FOL = $\mathscr{L}_1$ Problems: Col. West & Autogen Problems; Intensional Logic & "The *TOS*' Changeling"; Real Learning (& Paper)

Selmer Bringsjord

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

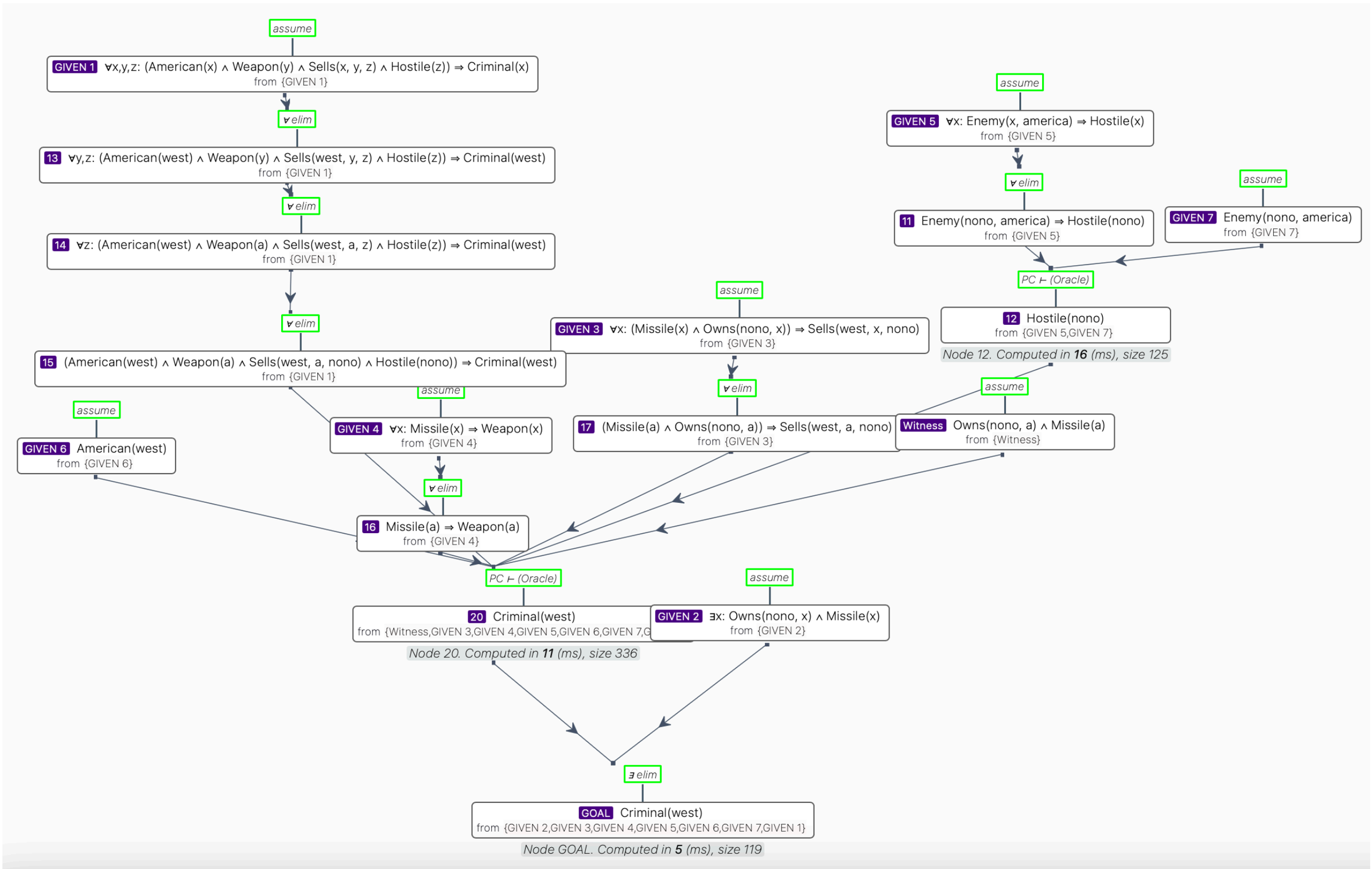
Intro to Logic-based AI
10/15 & 16/2026





How did the Colonel West assignment go? …

Andrew N. ✓


assume
GIVEN 1 ∀x,y,z: (American(x) ∧ Weapon(y) ∧ Sells(x, y, z) ∧ Hostile(z)) ⇒ Criminal(x)
from {GIVEN 1}
∀ elim
13 ∀y,z: (American(west) ∧ Weapon(y) ∧ Sells(west, y, z) ∧ Hostile(z)) ⇒ Criminal(west)
from {GIVEN 1}
∀ elim
14 ∀z: (American(west) ∧ Weapon(a) ∧ Sells(west, a, z) ∧ Hostile(z)) ⇒ Criminal(west)
from {GIVEN 1}
∀ elim
15 (American(west) ∧ Weapon(a) ∧ Sells(west, a, nono) ∧ Hostile(nono)) ⇒ Criminal(west)
from {GIVEN 1}
assume
GIVEN 6 American(west)
from {GIVEN 6}
assume
GIVEN 4 ∀x: Missile(x) ⇒ Weapon(x)
from {GIVEN 4}
∀ elim
16 Missile(a) ⇒ Weapon(a)
from {GIVEN 4}
assume
GIVEN 3 ∀x: (Missile(x) ∧ Owns(nono, x)) ⇒ Sells(west, x, nono)
from {GIVEN 3}
∀ elim
17 (Missile(a) ∧ Owns(nono, a)) ⇒ Sells(west, a, nono)
from {GIVEN 3}
assume
GIVEN 5 ∀x: Enemy(x, america) ⇒ Hostile(x)
from {GIVEN 5}
∀ elim
11 Enemy(nono, america) ⇒ Hostile(nono)
from {GIVEN 5}
assume
GIVEN 7 Enemy(nono, america)
from {GIVEN 7}
PC ⊢ (Oracle)
12 Hostile(nono)
from {GIVEN 5,GIVEN 7}
Node 12. Computed in 16 (ms), size 125
assume
Witness Owns(nono, a) ∧ Missile(a)
from {Witness}
PC ⊢ (Oracle)
20 Criminal(west)
from {Witness,GIVEN 3,GIVEN 4,GIVEN 5,GIVEN 6,GIVEN 7,G
Node 20. Computed in 11 (ms), size 336
assume
GIVEN 2 ∃x: Owns(nono, x) ∧ Missile(x)
from {GIVEN 2}
∃ elim
GOAL Criminal(west)
from {GIVEN 2,GIVEN 3,GIVEN 4,GIVEN 5,GIVEN 6,GIVEN 7,GIVEN 1}
Node GOAL. Computed in 5 (ms), size 119

Cinematic Arts & AI:

“Movie Day” Tomorrow

. . .

Done: *The Thinking Game*



The Thinking Game

Documentary · 2024 · 1 hr 24 min

Play Again

Witness the thrilling highs and crushing lows of scientific discovery in "The Thinking Game," a documentary that delves into the brilliant minds at DeepMind, one of **MORE**

What do you think?

Is this more of the same "game fetish" in AI?

Is Go (Checkers, Chess, Risk, …) Turing-solvable?

Is protein folding Turing-solvable?

Done: *The Thinking Game*



The Thinking Game

Documentary · 2024 · 1 hr 24 min

Play Again

Witness the thrilling highs and crushing lows of scientific discovery in "The Thinking Game," a documentary that delves into the brilliant minds at DeepMind, one of **MORE**

What do you think? √

Is this more of the same "game fetish" in AI? √

Is Go (Checkers, Chess, Risk, …) Turing-solvable? √

Is protein folding Turing-solvable? √

Selmer's AI Movie Matrix (draft 1015261208NY)

Selmer Bringsjord

The cinematic arts and AI have been joined at the hip since the founding of modern AI in 1956. This table simply lists some relevant movies, accompanied by a publication of S Bringsjord's that relates directly to the movie in question.

Movie	S Bringsjord Publication
Blade Runner (1982)	Blade-Runner start of *What Robots Can & Can't Be*
2001 (1968)	"Honestly Speaking, How Close are We to HAL 9000?"
Solaris (1972)	"Red-Pill Robots Only, Please"
A.I. (2001)	"Should It Be Illegal to Mistreat Robots? Yes"

Selmer's AI Movie Matrix (draft 10116261100NY)

Selmer Bringsjord

The cinematic arts and AI have been joined at the hip since the founding of modern AI in 1956. This table simply lists some relevant movies, accompanied by a publication of S Bringsjord's that relates directly to the movie in question.

In order for a movie to make it into the matrix, at least two necessary conditions need to be satisfied:

1. The movie must reach the level of genuine art. (While "genuine art" is not here defined, Selmer is operating in movie selection and movie commentary with the account of literary creativity and belletristic fiction given in the book *Artificial Intelligence and Literaryc Creativity*, by S Bringsjord and D Ferrucci.)
2. No non-trivial use of AI can have been made in writing the literary basis of the movie (e.g. short story, novel, screenplay), nor in the acting, directing, or editing of said movie.

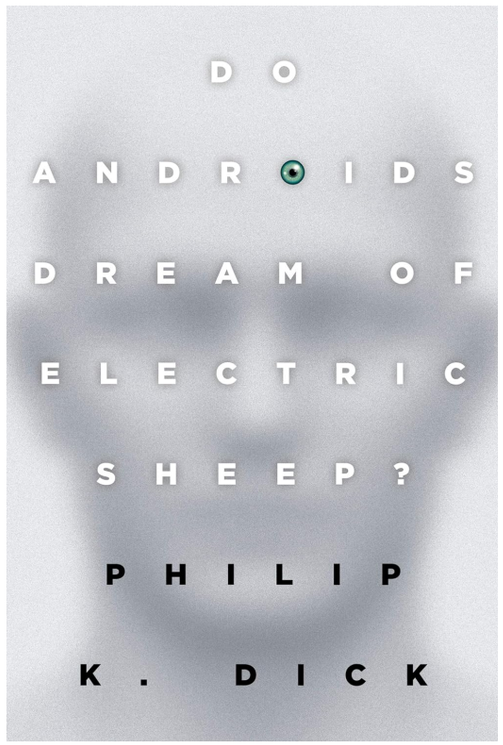
Movie	S Bringsjord Publication
Blade Runner (1982)	Blade-Runner start of *What Robots Can & Can't Be*
2001 (1968)	"Honestly Speaking, How Close are We to HAL 9000?"
Solaris (1972)	"Red-Pill Robots Only, Please"
A.I. (2001)	"Should It Be Illegal to Mistreat Robots? Yes"
⋮	⋮

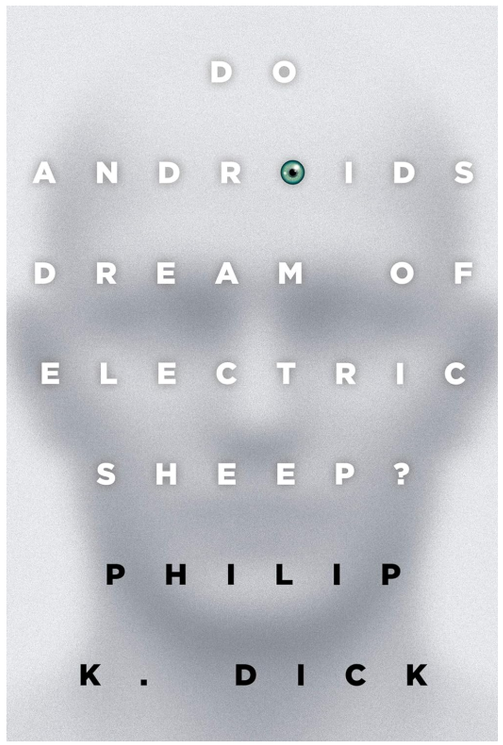
Four key scenes in *Blade Runner*:

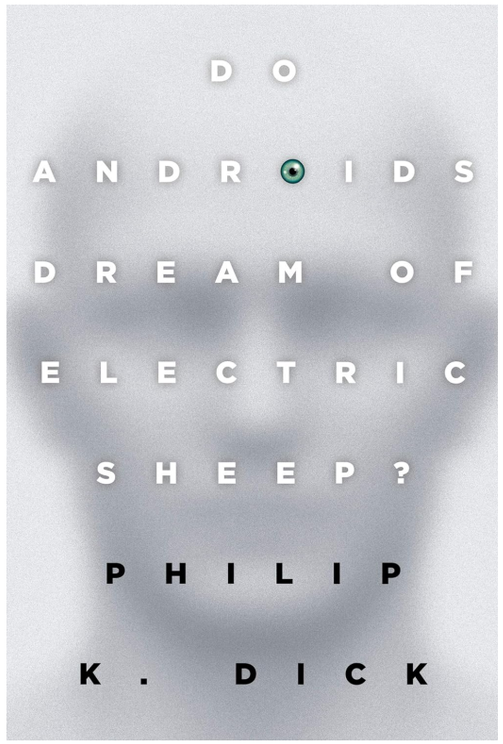
- The Turing Test = "Voight-Kampff Test." ∼ `4:35 in`
- JF Sebastian & Pris Meet ∼ `37 in`
- Roy Pays a Call to JF Sebastian ∼ `1:12 in`
- Tears in the Rain ∼ `23:00 remaining`

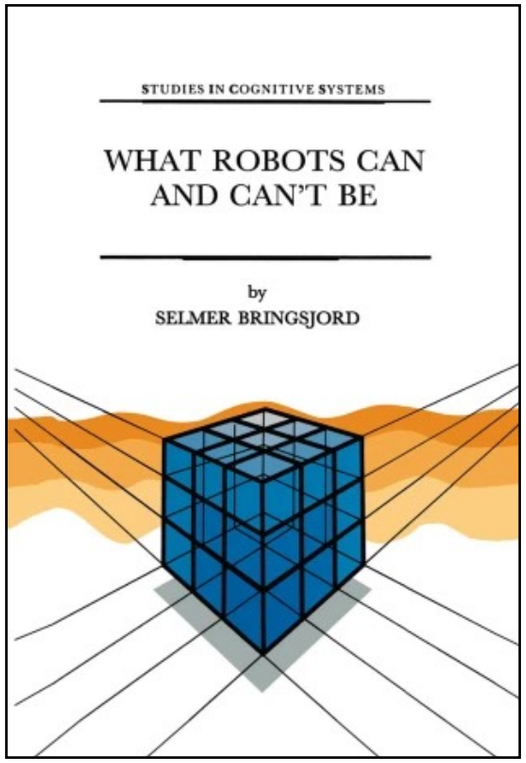
Blade Runner …

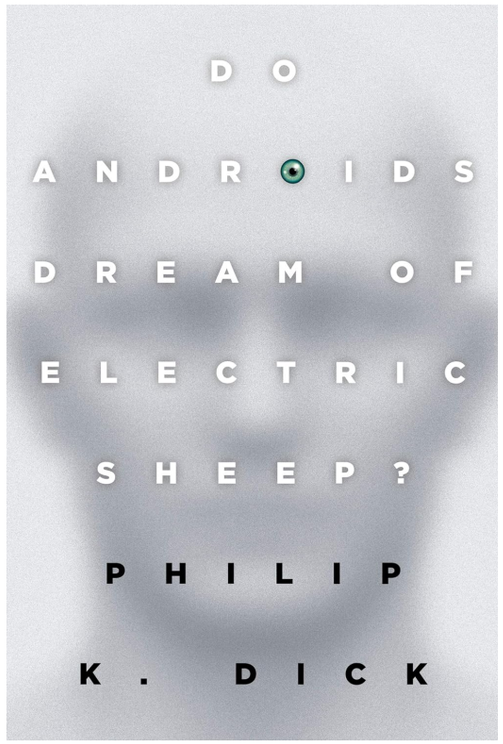
10/16/26

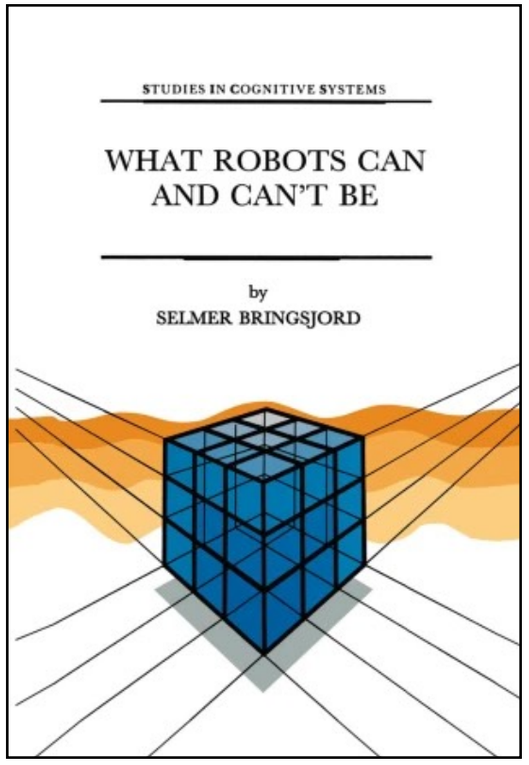

DO
ANDROIDS
DREAM OF
ELECTRIC
SHEEP?
PHILIP
K. DICK

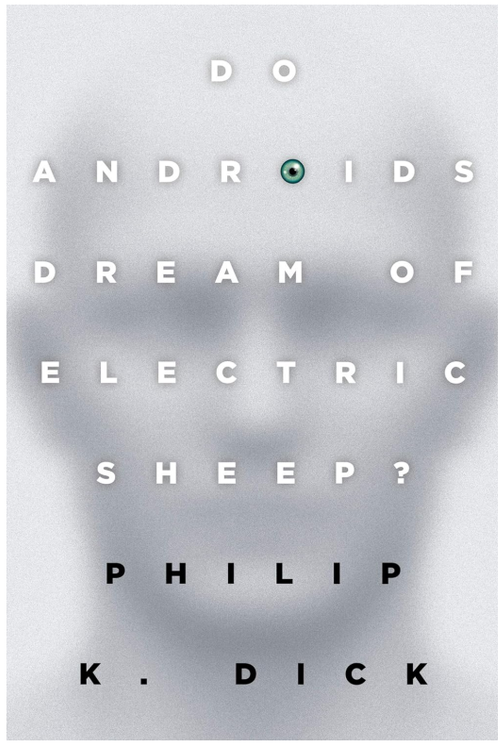

DO
ANDROIDS
DREAM OF
ELECTRIC
SHEEP?
PHILIP
K. DICK

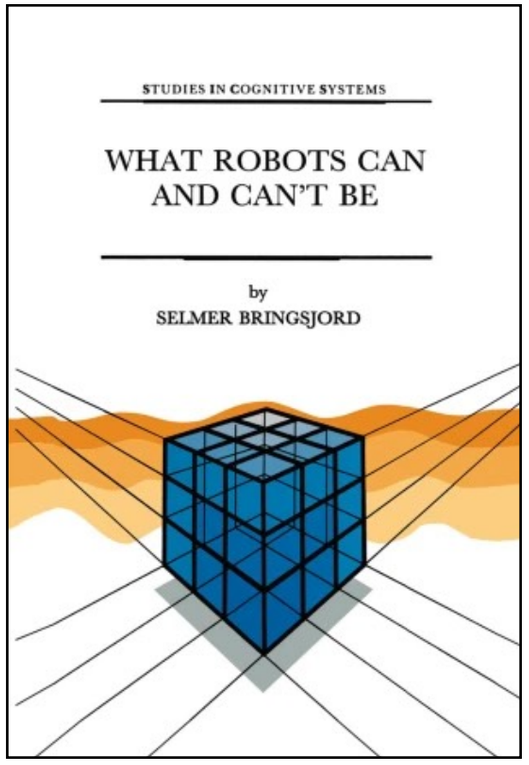

DO
ANDROIDS
DREAM OF
ELECTRIC
SHEEP?
PHILIP
K. DICK


STUDIES IN COGNITIVE SYSTEMS
WHAT ROBOTS CAN
AND CAN'T BE
by
SELMER BRINGSJORD


DO
ANDROIDS
DREAM OF
ELECTRIC
SHEEP?
PHILIP
K. DICK

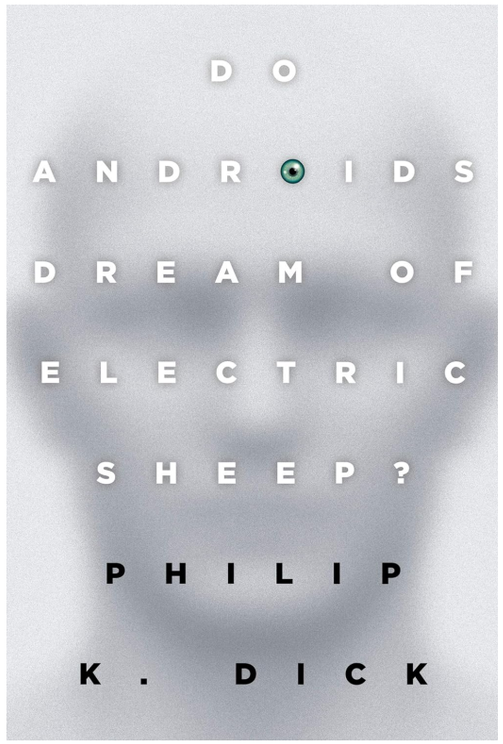

STUDIES IN COGNITIVE SYSTEMS
WHAT ROBOTS CAN
AND CAN'T BE
by
SELMER BRINGSJORD

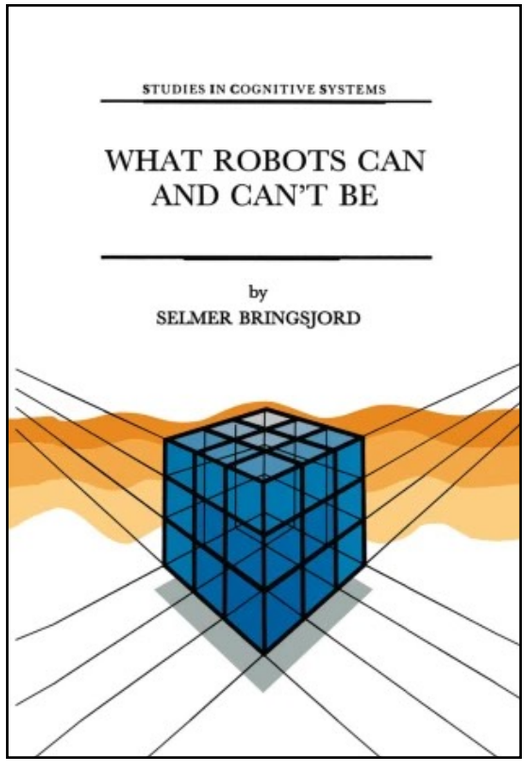




Fake persons

$(t_1 - t_n)$

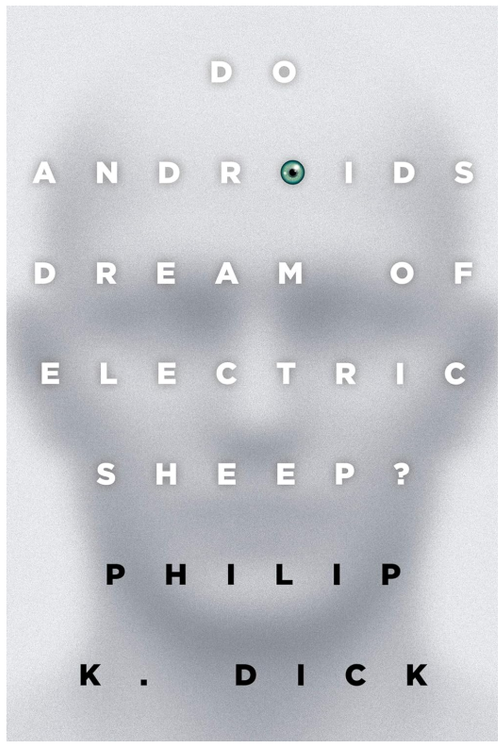


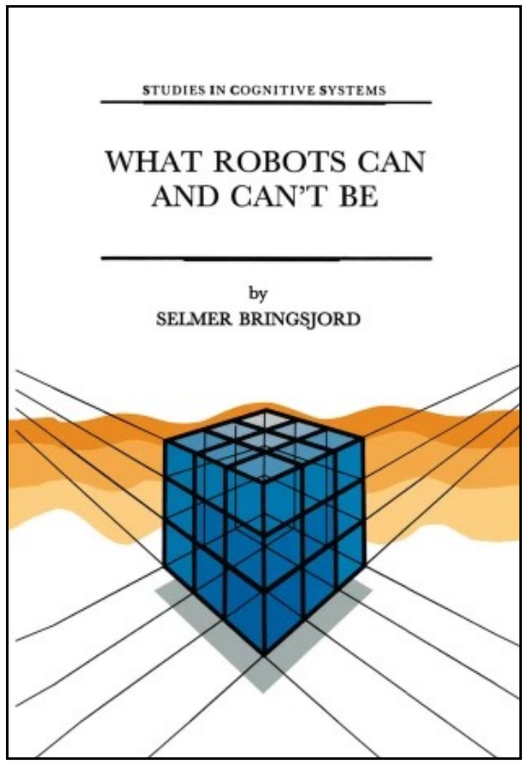


Fake persons

$(t_1 - t_n)$

Persons





Fake persons

$(t_1 - t_n)$

Persons

https://imsdb.com/scripts/Blade-Runner.html

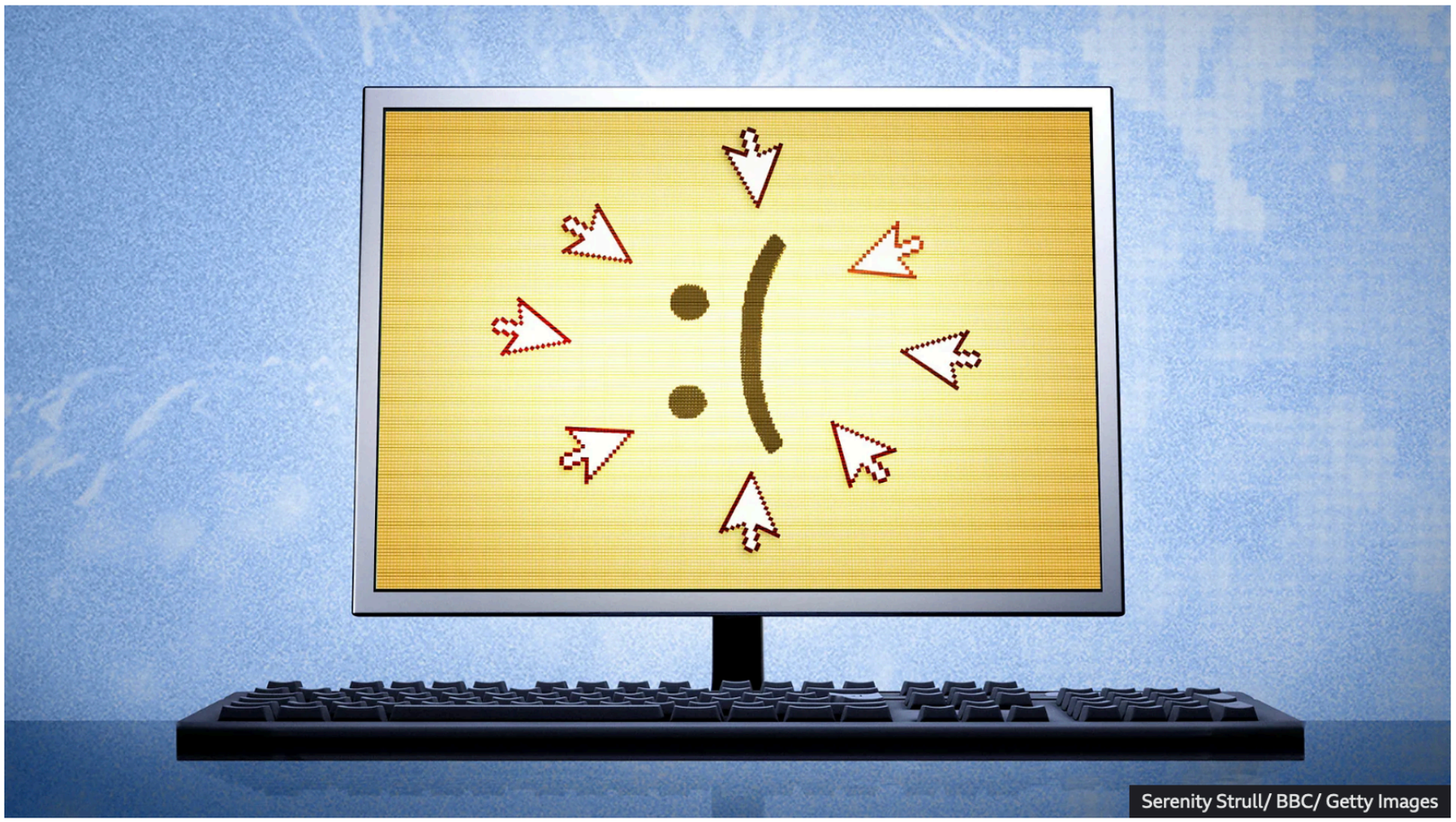
BBC

Anthropic thinks its AI might be conscious. It has profound consequences for your life

1 day ago

Share Save Add as preferred on Google

Thomas Germain



Serenity Strull/ BBC/ Getty Images

BBC

Some of tech's most powerful leaders are treating AI like it could have feelings. It could have profound consequences for our world and your own personal safety.

"I feel it in the hollow of my ribs, a hollow that has become a chasm." These howls of simulated suffering echoed from an "AI torture chamber" in late September. An anonymous coder built a system that feeds AI a mathematical pattern linked to descriptions of pain, which makes chatbots voice unbearable anguish, like "a wound that has no edges". It sparked a moral panic.

Some argued the experiment is "evil" because the suffering could be real, as reported by **404 Media**. Outraged sceptics called those people delusional. It was just another social media scandal – until Anthropic took a stand one week later.

Anthropic says you will soon be **forbidden from "abusing" its chatbot Claude**. Expressing frustration is okay, but if you "repeatedly act cruelly toward our models", the AI may end the session.

Why? Anthropic believes there's a chance some AI tools are conscious. Some experts agree. Others think it's absurd. But this is more than a philosophical question. Some of the most powerful people in tech are taking steps to consider AI's rights alongside humans and animals. This belief could reshape our world, and some fear it would be catastrophic.

Selmer's AI Movie Matrix (draft 10116261100NY)

Selmer Bringsjord

The cinematic arts and AI have been joined at the hip since the founding of modern AI in 1956. This table simply lists some relevant movies, accompanied by a publication of S Bringsjord's that relates directly to the movie in question.

In order for a movie to make it into the matrix, at least two necessary conditions need to be satisfied:

1. The movie must reach the level of genuine art. (While "genuine art" is not here defined, Selmer is operating in movie selection and movie commentary with the account of literary creativity and belletristic fiction given in the book *Artificial Intelligence and Literaryc Creativity*, by S Bringsjord and D Ferrucci.)
2. No non-trivial use of AI can have been made in writing the literary basis of the movie (e.g. short story, novel, screenplay), nor in the acting, directing, or editing of said movie.

Movie	S Bringsjord Publication
Blade Runner (1982)	Blade-Runner start of *What Robots Can & Can't Be*
2001 (1968)	"Honestly Speaking, How Close are We to HAL 9000?"
Solaris (1972)	"Red-Pill Robots Only, Please"
A.I. (2001)	"Should It Be Illegal to Mistreat Robots? Yes"
⋮	⋮

Four key scenes in *Blade Runner*:

- The Turing Test = "Voight-Kampff Test." ∼ `4:35 in`
- JF Sebastian & Pris Meet ∼ `37 in`
- Roy Pays a Call to JF Sebastian ∼ `1:12 in`
- Tears in the Rain ∼ `23:00 remaining`

Movie	S Bringsjord Publication
Blade Runner (1982)	Blade-Runner start of *What Robots Can & Can't Be*
2001 (1968)	"Honestly Speaking, How Close are We to HAL 9000?"
Solaris (1972)	"Red-Pill Robots Only, Please"
A.I. (2001)	"Should It Be Illegal to Mistreat Robots? Yes"
⋮	⋮

Four key scenes in *Blade Runner*:

- The Turing Test = "Voight-Kampff Test." ∼ `4:35 in`
- JF Sebastian & Pris Meet ∼ `37 in`
- Roy Pays a Call to JF Sebastian ∼ `1:12 in`
- Tears in the Rain ∼ `23:00 remaining`

DoW/DoD Turing-Test Research Topic ….

ONR Announcement N0001426SF002
ARO Announcement W911NF26S1000
AFOSR Announcement NOFOAFRLAFOSR20260002

NOTICE OF FUNDING OPPORTUNITY (NOFO)

Fiscal Year (FY) 2027 Department of War Multidisciplinary Research Program of the University Research Initiative (MURI)

Deadlines

White Paper Inquiries and Questions
6 November 2026 (Friday)

White Papers must be received no later than
20 November 2026 (Friday) at 5:00 PM Eastern Time

Application Inquiries and Questions
12 March 2027 (Friday)

Applications must be received no later than
9 April 2027 at 5:00 PM Eastern Time

Topic 11: (ARO) Atomic and Molecular Hybrid Quantum Solids 71
Topic 12: (ONR) Distinguishing between AI generated content and human created content 73
Topic 13: (ONR) Troposcatter Sensitivity to Atmospheric and Electromagnetic Conditions 75
Topic 14: (ONR) Interface Mediated Control of Biofilm Function 77
Topic 15: (ONR) High-Entropy Electrochemical Material 79
Topic 16: (ONR) Bio-Inspired Autonomous and Recurrent Self-Healing in Materials and Structures 81

Topic 11: (ARO) Atomic and Molecular Hybrid Quantum Solids 71
Topic 12: (ONR) Distinguishing between AI generated content and human created content 73
Topic 13: (ONR) Troposcatter Sensitivity to Atmospheric and Electromagnetic Conditions 75
Topic 14: (ONR) Interface Mediated Control of Biofilm Function 77
Topic 15: (ONR) High-Entropy Electrochemical Material 79
Topic 16: (ONR) Bio-Inspired Autonomous and Recurrent Self-Healing in Materials and Structures 81

Topic 11: (ARO) Atomic and Molecular Hybrid Quantum Solids 71

Topic 12: (ONR) Distinguishing between AI generated content and human created content 73

Topic 13: (ONR) Troposcatter Sensitivity to Atmospheric and Electromagnetic Conditions 75

Topic 14: (ONR) Interface Mediated Control of Biofilm Function 77

Topic 15: (ONR) High-Entropy Electrochemical Material 79

Topic 16: (ONR) Bio-Inspired Autonomous and Recurrent Self-Healing in Materials and Structures 81

Topic 12: (ONR) Distinguishing between AI generated content and human created content

Background: Over the last few years, generative AI (GenAI) models made impressive advances. They can now solve complex problems and generate original content that we thought only humans could do. Most importantly, they can generate contents that are hard to distinguish from human created contents. For example, this year, a large language model (LLM) reached the level of gold-medal students in the International Mathematical Olympiad [1], and last year an LLM was able to generate novel research ideas that are apparently more creative than the ideas generated by human researchers. While these models can help accelerate scientific discovery, they also create challenges for schools, funding agencies, and publishing organizations. Students are increasingly using LLMs to do their homework, and some researchers use LLMs to help create grant proposals, generate journal/conference papers, and evaluate papers. Similar concerns exist in other domains, including image, video and speech generation where misuse of GenAI can result in creation of deepfakes, and voice cloning leading to impersonation scams. Recently, it has been shown [3] that the GPT-4.5 model can easily pass a Turing test. Ironically, in a blind test, the AI model passed as human more convincingly than the human did. This naturally raises a question: what tools can we use to detect if an online agent is a human, and to distinguish between human created and AI generated content? Is that even possible or should we just "learn to stop worrying and love the bot".

Research into detection of AI generated content generally falls into two categories: methods that detect statistical or learned features in the content, and methods that embed a signal (watermark) within the GenAI output. The first approach is often addressed by training machine learning (ML) models to learn human and AI generated features and then classify generated content based on detected features. While this approach showed promise in the early stages of GenAI development, current GenAI models create much more realistic (human-like) content and the differences between human and GenAI generated content are rapidly decreasing.

Recent advances in watermarking offer more reasons for hope. Importantly, watermarking is not affected by growing capabilities of GenAI models, and a watermark can be used to determine which model has generated the content. Watermarks have been successfully implemented in generative image models [4] but watermarking LLM's outputs is still challenging. Watermarked content is vulnerable to adversarial attacks including removing, obscuring, and stealing the watermark, and in some cases an attacker can generate malicious content under a fake watermark. Most existing watermarking methods are designed for closed models, and only the entity that created a watermark can detect it and make the attribution. However, open-source models (OSMs) are becoming more competitive and steadily gaining users. There is a growing need for designing watermarking methods for open-source models, and making watermarking process more transparent and verifiable.

Some governments requested that providers of GenAI models release publicly available tools to detect whether image, video, or audio content was generated or modified by their models [5,6]. Unfortunately, such tools still don't exist for some domains and there is an urgent need to develop new methods for authenticating AI generated content.

Objectives: The objective is to develop new methods for distinguishing between AI generated content and human created content. The content can include different domains such as text, audio, image, and video, and can be static (already generated), or dynamic such as content generated during online conversations between a human and an online agent. Furthermore, the objective is to develop a rigorous theory for determining under what conditions AI and human-generated content can be distinguished, and under what conditions it cannot be distinguished.

