resulting epistemic state

- Example:
 - Initial state: I don't know the time 12:00
(possibilities: 12:00, 13:00, 14:00...)
 - Learning event: Looking at a clock
 - Resulting state: I know the time is 12:00
(possibilities: 12:00, ~~13:00~~, ~~14:00~~...)



Clock Tower

Introspection

- Motivating question: What does the agent know about the their own learning process? Why is that important?

Introspection

- Motivating question: What does the agent know about the their own learning process? Why is that important?
- I distinguish between *static* and *dynamic* introspection.

Introspection

- Motivating question: What does the agent know about the their own learning process? Why is that important?
- I distinguish between *static* and *dynamic* introspection.
- My complaint: Epistemologists mostly focus on static introspection.
- This project: study the logical and epistemological concept of *dynamic* introspection.

Introspection

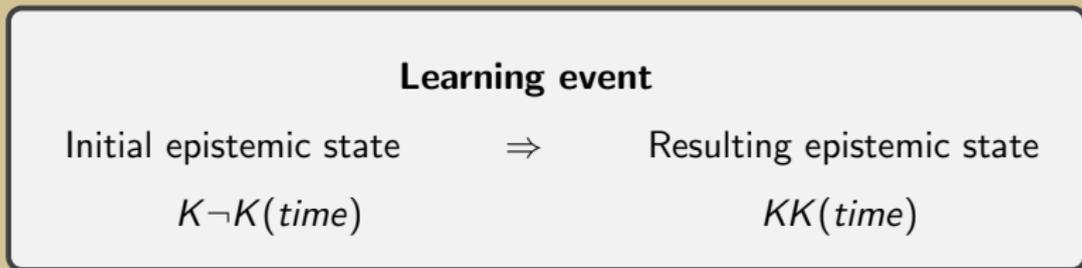
- Motivating question: What does the agent know about the their own learning process? Why is that important?
- I distinguish between *static* and *dynamic* introspection.
- My complaint: Epistemologists mostly focus on static introspection.
- This project: study the logical and epistemological concept of *dynamic* introspection.
- I argue that dynamic introspection is central in epistemology.
- Formally studying dynamic introspection offers a new bridge to traditional epistemological questions.

(Static) Introspection

- Static (or synchronic) introspection: Knowledge that the *agent* has about the epistemic state at a given moment.

(Static) Introspection

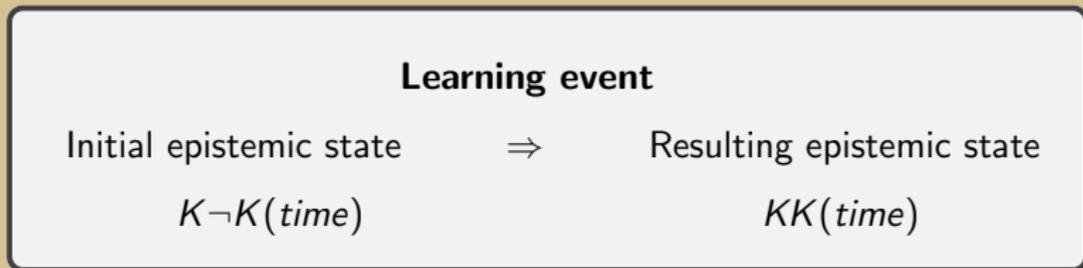
- Static (or synchronic) introspection: Knowledge that the *agent* has about the epistemic state at a given moment.
- Example:



Example: In the initial state, the agent knows that they don't know the time.

(Static) Introspection

- Static (or synchronic) introspection: Knowledge that the *agent* has about the epistemic state at a given moment.
- Example:



Example: In the initial state, the agent knows that they don't know the time.

- Important in the interface of epistemology and (static) epistemic logic

Dynamic Introspection (and ignorance)

- What about our knowledge (and ignorance) about the process?

Dynamic Introspection (and ignorance)

- What about our knowledge (and ignorance) about the process?

Question: How to think about it? Modify the simple bare-bones model

Dynamic Introspection (and ignorance)

- What about our knowledge (and ignorance) about the process?

