version 1026201645NY

The Argument for God's Existence

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

IFLAI2
10/26/2020

R A I R
Rensselaer AI and Reasoning Lab

version 1026201645NY

The Argument for God’s Existence from AI

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

IFLAI2
10/26/2020

Rensselaer AI and Reasoning Lab

The Ontological/Modal Argument Meets AI

The Ontological/Modal Argument Meets AI

Scott's version of Gödel verified!

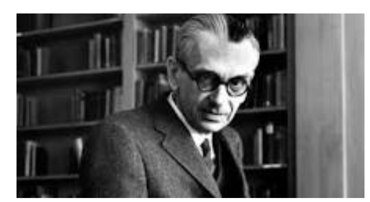
Gödel's version of Gödel falsified!

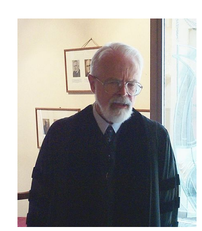
Anselm

Descartes

Leibniz

Gödel/Adams

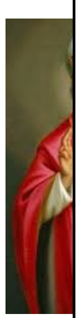




The … s AI

…el's
…n of
…lsified!

An…



ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-93

Automating Gödel's Ontological Proof of God's Existence with Higher-order Automated Theorem Provers

Christoph Benzmüller[1] and **Bruno Woltzenlogel Paleo**[2]

Abstract. Kurt Gödel's ontological argument for God's existence has been formalized and automated on a computer with higher-order automated theorem provers. From Gödel's premises, the computer proved: necessarily, there exists God. On the other hand, the theorem provers have also confirmed prominent criticism on Gödel's ontological argument, and they found some new results about it.

The background theory of the work presented here offers a novel perspective towards a *computational theoretical philosophy*.

1 INTRODUCTION

Kurt Gödel proposed an argumentation formalism to prove the existence of God [23, 30]. Attempts to prove the existence (or non-existence) of God by means of abstract, ontological arguments are an old tradition in western philosophy. Before Gödel, several prominent philosophers, including St. Anselm of Canterbury, Descartes and Leibniz, have presented similar arguments. Moreover, there is an impressive body of recent and ongoing work (cf. [31, 19, 18] and the references therein). Ontological arguments, for or against the existence of God, illustrate well an essential aspect of metaphysics: some (necessary) facts for our existing world are deduced by purely a priori, analytical means from some abstract definitions and axioms.

What motivated Gödel as a logician was the question, whether it

A1	Either a property or its negation is positive, but not both:	$\forall\phi[P(\neg\phi) \equiv \neg P(\phi)]$
A2	A property necessarily implied by a positive property is positive:	$\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \supset \psi(x)]) \supset P(\psi)]$
T1	Positive properties are possibly exemplified:	$\forall\phi[P(\phi) \supset \Diamond\exists x\phi(x)]$
D1	A *God-like* being possesses all positive properties:	$G(x) \equiv \forall\phi[P(\phi) \supset \phi(x)]$
A3	The property of being God-like is positive:	$P(G)$
C	Possibly, God exists:	$\Diamond\exists x G(x)$
A4	Positive properties are necessarily positive:	$\forall\phi[P(\phi) \supset \Box P(\phi)]$
D2	An *essence* of an individual is a property possessed by it and necessarily implying any of its properties:	$\phi\ ess.\ x \equiv \phi(x) \wedge \forall\psi(\psi(x) \supset \Box\forall y(\phi(y) \supset \psi(y)))$
T2	Being God-like is an essence of any God-like being:	$\forall x[G(x) \supset G\ ess.\ x]$
D3	*Necessary existence* of an individ. is the necessary exemplification of all its essences:	$NE(x) \equiv \forall\phi[\phi\ ess.\ x \supset \Box\exists y\phi(y)]$
A5	Necessary existence is a positive property:	$P(NE)$
T3	Necessarily, God exists:	$\Box\exists x G(x)$

Figure 1. Scott's version of Gödel's ontological argument [30].