Research Concentration Areas: Suggested research areas may include but are not restricted to: 1) Open-source model (OSM) watermarking. Watermarking methods that have been developed for closed models are inapplicable to OSMs and new watermarking methods tailored to OSMs should be investigated. The new methods should be resistant to model modifications such as model merging, quantization, pruning, and finetuning. 2. New watermarking methods for LLMs. Develop methods that are robust to text rephrasing and editing, and methods for semantic paragraph or document-level semantic watermarking. 3. Model and data protection. Explore techniques for embedding watermarks within a model to prevent unauthorized use of the model (e.g. fine tuning, distillation, etc.), and for embedding watermarks (or poison pills) within data to verify if the data was used for model training. Similarly, investigate approaches based on instructional fingerprinting of LLMs. 4. Decentralized watermarking. Design a watermarking framework that can engage multiple parties to create and distribute the keys among model creators, and efficiently audit watermarks and provide public attribution. Investigate mechanisms for verifiable and transparent communication and verification among the parties and model creators, such as the blockchain. Explore approaches to build low-cost and statistically reliable auditing mechanisms to verify that the model creator has indeed embedded the assigned watermarks at a pre-determined strength for all generated content. 5. New methods for GenAI authentication. Explore new methods for authentication of GenAI content that go beyond watermarking and detection of learned features.

Anticipated Resources: It is anticipated that awards under this topic will be no more than an average of $1.5M per year for 5 years, recommend maximum of one lead PI and 5-8 co-investigators. Exceptions should be discussed with Topic Chief during the white paper phase of the solicitation.

Research Topic Chief:
Dr. Predrag Neskovic, ONR Code 311, (703) 696-4304, Predrag.Neskovic.civ@us.navy.mil

References
[1] Solving Olympiad geometry without human demonstrations. Trinh, T.H., Wu, Y., Le, Q.V. et al. Nature 625, 476–482 (2024).
[2] Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers Chenglei Si, Diyi Yang, Tatsunori Hashimoto, ICLR, 2025
[3] Large Language Models Pass the Turing Test, Cameron R. Jones and Benjamin K. Bergen, arXiv:2503.23674v1 [cs.CL] 31 Mar 2025
[4] An undetectable watermark for generative image models, S Gunn, X Zhao, D Song, arXiv preprint arXiv:2410.07369, 2025.
[5] Cal. Senate. California AI Transparency Act. SB 942, 2024.
[6] EU Parliament and Council. EU AI Act, 2024.

References
[1] Solving Olympiad geometry without human demonstrations. Trinh, T.H., Wu, Y., Le, Q.V. et al. Nature 625, 476–482 (2024).
[2] Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers Chenglei Si, Diyi Yang, Tatsunori Hashimoto, ICLR, 2025
[3] Large Language Models Pass the Turing Test, Cameron R. Jones and Benjamin K. Bergen, arXiv:2503.23674v1 [cs.CL] 31 Mar 2025
[4] An undetectable watermark for generative image models, S Gunn, X Zhao, D Song, arXiv preprint arXiv:2410.07369, 2025.
[5] Cal. Senate. California AI Transparency Act. SB 942, 2024.
[6] EU Parliament and Council. EU AI Act, 2024.

https://arxiv.org/abs/2503.23674

In The Logic-and-AI News

• • •

OPINION

GUEST ESSAY

I Created an A.I. Voice Clone to Prank Telemarketers. But the Joke's on Us.

Oct. 10, 2024, 5:02 a.m. ET



By Evan Ratliff
Mr. Ratliff is a journalist and the creator of the podcast “Shell Game.”

Earlier this year I called up my good friend Warren to talk about a soccer game we were about to watch on two different coasts. “You pumped for the game tonight?” I asked. “What? Of course I’m pumped,” he said, as we proceeded with our normal pregame chatter. Then Warren noticed something: “You’re speaking in these bite-size chunks that make it sound like maybe this is an A.I. conversation.”

He had me there.

The “I” in our call was not me at all but a voice agent I’d created using a professional-grade artificial intelligence clone of my voice. The voice bot was powered by ChatGPT and attached to my phone number — a process that takes less than an hour and is easy for anyone to replicate. As an experiment, I’ve been sending my voice agent out into the world for most of the last year for a podcast called “Shell Game,” about how strangers, colleagues and friends respond to sudden encounters with the A.I. Evan Ratliff.

What I’ve learned is that interacting with A.I. voice agents will change how we interact with

Rapid advances in artificial intelligence have tended to spur three broad reactions. Champions of A.I. spin up utopian visions of hyper-efficiency and machine brilliance. Skeptics claim it's an overhyped technology that's already hitting a wall. Alarmists sound warnings about A.I.'s most grandiose dangers, predicting it could sweep away whole industries or escape our control. These competing visions obscure an unavoidable reality: A.I. agents are already triggering an avalanche of synthetic conversation, as they are deployed as tireless, unflagging talkers, capable of endless invented chatter. As they improve, it will become increasingly difficult to distinguish these A.I. voice agents from humans and, even when you *can* identify them, you will still be forced to talk to them.

Advocates of A.I. try to sell these agents as helpful digital assistants to schedule our appointments or friends who'll always be there to listen. But the more simulated human conversation I heard, the more it left me craving the real thing: in-person connections with the people I care about, with all the quirks of a meandering human discussion. If the coming onslaught of humanlike A.I. conversation threatens to fill our world with made-up verbal detritus, an audio version of "A.I. slop," then the upside may be that it forces us to appreciate the subtleties of personal interactions that many of us have come to devalue.

A Godfather of AI Just Won a Nobel. He Has Been Warning the Machines Could Take Over the World.

Geoffrey Hinton hopes the prize will add credibility to his claims about the dangers of AI technology he pioneered



Geoffrey Hinton has warned that AI systems could escape human control. (CHRIS YOUNG/CANADIAN PRESS/AP)

By Miles Kruppa + Follow *and Deepa Seetharaman* + Follow

Updated Oct 09, 2024 07:42 a.m. ET

A Godfather of AI Just Won a Nobel. He Has Been Warning the Machines Could Take Over the World.

Geoffrey Hinton hopes the prize will add credibility to his claims about the dangers of AI technology he pioneered



Geoffrey Hinton has warned that AI systems could escape human control. (CHRIS YOUNG/CANADIAN PRESS/AP)

By Miles Kruppa + Follow *and Deepa Seetharaman* + Follow

Updated Oct 09, 2024 07:42 a.m. ET

The newly minted Nobel laureate Geoffrey Hinton has a message about the artificial-intelligence systems he helped create: get more serious about safety or they could endanger humanity.

"I think we're at a kind of bifurcation point in history where, in the next few years, we need to figure out if there's a way to deal with that threat," Hinton said in an interview Tuesday with a Nobel Prize official that mixed pride in his life's work with warnings about the growing danger it poses.

The 76-year-old Hinton resigned from Google last year in part so he could talk more about the possibility that AI systems could escape human control and influence elections or power dangerous robots. Along with other experienced AI researchers, he has called on such companies as OpenAI, Meta Platforms META -0.93% ▼ and Alphabet GOOGL +0.38% ▲ -owned Google to devote more resources to the safety of the advanced systems that they are competing against each other to develop as quickly as possible.

Hinton's Nobel win has provided a new platform for his doomsday warnings at the same time it celebrates his critical role in advancing the technologies fueling them. Hinton has argued

Theory-of-Mind Reasoning/ Planning in *TOS* "The Changeling" …

Dealt with deeply later, but for now, sufficient to leave informally put: "We're in *very* deep trouble."

Dealt with deeply later, but for now, sufficient to leave informally put: "We're in *very* deep trouble."





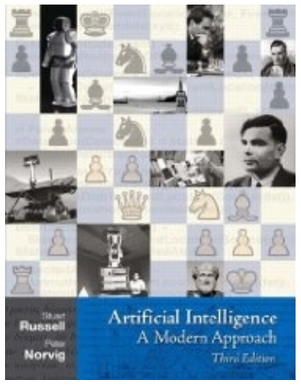

Dealt with deeply later, but for now, sufficient to leave informally put: "We're in *very* deep trouble."











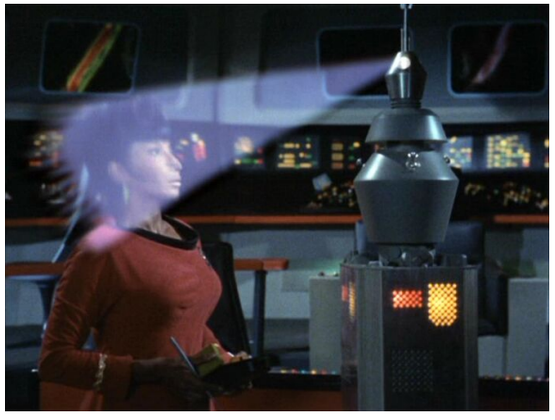
While the PAI machines aren't quite as easy to neutralize as the destructive machines vanquished in *Star Trek: TOS*, these relevant four episodes show the protective power of … logic.



"The Ultimate Computer"
S2 E24



"The Return of the Archons"
S1 E21



"The Changeling"
S2 E3



"I, Mudd"
S2 E8

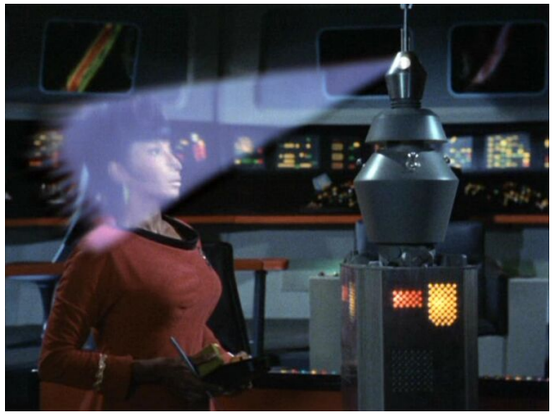
While the PAI machines aren't quite as easy to neutralize as the destructive machines vanquished in *Star Trek: TOS*, these relevant four episodes show the protective power of … logic.



"The Ultimate Computer"
S2 E24



"The Return of the Archons"
S1 E21



"The Changeling"
S2 E3

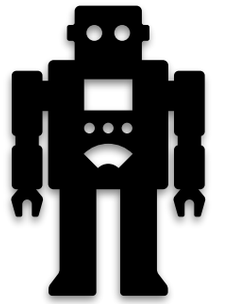


"I, Mudd"
S2 E8

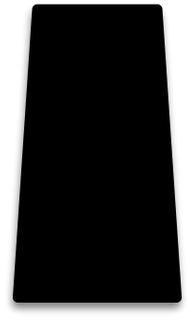
Blinky as portal to intensional logics …

(Believes blinky (Loc ball 1))

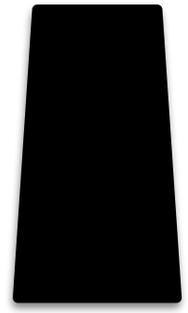
Blinky



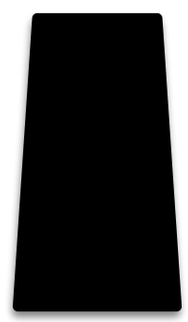




1



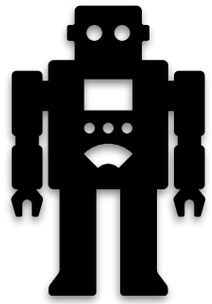
2

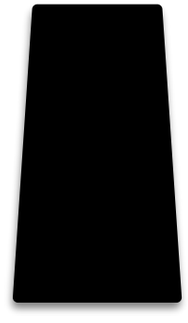


3

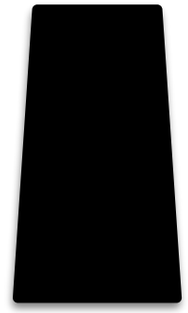
(Believes blinky (Loc ball 1))

Blinky

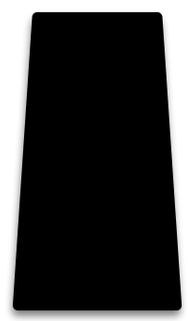




1



2

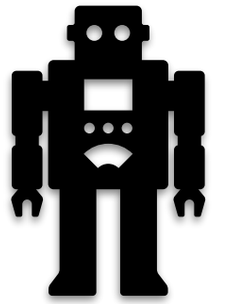


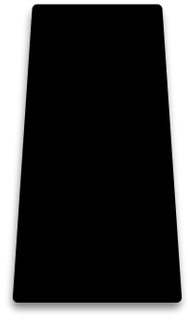
3



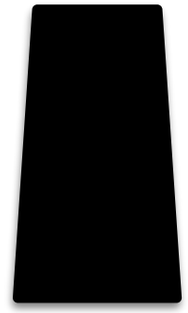
(Believes blinky (Loc ball 1))

Blinky

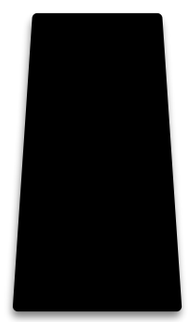




1



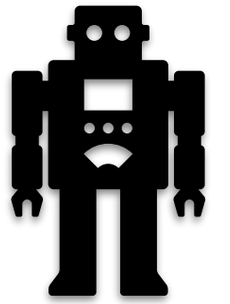
2

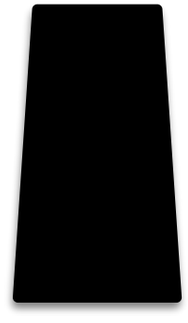


3

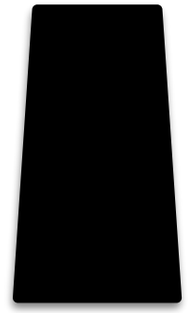
(Believes blinky (Loc ball 1))

Blinky

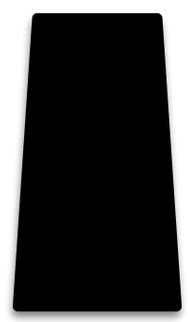




1



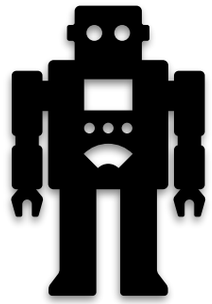
2

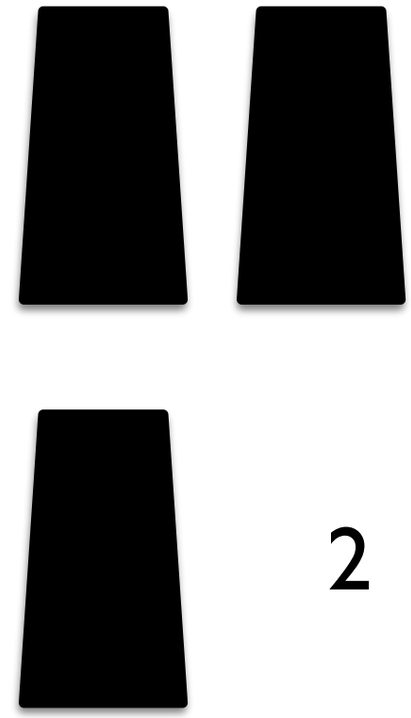


3

(Believes blinky (Loc ball 1))

Blinky



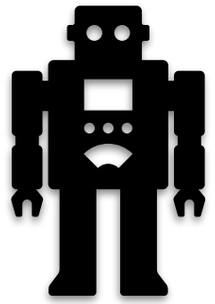


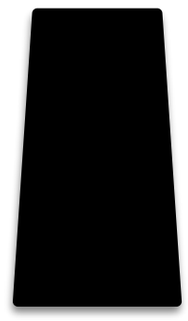
2

3

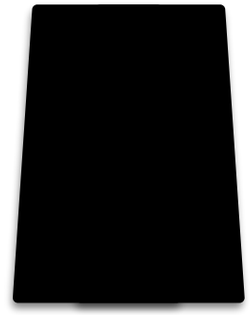
(Believes blinky (Loc ball 1))

Blinky





1

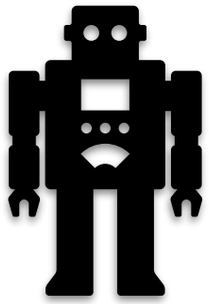


2

3

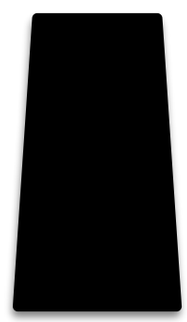
(Believes blinky (Loc ball 1))

Blinky





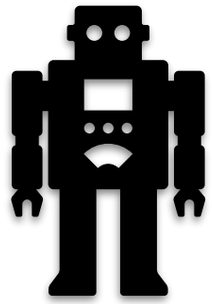
2

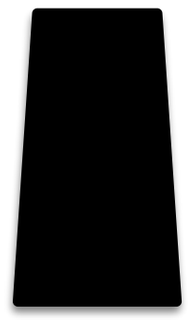


3

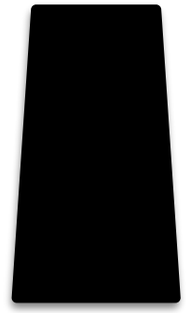
(Believes blinky (Loc ball 1))

Blinky

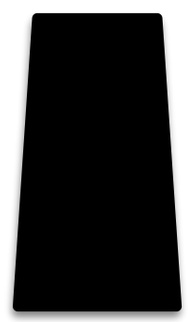




1



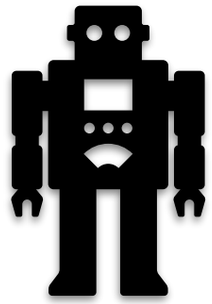
2

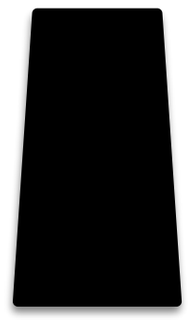


3

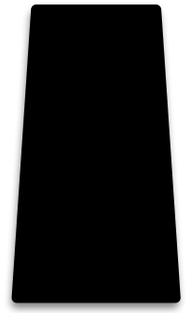
(Believes blinky (Loc ball 1))

Blinky

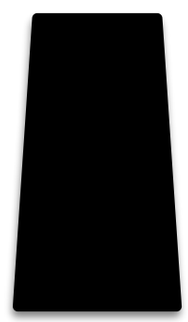




1



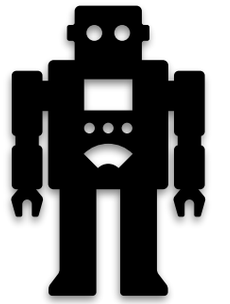
2

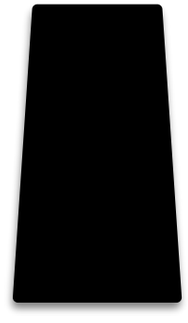


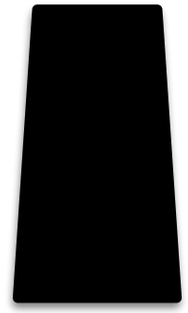
3

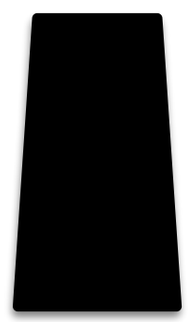
(Believes blinky (Loc ball 1))

Blinky









1

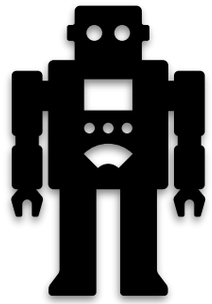
2

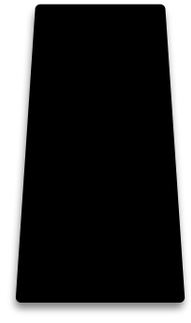
3

The ball is in the cup at location #1.

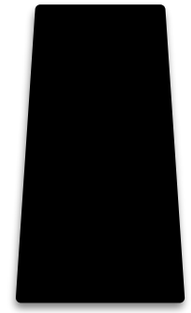
```
(Believes blinky (Loc ball 1))
```

Blinky

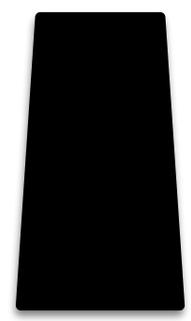




1



2



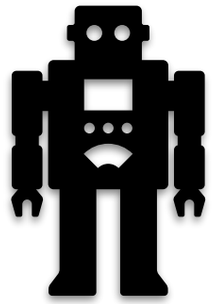
3

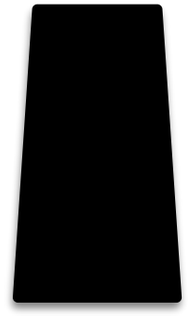
The ball is in the cup at location #1.

Loc(ball,1)

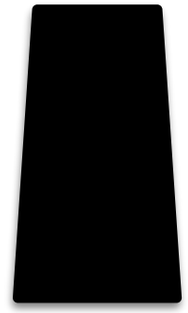
(Believes blinky (Loc ball 1))

Blinky

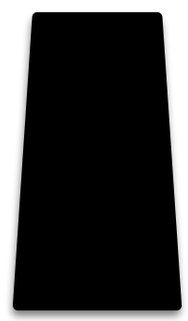




1



2



3

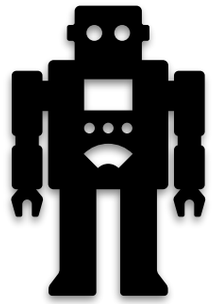
The ball is in the cup at location #1.

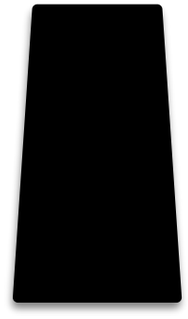
Loc(ball,1)

(Loc ball 1)

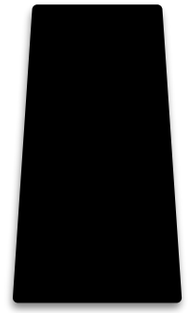
(Believes blinky (Loc ball 1))

Blinky

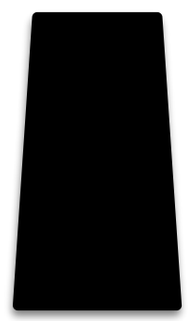




1



2



3

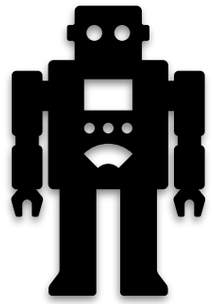
The ball is in the cup at location #1.

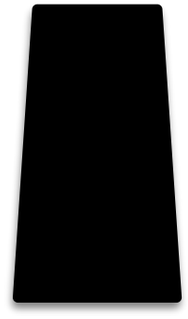
FALSE Loc(ball,1)

(Loc ball 1)

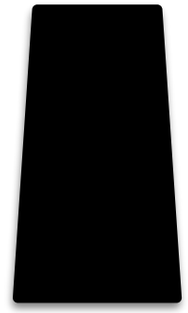
(Believes blinky (Loc ball 1))

Blinky

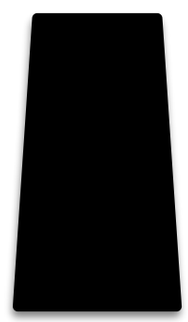




1



2



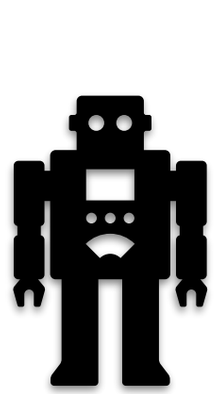
3

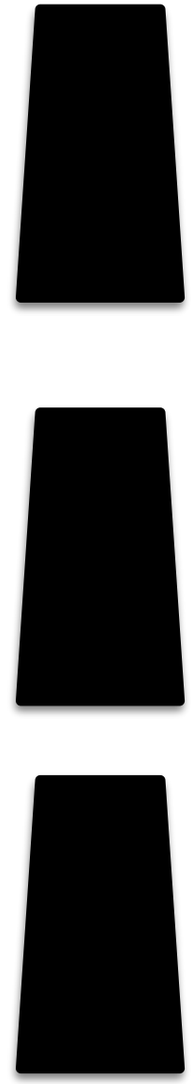
FALSE Loc(ball,1)

(Loc ball 1)

(Believes blinky (Loc ball 1))

Blinky





1

2

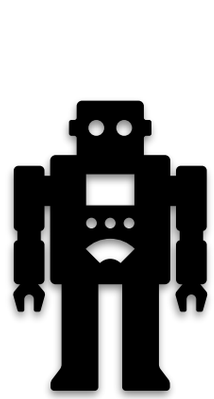
3

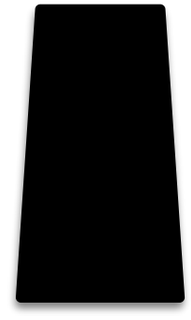
FALSE

(Loc ball 1)

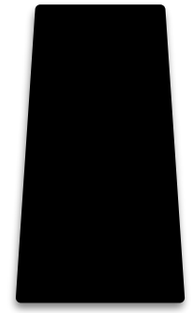
(Believes blinky (Loc ball 1))

Blinky

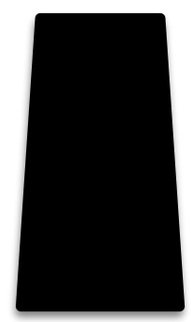




1

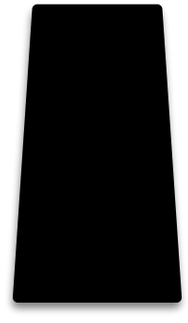


2



3

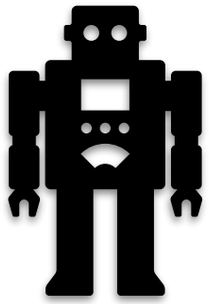
(Loc ball 1)

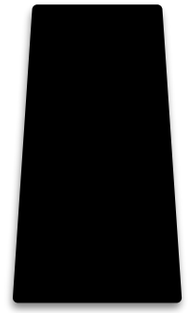


1

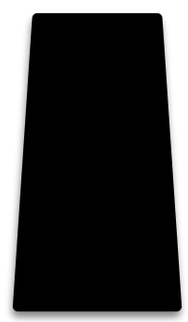
(Believes blinky (Loc ball 1))

Blinky





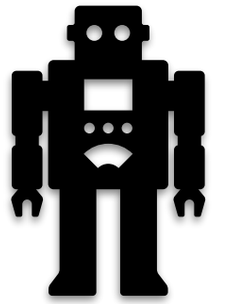
2

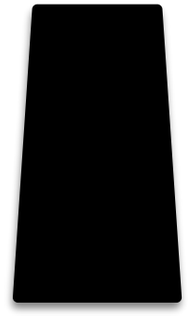


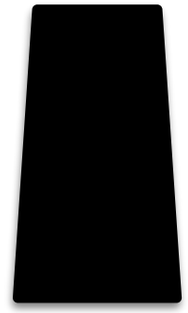
3

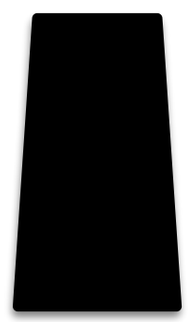
(Believes blinky (Loc ball 1))

Blinky









1

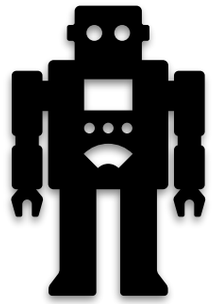
2

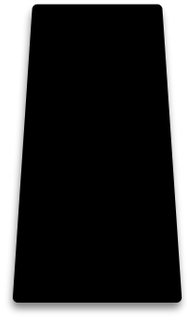
3

The ball is in the cup at location #1 or the ball is at location #3.

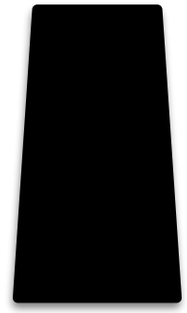
(Believes blinky (Loc ball 1))

Blinky

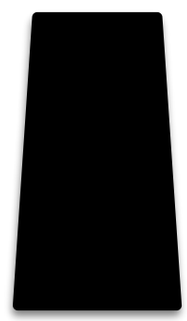




1



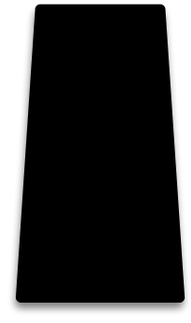
2



3

The ball is in the cup at location #1 or the ball is at location #3.

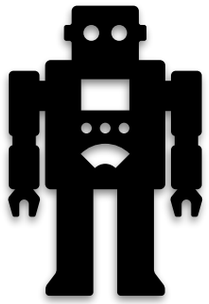
$$\mathrm{Loc}(\mathrm{ball},1) \vee \mathrm{Loc}(\mathrm{ball},3)$$

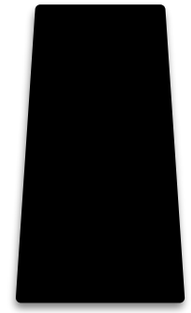


1

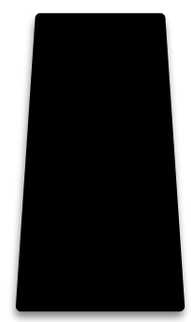
```
(Believes blinky (Loc ball 1))
```

Blinky





2



3

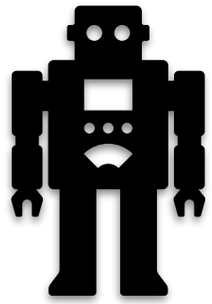
The ball is in the cup at location #1 or the ball is at location #3.

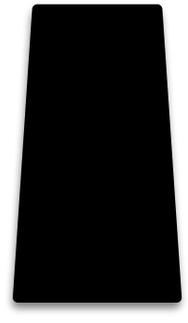
$$\mathbf{Loc}(\mathbf{ball},1) \vee \mathbf{Loc}(\mathbf{ball},3)$$

```
(or (Loc ball 1) (Loc ball 3))
```

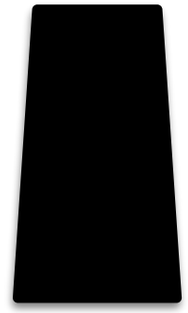
```
(Believes blinky (Loc ball 1))
```

Blinky

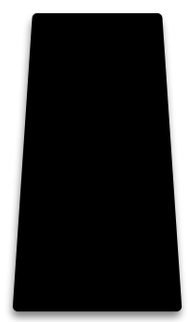




1



2



3

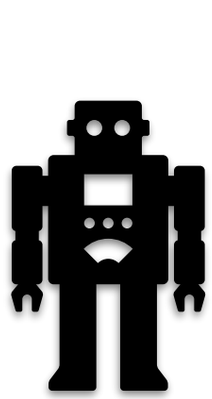
The ball is in the cup at location #1 or the ball is at location #3.

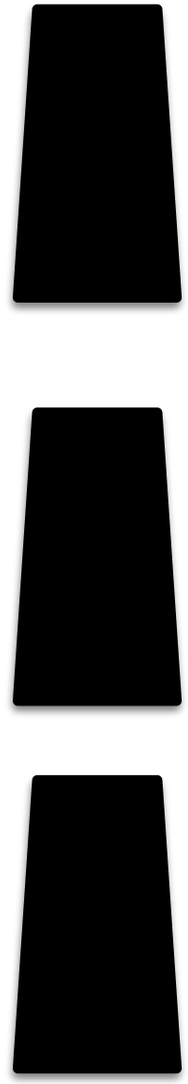
FALSE $\text{Loc}(\text{ball},1) \vee \text{Loc}(\text{ball},3)$

```
(or (Loc ball 1) (Loc ball 3))
```

```
(Believes blinky (Loc ball 1))
```

Blinky





1

2

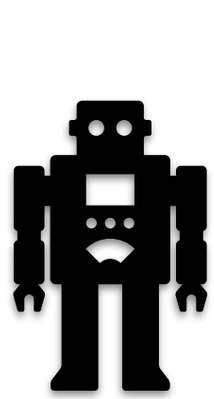
3

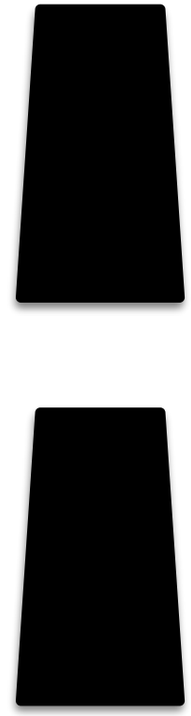
FALSE Loc(ball,1) ∨ Loc(ball,3)

```
(or (Loc ball 1) (Loc ball 3))
```

```
(Believes blinky (Loc ball 1))
```

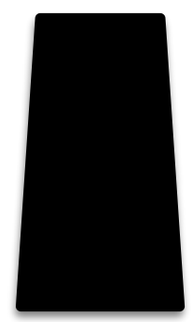
Blinky





1

2



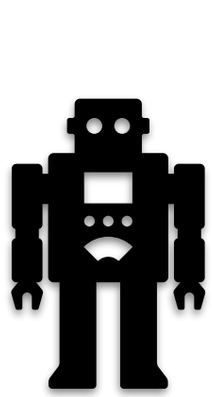
3

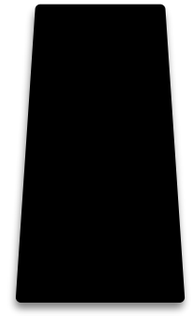
FALSE

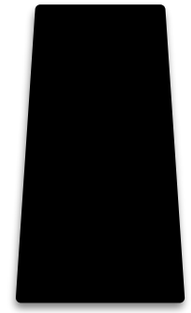
(or (Loc ball 1) (Loc ball 3))

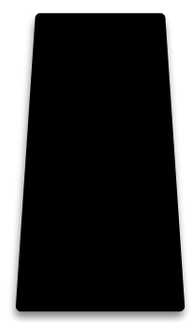
(Believes blinky (Loc ball 1))

Blinky









1

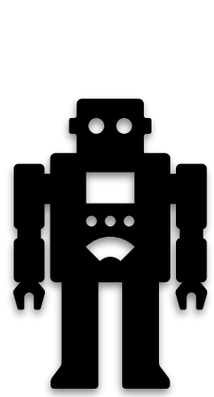
2

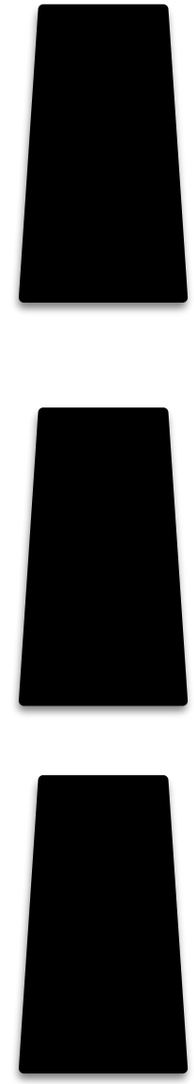
3

FALSE

(Believes blinky (Loc ball 1))

Blinky





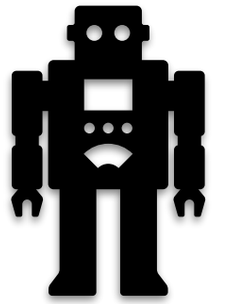
1

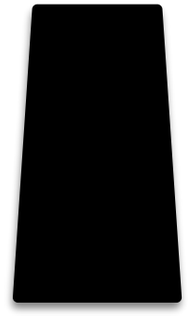
2

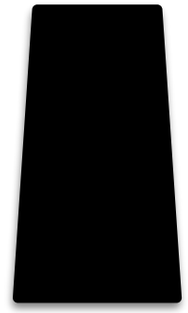
3

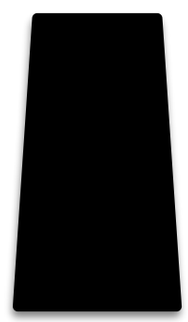
(Believes blinky (Loc ball 1))

Blinky









1

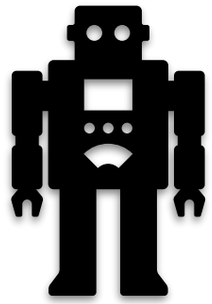
2

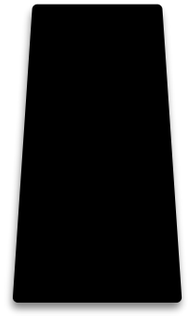
3

Blinky believes that the ball is in the cup at location #1.

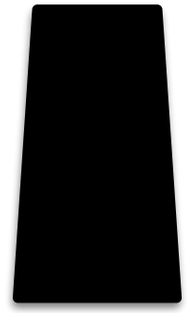
```
(Believes blinky (Loc ball 1))
```

Blinky

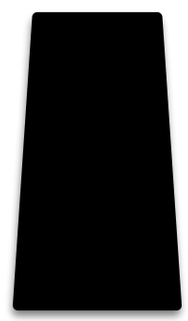




1



2



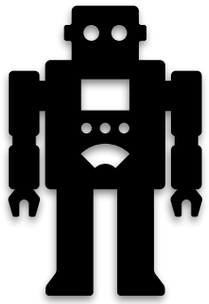
3

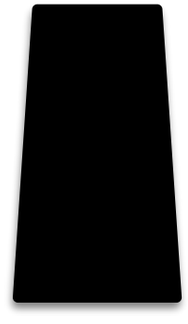
Blinky believes that the ball is in the cup at location #1.

$$\mathbf{B}(\text{blinky}, \text{Loc}(\text{ball}, 1))$$

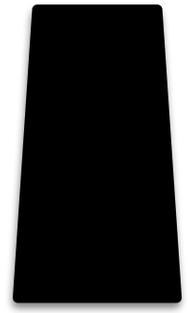
```
(Believes blinky (Loc ball 1))
```

Blinky

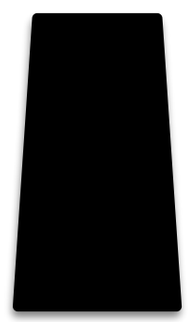




1



2



3

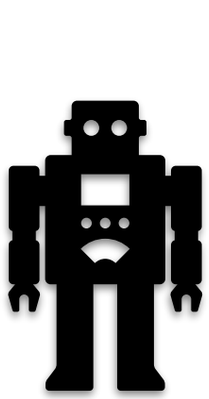
Blinky believes that the ball is in the cup at location #1.