Question: How to think about it? Modify the simple bare-bones model

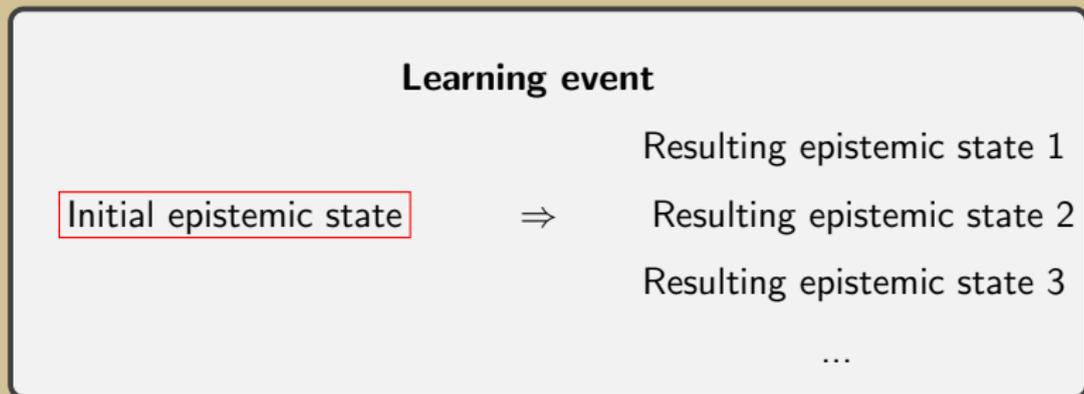
Learning event

Initial epistemic state \Rightarrow Resulting epistemic state

Dynamic Introspection (and ignorance)

- What about our knowledge (and ignorance) about the process?

Question: How to think about it? Modify the simple bare-bones model



Forward looking dynamic ignorance

- Dynamic ignorance: *Where am I going from here?*

Learning event

Initial epistemic state

⇒

Resulting epistemic state 1

Resulting epistemic state 2

Resulting epistemic state 3

Forward looking dynamic ignorance

- Dynamic ignorance: *Where am I going from here?*

Learning event

Initial epistemic state \Rightarrow Resulting epistemic state 1
Resulting epistemic state 2
Resulting epistemic state 3

Looking at Clock

Not knowing the time $(reliable) \Rightarrow$ Know time
 $(unreliable) \Rightarrow$ Don't know time

Backward looking dynamic ignorance

- Backward looking introspection: *How did I get here?*

Backward looking dynamic ignorance

- Backward looking introspection: *How did I get here?*

Initial state, learning event 1 ↘

Initial state, learning event 2 → Resulting epistemic state

Initial state, learning event 3 ↗

- Sometimes we don't know *how* we know:
what epistemic event brought me to my current situation
- Simple example: forgetting how you came to know the time

Dynamic Introspection

- The questions:
 - Forward looking: *Where am I going?*
 - Backwards looking: *How did I get here?*
- When the agent can answer them: *full dynamic introspection*.
- Otherwise: *dynamic ignorance*.

Dynamic Introspection

- The questions:
Forward looking: *Where am I going?*
Backwards looking: *How did I get here?*
- When the agent can answer them: *full dynamic introspection*.
- Otherwise: *dynamic ignorance*.

My terminology

- introspection = transparency
- No dynamic introspection = learning event, process, update, is opaque

Dynamic Introspection

- The questions:
Forward looking: *Where am I going?*
Backwards looking: *How did I get here?*
- When the agent can answer them: *full dynamic introspection*.
- Otherwise: *dynamic ignorance*.

My terminology

- introspection = transparency
- No dynamic introspection = learning event, process, update, is **opaque**
- How can these ideas be formalized?