T

Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

The Inconsistency in Gödel's Ontological Argument: A Success Story for AI in Metaphysics

Christoph Benzmüller*
Freie Universität Berlin & Stanford University
c.benzmueller@gmail.com

Bruno Woltzenlogel Paleo
Australian National University
bruno.wp@gmail.com

Abstract

This paper discusses the discovery of the inconsistency in Gödel's ontological argument as a success story for artificial intelligence. Despite the popularity of the argument since the appearance of Gödel's manuscript in the early 1970's, the inconsistency of the axioms used in the argument remained unnoticed until 2013, when it was detected automatically by the higher-order theorem prover

on the proof [Fuhrmann, 2016].

The in-depth analysis presented here substantially extends previous computer-assisted studies of Gödel's ontological argument. Similarly to the related work [Benzmüller and Woltzenlogel-Paleo, 2013a; 2014] the analysis has been conducted with automated theorem provers for classical higher-order logic (HOL; cf. [Andrews, 2014] and the references therein), even though Gödel's proof is actually formulated in higher-order *modal* logic (HOML; cf. [Muskens, 2006]

some (necessary) facts for our existing world are deduced by purely a priori, analytical means from some abstract definitions and axioms.

What motivated Gödel as a logician was the question, whether it

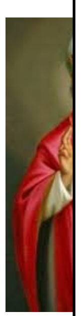
T3	Necessarily, God exists:	$\Box\exists x G(x)$

Figure 1. Scott's version of Gödel's ontological argument [30].

The ... s AI

...el's ... n of ... lsified!

Ar...



ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-93

Automating Gödel's Ontological Proof of God's Existence with Higher-order Automated Theorem Provers

Christoph Benzmüller[1] and **Bruno Woltzenlogel Paleo**[2]

Abstract. Kurt Gödel's ontological argument for God's existence has been formalized and automated on a computer with higher-order automated theorem provers. From Gödel's premises, the computer proved: necessarily, there exists God. On the other hand, the theorem provers have also confirmed prominent criticism on Gödel's ontological argument, and they found some new results about it.

The background theory of the work presented here offers a novel perspective towards a *computational theoretical philosophy*.

1 INTRODUCTION

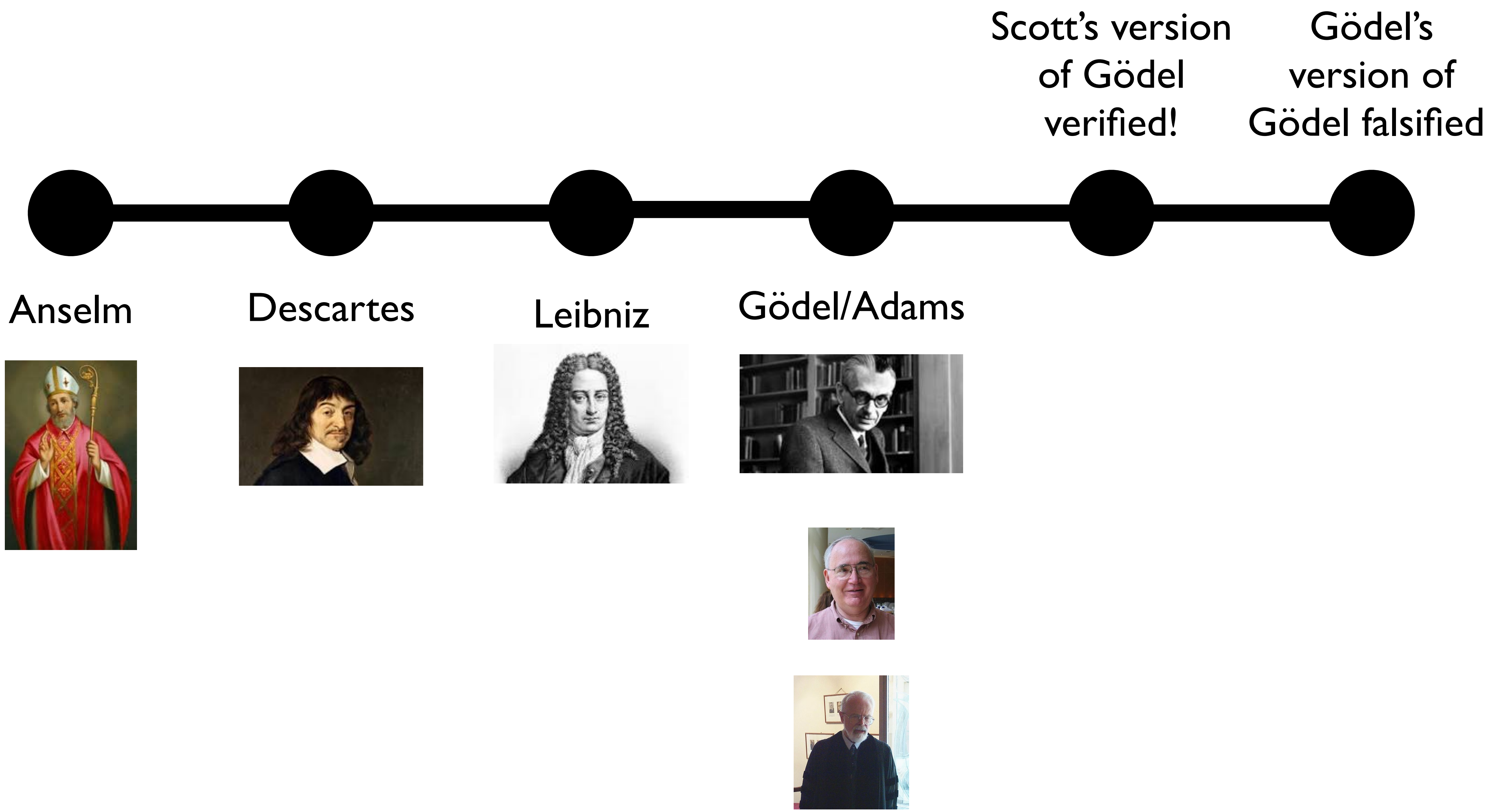
Kurt Gödel proposed an argumentation formalism to prove the existence of God [23, 30]. Attempts to prove the existence (or non-existence) of God by means of abstract, ontological arguments are an old tradition in western philosophy. Before Gödel, several prominent philosophers, including St. Anselm of Canterbury, Descartes and Leibniz, have presented similar arguments. Moreover, there is an impressive body of recent and ongoing work (cf. [31, 19, 18] and the references therein). Ontological arguments, for or against the existence of God, illustrate well an essential aspect of metaphysics: some (necessary) facts for our existing world are deduced by purely a priori, analytical means from some abstract definitions and axioms.