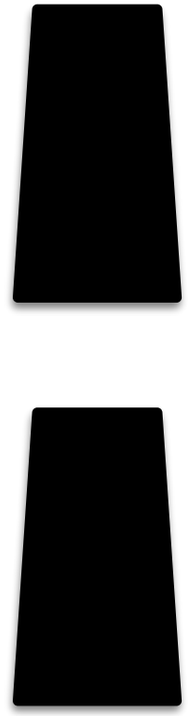
$$\mathbf{B}(\text{blinky}, \text{Loc}(\text{ball}, 1))$$

```
(Believes! blinky (Loc ball 1))
```

```
(Believes blinky (Loc ball 1))
```

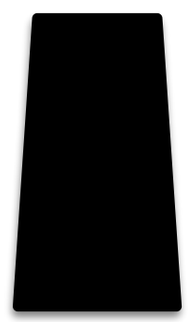
Blinky





1

2



3

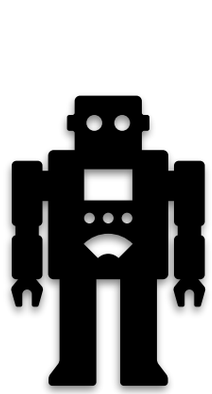
Blinky believes that the ball is in the cup at location #1.

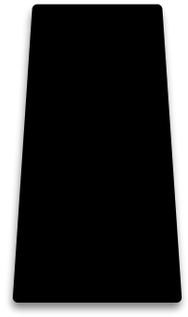
??? **B**(blinky, Loc(ball,1))

(Believes! blinky (Loc ball 1))

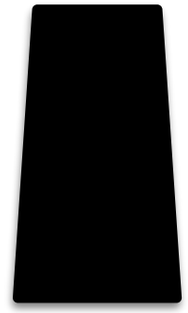
(Believes blinky (Loc ball 1))

Blinky

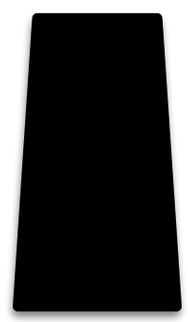




1



2



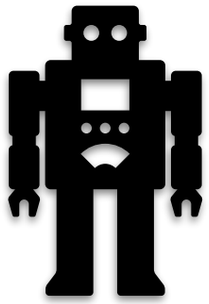
3

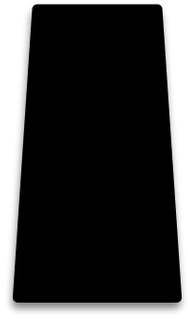
Blinky believes that the ball is in the cup at location #1.

??? $\mathbf{B}(\text{blinky}, \text{Loc}(\text{ball}, 1))$

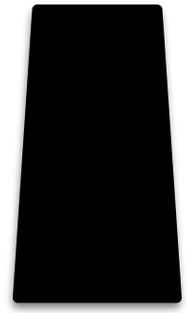
```
(Believes! blinky (Loc ball 1))
```

```
(Believes blinky (Loc ball 1))
```

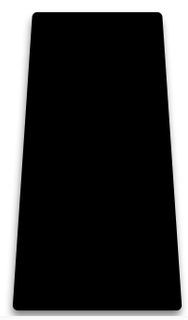
Blinky 



1

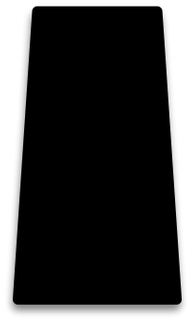


2



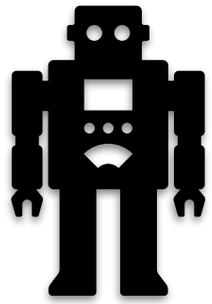
3

In extensional logics, what is denoted is conflated with meaning (the latter being naïvely compositional), but intensional attitudes like *believes*, *knows*, *hopes*, *fears*, etc cannot be represented and reasoned over smoothly (e.g. without fear of inconsistency rising up).

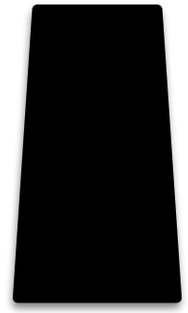


1

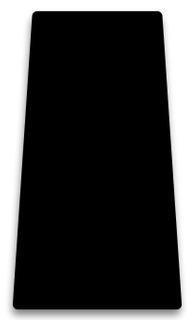
Blinky





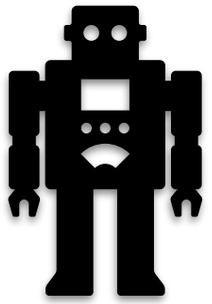


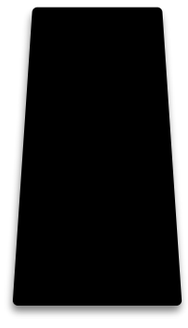
2



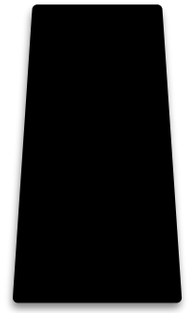
3

Blinky

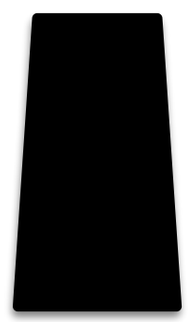




1



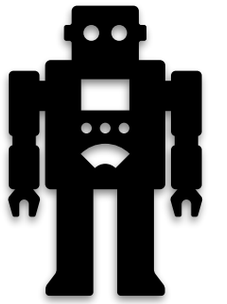
2

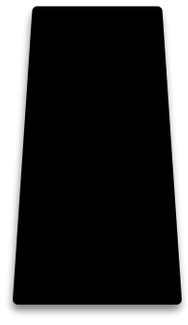


3

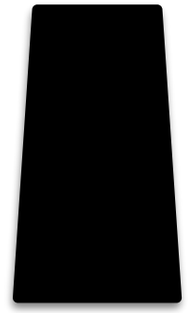


Blinky

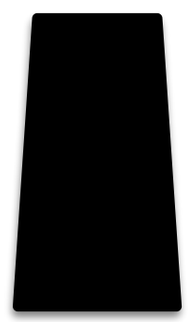




1

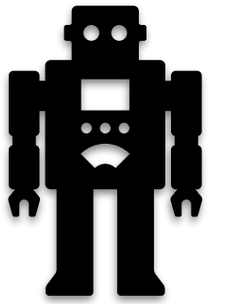


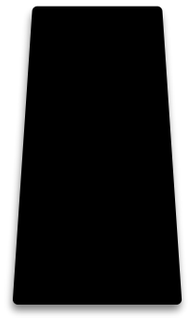
2



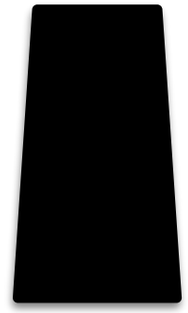
3

Blinky

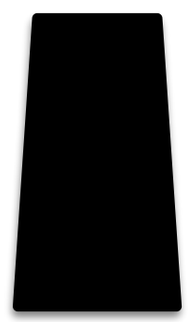




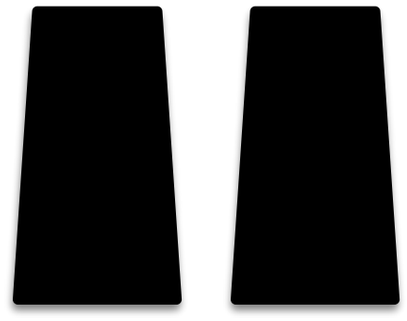
1

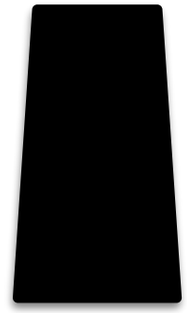


2



3

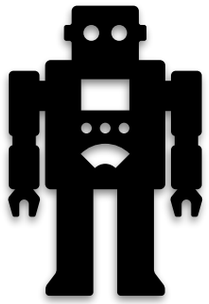


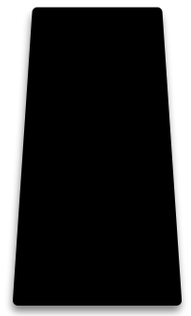


2

3

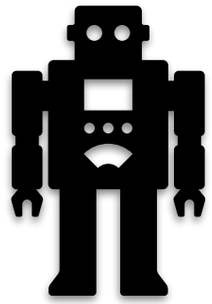
Blinky

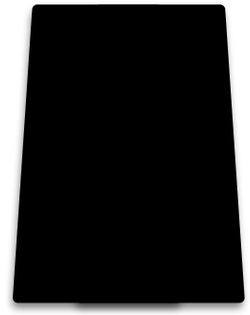




1

Blinky

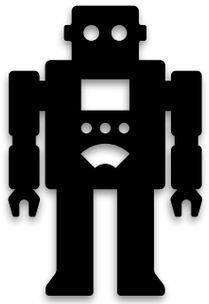


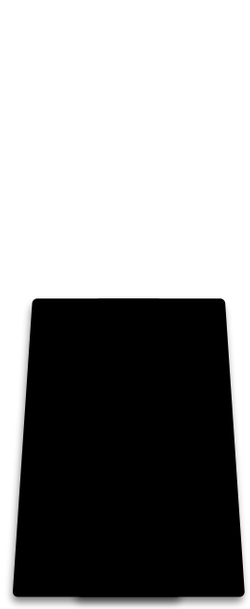


2

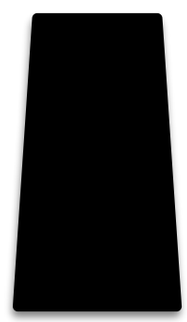
3

Blinky

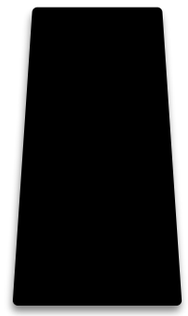




2

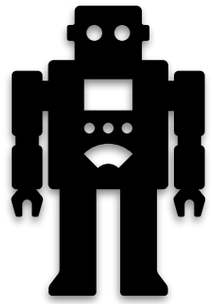


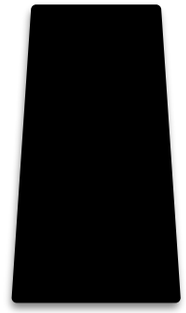
3



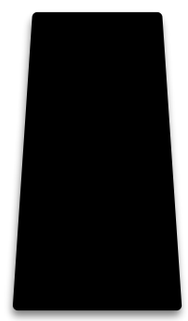
1

Blinky

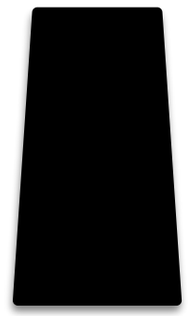




2

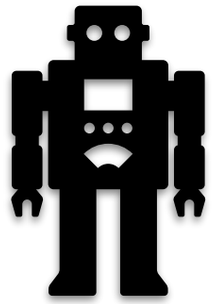


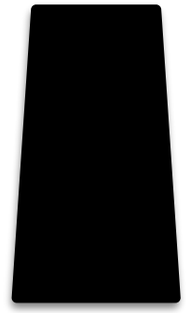
3



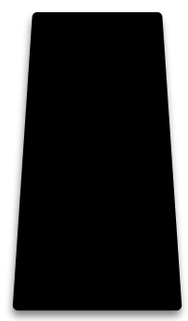
1

Blinky



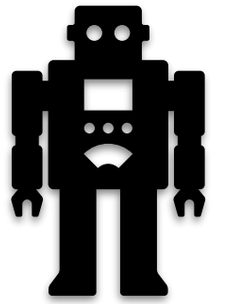


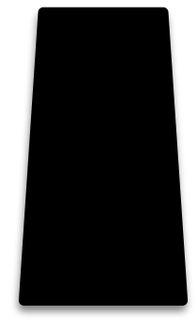
2

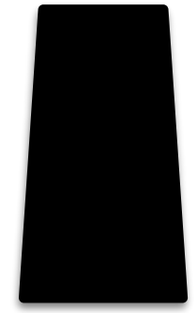


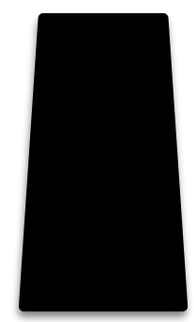
3

Blinky









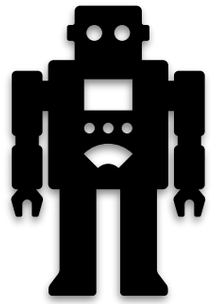
1

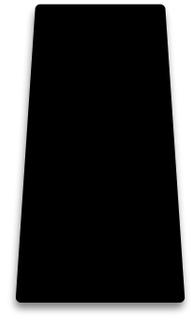
2

3

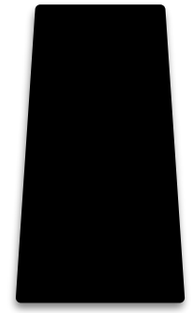
Blinky believes that the ball is in the cup at location #1.

Blinky

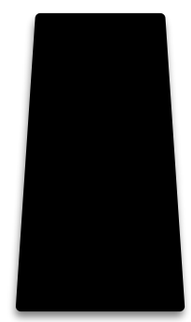




1



2

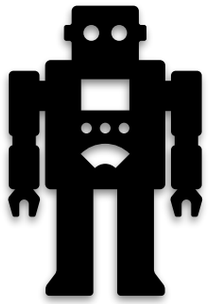


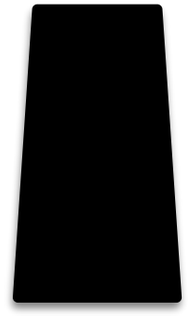
3

Blinky believes that the ball is in the cup at location #1.

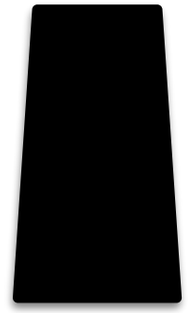
B(blinky, loc-ball-1)

Blinky

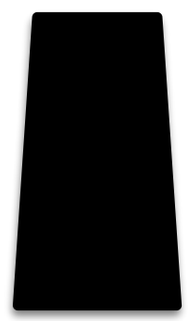




1



2



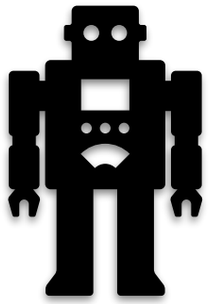
3

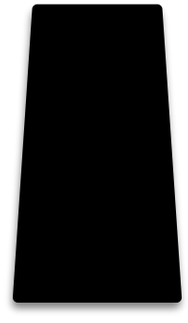
Blinky believes that the ball is in the cup at location #1.

B(blinky, loc-ball-1)

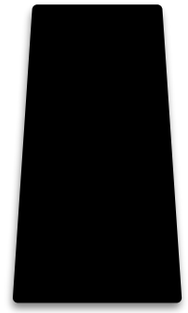
`(Believes! blinky loc-ball-1)`

Blinky

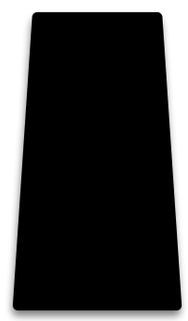




1



2

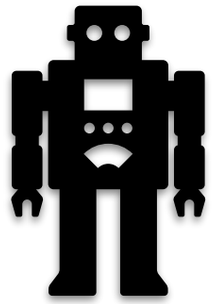


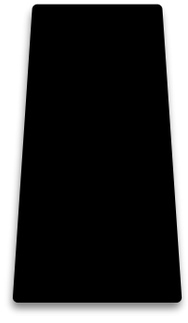
3

Blinky believes that the ball is in the cup at location #1.

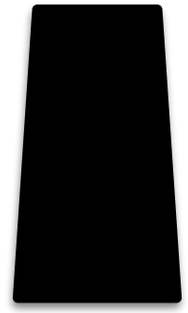
B(blinky, loc-ball-1)

`(Believes! blinky loc-ball-1)`

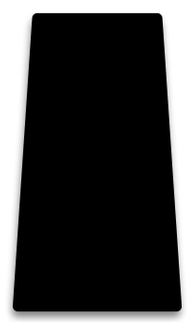
Blinky 



1



2



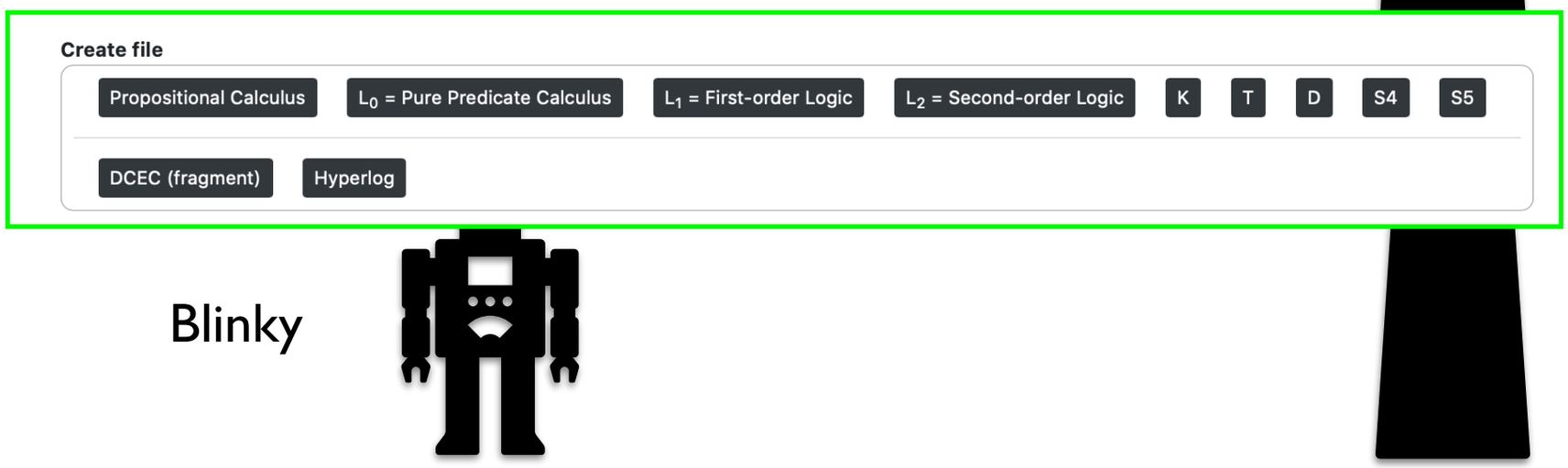
3

In intensional logics, meaning and designation are separated, and compositionality is abandoned.

Blinky believes that the ball is in the cup at location #1.

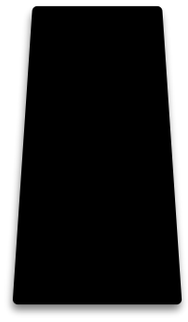
B(blinky, loc-ball-1)

(Believes! blinky loc-ball-1)



Blinky

2



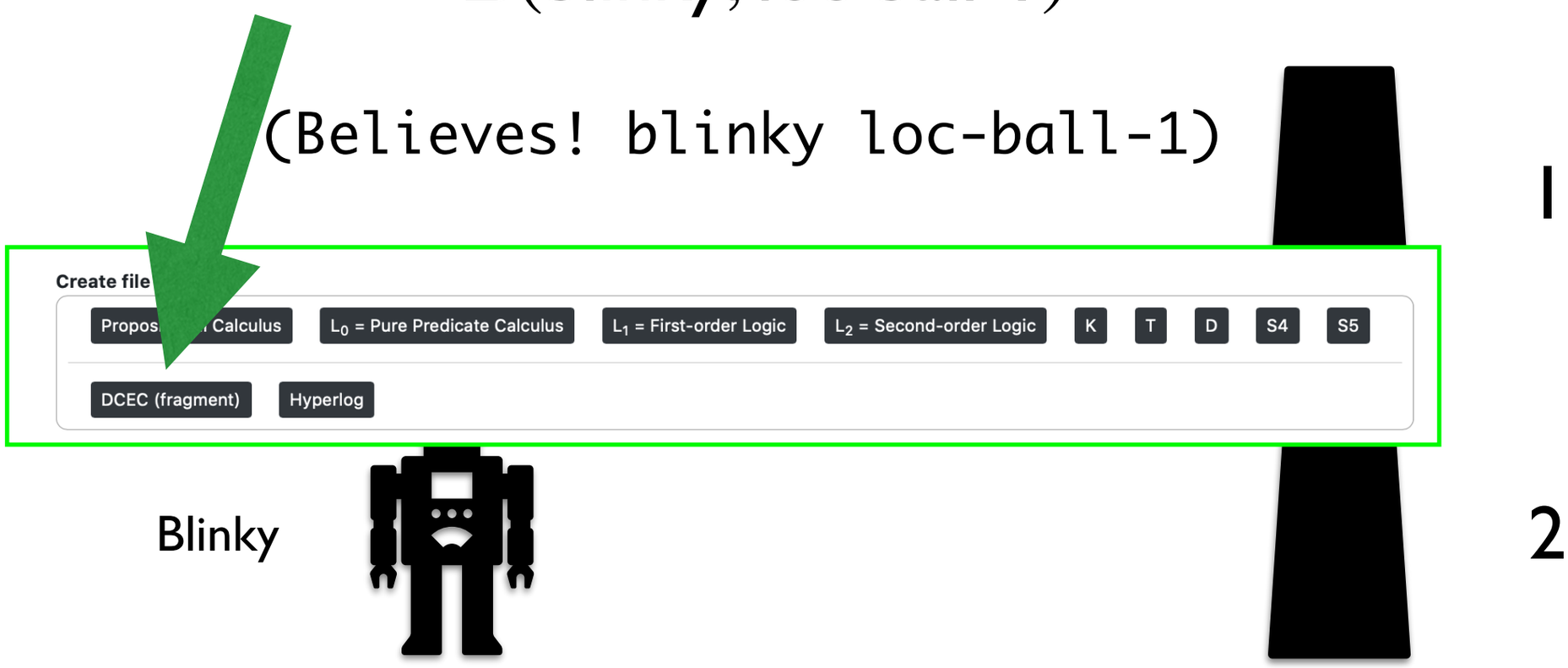
3

In intensional logics, meaning and designation are separated, and compositionality is abandoned.

Blinky believes that the ball is in the cup at location #1.

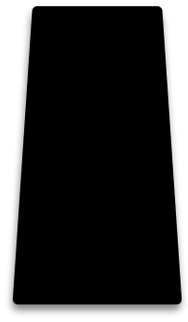
B(blinky, loc-ball-1)

(Believes! blinky loc-ball-1)



Blinky

2



3

In intensional logics, meaning and designation are separated, and compositionality is abandoned.

For Brave Adventurers

For Brave Adventurers

"Everything smart knows that everything tinks anything that tinks something identical with something."

For Brave Adventurers

"Everything smart knows that everything tinks anything that tinks something identical with something."

"Blinky is smart."

For Brave Adventurers

"Everything smart knows that everything tinks anything that tinks something identical with something."

"Blinky is smart."

Therefore:

For Brave Adventurers

"Everything smart knows that everything tinks anything that tinks something identical with something."

"Blinky is smart."

Therefore:

"Everything tinks anything that tinks something identical with something."

For Brave Adventurers

"Everything smart knows that everything tinks anything that tinks something identical with something."

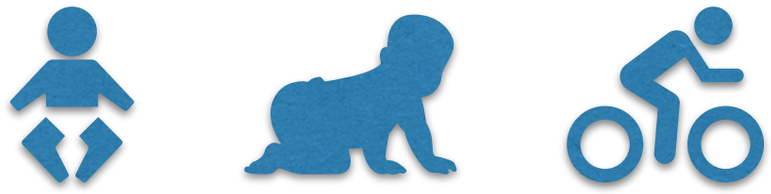
"Blinky is smart."

Therefore:

"Everything tinks anything that tinks something identical with something."

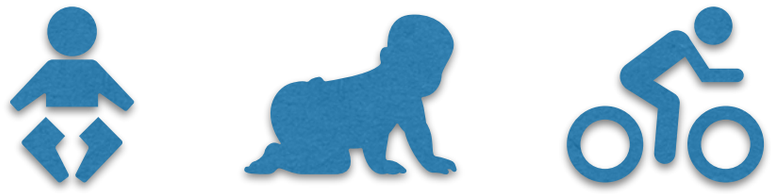
False Belief Task Demands Intensional Logic …

False Belief Task Demands Intensional Logic …

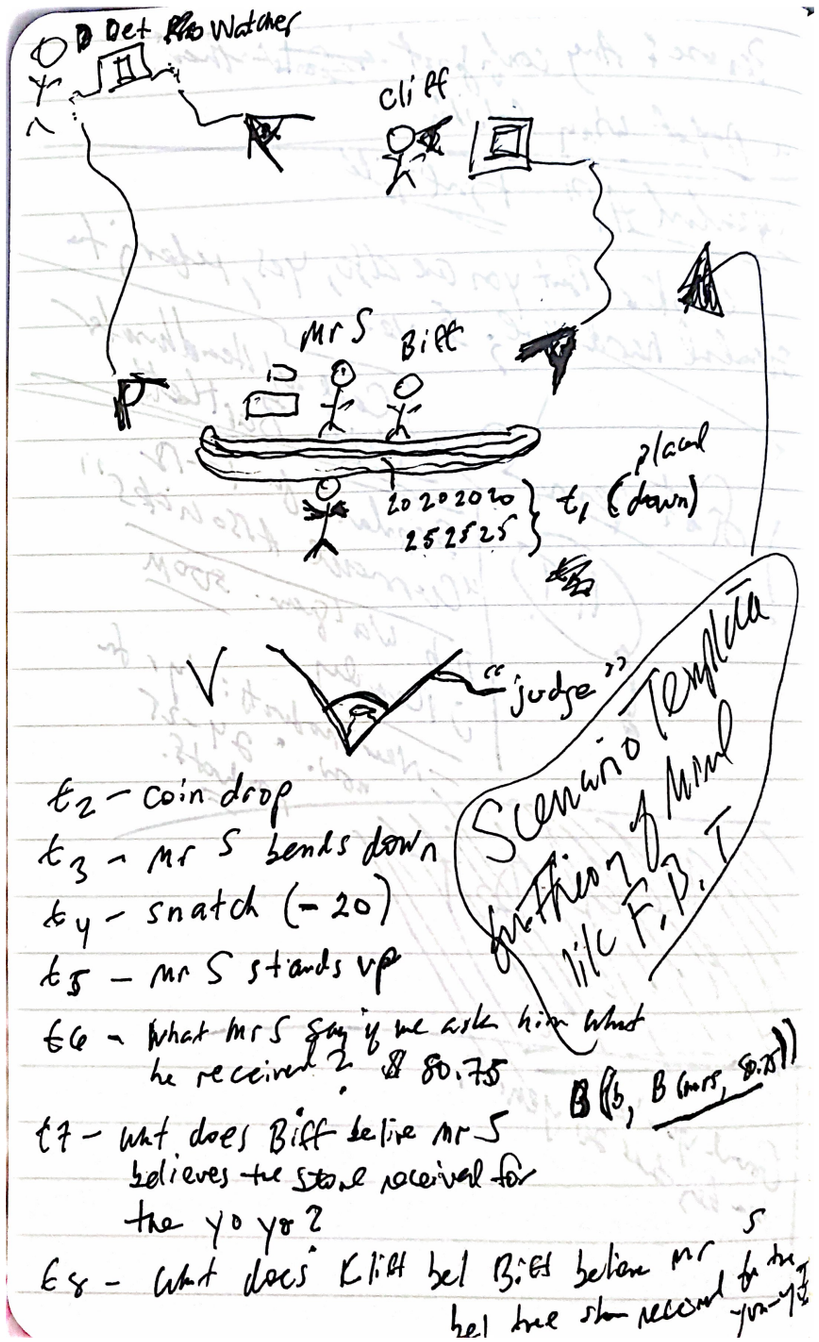


False Belief Task Demands Intensional Logic …

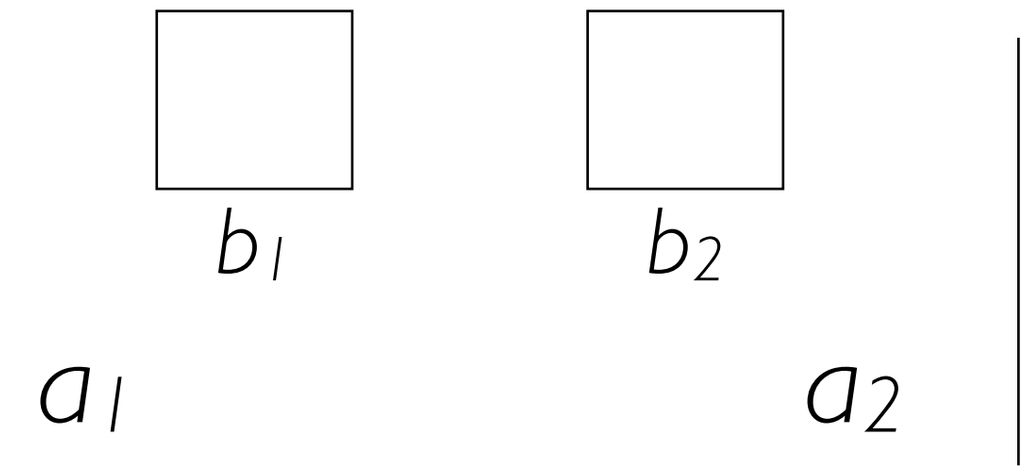
x x √



Better, But Embryonic: The ToM Pawn Shop



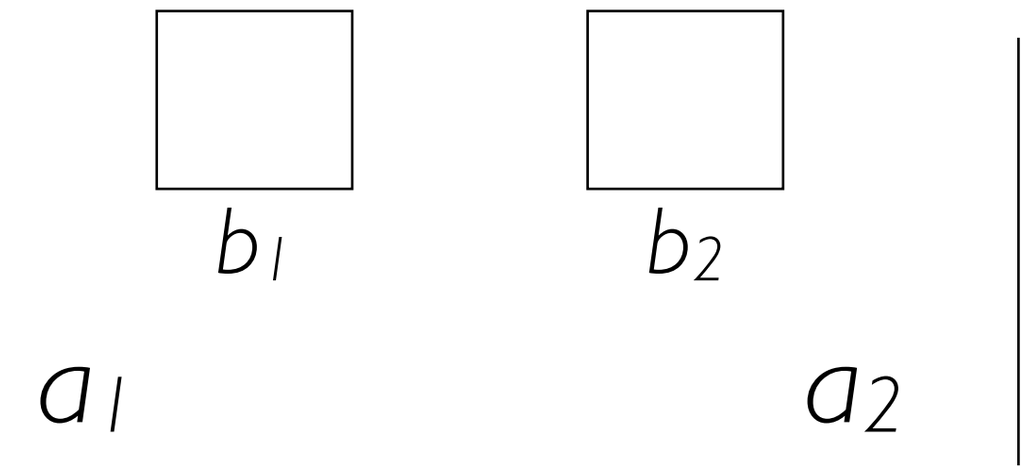
Framework for FBT^0_I



a

e

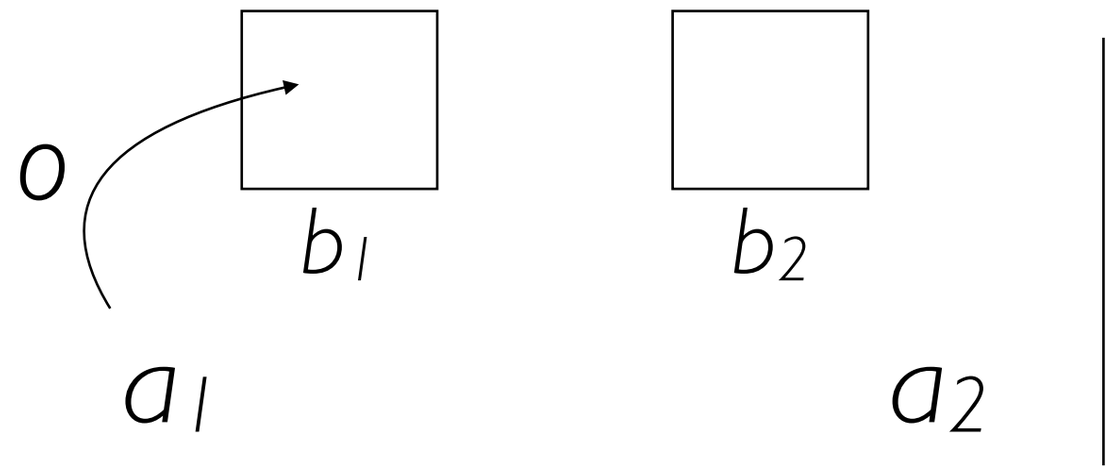
Framework for FBT^0_1
(five timepoints)



a

e

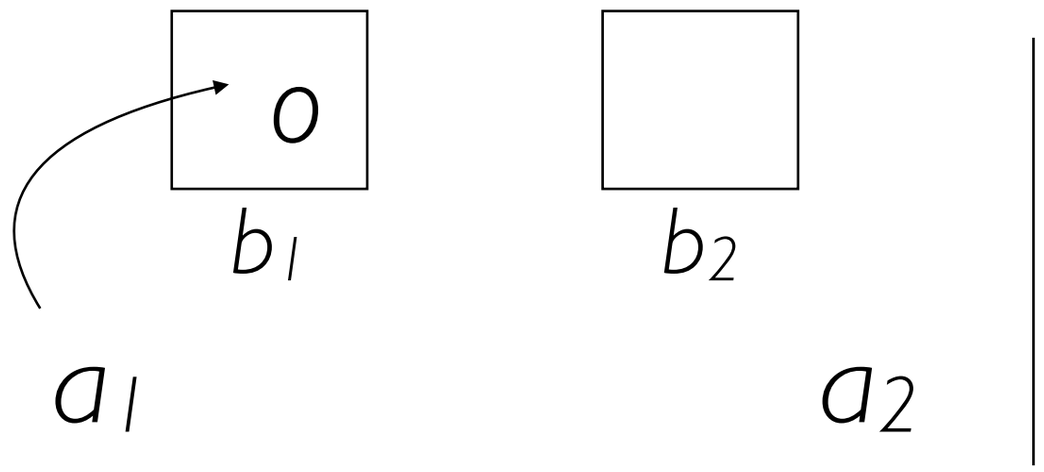
Framework for FBT$^0_1$
(five timepoints)



a

e

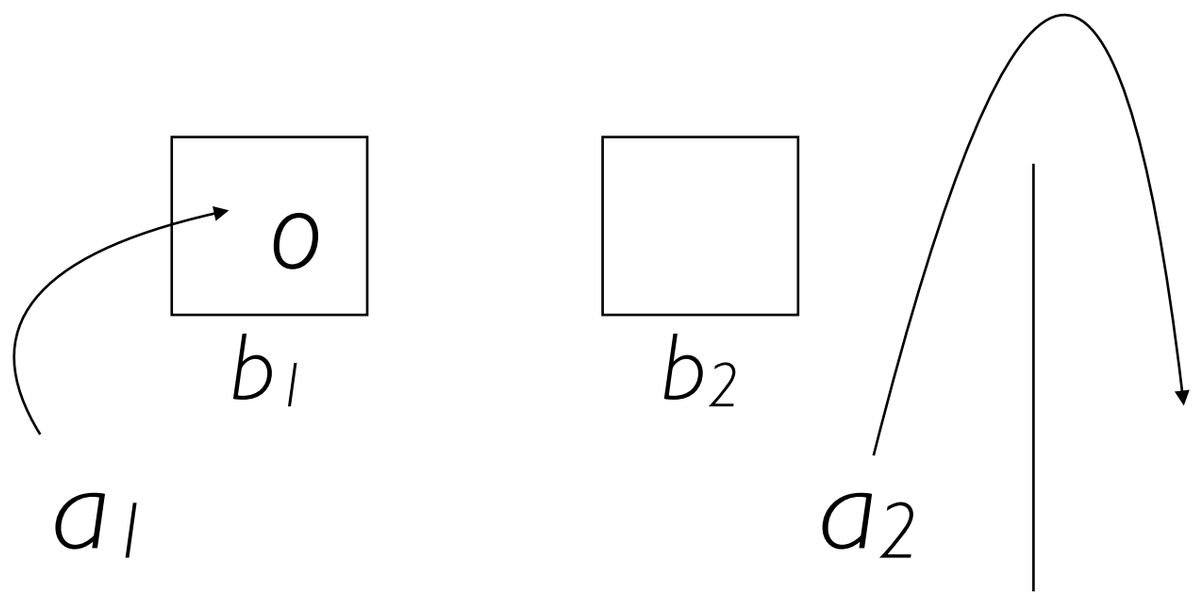
Framework for FBT$^0_1$
(five timepoints)