Dynamic epistemic logic

- I develop a version of dynamic epistemic logic for that:¹
 - Propositional logic
 - propositional Knowledge operator: $K\varphi$
 - epistemic events as propositional operators: $[epistemic_event]\varphi$

¹ Cohen, Michael (2021). Opaque Updates. *Journal of Philosophical Logic*

Dynamic epistemic logic

- I develop a version of dynamic epistemic logic for that:¹
 - Propositional logic
 - propositional Knowledge operator: $K\varphi$
 - epistemic events as propositional operators: $[epistemic_event]\varphi$
- [looking at clock] K (time is 12:00)

¹ Cohen, Michael (2021). Opaque Updates. *Journal of Philosophical Logic*

Dynamic epistemic logic

- I develop a version of dynamic epistemic logic for that:¹
 - Propositional logic
 - propositional Knowledge operator: $K\varphi$
 - epistemic events as propositional operators: $[epistemic_event]\varphi$
- [looking at clock] K (time is 12:00)
- Existing dynamic epistemic logic, two axioms: *agents are able to:*
 - **No-Miracles:** $\langle event \rangle K\varphi \rightarrow K[event]\varphi$ a-priori predict the effect of events
 - **Perfect-Recall:** $K[event]\varphi \rightarrow [event]K\varphi$ a-posteriori tell how they got there
- In my framework, these are no longer axioms

¹ Cohen, Michael (2021). Opaque Updates. *Journal of Philosophical Logic*

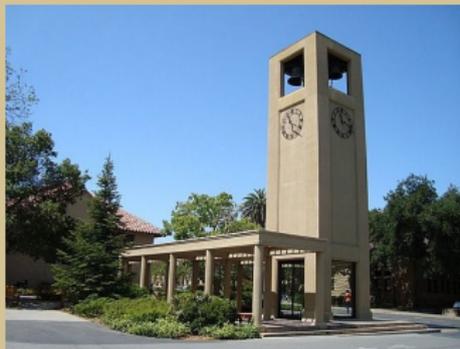
The Semantic idea

- A standard possible worlds model for the initial ignorance of the agent.

The Semantic idea

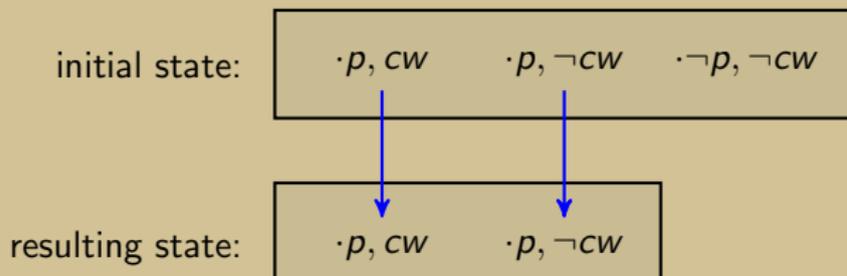
- A standard possible worlds model for the initial ignorance of the agent.
- p = the time is 12:00
- cw = the clock mechanism working

$\cdot p, cw$ $\cdot p, \neg cw$ $\cdot \neg p, \neg cw$



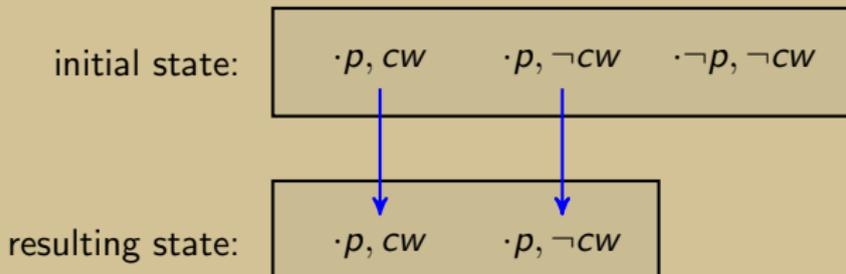
The received conception of updates

- The received idea:
learning p (the time is 12:00) = eliminating all the not- p states from the model.



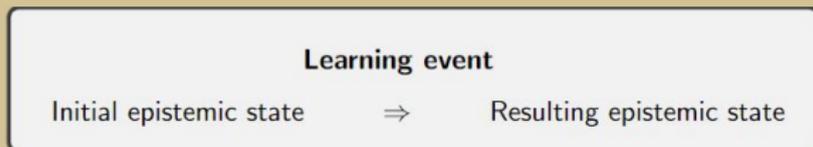
The received conception of updates

- The received idea:
learning p (the time is 12:00) = eliminating all the not- p states from the model.