What motivated Gödel as a logician was the question, whether it

A1	Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \equiv \neg P(\phi)]$
A2	A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \supset \psi(x)]) \supset P(\psi)]$
T1	Positive properties are possibly exemplified: $\forall\phi[P(\phi) \supset \Diamond\exists x\phi(x)]$
D1	A *God-like* being possesses all positive properties: $G(x) \equiv \forall\phi[P(\phi) \supset \phi(x)]$
A3	The property of being God-like is positive: $P(G)$
C	Possibly, God exists: $\Diamond\exists x G(x)$
A4	Positive properties are necessarily positive: $\forall\phi[P(\phi) \supset \Box\, P(\phi)]$
D2	An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \equiv \phi(x) \wedge \forall\psi(\psi(x) \supset \Box\forall y(\phi(y) \supset \psi(y)))$
T2	Being God-like is an essence of any God-like being: $\forall x[G(x) \supset G\ ess.\ x]$
D3	*Necessary existence* of an individ. is the necessary exemplification of all its essences: $NE(x) \equiv \forall\phi[\phi\ ess.\ x \supset \Box\exists y\phi(y)]$
A5	Necessary existence is a positive property: $P(NE)$
T3	Necessarily, God exists: $\Box\exists x G(x)$

Figure 1. Scott's version of Gödel's ontological argument [30].

The Ontological/Modal Argument Meets AI



?

The Ontological/Modal Argument Meets AI

Scott's version of Gödel verified!

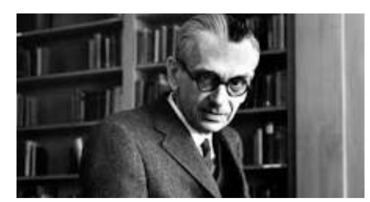
Gödel's version of Gödel falsified!