a

e

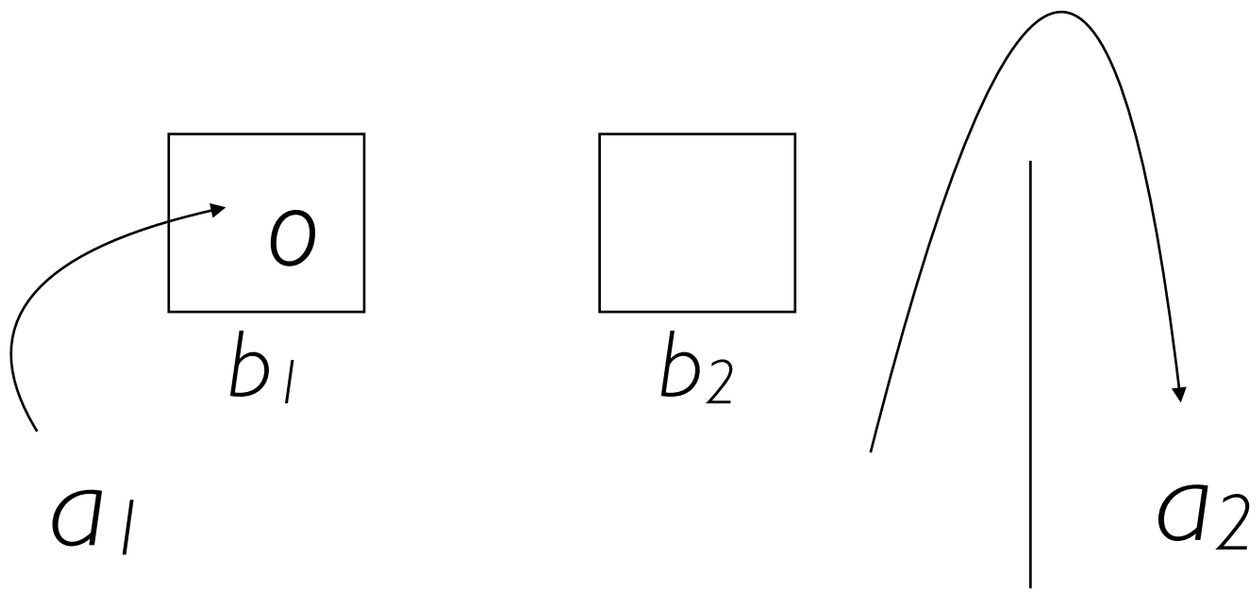
Framework for FBT^0_1
(five timepoints)



a

e

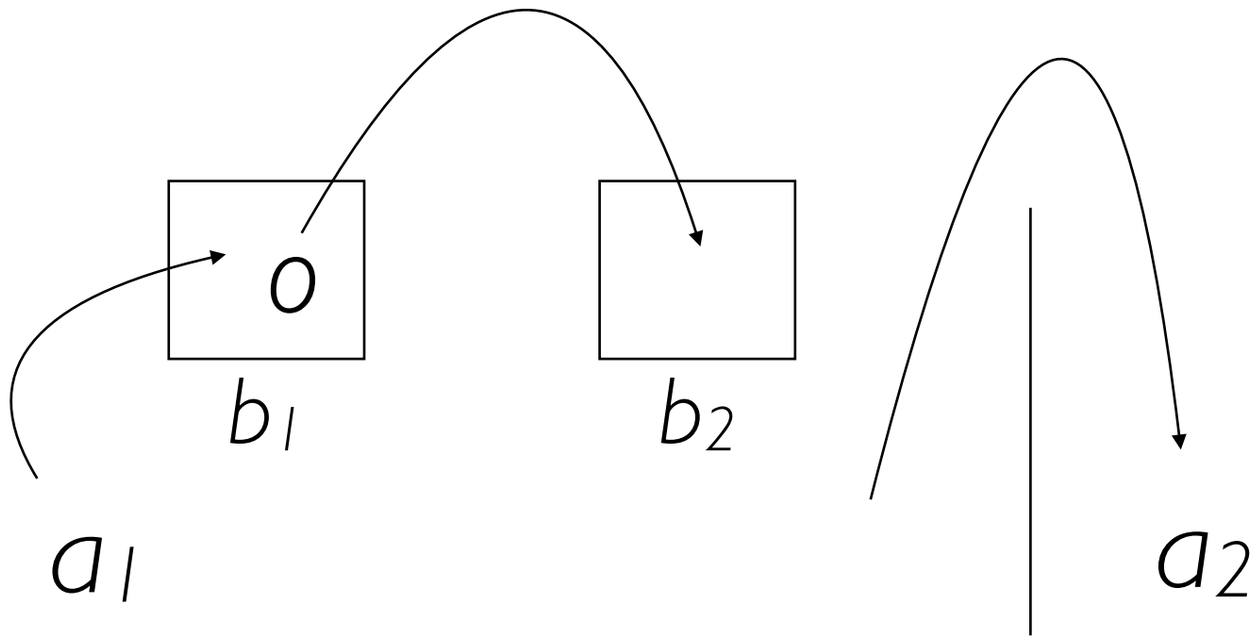
Framework for FBT$^0_1$
(five timepoints)



a

e

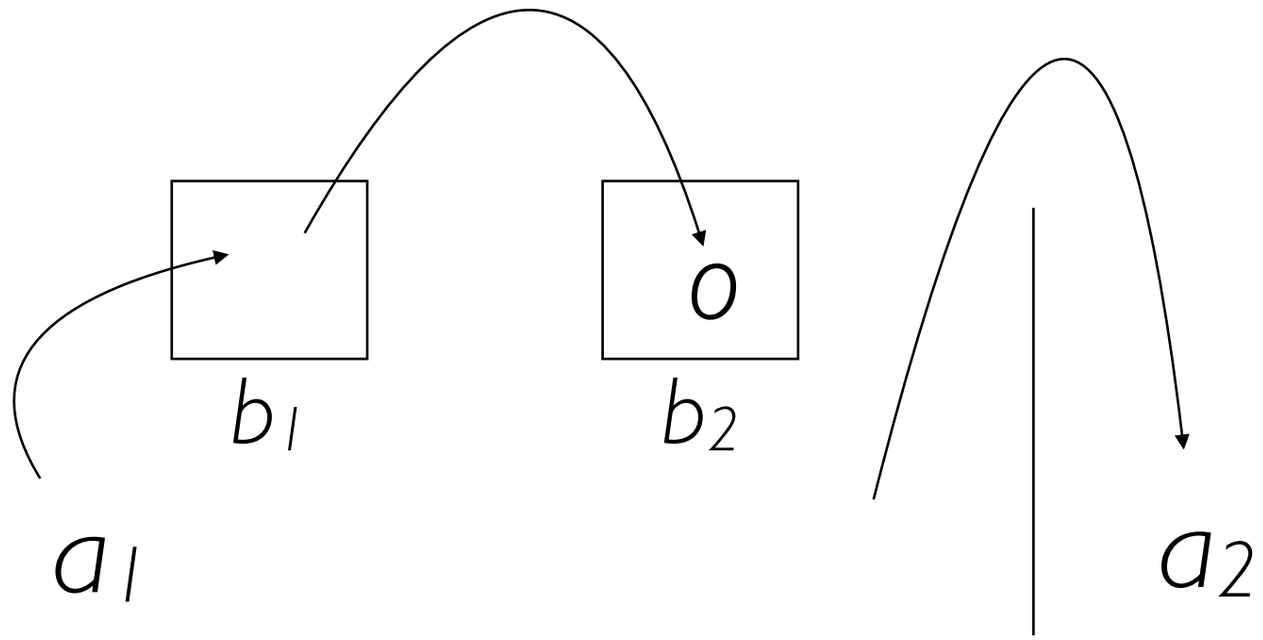
Framework for FBT$^0_1$
(five timepoints)



a

e

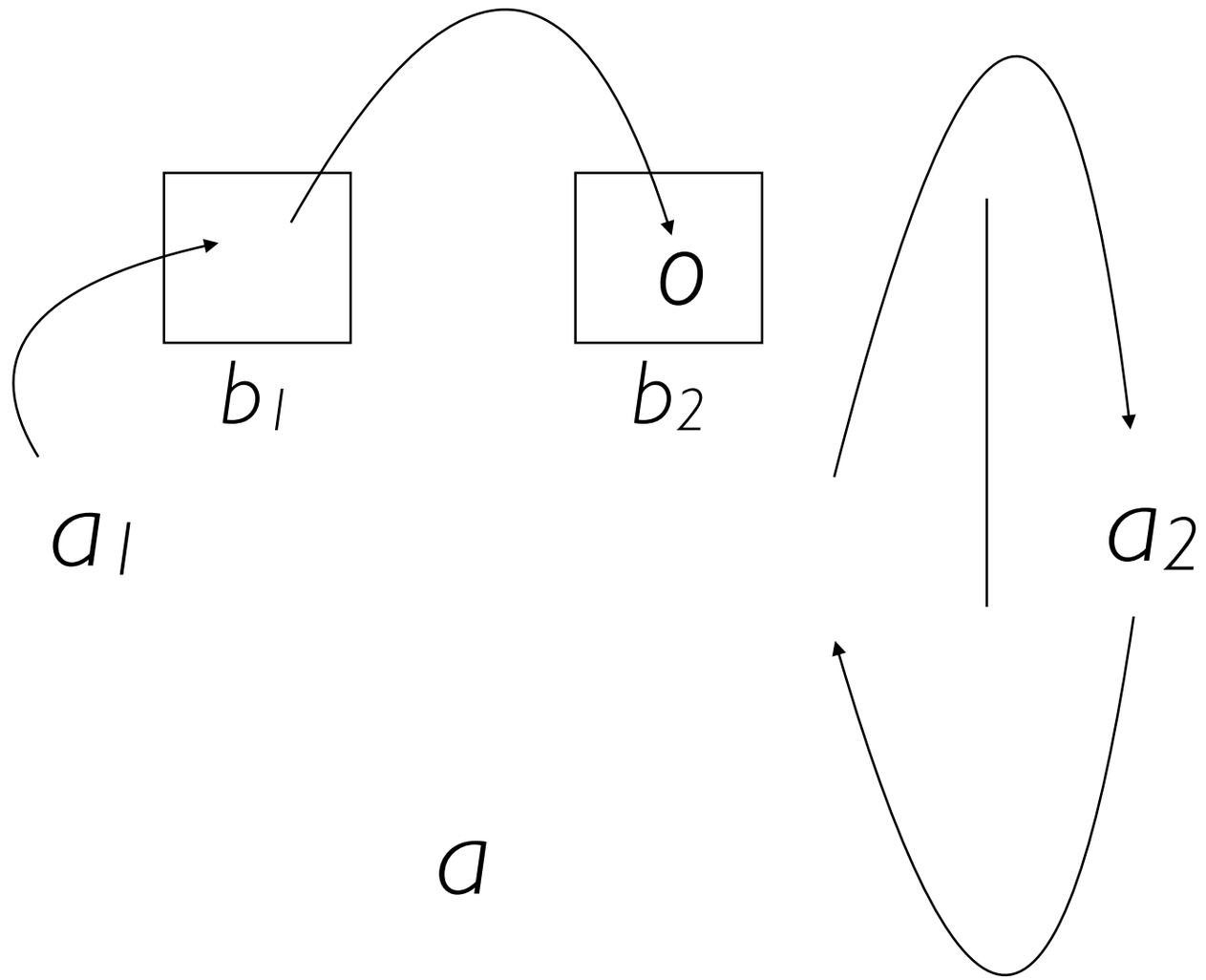
Framework for FBT$^0_1$
(five timepoints)



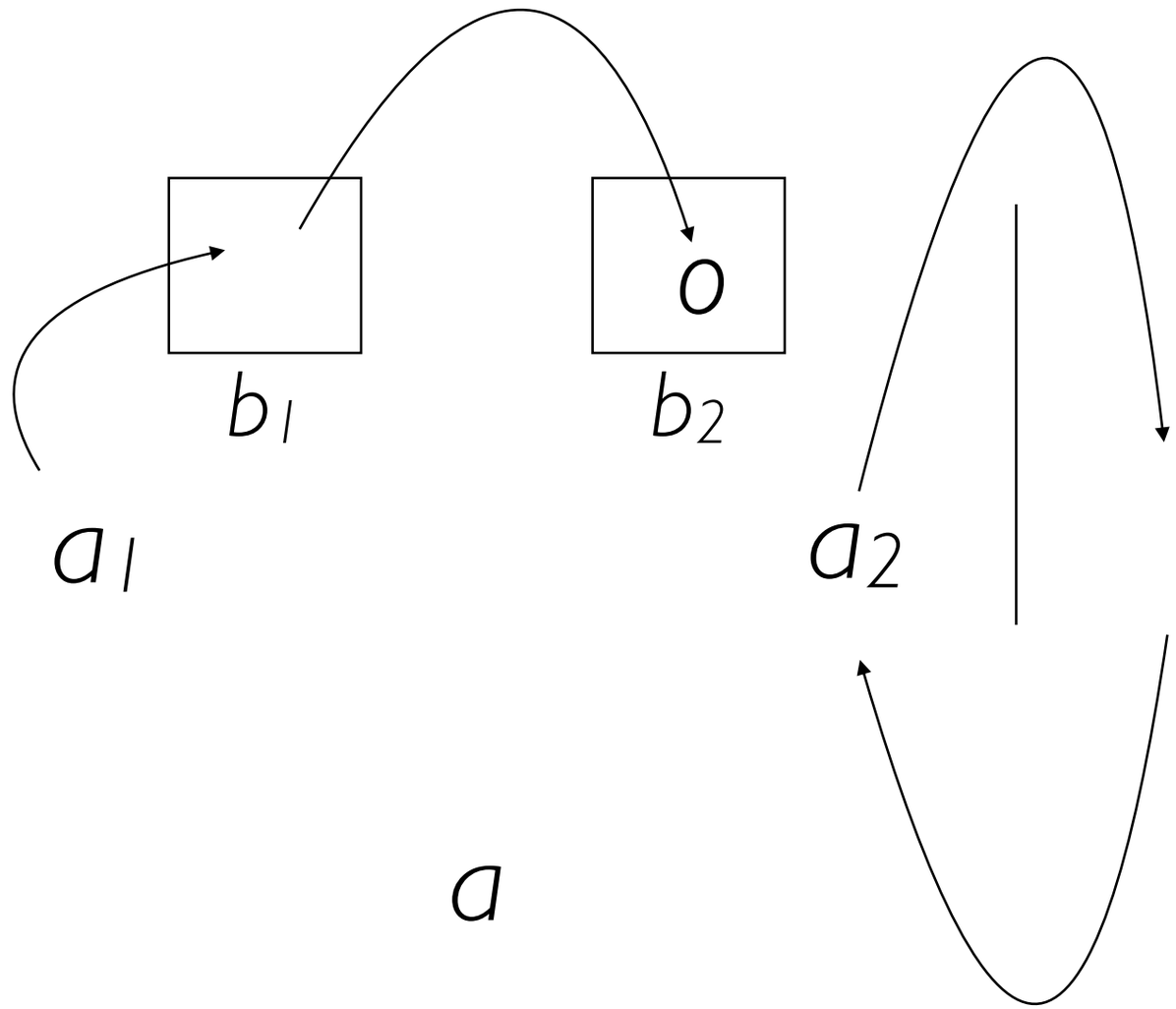
a

e

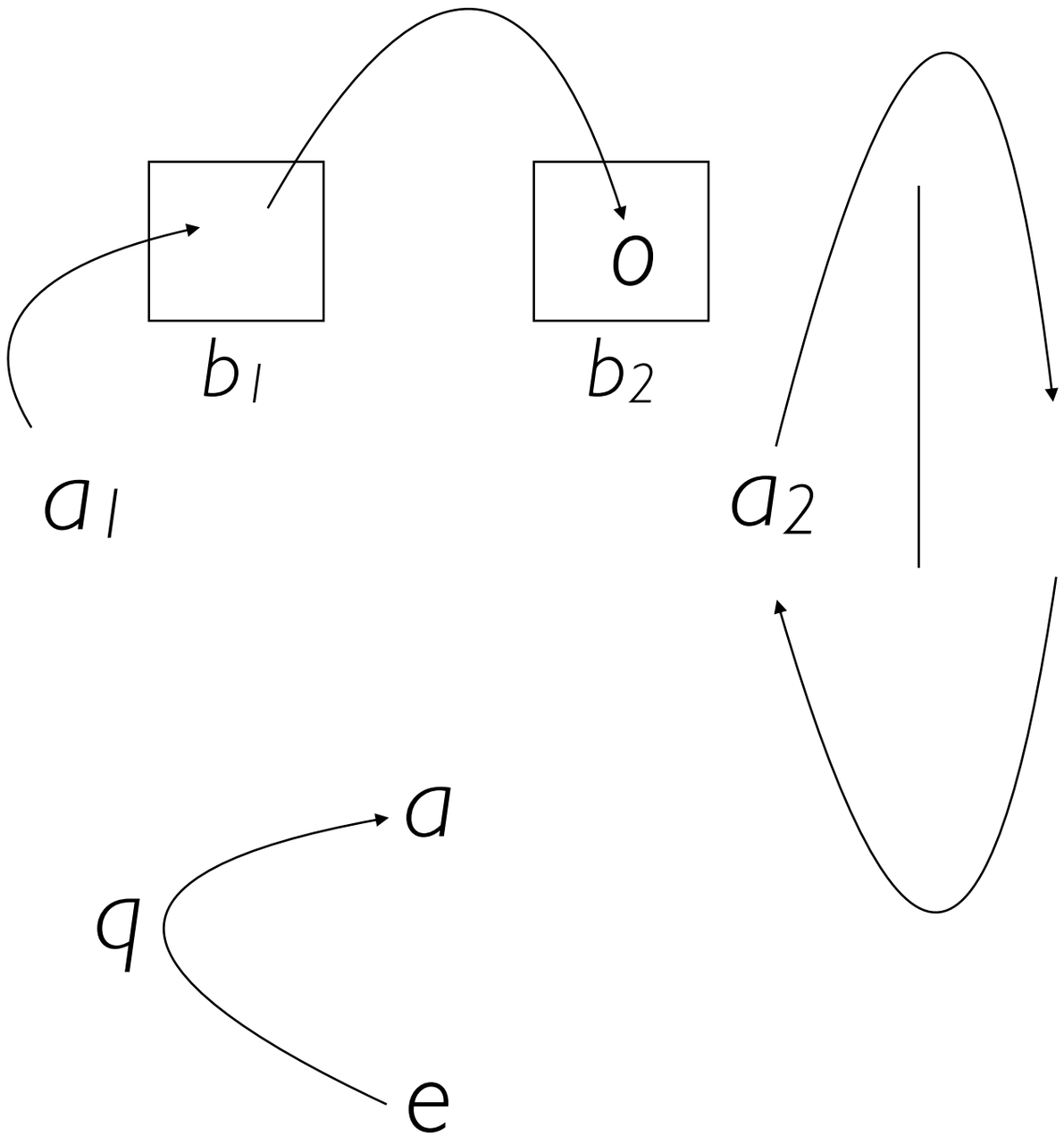
Framework for FBT^0_1
(five timepoints)



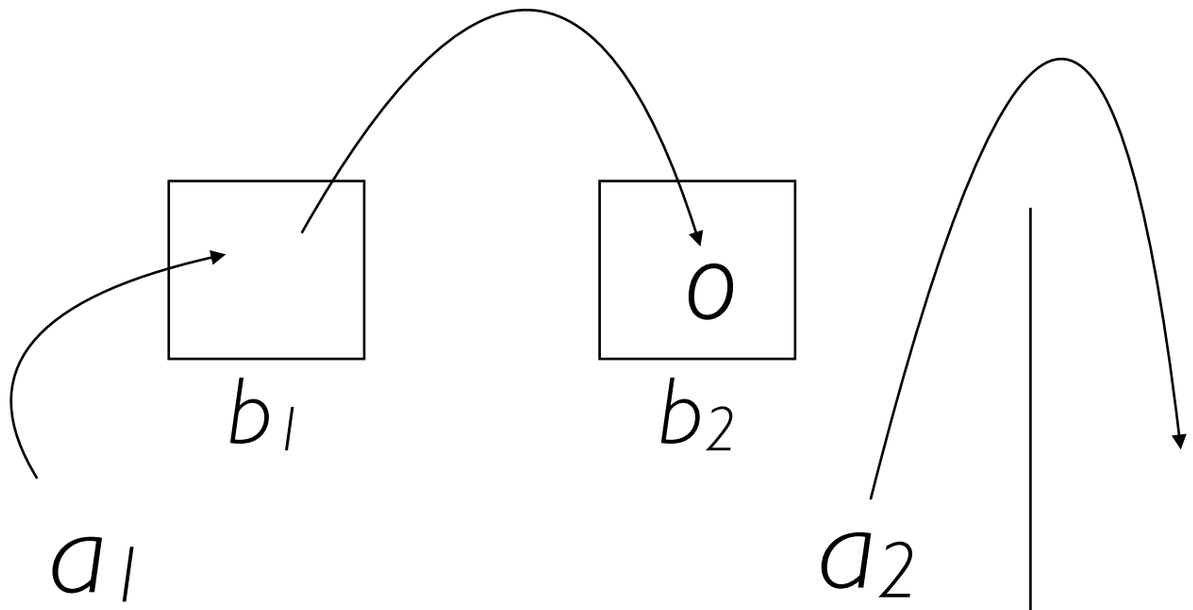
Framework for FBT^0_1
(five timepoints)



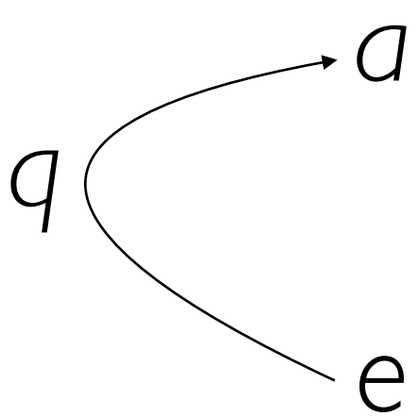
Framework for FBT^0_1
(five timepoints)



Framework for FBT^0_1
(five timepoints)

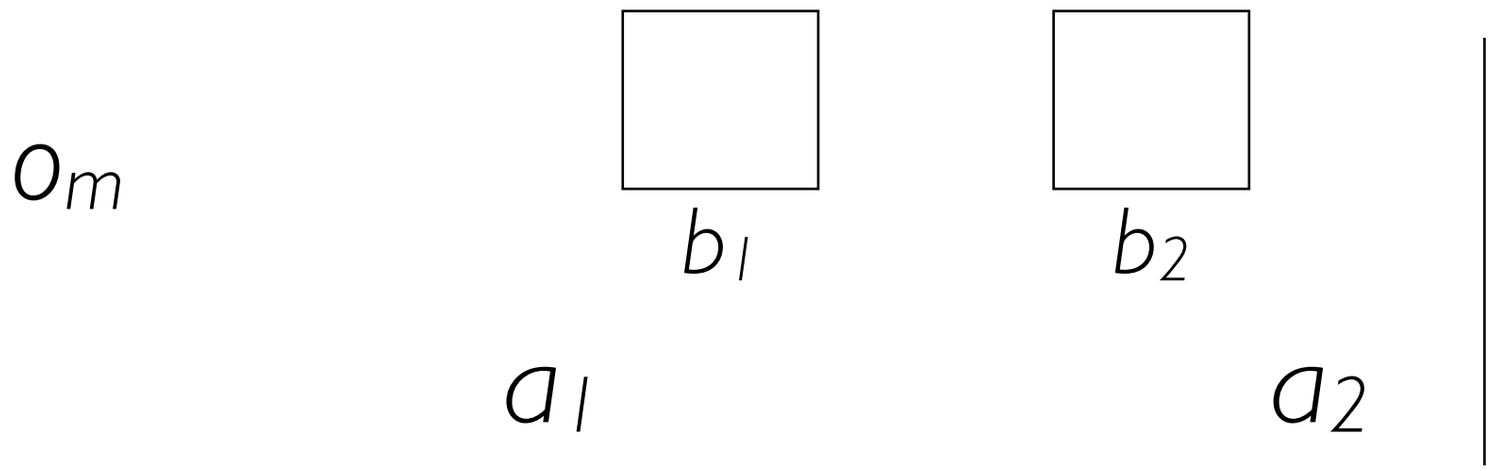


q a formula in modal $\mathscr{L}^n$



Framework for FBT$^{I}_{I}$

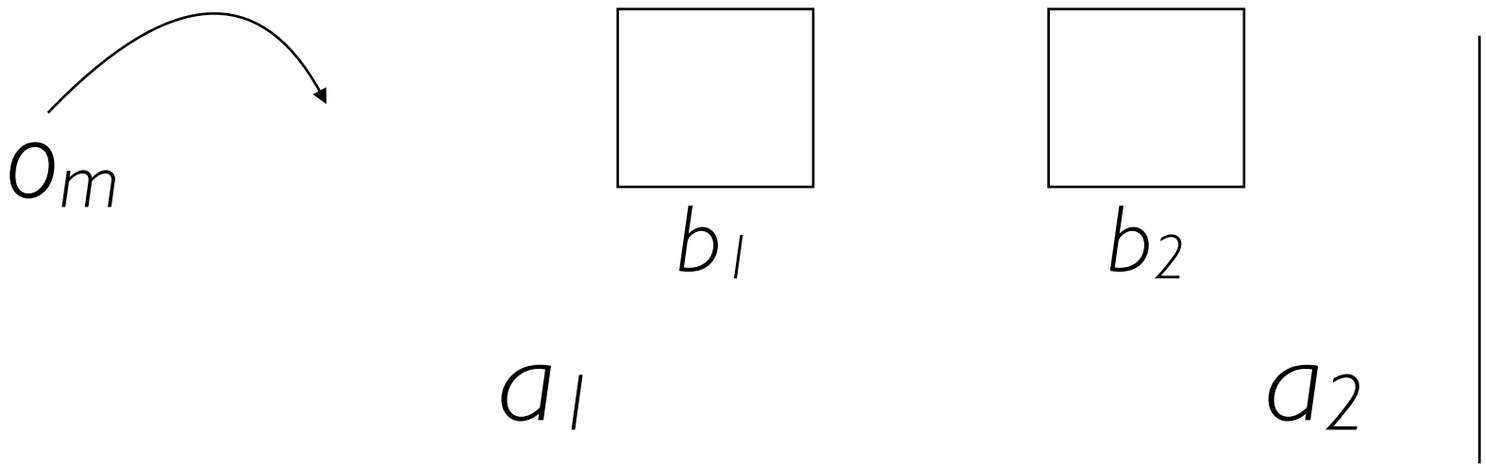
Framework for FBT[I]
(six timepoints)



a

e

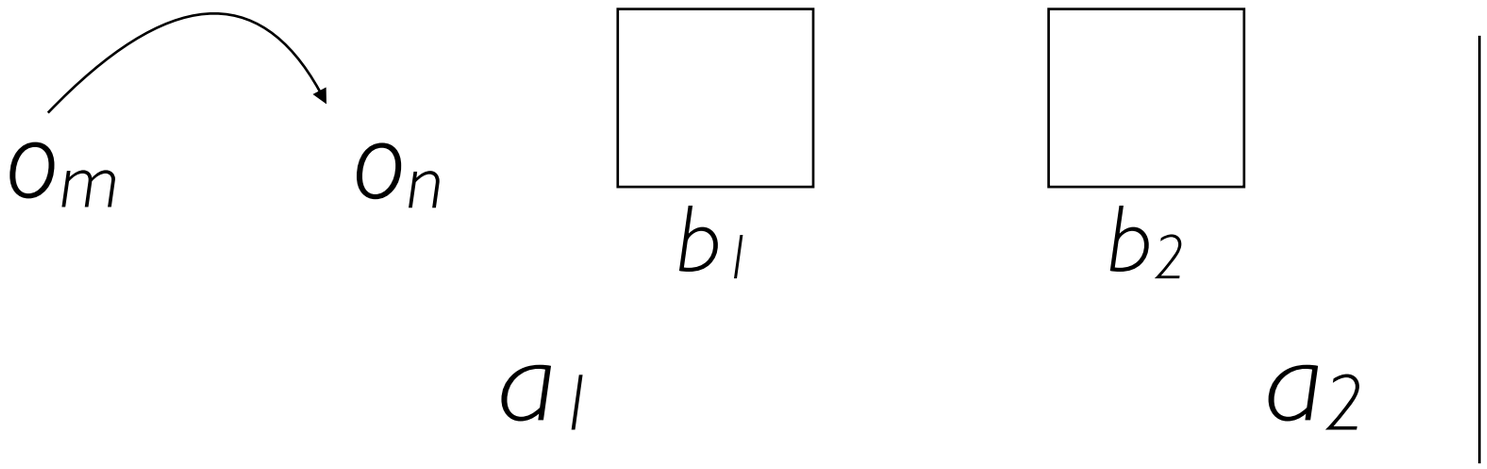
Framework for FBT[1]
(six timepoints)



a

e

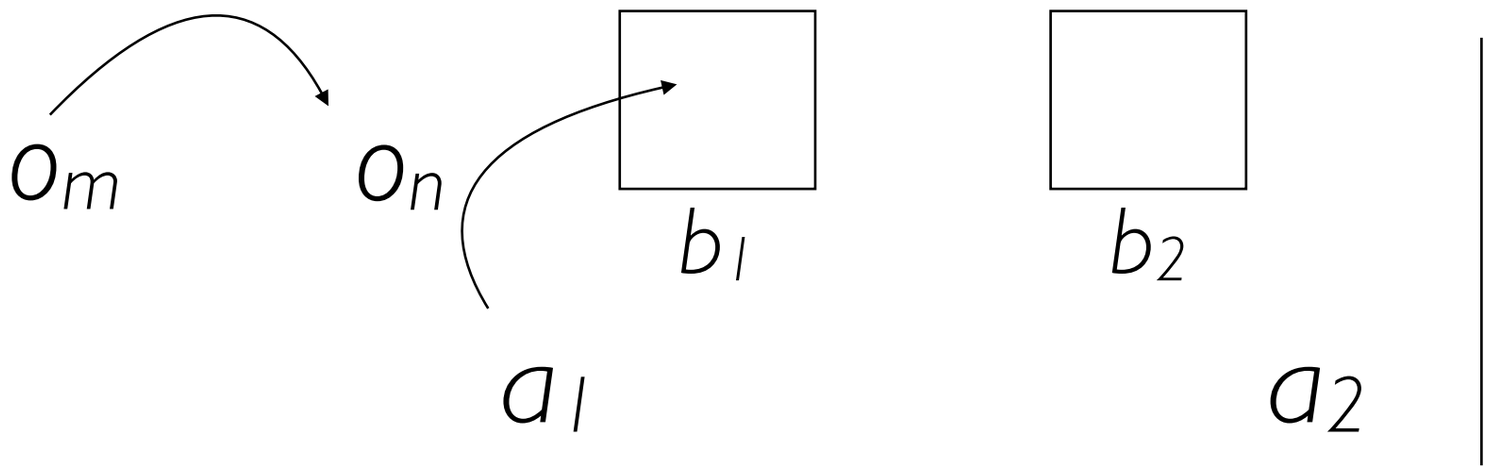
Framework for FBT[I]
(six timepoints)



a

e

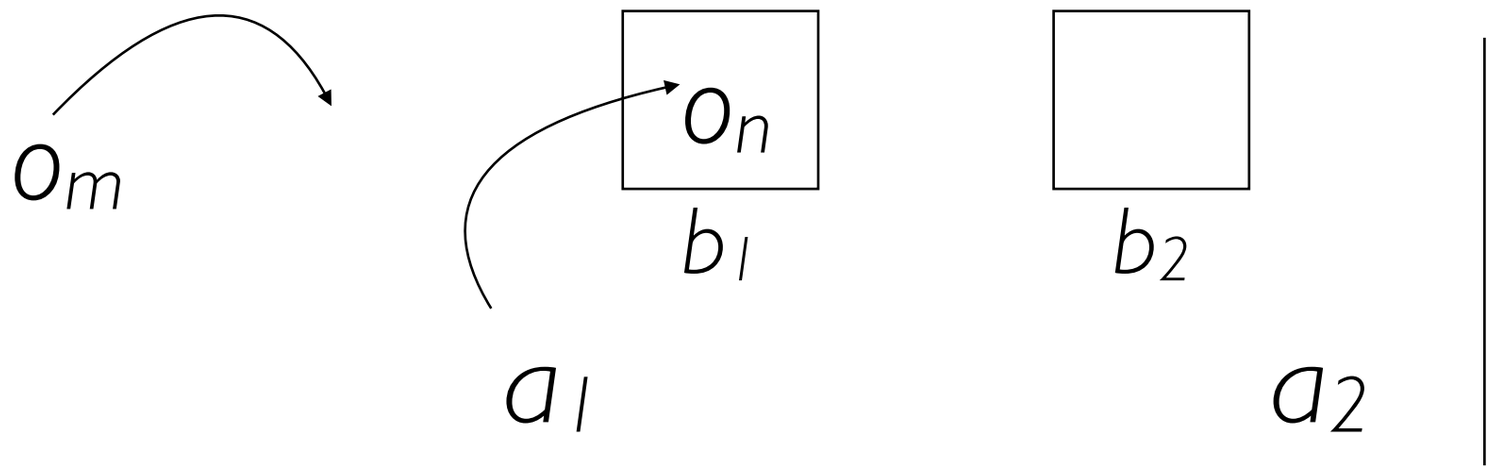
Framework for FBT[I]
(six timepoints)



a

e

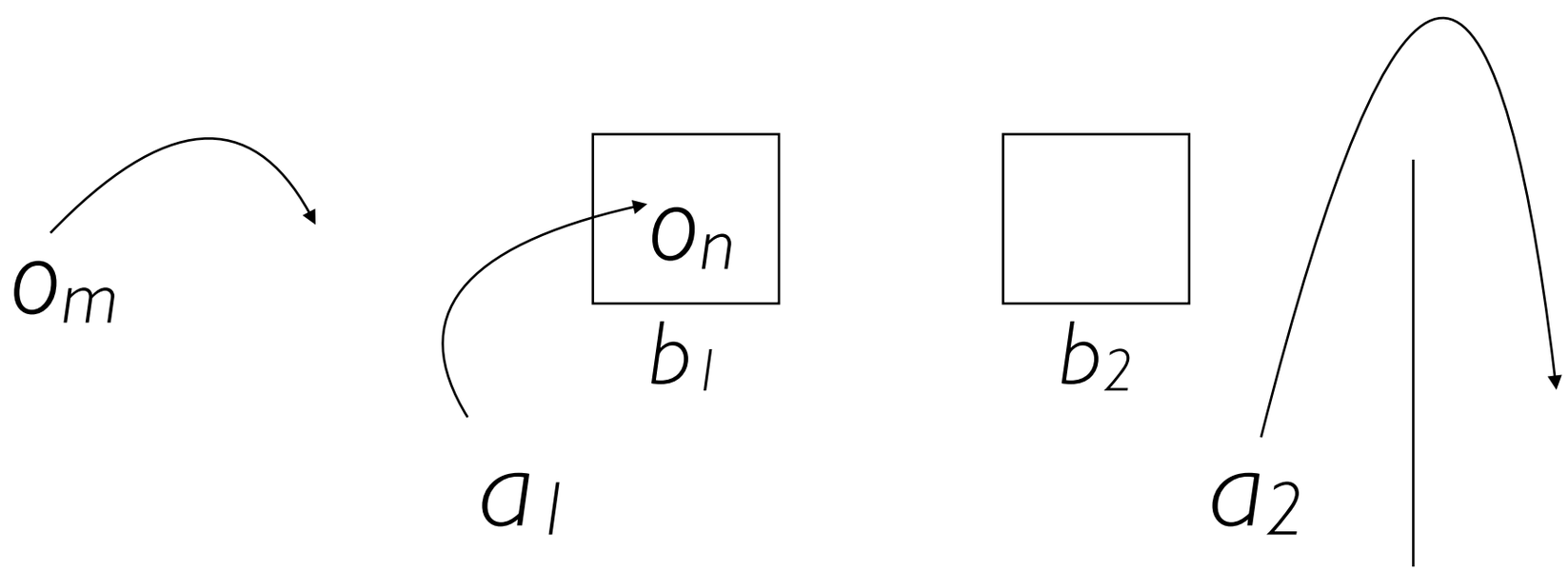
Framework for FBT[I]
(six timepoints)