Assumptions

- Such an update validates **No-Miracles, Perfect-Recall**, assume full dynamic introspection



My opaque conception of updates

- My approach: epistemic events are not transparent; behave differently at different possibilities

My opaque conception of updates

- My approach: epistemic events are not transparent; behave differently at different possibilities
- At one world - results in knowledge of p (good state)
- At a different world - results in no knowledge of p (bad state)

initial state:

$\cdot p, CW$

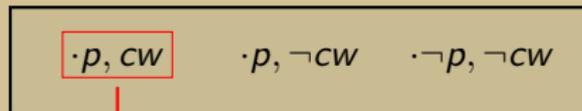
$\cdot p, \neg CW$

$\cdot \neg p, \neg CW$

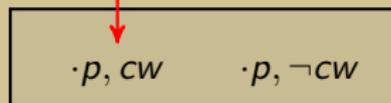
My opaque conception of updates

- My approach: epistemic events are not transparent; behave differently at different possibilities
- At one world - results in knowledge of p (good state)
- At a different world - results in no knowledge of p (bad state)

initial state:

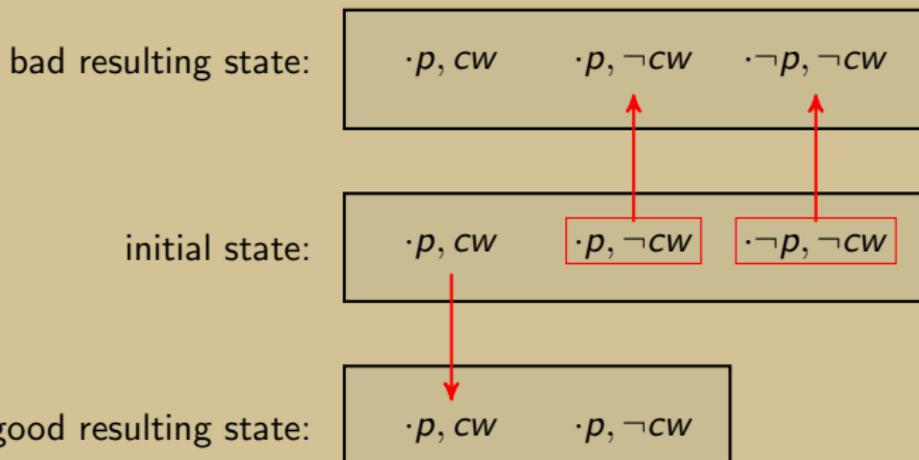


good resulting state:



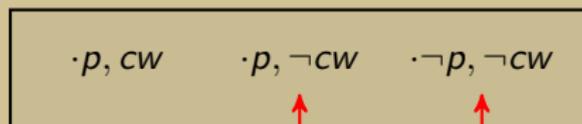
My opaque conception of updates

- My approach: epistemic events are not transparent; behave differently at different possibilities
- At one state - results in knowledge of p (good state)
- At a different state - results in no knowledge of p (bad state)

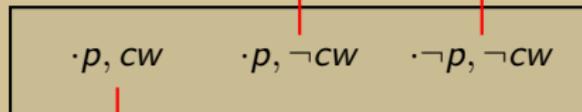


My opaque conception of updates

bad resulting state:



initial state:

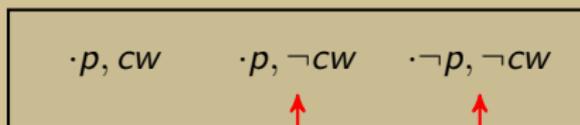


good resulting state:

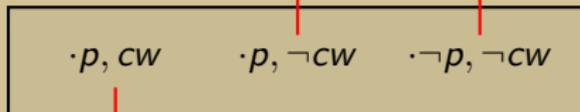


My opaque conception of updates

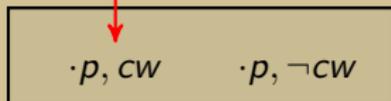
bad resulting state:



initial state:



good resulting state:



Learning event

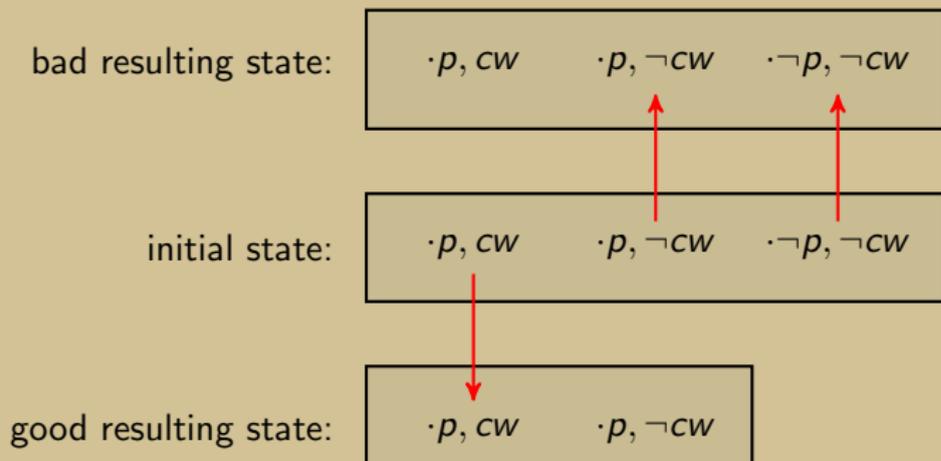
Initial epistemic state

⇒

Bad resulting state

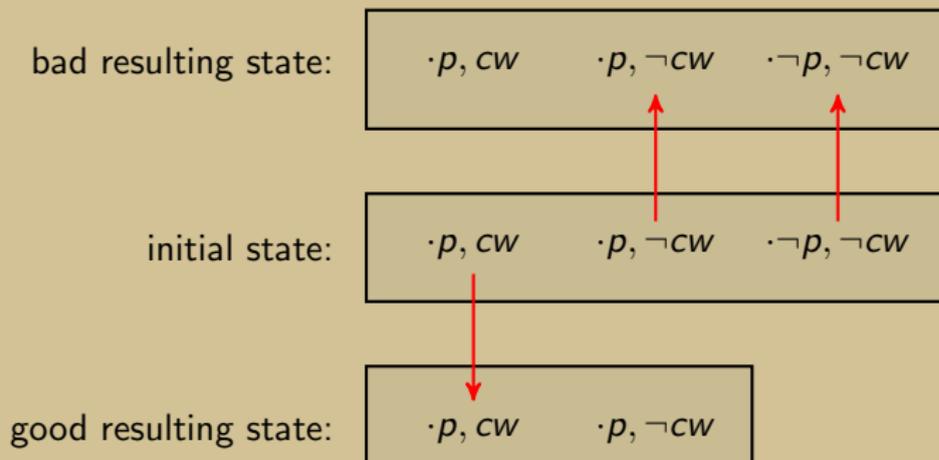
Good resulting state

My opaque conception of updates



- Updates are composite programs or actions (PDL), not propositions.

My opaque conception of updates



- Updates are composite programs or actions (PDL), not propositions.
- Epistemic program:
If cw is the case, update with p , otherwise, update with tautology.
- $\pi := (?cw; !p) \cup (? \neg cw; !T)$
- the operator $[\pi]$ does not validate **No-Miracles**, **Perfect-Recall**, it's opaque

Application to epistemology: skeptical reasoning

- Dynamic introspection is often *implicitly* assumed in epistemology. Worthwhile to make it explicit. Skeptical reasoning as an example.

Application to epistemology: skeptical reasoning

- Dynamic introspection is often *implicitly* assumed in epistemology. Worthwhile to make it explicit. Skeptical reasoning as an example.
- the skeptic argues that since we do not know that the source of information is reliable (clock is working properly), we can't get knowledge from the source (come to know the time). Applies to: perception, memory