a

e

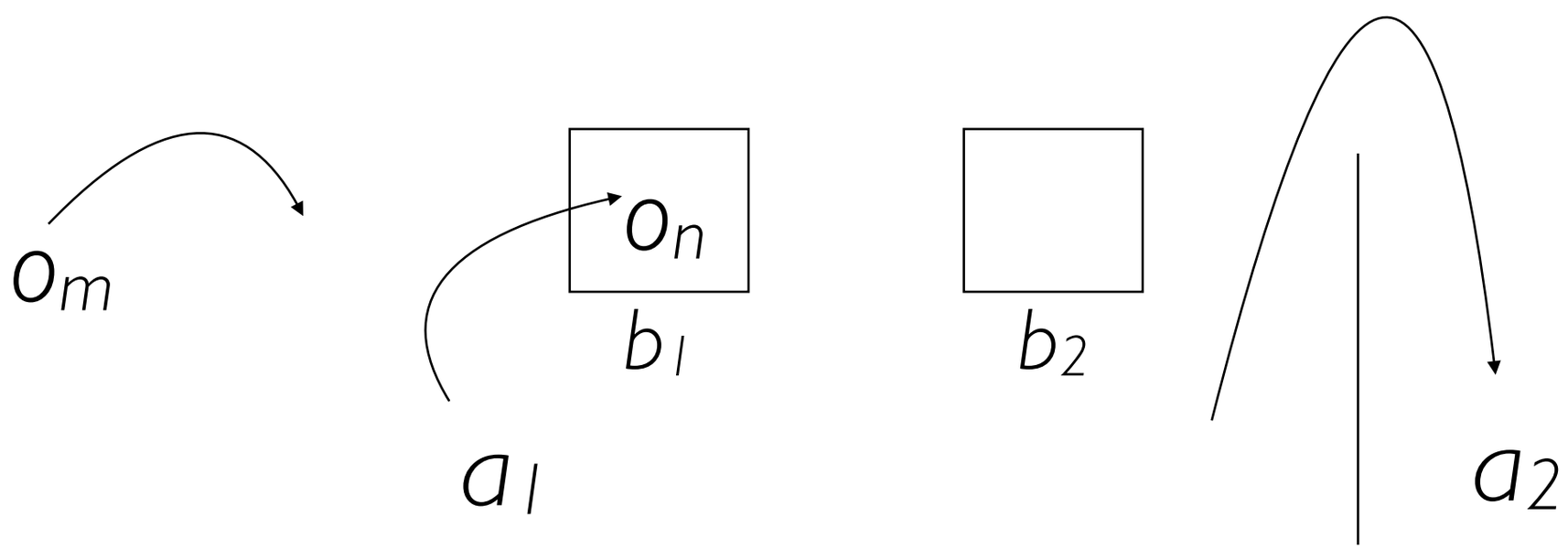
Framework for FBT[I]
(six timepoints)



a

e

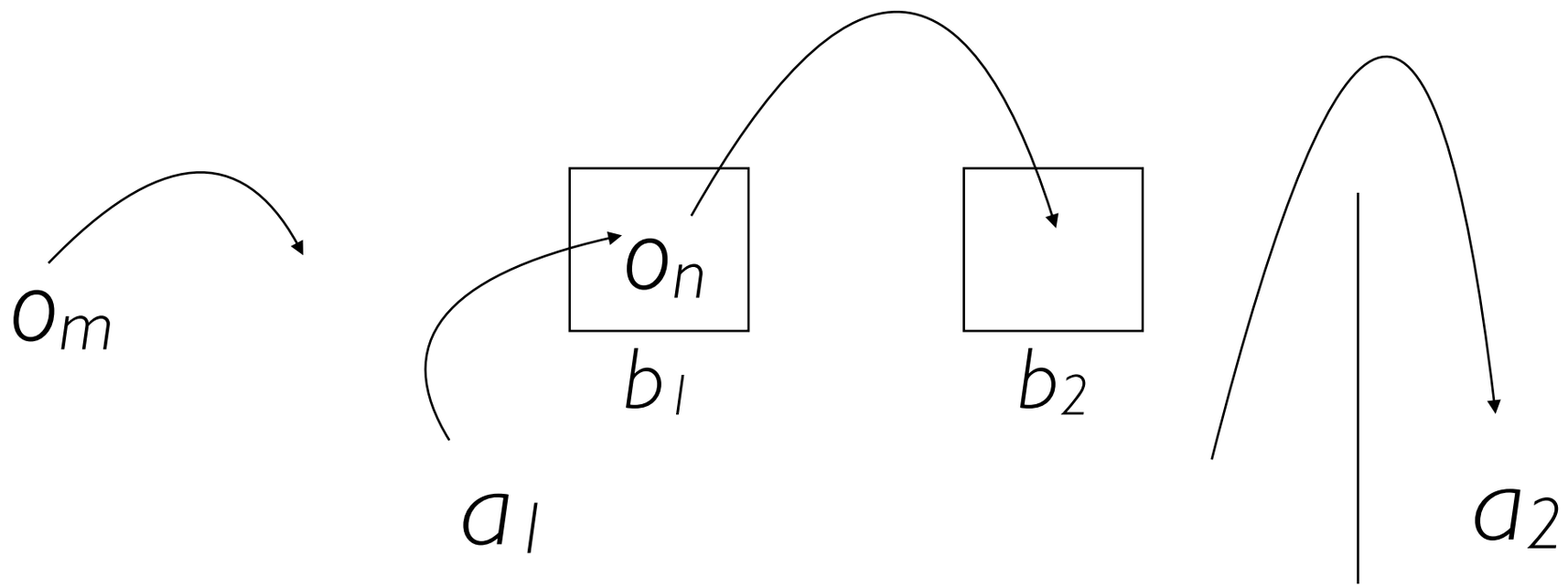
Framework for FBT[I]
(six timepoints)



a

e

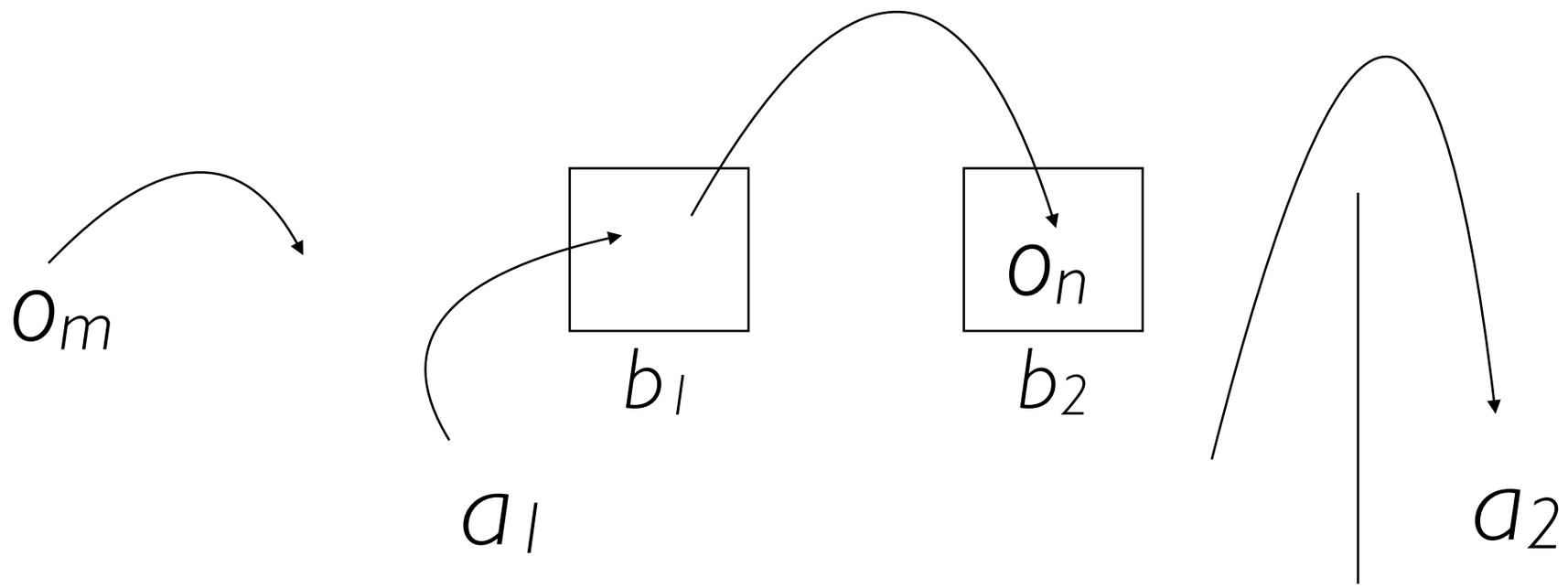
Framework for FBT[1]
(six timepoints)



a

e

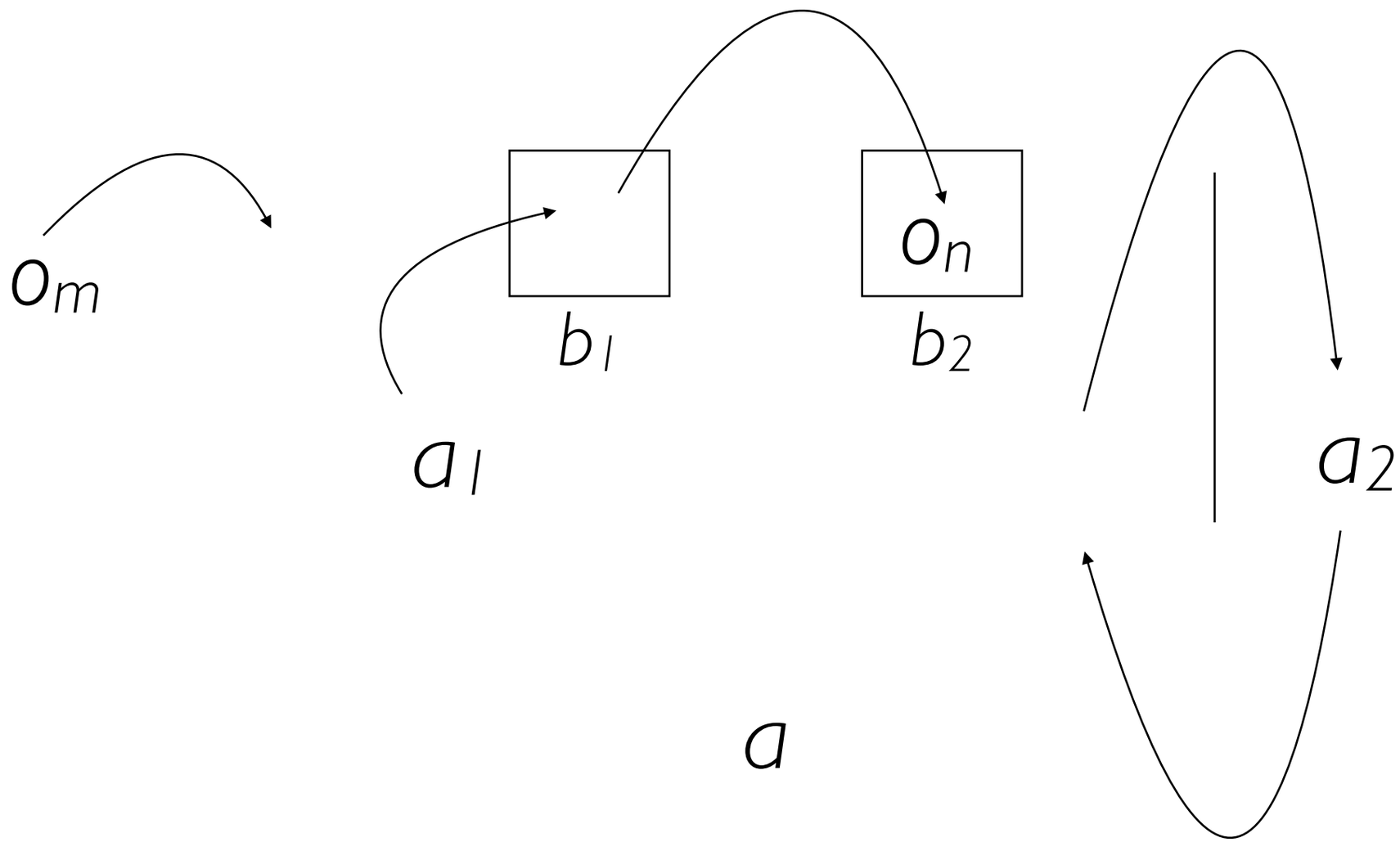
Framework for FBT[I]
(six timepoints)



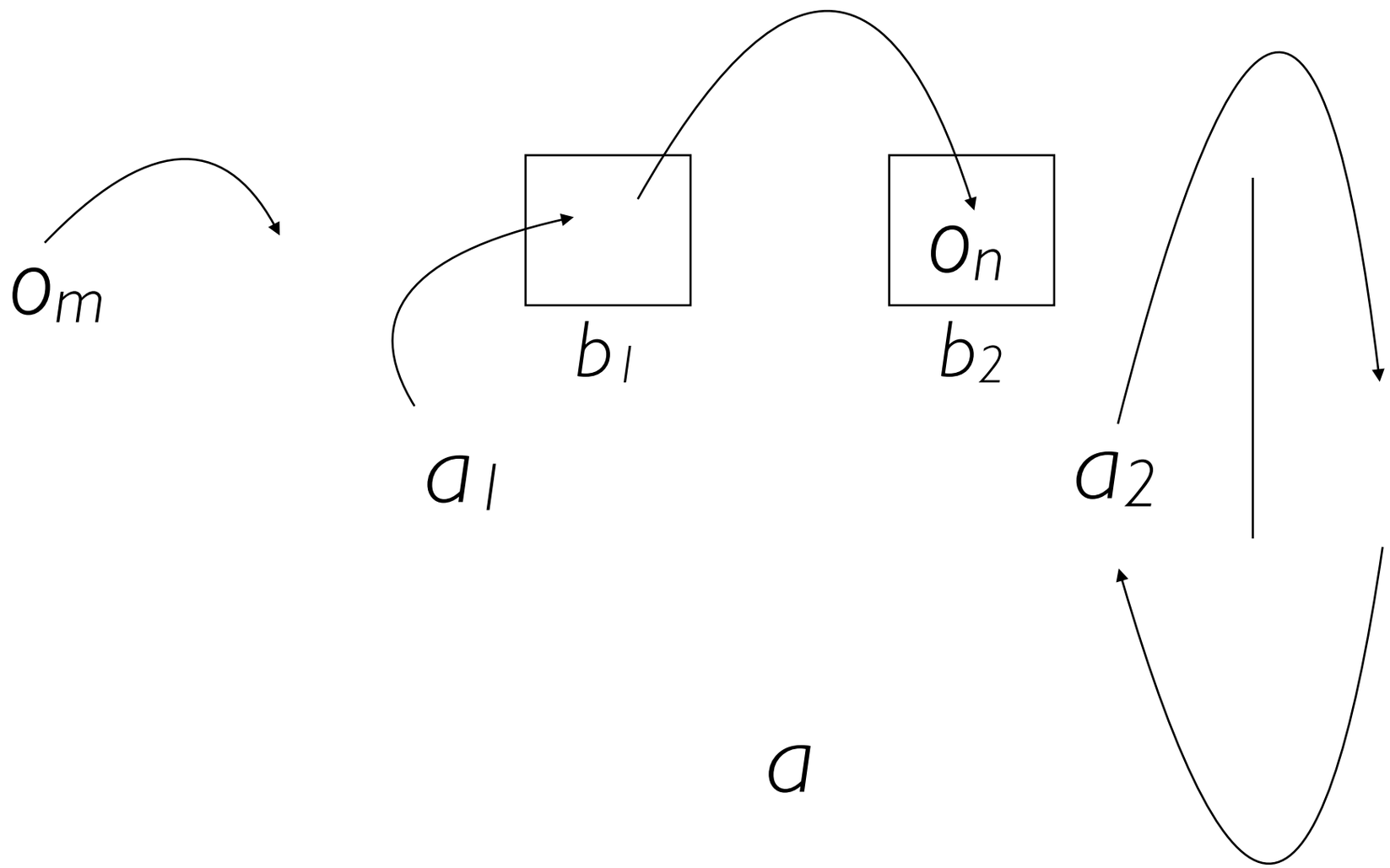
a

e

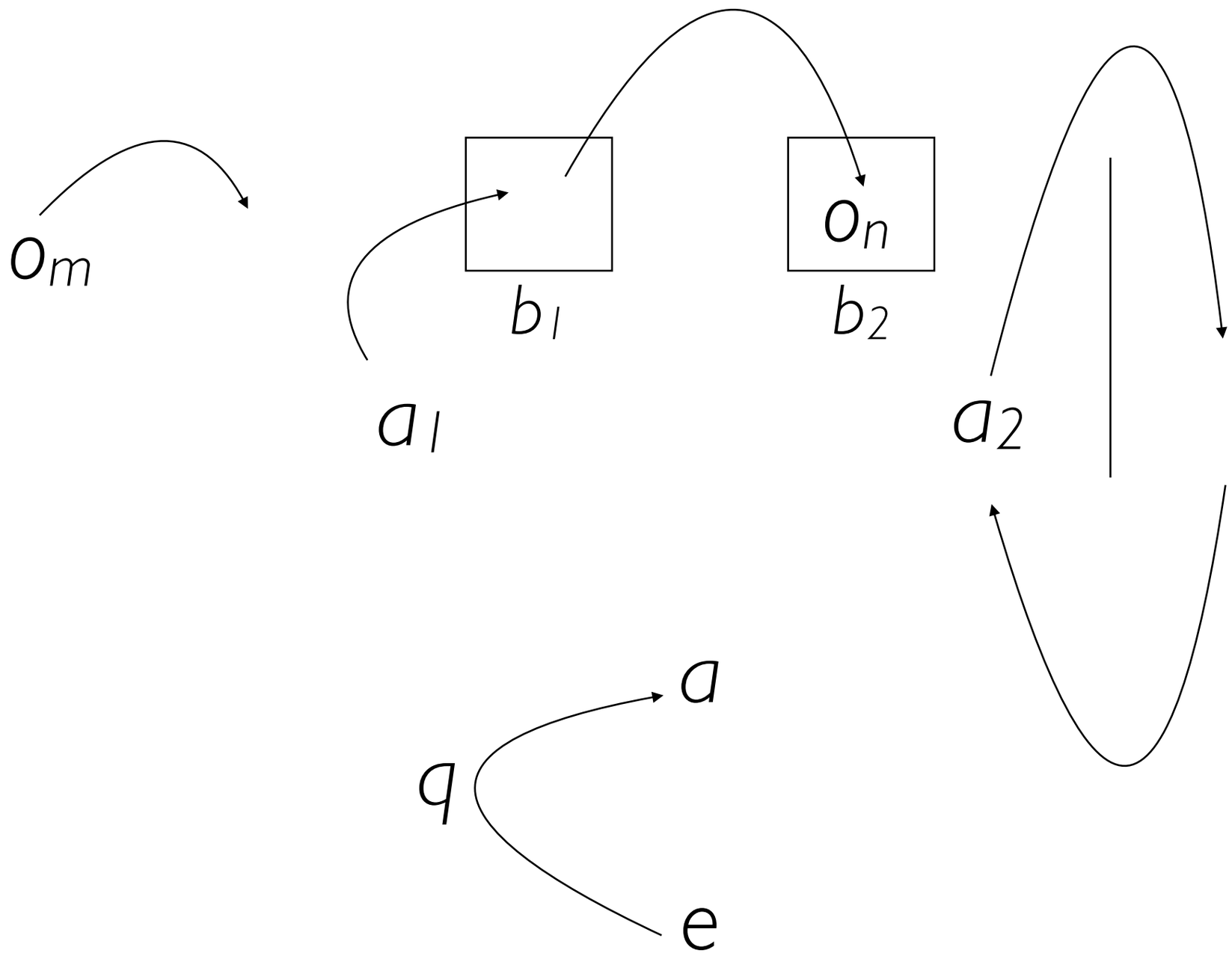
Framework for FBT[I]
(six timepoints)



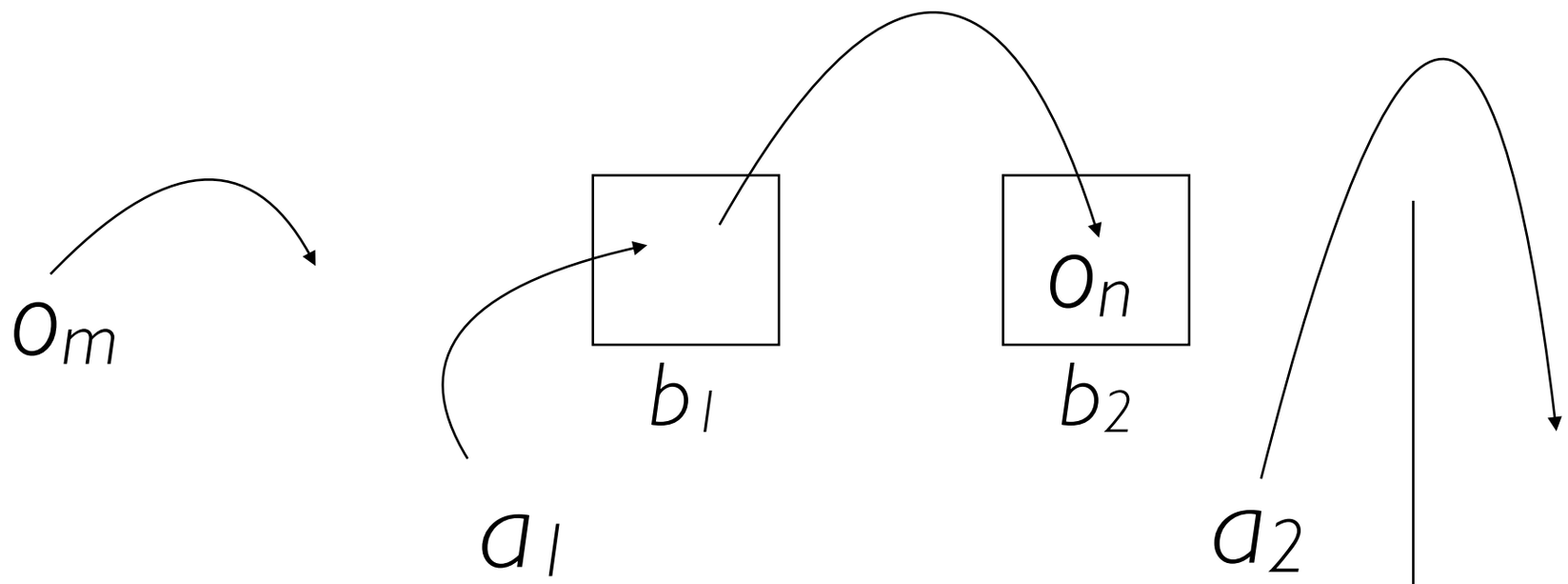
Framework for FBT[1]
(six timepoints)



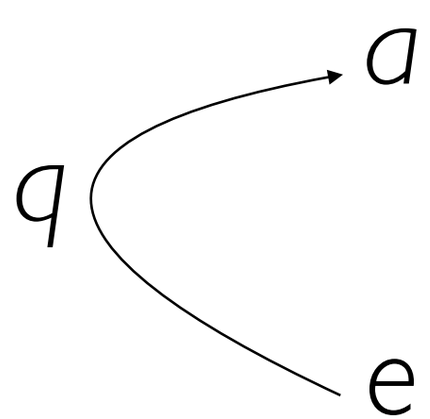
Framework for FBT[I]
(six timepoints)



Framework for FBT$^{I}_{|}$
(six timepoints)



q a formula in modal $\mathscr{L}^n$



Done, a Decade Ago, Formally & Implementation/Simulation

Arkoudas, K. & Bringsjord, S. (2009) "Propositional Attitudes and Causation" *International Journal of Software and Informatics* **3.1**: 47–65.

http://kryten.mm.rpi.edu/PRICAI_w_sequentcalc_041709.pdf



Propositional attitudes and causation

Konstantine Arkoudas and Selmer Bringsjord

Cognitive Science and Computer Science Departments, RPI
arkouk@rpi.edu, brings@rpi.edu

Abstract. Predicting and explaining the behavior of others in terms of mental states is indispensable for everyday life. It will be equally important for artificial agents. We present an inference system for representing and reasoning about mental states, and use it to provide a formal analysis of the false-belief task. The system allows for the representation of information about events, causation, and perceptual, doxastic, and epistemic states (vision, belief, and knowledge), incorporating ideas from the event calculus and multi-agent epistemic logic. Unlike previous AI formalisms, our focus here is on mechanized proofs and proof programmability, not on metamathematical results. Reasoning is performed via relatively cognitively plausible inference rules, and a degree of automation is achieved by general-purpose inference methods and by a syntactic embedding of the system in first-order logic.

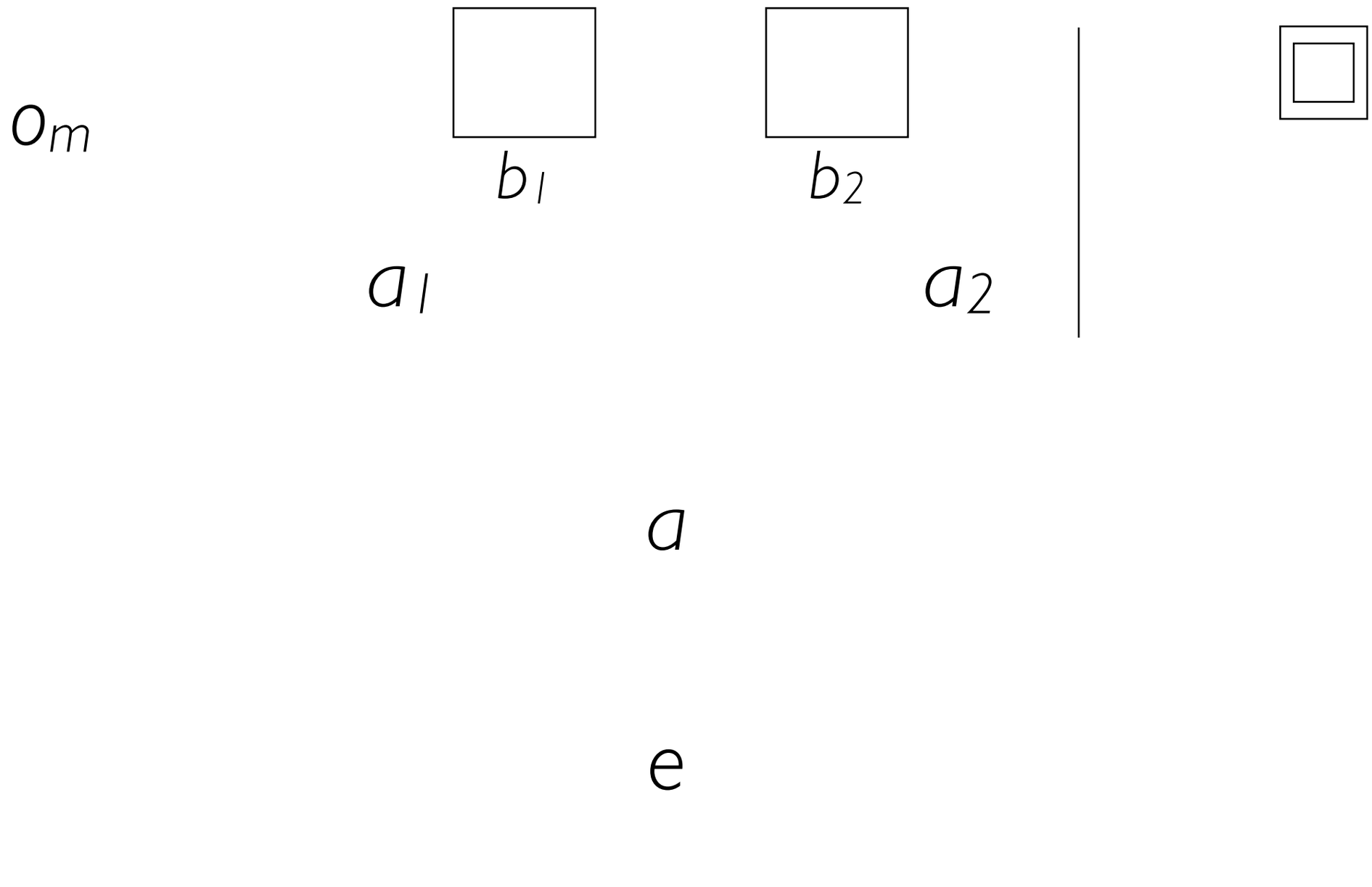
1 Introduction

Interpreting the behavior of other people is indispensable for everyday life. It is something that we do constantly, on a daily basis, and it helps us not only to make sense of human behavior, but also to predict it and—to a certain extent—to control it. How exactly do we manage that? That is not currently known, but many have argued that the ability to ascribe mental states to others and to reason about such mental states is a key component of our capacity to understand human behavior. In particular, all social transactions, from engaging in commerce and negotiating to making jokes and empathizing with other people's pain or joy, appear to require at least a rudimentary grasp of common-sense psychology (CSP), i.e., a large body of truisms such as the following: When an agent a (1) wants to achieve a certain state of affairs p, and (2) believes that some action c can bring about p, and (3) a knows how to carry out c; then, ceteris paribus,[1] a will carry out c; when a sees that p, a knows that p; when a fears that p and a discovers that p is the case, a is disappointed; and so on.

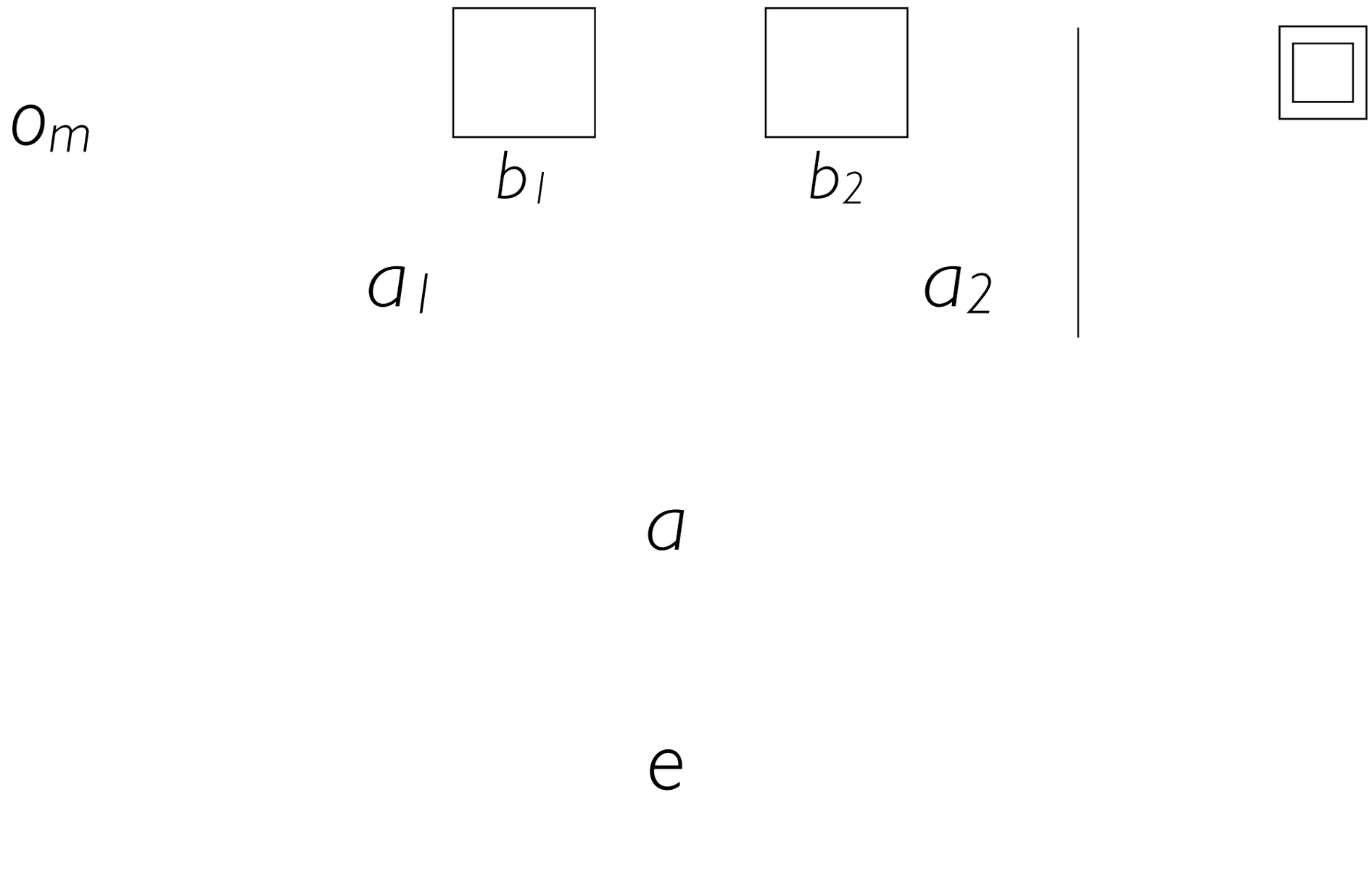
Artificial agents without a mastery of CSP would be severely handicapped in their interactions with humans. This could present problems not only for artificial agents trying to interpret human behavior, but also for artificial agents trying to interpret the behavior of one another. When a system exhibits a complex but rational behavior, and detailed knowledge of its internal structure is not

[1] Assuming that a is able to carry out c, that a has no conflicting desires that override his goal that p; and so on.

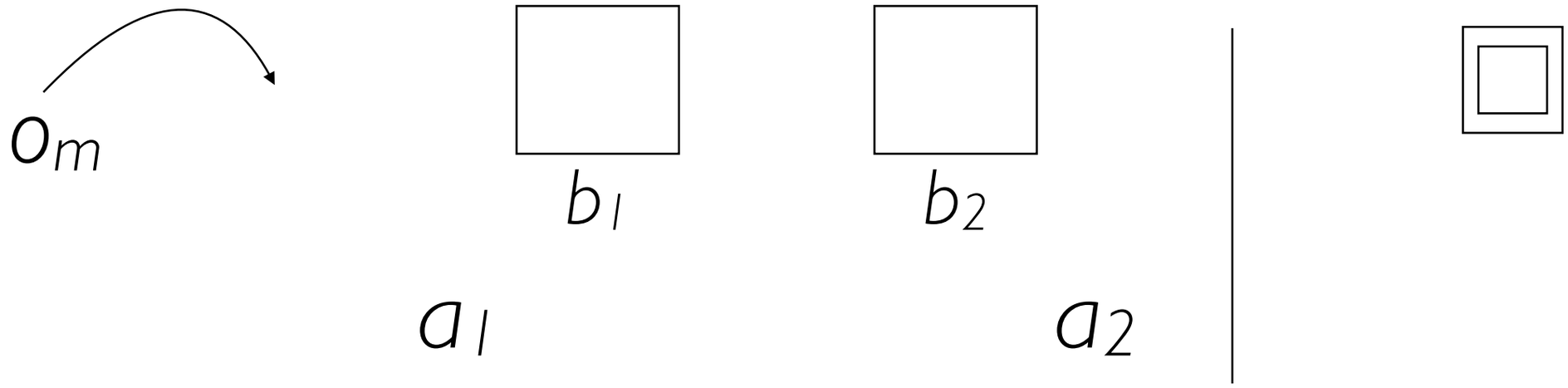
Framework for FBT^{I}_{2}



Framework for FBT$^{I}_{2}$
(seven timepoints)



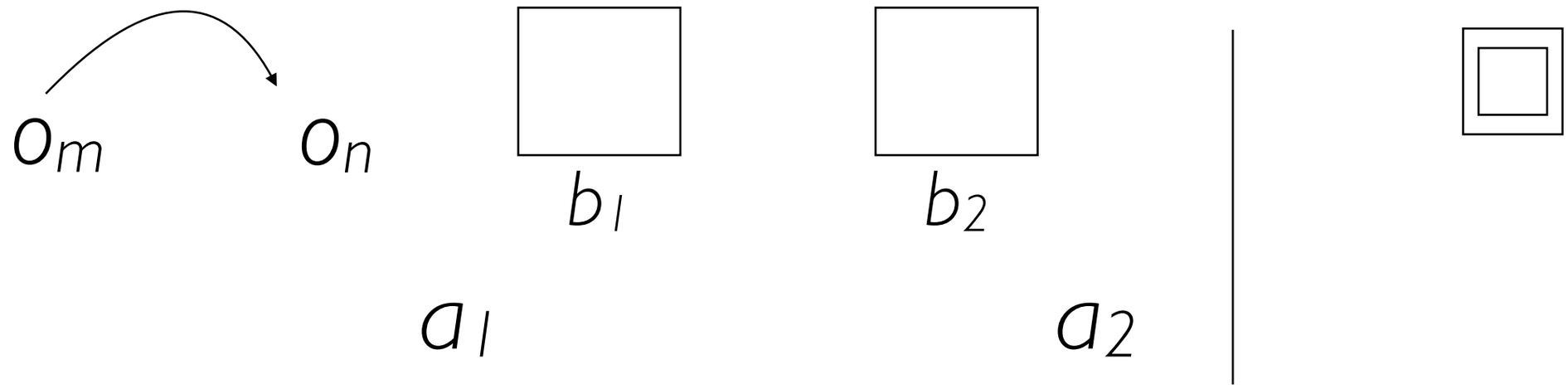
Framework for FBT$^{I}_{2}$
(seven timepoints)



a

e

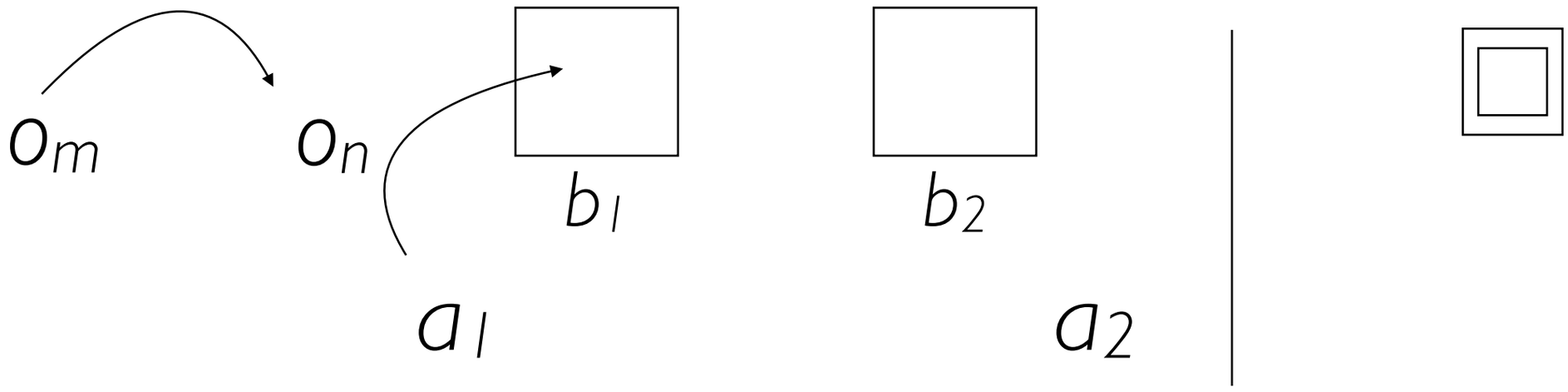
Framework for FBT$^{I}_{2}$
(seven timepoints)



a

e

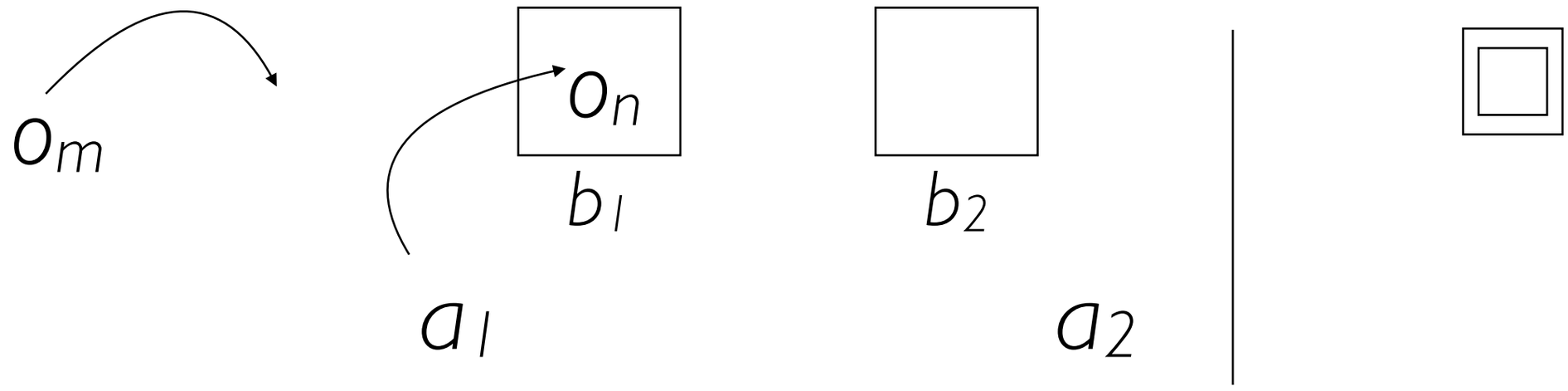
Framework for FBT$^{I}_{2}$
(seven timepoints)



a

e

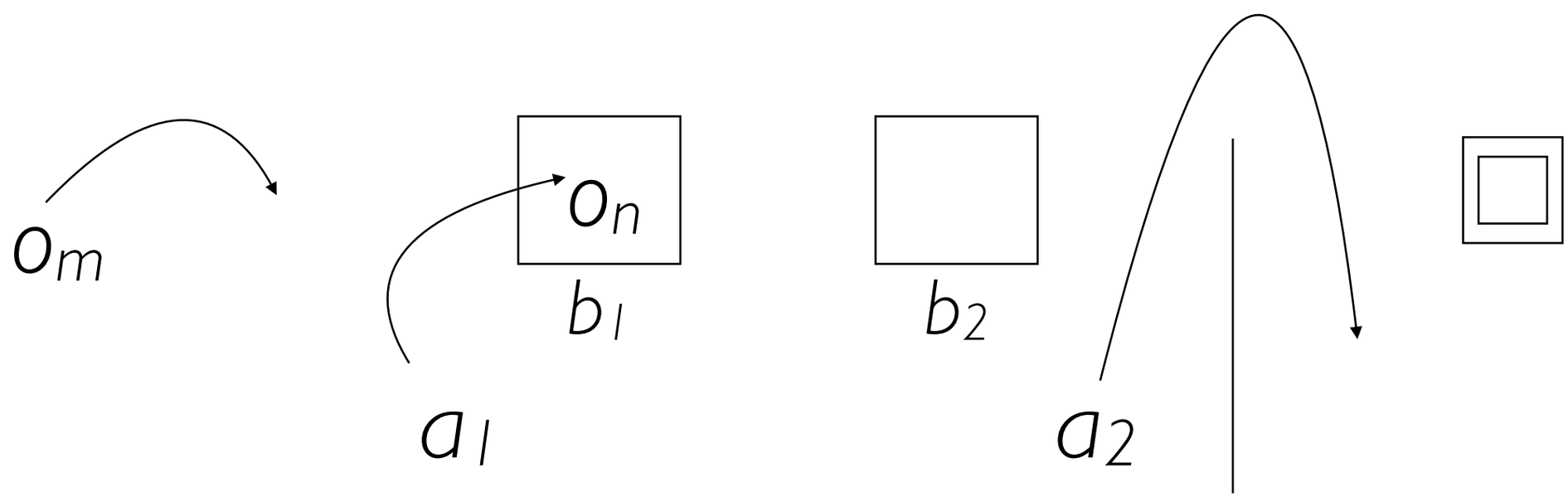
Framework for FBT$^{I}_{2}$
(seven timepoints)



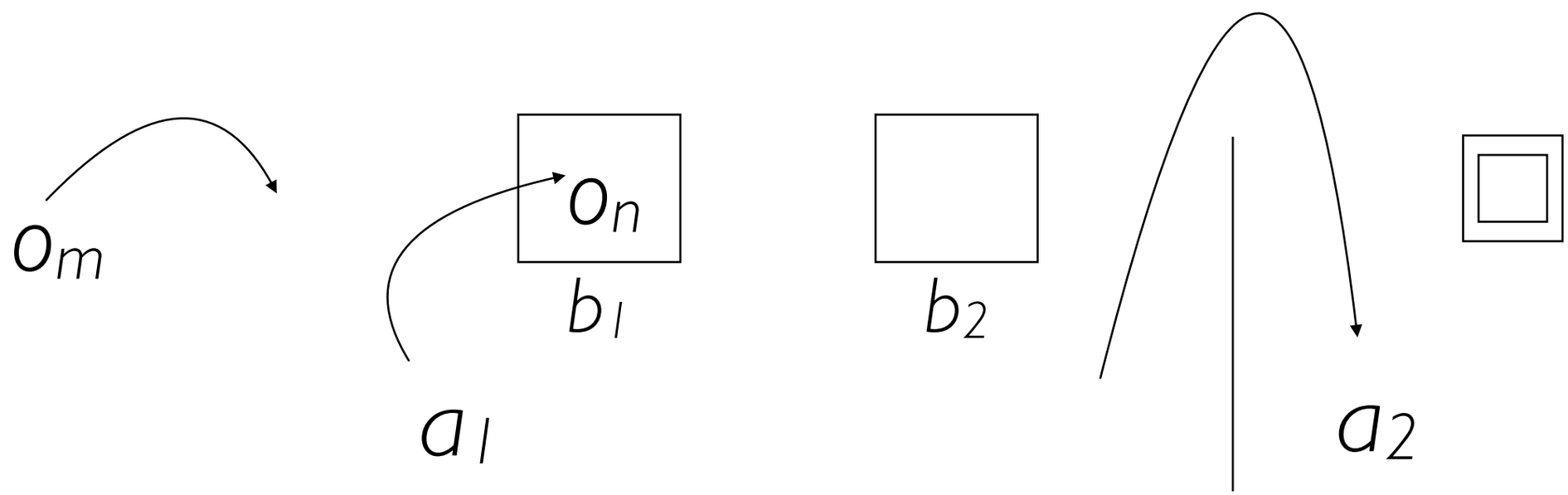
a

e

Framework for FBT$^{I}_{2}$
(seven timepoints)



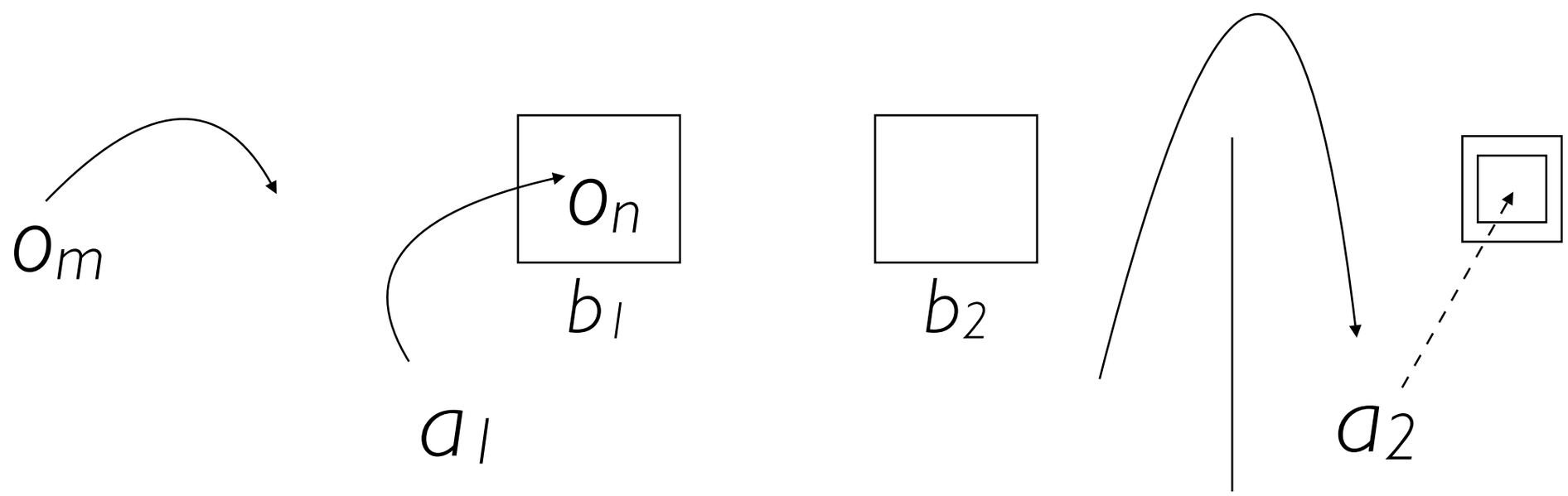
Framework for FBT^{I}_{2}
(seven timepoints)



a

e

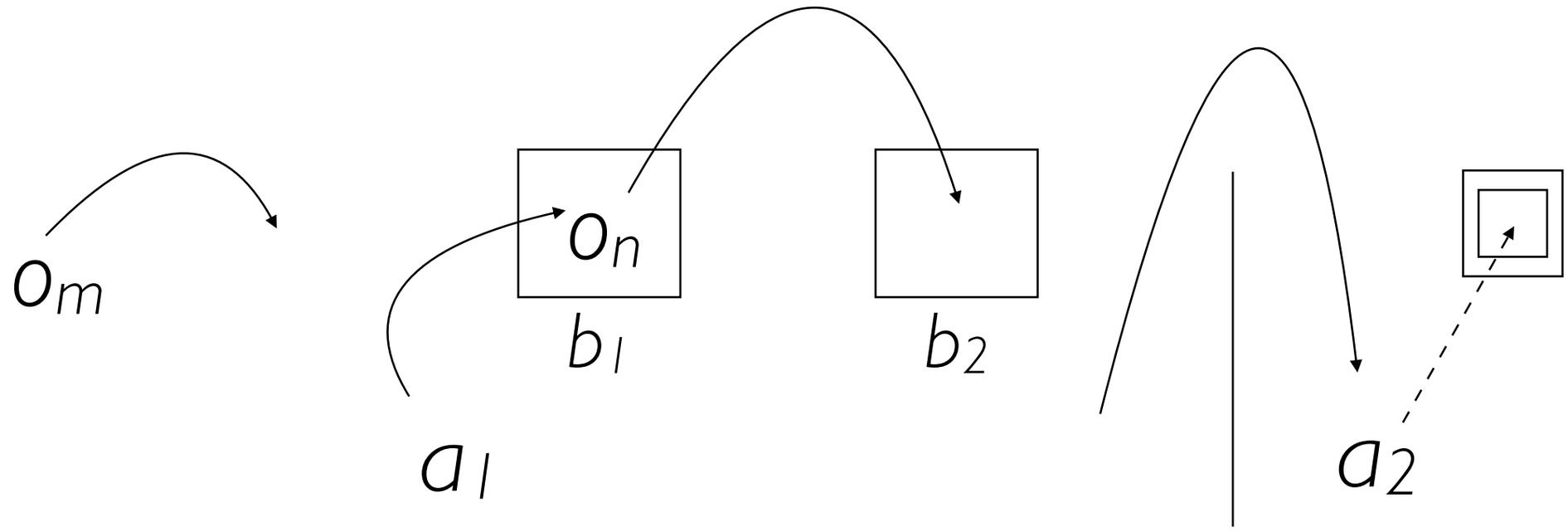
Framework for FBT$^{I}_{2}$
(seven timepoints)



a

e

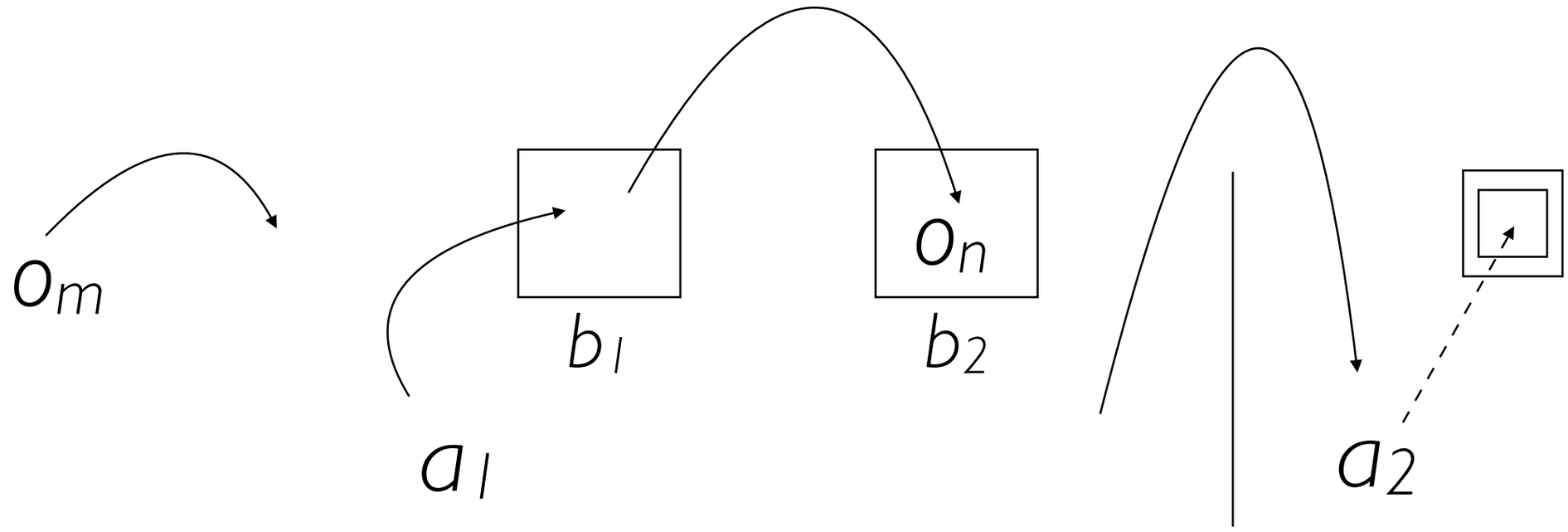
Framework for FBT$^{I}_{2}$
(seven timepoints)



a

e

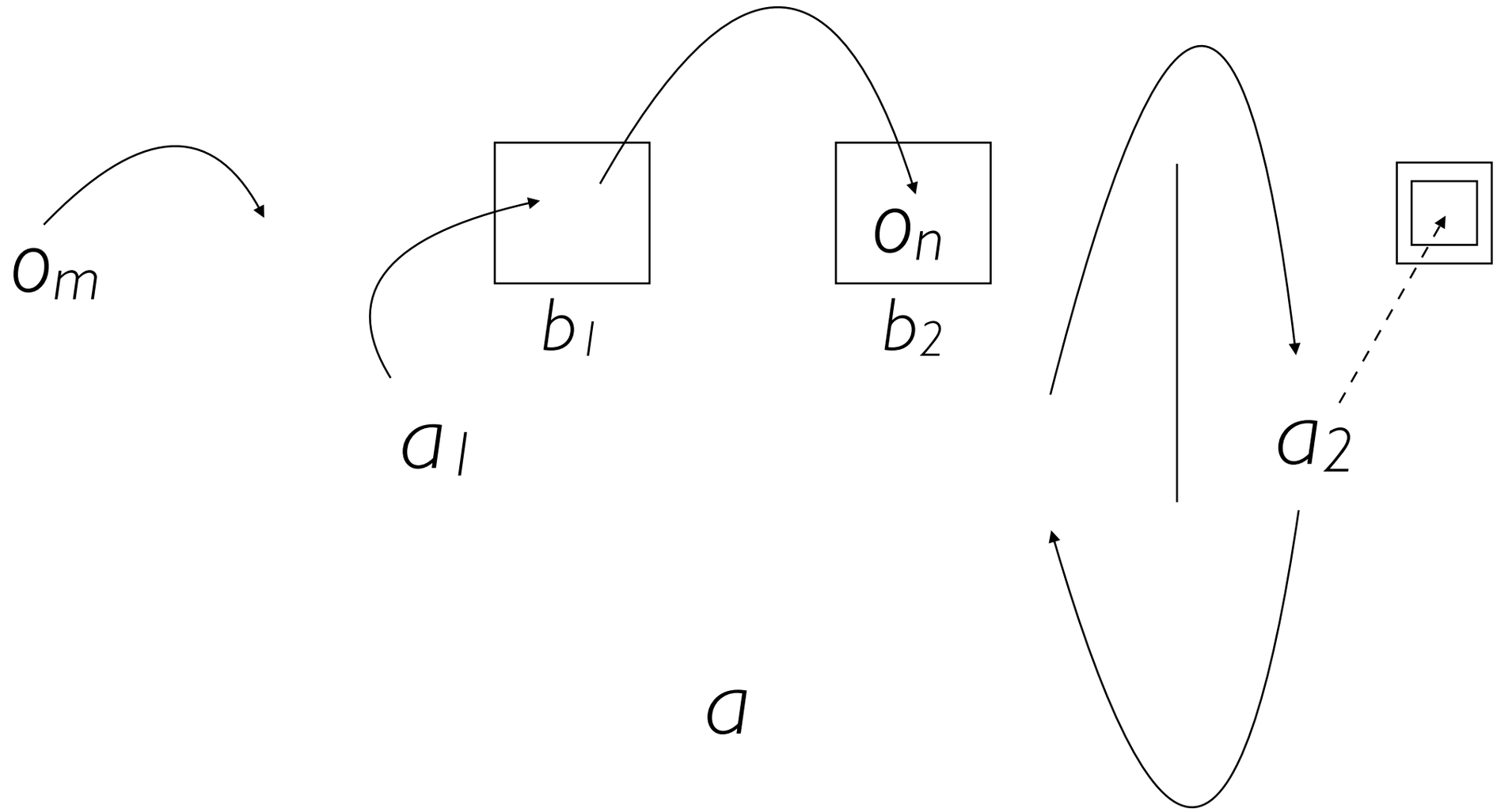
Framework for FBT$^{I}_{2}$
(seven timepoints)



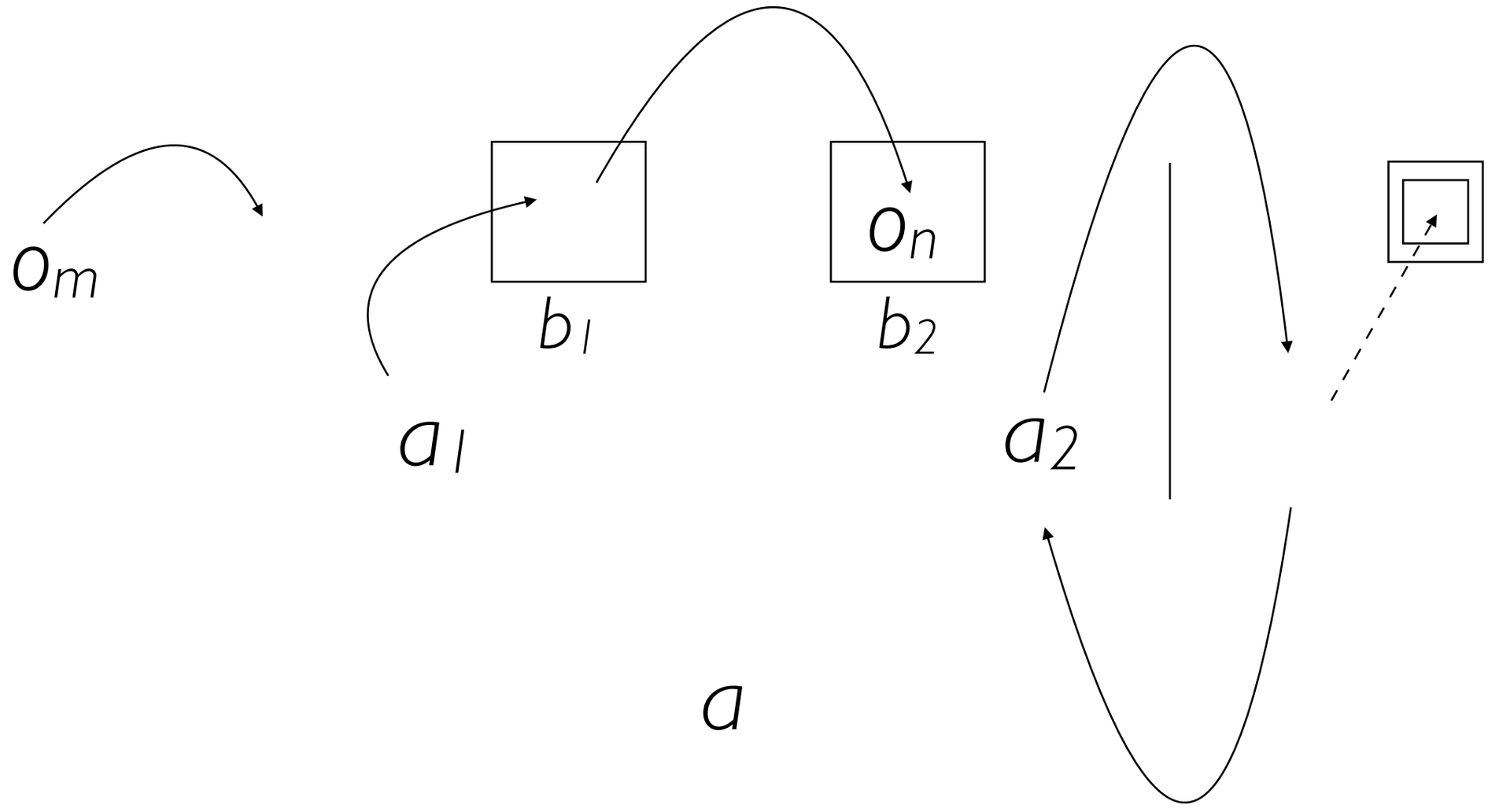
a

e

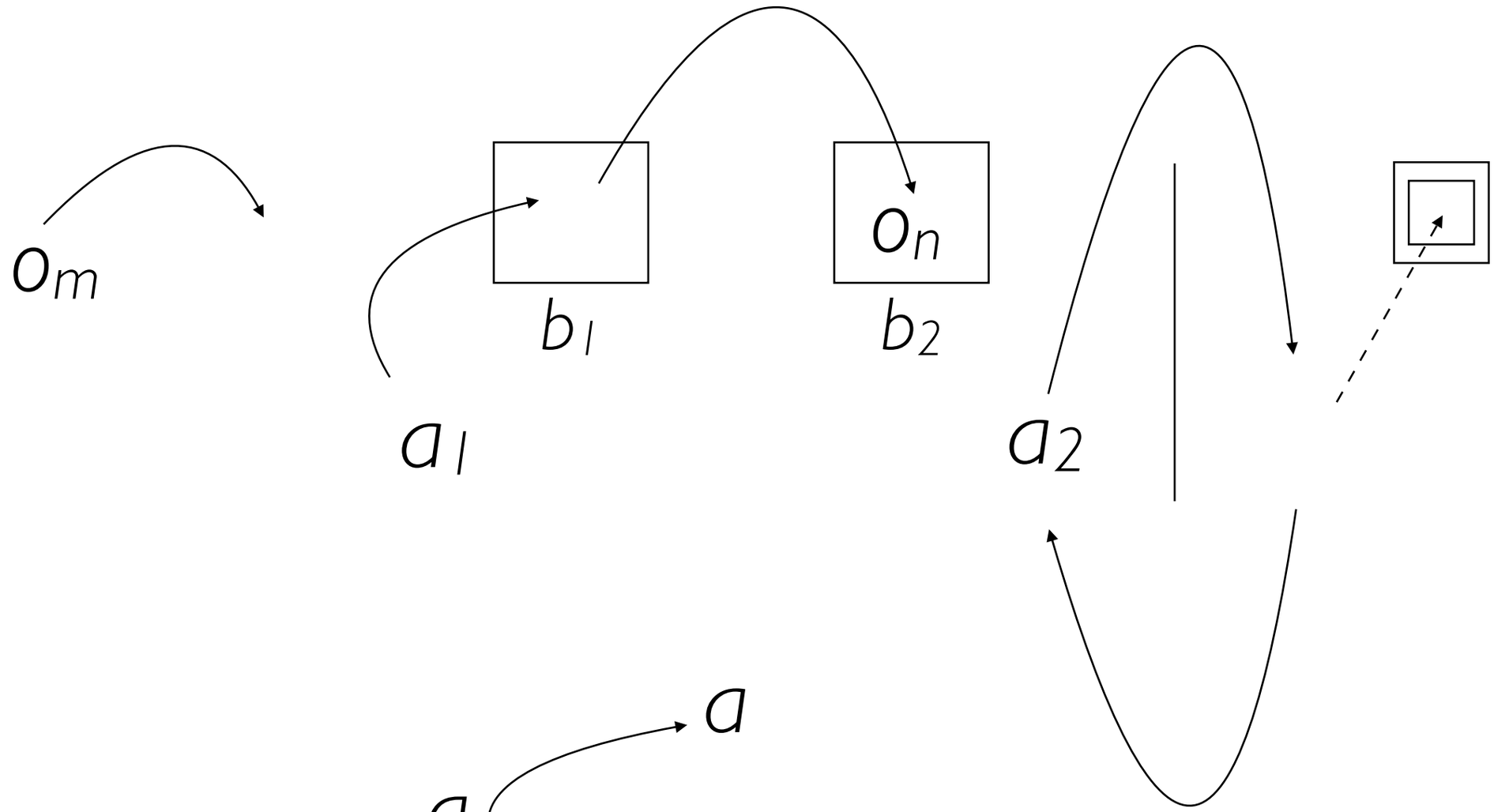
Framework for FBT$^{I}_{2}$
(seven timepoints)



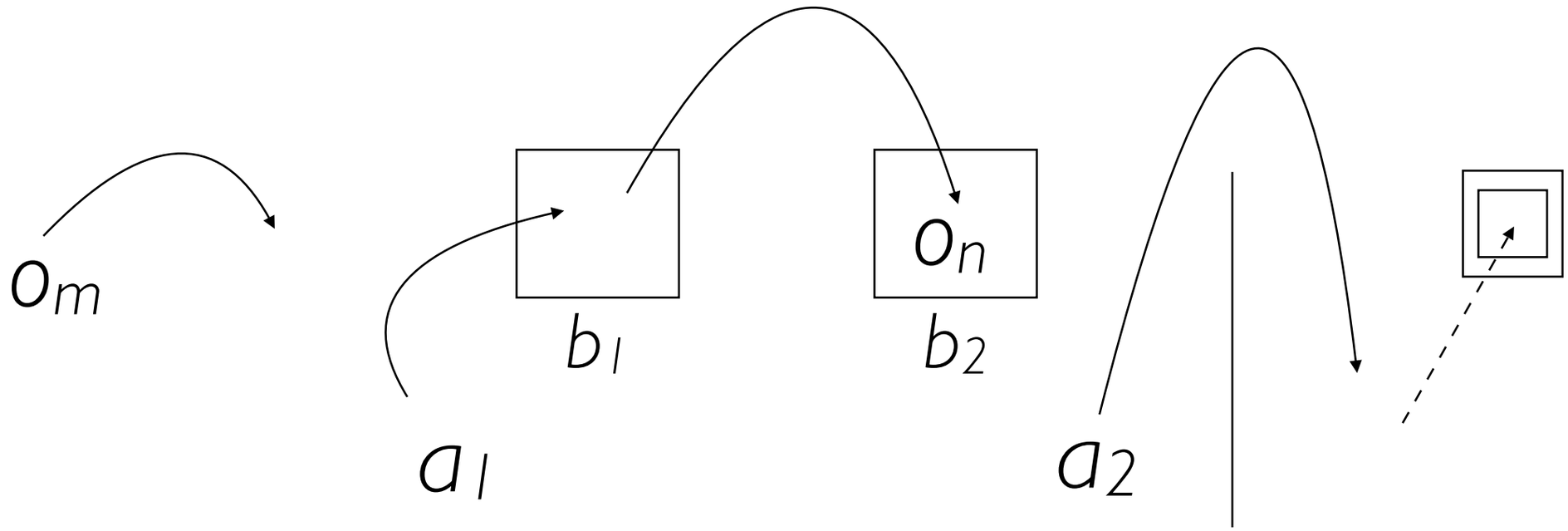
Framework for FBT$^{I}_{2}$
(seven timepoints)



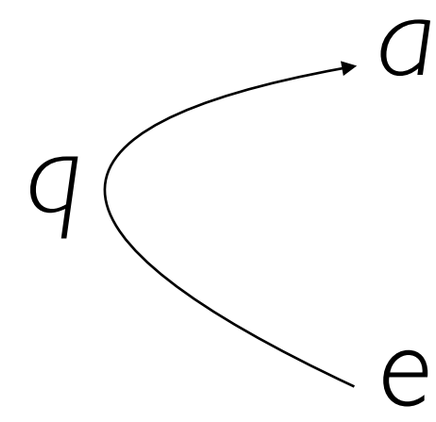
Framework for FBT$^{I}_{2}$
(seven timepoints)



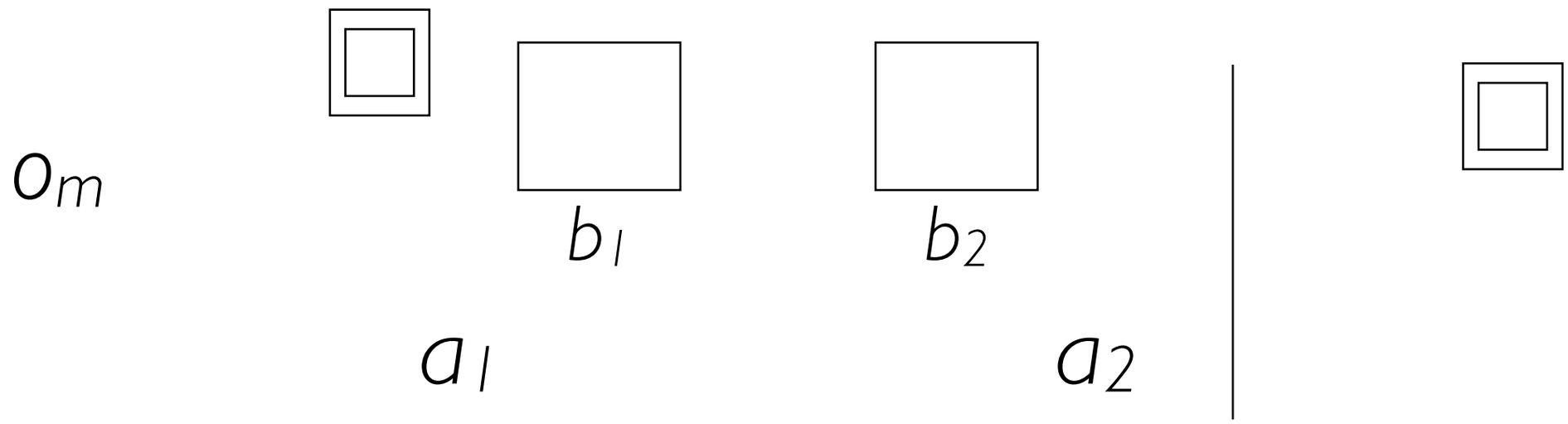
Framework for FBT^{I}_{2}
(seven timepoints)



q a formula in modal $\mathscr{L}^n$



Framework for FBT^{I}_{3}

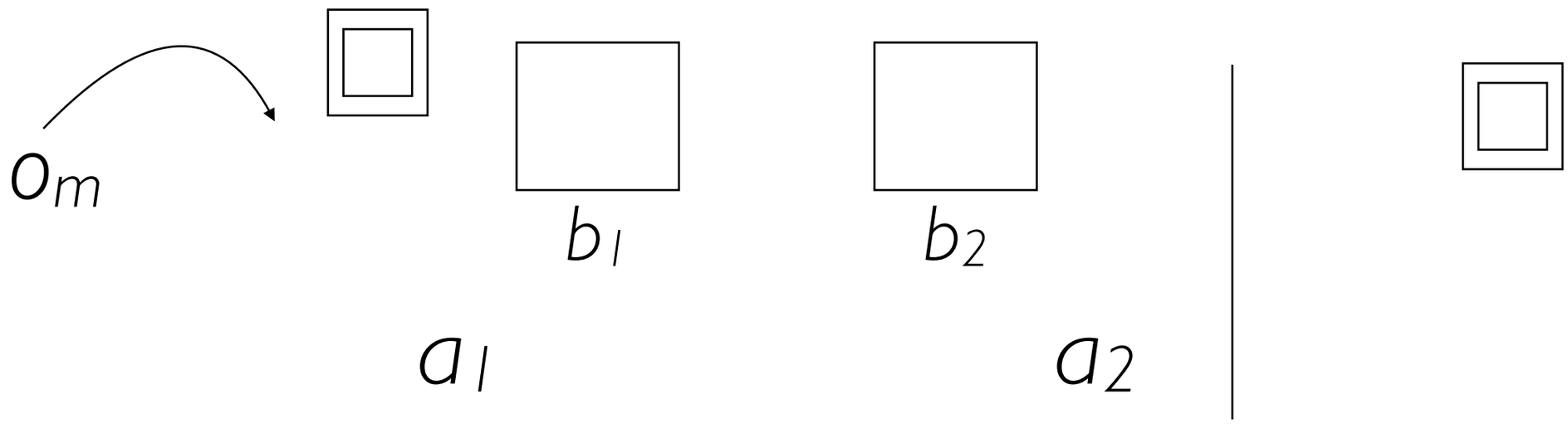


a

e

Framework for FBT$^{I}_{3}$
(eight timepoints)

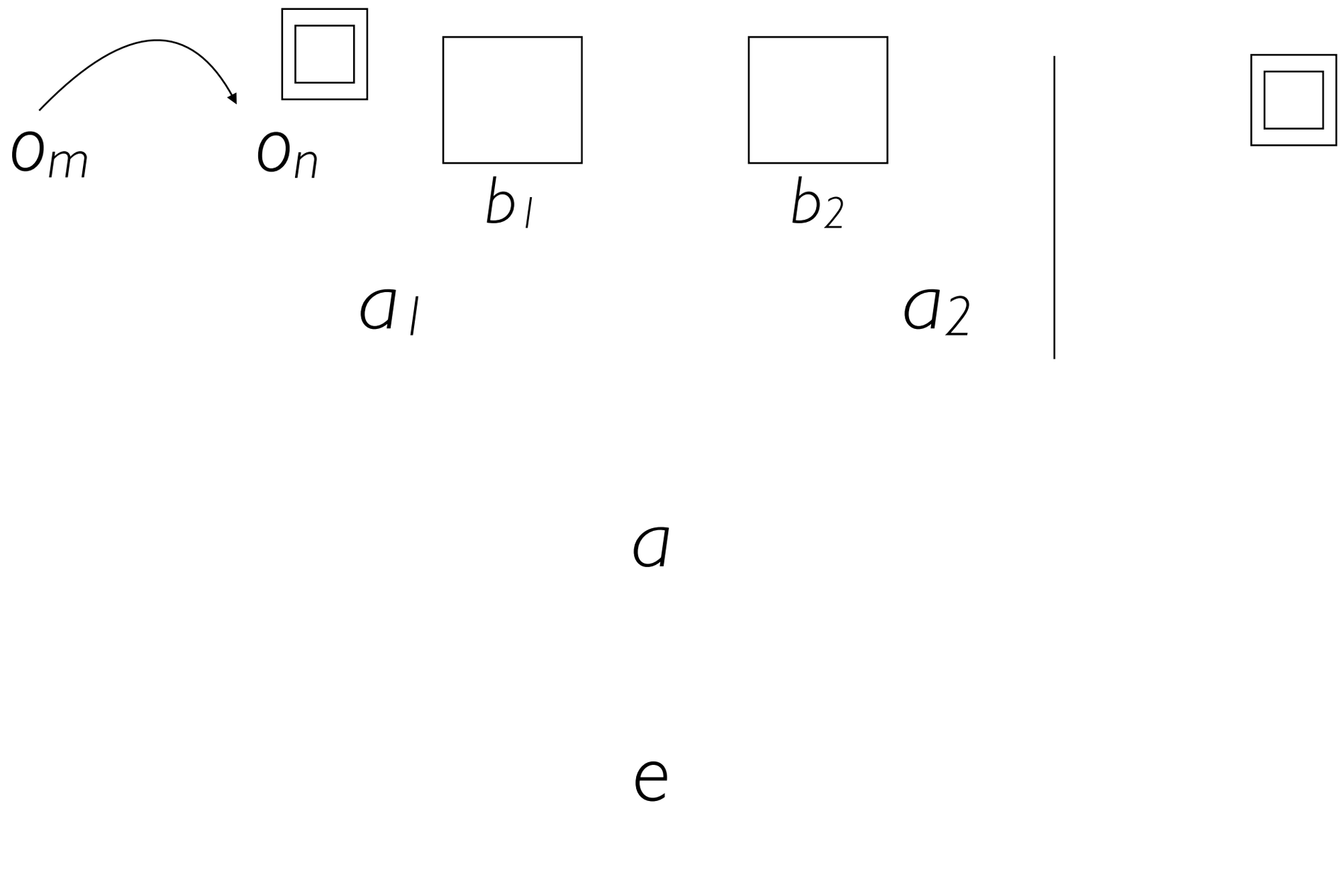
Framework for FBT$^{I}_{3}$
(eight timepoints)