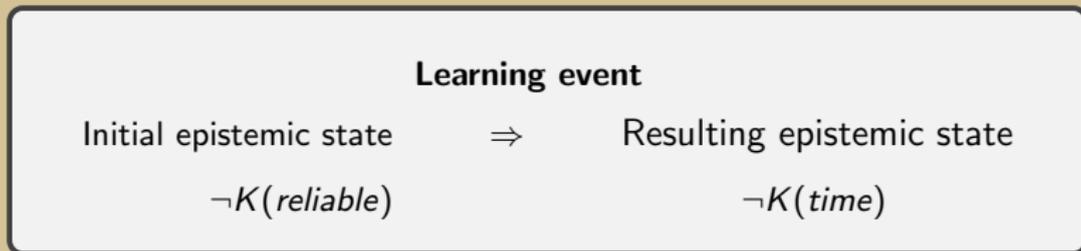
Learning event

Initial epistemic state	⇒	Resulting epistemic state
$\neg K(\text{reliable})$		$\neg K(\text{time})$

²Cf. Lyons (2016) Epistemological problems of perception.

Application to epistemology: skeptical reasoning

- Dynamic introspection is often *implicitly* assumed in epistemology. Worthwhile to make it explicit. Skeptical reasoning as an example.
- the skeptic argues that since we do not know that the source of information is reliable (clock is working properly), we can't get knowledge from the source (come to know the time). Applies to: perception, memory



- The implicit assumption of the skeptic:²
In order to get knowledge from the source, you have to antecedently know that the source is reliable

²Cf. Lyons (2016) Epistemological problems of perception.

The skeptic and no-miracles

- I argue: The skeptic implicitly assumes an instance of the **No-Miracles** principle (forward looking transparency):

$$[event]K\varphi \rightarrow K[event]\varphi$$

$$[\text{clock says 12}]K(\text{time is 12}) \rightarrow K[\text{clock says 12}](\text{time is 12})$$

The skeptic and no-miracles

- I argue: The skeptic implicitly assumes an instance of the **No-Miracles** principle (forward looking transparency):

$$[event]K\varphi \rightarrow K[event]\varphi$$

$$[\text{clock says 12}]K(\text{time is 12}) \rightarrow K[\text{clock says 12}](\text{time is 12})$$

- In contra-positive:

$$\neg K[event]\varphi \rightarrow \neg[event]K\varphi$$

$$\neg K[\text{clock says 12}](\text{time is 12}) \rightarrow \neg[\text{clock says 12}]K(\text{time is 12})$$

The skeptic and no-miracles

- I argue: The skeptic implicitly assumes an instance of the **No-Miracles** principle (forward looking transparency):

$$[event]K\varphi \rightarrow K[event]\varphi$$

$$[\text{clock says 12}]K(\text{time is 12}) \rightarrow K[\text{clock says 12}](\text{time is 12})$$

- In contra-positive:

$$\neg K[event]\varphi \rightarrow \neg[event]K\varphi$$

$$\neg K[\text{clock says 12}](\text{time is 12}) \rightarrow \neg[\text{clock says 12}]K(\text{time is 12})$$

- The skeptical argument:

P1: $\neg K[\text{clock says 12}]\text{time is 12}$

P2: $[\text{clock says 12}]K(\text{time is 12}) \rightarrow K[\text{clock says 12}](\text{time is 12})$

C: $\neg[\text{clock says 12}]K(\text{time is 12})$

By modus tollens

Reevaluation of the skeptical reasoning

- Dynamic introspection is central to the traditional debate about skepticism. Skeptic assumes transparency
- The skeptical argument is persuasive because it implicitly assumes **no-miracles**, and we did not have the conceptual tools to reject it.

³ Cohen, M. (2000) The problem of perception and the no-miracles principle. *Synthese*

Reevaluation of the skeptical reasoning

- Dynamic introspection is central to the traditional debate about skepticism. Skeptic assumes transparency
- The skeptical argument is persuasive because it implicitly assumes **no-miracles**, and we did not have the conceptual tools to reject it.
- Further, by *adding* dynamic ignorance, the sceptical argument is blocked.³

Looking at Clock

(*reliable*) ⇒ Know time

Not knowing the time

(*unreliable*) ⇒ Don't know time

- This line of reasoning is compatible with traditional responses to the skeptical problem, like externalism and disjunctivism.

³ Cohen, M. (2000) The problem of perception and the no-miracles principle. *Synthese*

Thank you