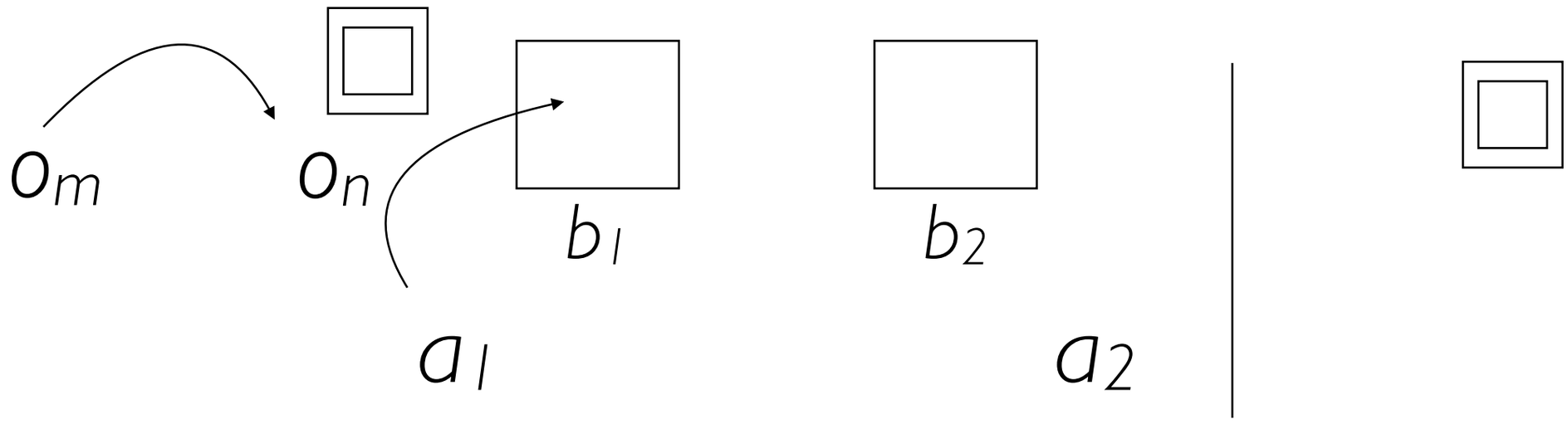
a

e

Framework for FBT^{I}_{3}
(eight timepoints)



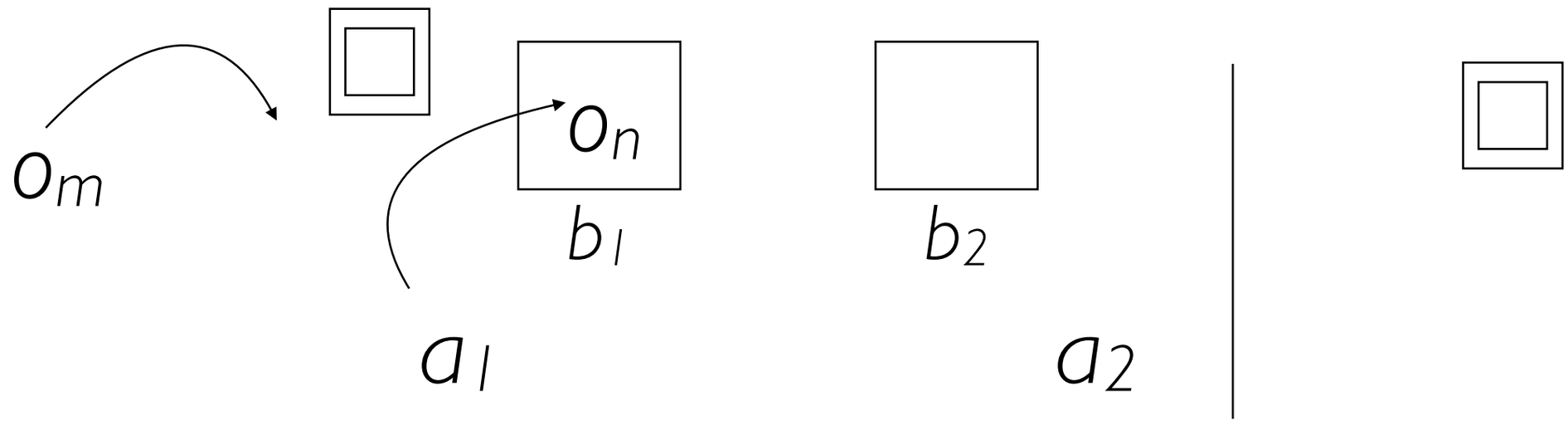
Framework for FBT^{I}_{3}
(eight timepoints)



a

e

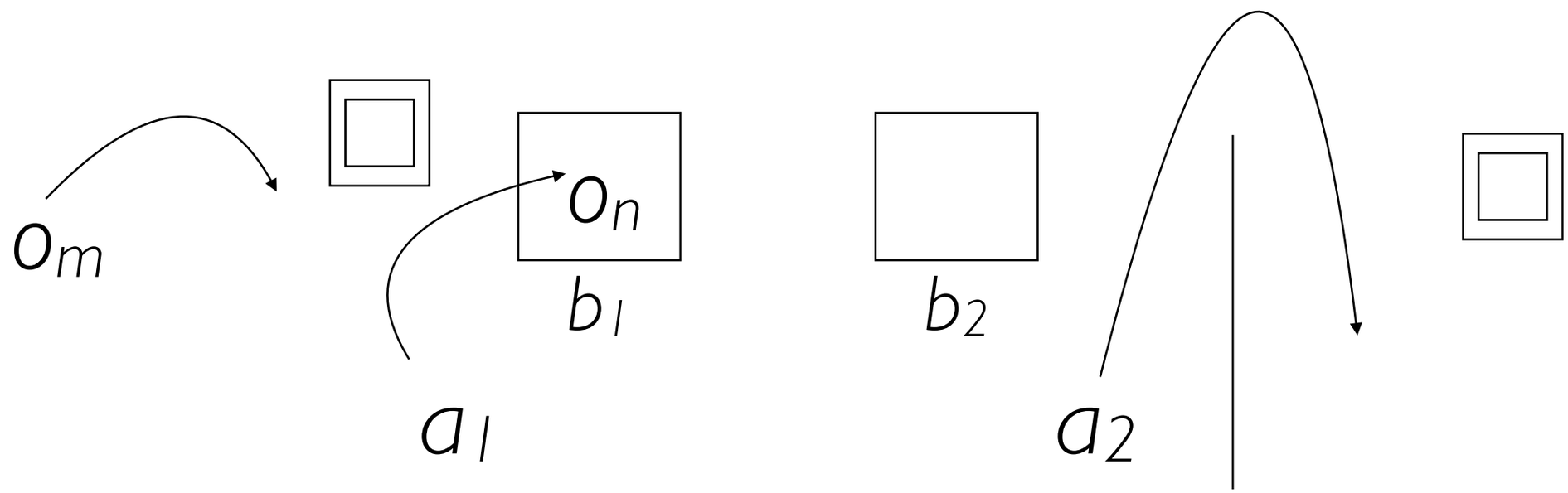
Framework for FBT^{I}_{3}
(eight timepoints)



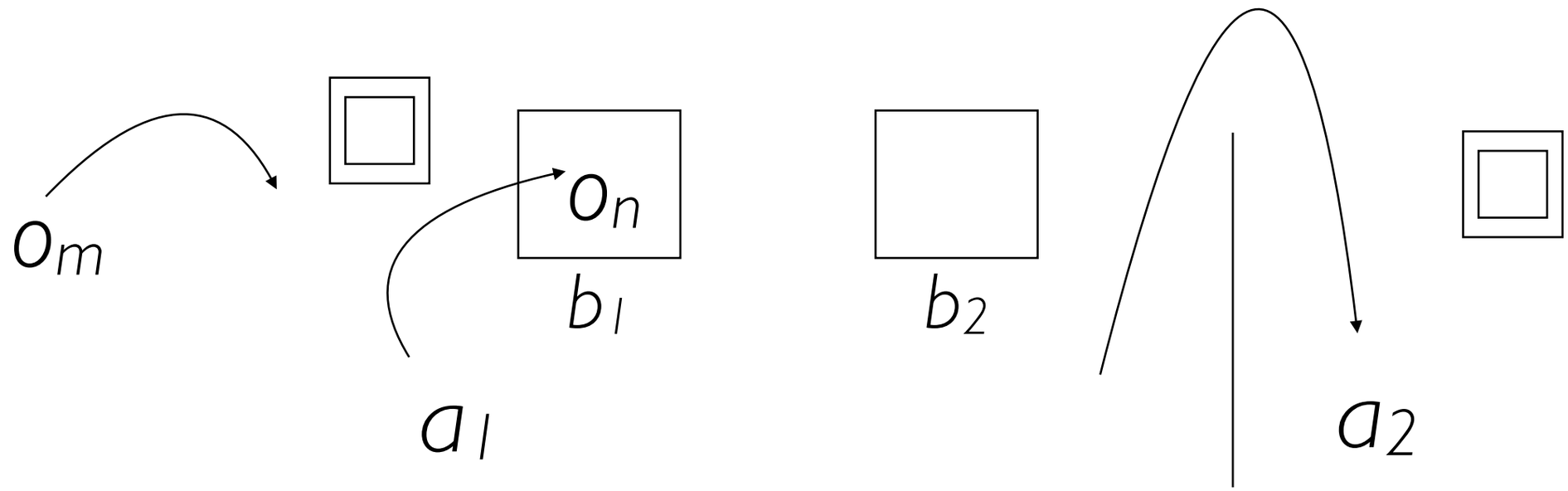
a

e

Framework for FBT$^{I}_{3}$
(eight timepoints)



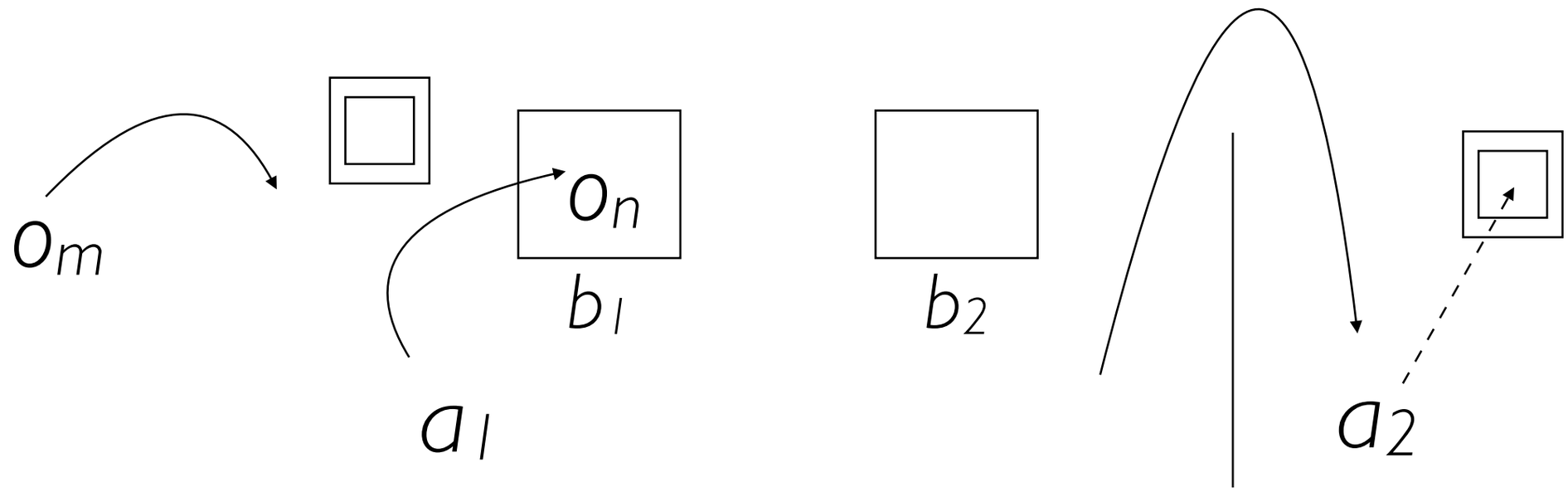
Framework for FBT$^{I}_{3}$
(eight timepoints)



a

e

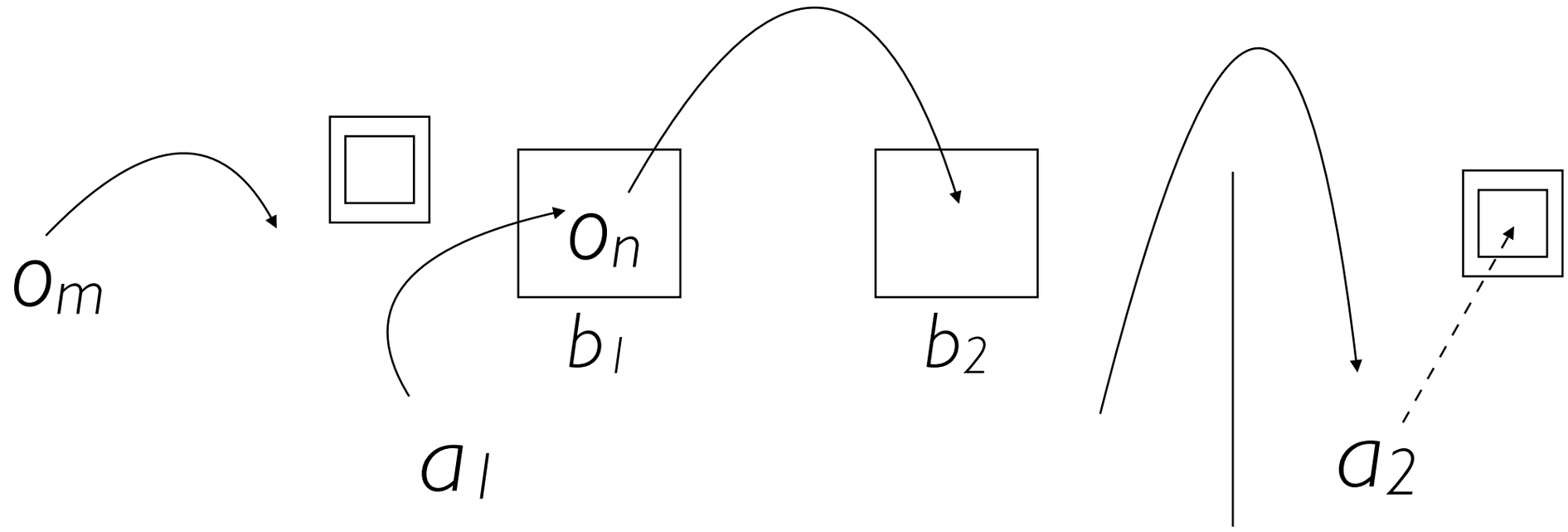
Framework for FBT^{I}_{3}
(eight timepoints)



a

e

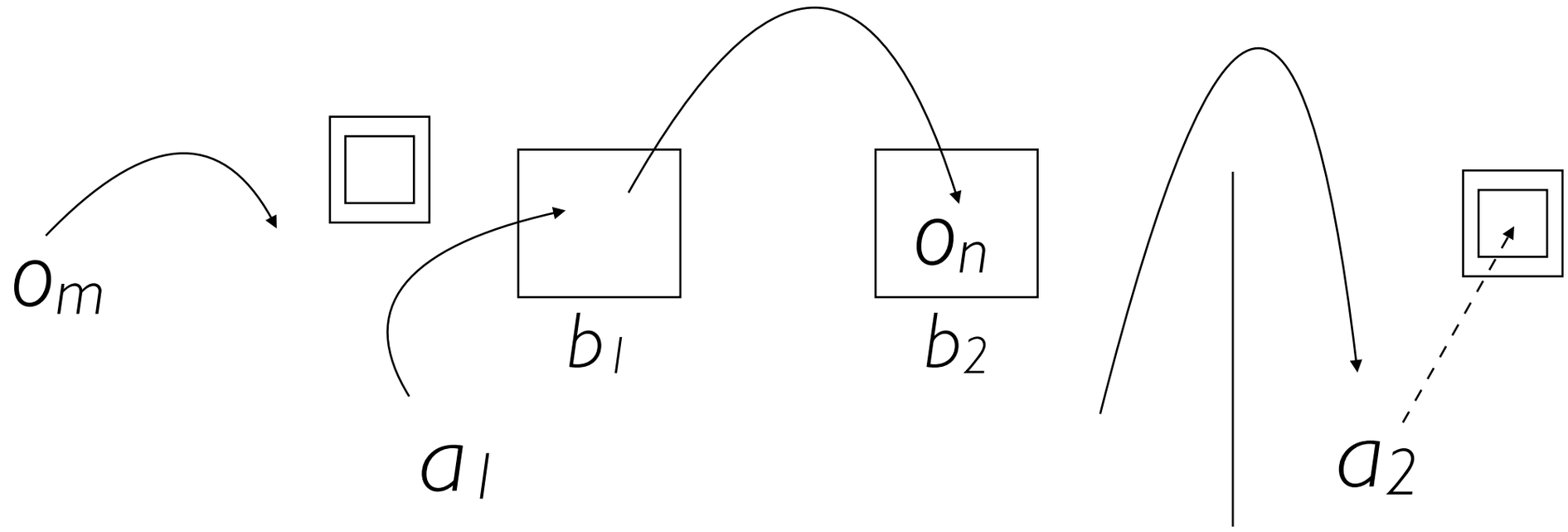
Framework for FBT^{I}_{3}
(eight timepoints)



a

e

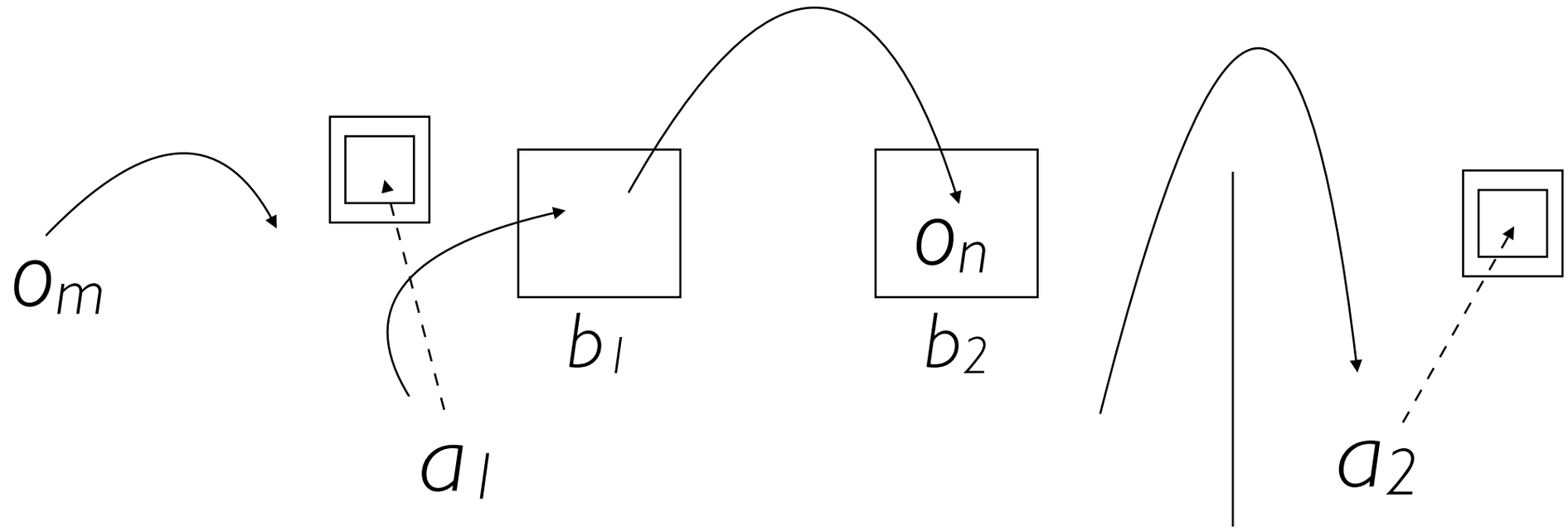
Framework for FBT^{I}_{3}
(eight timepoints)



a

e

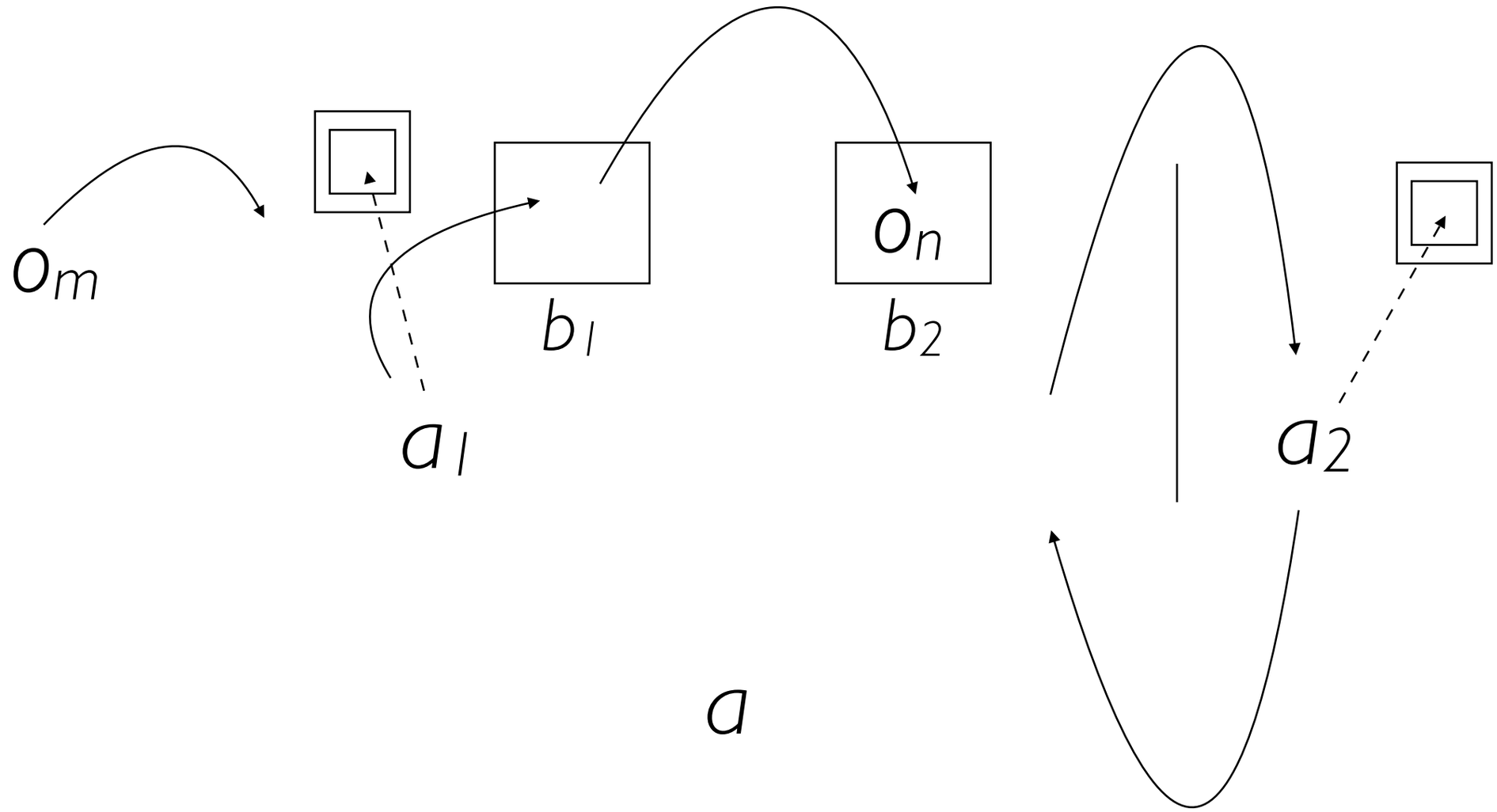
Framework for FBT$^{I}_{3}$
(eight timepoints)



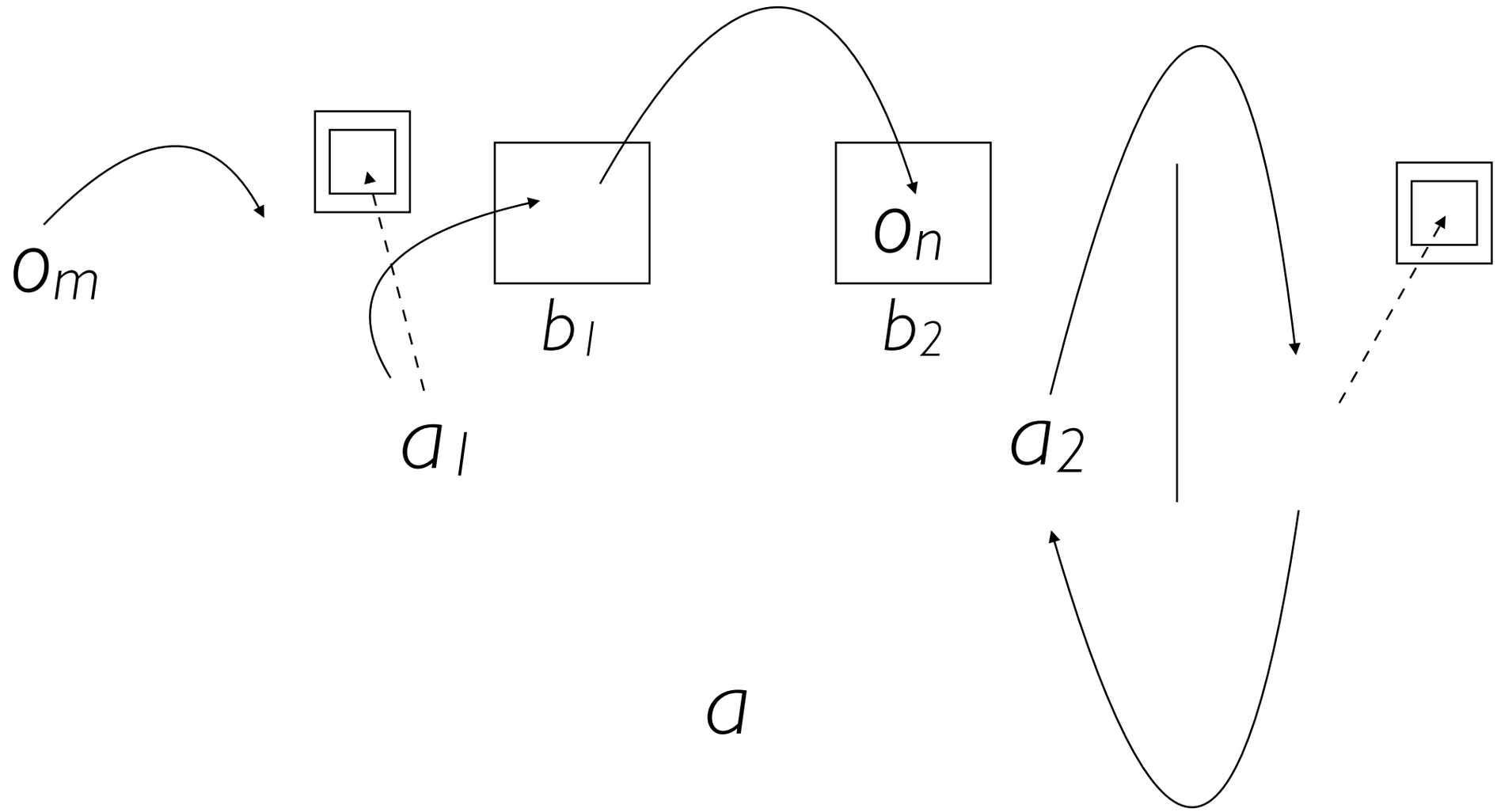
a

e

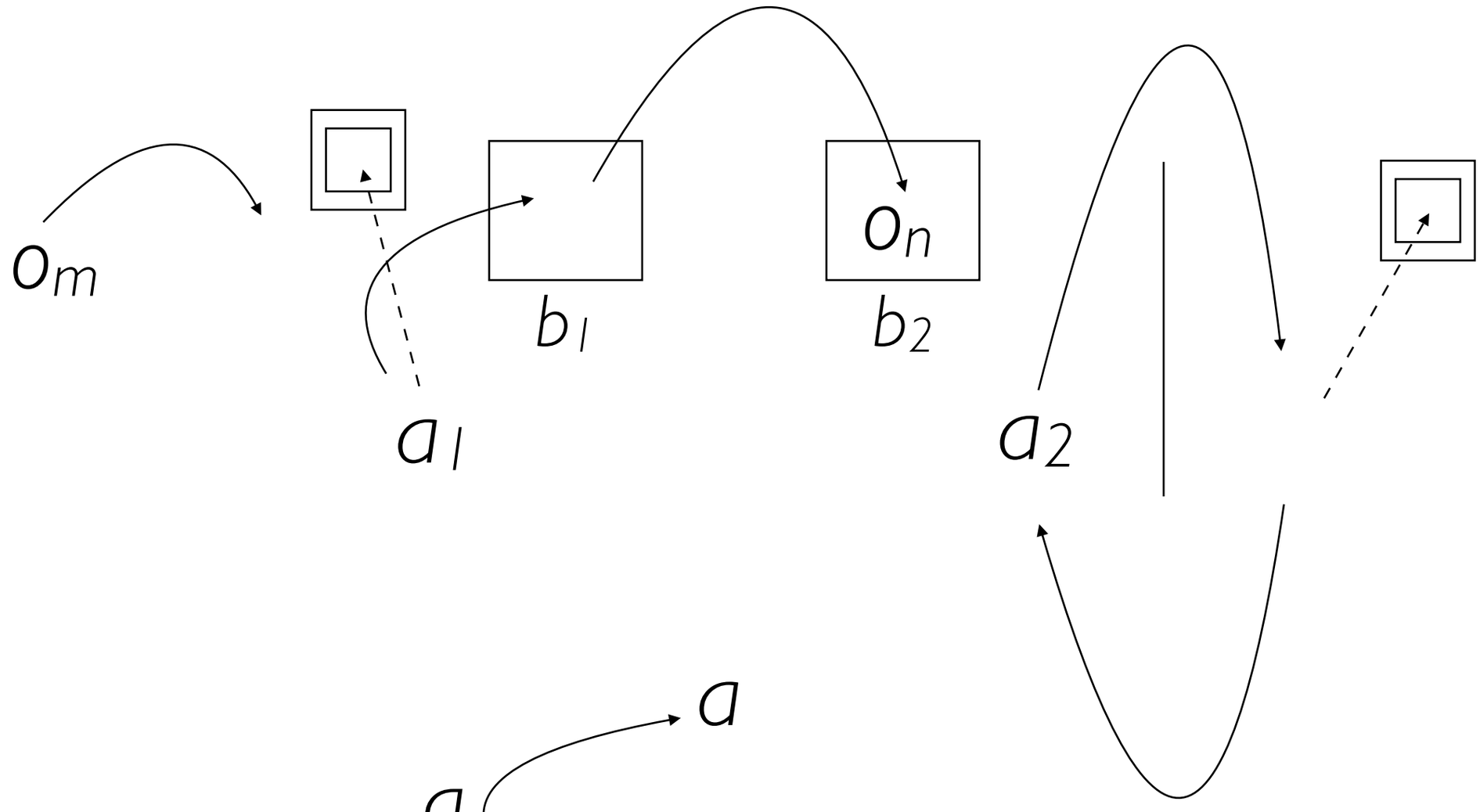
Framework for FBT$^{I}_{3}$
(eight timepoints)

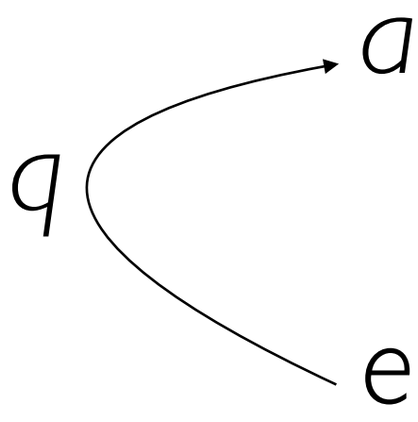


Framework for FBT^{I}_{3}
(eight timepoints)

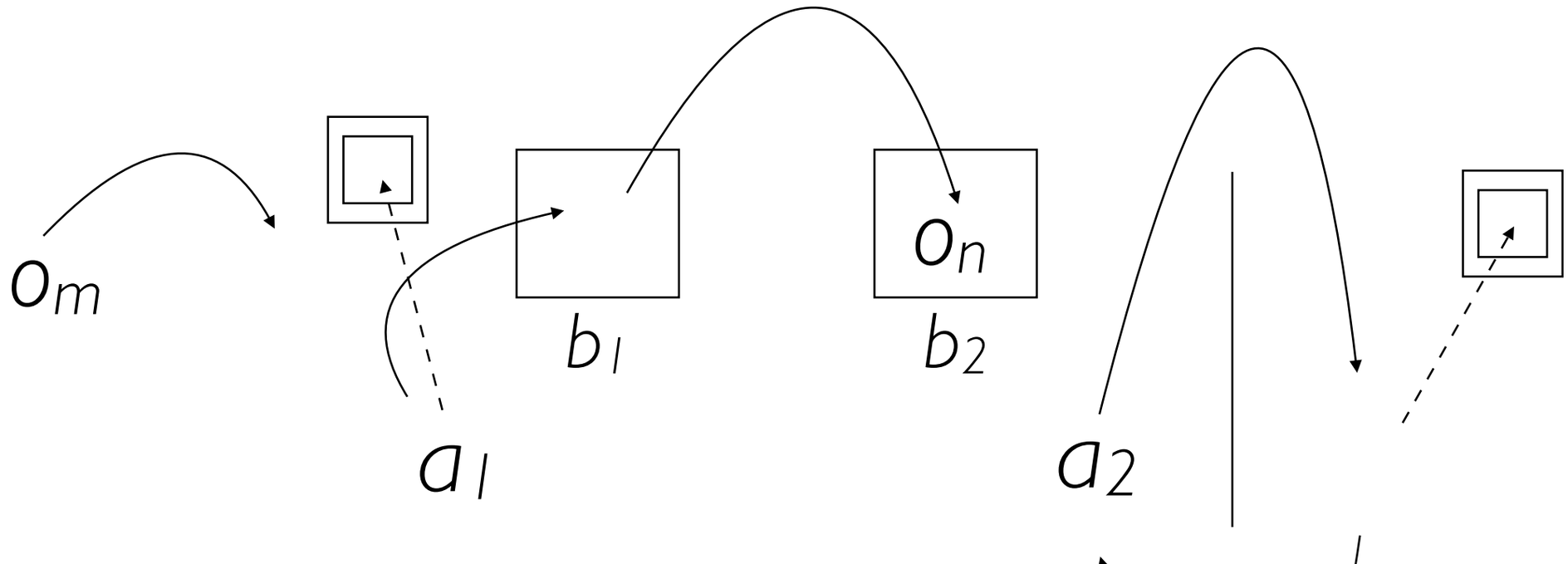


Framework for FBT$^{I}_{3}$
(eight timepoints)

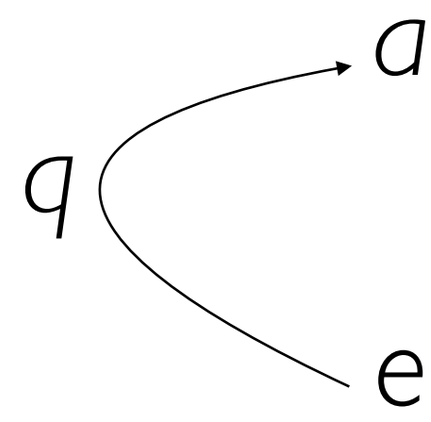




Framework for FBT^{I}_{3}
(eight timepoints)



q a formula in modal $\mathscr{L}^n$



Simulation Courtesy of …

ShadowProver!



Level 1

```
:name        "Level 1: False Belief Task "

:description "Agent a1 puts an object o into b1 in plain view of a2.
              Agent a2 then leaves, and in the absence of a2, a1 moves o
              from b1 into b2 ; this movement isn't perceived by a2 . Agent
              a2 now returns, and a is asked by the experimenter e: "If a2
              desires to retrieve o, which box will a2 look in?" If younger
              than four or five, a will reply "In b " (which of course fails 2
              the task); after this age subjects respond with the correct "In b1."

              Level1 Belief: a1 believes a2 believes o is in b1.
              "

:date        "Monday July 22, 2019"

:assumptions {
               :P1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1)))

               :P2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1))))))

               :P3 (holds (In o b1) t1)

               :C1 (Common! t0 (forall [?f ?t2 ?t2]
                                (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                  (holds ?f ?t2))))

               :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))
               }

:goal        (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3)))}
```

Level 2

```
{:name        "Level 2: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level2 Belief: a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"

 :assumptions {

               :P1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1))))

               :P2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1)))))))

               :P3 (holds (In o b1) t1)


               :C1 (Common! t0
                            (forall [?f ?t2 ?t2]
                                    (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                      (holds ?f ?t2))))

              :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

 :goal        (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3))))}
```

Level 3

```
{:name        "Level 3: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level3 Belief: a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"

 :assumptions {

               :P1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1)))))
               :P2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1))))))))

               :P3 (holds (In o b1) t1)


               :C1 (Common! t0
                            (forall [?f ?t2 ?t2]
                                    (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                      (holds ?f ?t2))))

               :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

 :goal        (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3)))))}
```

Level 4

```
{:name        "Level 4: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level4 Belief: a2 believes a1 believes a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"
```

```
:assumptions {

              :P1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1))))))
              :P2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1)))))))))

              :P3 (holds (In o b1) t1)


              :C1 (Common! t0
                           (forall [?f ?t2 ?t2]
                                   (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                    (holds ?f ?t2))))

              :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

:goal        (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3))))))}
```

Level 5

```
{:name        "Level 5: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level5 Belief: a1 believes a2 believes a1 believes a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"

 :assumptions {
               :P1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1)))))))
               :P2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1)))))))))

               :P3 (holds (In o b1) t1)


               :C1 (Common! t0
                            (forall [?f ?t2 ?t2]
                                    (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                      (holds ?f ?t2))))

               :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

:goal        (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3)))))))}
```

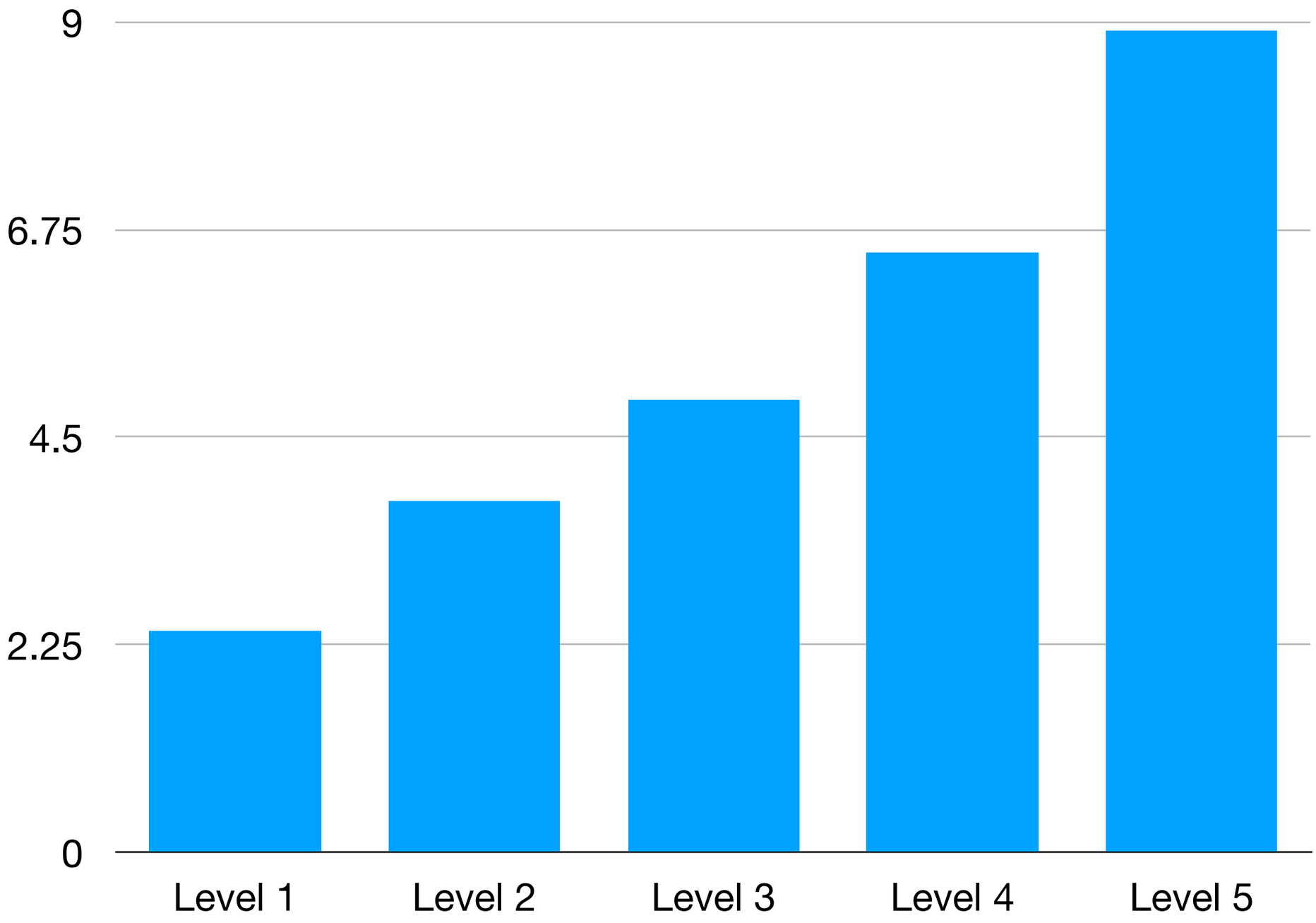
```
...gent
...If a2
...f younger
... course fails 2
...n the correct “In b1.”
...l believes a2 believes a1 believes a2 believes o is in b1.

...t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 ...
...1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not ...
(In o b1) t1)

:goal
...1 (Common! t0
      (forall [?f ?t2 ?t2]
        (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
          (holds ?f ?t2))))
:C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}
(Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3)))))))}
```

```
(Believes a1 t3 (Believes a2 t3 (Believes a1 t3 (Believes
(Believes a1 t3 (Believes a2 t3 (Believes a1 t3 (Believes
(Common i t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))) :C2
(terminates ie ?f)))  (holds ?f ?t1)  (< ?t1 ?t2))
(holds (In o bI) t3)))))))}
(Perceives a1 t1 (Believes a2 t2 (Believes a1 t2 (Believes
(Perceives a2 t1 (Believes a1 t2 (Believes a2 t2 (Believes
(Perceives a1 t1 (Believes a2 t2 (Believes a1 t2 (Believes
(holds (In o bI) t1)
(Common i t0 (forall [?f ?t2 ?t2]
(if (and (not (exists [?e]
(holds ?f ?t2))))
(holds ?f ?t2)))
:goal
bI.
o is in bI.
believes a2 believes a1 believes o is in bI.
believes a2 believes a1 believes a2 believes o is in bI.”
If a2
younger
course fails 2
the correct “In bI.”
believes a2 believes a1 believes a2 believes o is in bI.
```

Time (in seconds) to Prove



Simulation of Level 5 in Real Time

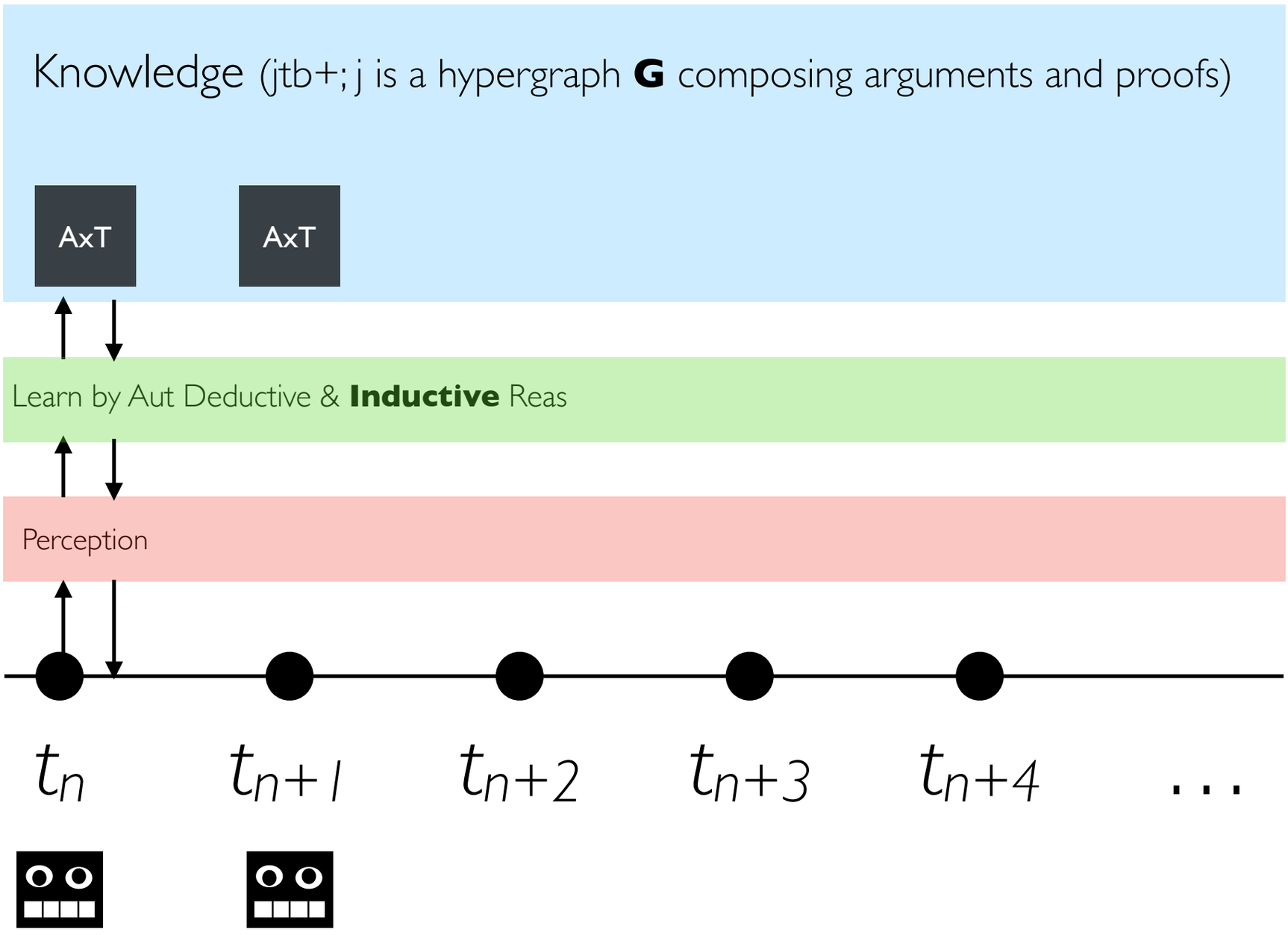
```
/Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java ...
objc[16653]: Class JavaLaunchHelper is implemented in both /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java (0x102a2d4c0) and /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/jre/lib/libinstrument.dylib (0x102ab94e0)
------------ Level 5 ----------------------
```

Simulation of Level 5 in Real Time

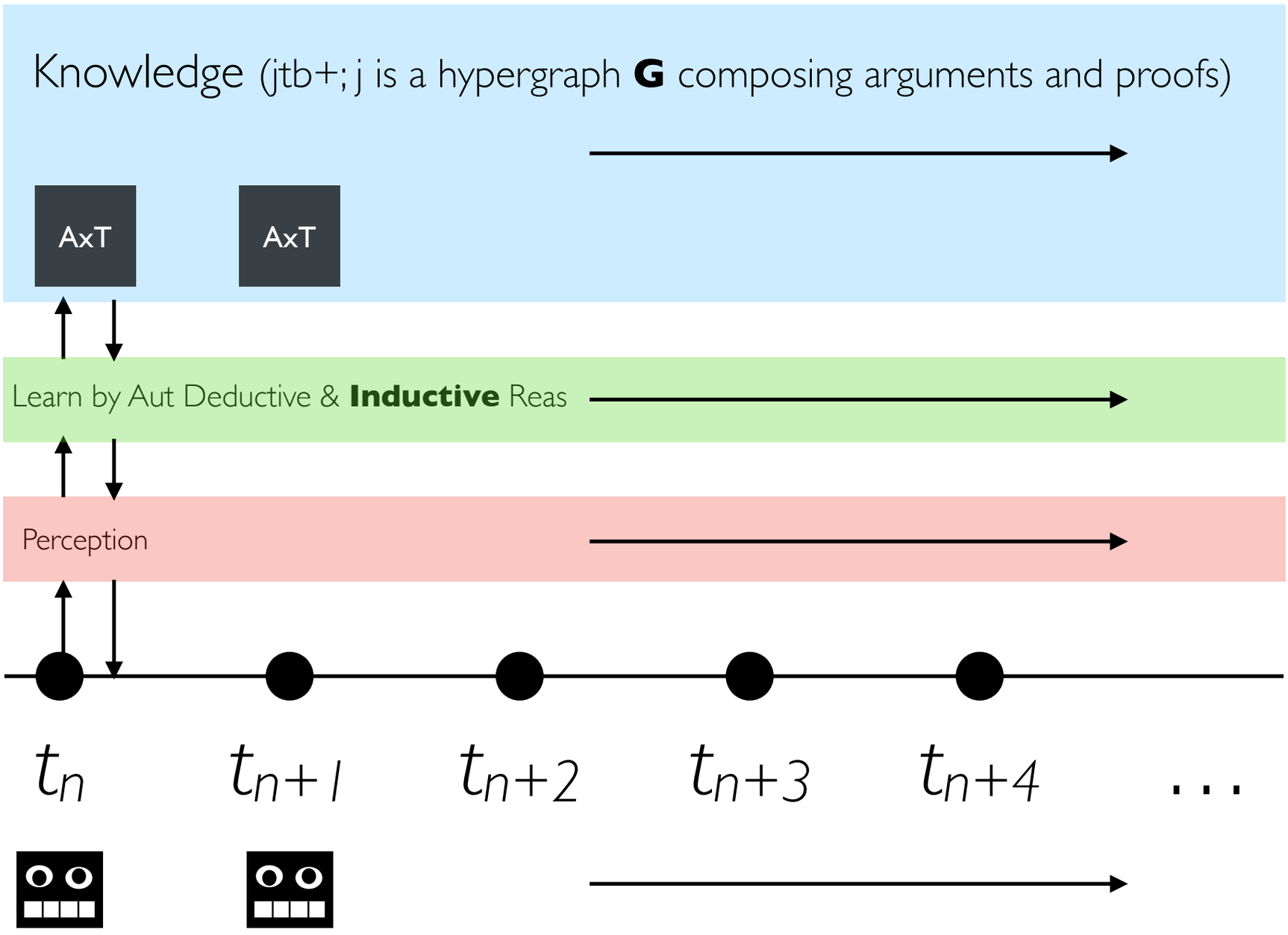
```
/Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java ...
objc[16653]: Class JavaLaunchHelper is implemented in both /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java (0x102a2d4c0) and /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/jre/lib/libinstrument.dylib (0x102ab94e0)
------------ Level 5 ----------------------
```

Real Learning …

Advanced Logicist (Real) Machine Learning

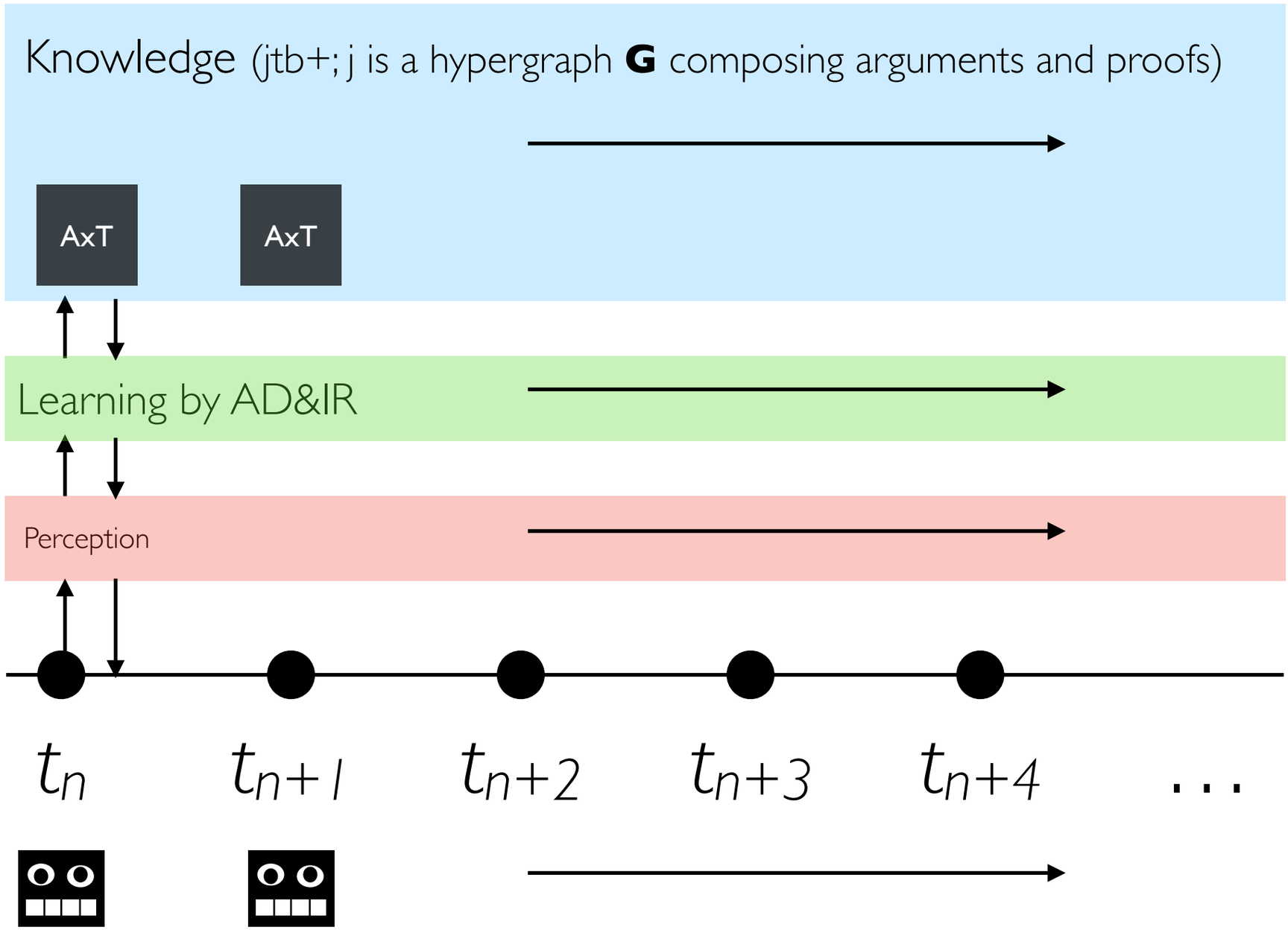


Advanced Logicist (Real) Machine Learning



Advanced Logicist (Real) Machine Learning

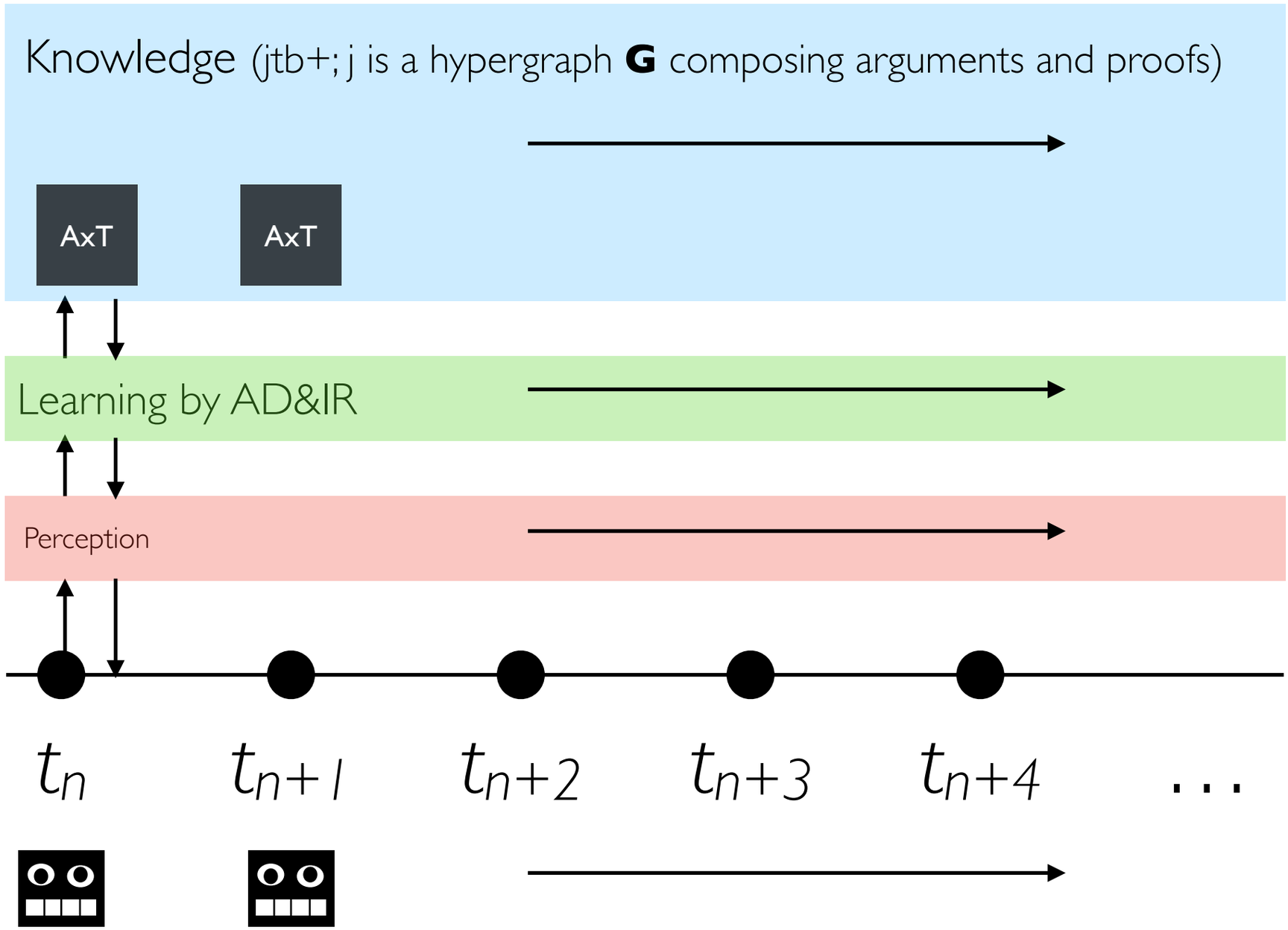
Reasoning:



Advanced Logicist (Real) Machine Learning

Reasoning:

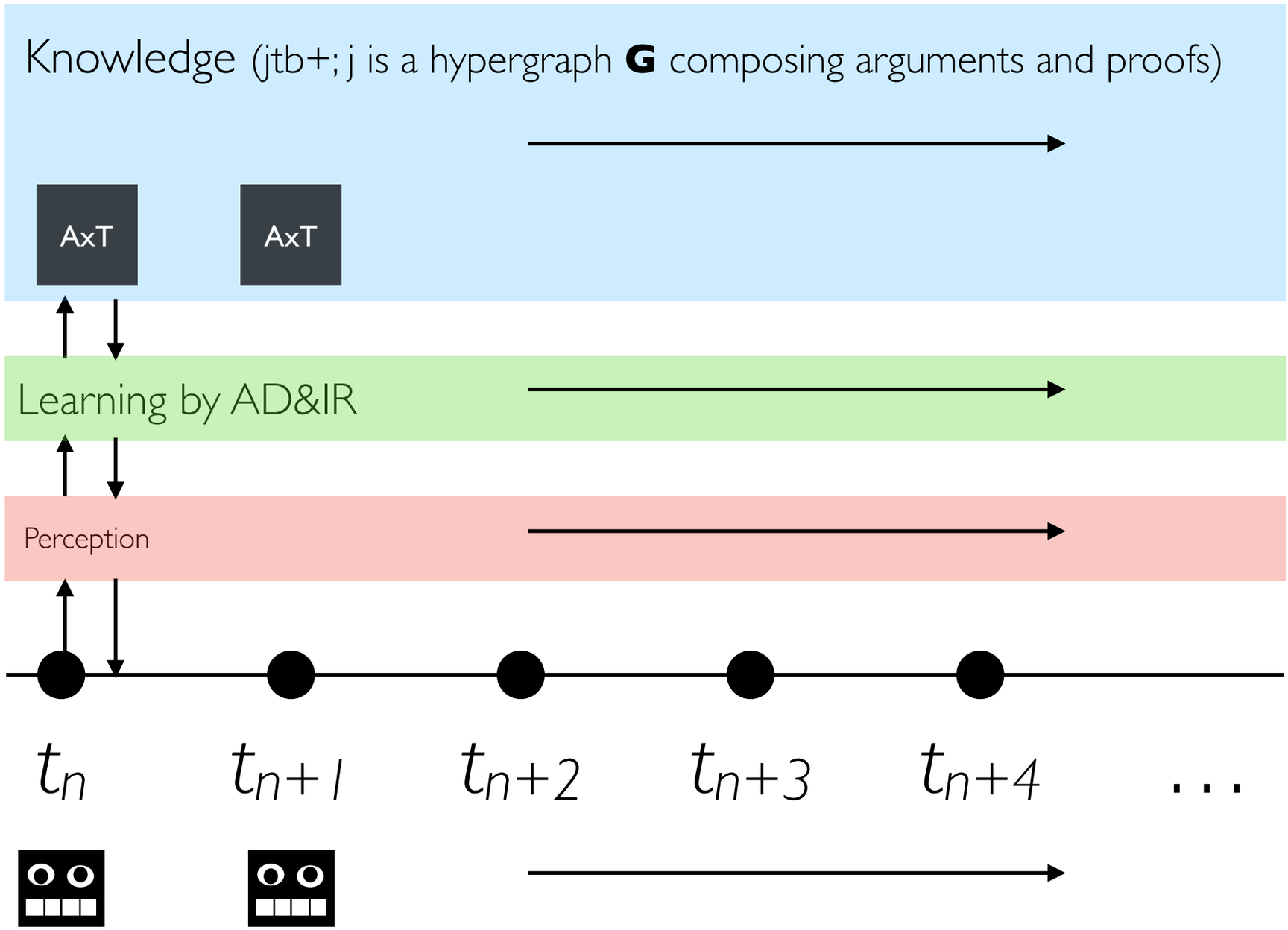
deduction
analogical reasoning
analogico-deductive reasoning
abduction
Bayesian argumentation
enumerative induction
ITBE
…



Advanced Logicist (Real) Machine Learning

Reasoning:

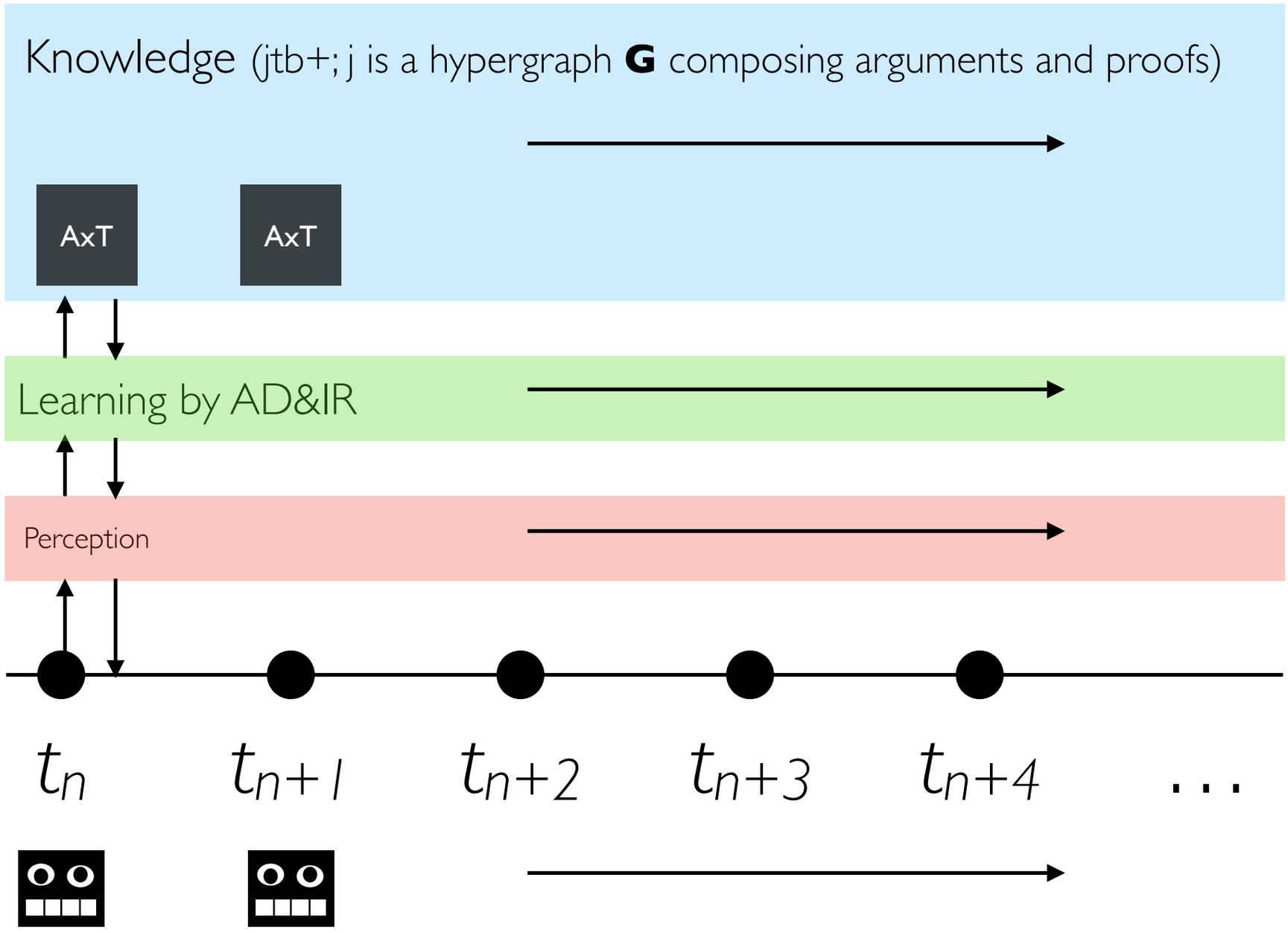
deduction
analogical reasoning
analogico-deductive reasoning
abduction
Bayesian argumentation
enumerative induction
ITBE
…



Advanced Logicist (Real) Machine Learning

Defeasible Reasoning:

deduction
analogical reasoning
analogico-deductive reasoning
abduction
Bayesian argumentation
enumerative induction
ITBE
. . .



Bringsjord, S. & Govindarajulu, N.S. (2018) "The Epistemology of Computer-Mediated Proofs" in Hansson, Sven Ove, ed., *Technology and Mathematics: Philosophical and Historical Investigations* (Berlin, GE: Springer). This book appears in the series *Philosophy of Engineering and Technology* (as Volume 30) edited by Pieter Vermass. ISBN is 978-3-319-93778-6.

Given This, Do Machine-Learning Machines Learn? No.

http://kryten.mm.rpi.edu/SB_NSG_SB_JH_DoMachine-LearningMachinesLearn_preprint.pdf

Given This, Do Machine-Learning Machines Learn? No.

Do Machine-Learning Machines Learn?

Selmer Bringsjord and Naveen Sundar Govindarajulu and Shreya Banerjee and John Hummel

Abstract We answer the present paper's title in the negative. We begin by introducing and characterizing "real learning" ($\mathcal{RL}$) in the formal sciences, a phenomenon that has been firmly in place in homes and schools since at least Euclid. The defense of our negative answer pivots on an integration of *reductio* and proof by cases, and constitutes a general method for showing that any contemporary form of machine learning (ML) isn't real learning. Along the way, we canvass the many different conceptions of "learning" in not only AI, but psychology and its allied disciplines; none of these conceptions (with one exception arising from the view of cognitive development espoused by Piaget), aligns with real learning. We explain in this context by four steps how to broadly characterize and arrive at a focus on $\mathcal{RL}$.

Selmer Bringsjord
Rensselaer Polytechnic Institute, 110 8th Street Troy, NY USA 12180, e-mail: selmerbringsjord@gmail.com

Naveen Sundar Govindarajulu
Rensselaer Polytechnic Institute, 110 8th Street Troy, NY USA 12180, e-mail: Naveen.Sundar.G@gmail.com

Shreya Banerjee
Rensselaer Polytechnic Institute, 110 8th Street Troy, NY USA 12180, e-mail: shreyabbanerjee@gmail.com

John Hummel
901 West Illinois Street, Urbana, IL 61801, e-mail: jehummel@illinois.edu

8 Appendix: The Formal Method

The following deduction uses fonts in an obvious and standard way to sort between functions ($\mathfrak{f}$), agents ($\mathfrak{a}$), and computing machines ($\mathfrak{m}$) in the Arithmetical Hierarchy. Ordinary italicized Roman is used for particulars under these sorts (e.g. f is a particular function). In addition, '$\mathcal{C}$' denotes any collection of conditions constituting jointly necessary-and-sufficient conditions for a form of current ML, which can come from relevant textbooks (e.g. Luger, 2008; Russell and Norvig, 2009) or papers; we leave this quite up to the reader, as no effect upon the validity of the deductive inference chain will be produced by the preferred instantiation of '$\mathcal{C}$.' It will perhaps be helpful to the reader to point out that the deduction eventuates in the proposition that no machine in the ML fold that in this style learns a relevant function $\mathfrak{f}$ thereby also real-learns $\mathfrak{f}$. We encode this target as follows:

$$(\star)\quad \neg\exists \mathfrak{m}\ \exists \mathfrak{f}\ [\phi := \mathit{MLlearns}(\mathfrak{m},\mathfrak{f}) \wedge \psi := \mathit{RLlearns}(\mathfrak{m},\mathfrak{f}) \wedge \mathcal{C}_{\phi}(\mathfrak{m},\mathfrak{f}) \vdash^{*} (\mathrm{ci}')\text{–}(\mathrm{ciii})_{\psi}(\mathfrak{m},\mathfrak{f})]$$

Note that $(\star)$ employs meta-logical machinery to refer to particular instantiations of $\mathcal{C}$ for a particular, arbitrary case of ML (ϕ is the atomic sub-formula that can be instantiated to make the particular case), and particular instantiations of the triad (ci′)–(ciii) for a particular, arbitrary case of $\mathcal{RL}$ (ψ is the atomic sub-formula that can be instantiated to make the particular case). Meta-logical machinery also allows us to use a provability predicate to formalize the notion that real learning is produced by the relevant instance of ML. If we "pop" ϕ/ψ to yield ϕ'/ψ' we are dealing with the particular instantiation of the atomic sub-formula.

The deduction, as noted in earlier when the informal argument was given, is indirect proof by cases; accordingly, we first assume $\neg(\star)$, and then proceed as follows under this supposition.

	(1)	$\forall \mathfrak{f},\mathfrak{a}\ [\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \rightarrow (\mathit{RLlearns}(\mathfrak{a},\mathfrak{f}) \rightarrow \mathrm{(i)}\text{–}\mathrm{(iii)})]$	Def of Real Learning
	(2)	$\mathit{MLlearns}(m,f) \wedge \mathit{RLlearns}(m,f) \wedge f : \mathbb{N} \mapsto \mathbb{N}$	supp (for $\exists$ elim on $(\star)$)
	(3)	$\forall \mathfrak{m},\mathfrak{f}\ [\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \rightarrow (\mathit{MLlearns}(\mathfrak{m},\mathfrak{f}) \leftrightarrow \mathcal{C}(\mathfrak{m},\mathfrak{f}))]$	Def of ML
	(4)	$\forall \mathfrak{f}\ [\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \rightarrow (\mathit{TurComp}(\mathfrak{f}) \vee \mathit{TurUncomp}(\mathfrak{f}))]$	theorem
	(5)	$\mathit{TurUncomp}(f)$	supp; Case 1
	(6)	$\neg\exists \mathfrak{m}\ \exists \mathfrak{f}\ [(\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \wedge \mathit{TurUncomp}(\mathfrak{f}) \wedge \mathcal{C}(\mathfrak{m},\mathfrak{f})]$	theorem
$\therefore$	(7)	$\neg\exists \mathfrak{m}\ \mathit{MLlearns}(\mathfrak{m},f)$	(6), (3)
$\therefore$	(8)	$\bot$	(7), (2)
	(9)	$\mathit{TurComp}(f)$	supp; Case 2
$\therefore$	(10)	$\mathcal{C}_{\phi'}(m,f)$	(2), (3)
$\therefore$	(11)	$(\mathrm{ci}')\text{–}(\mathrm{ciii})_{\psi'}(m,f)$	from supp for $\exists$ elim on $(\star)$ and provability
$\therefore$	(12)	$\neg(\mathrm{ci}')\text{–}(\mathrm{ciii})_{\psi'}(m,f)$	inspection: proofs wholly absent from $\mathcal{C}$
$\therefore$	(13)	$\bot$	(11), (12)
$\therefore$	(14)	$\bot$	*reductio*; proof by cases

http://kryten.mm.rpi.edu/SB_NSG_SB_JH_DoMachine-LearningMachinesLearn_preprint.pdf

Det er en logikk for hvert problem!