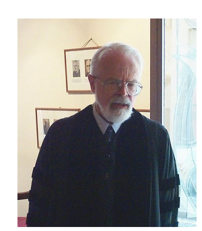
Anselm

Descartes

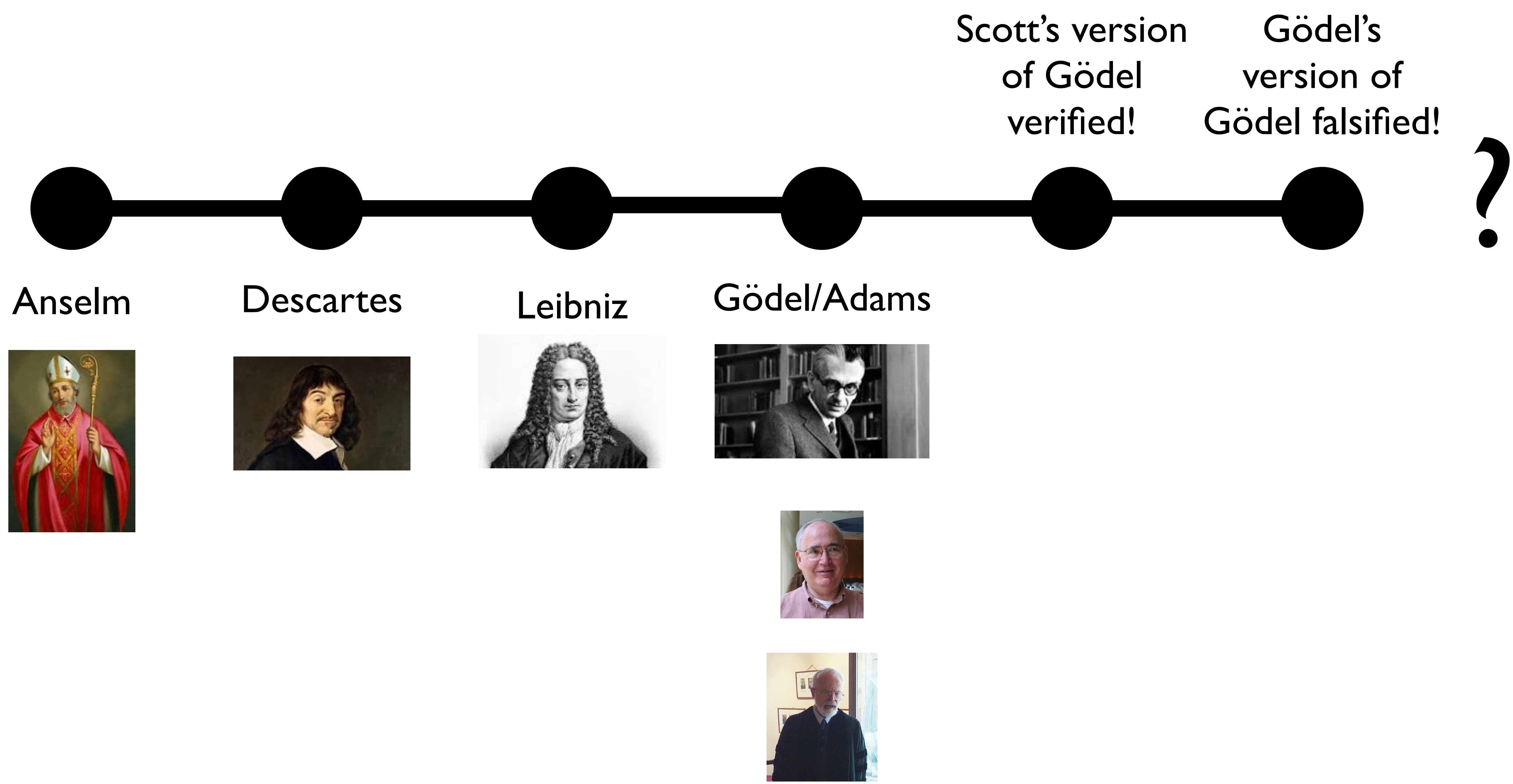
Leibniz

Gödel/Adams





The Ontological/Modal Argument Meets AI



This is argumentation for God's existence that *uses* AI; it's not *based* on AI.

This is argumentation for God's existence that *uses* AI; it's not *based* on AI.

Today, The Argument for God's Existence *from* AI …

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

gradual increase in cognitive powers

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

gradual increase in cognitive powers

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

gradual increase in cognitive powers

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

gradual increase in
cognitive powers

%% SELMER: Add in e.g. neanderthals.

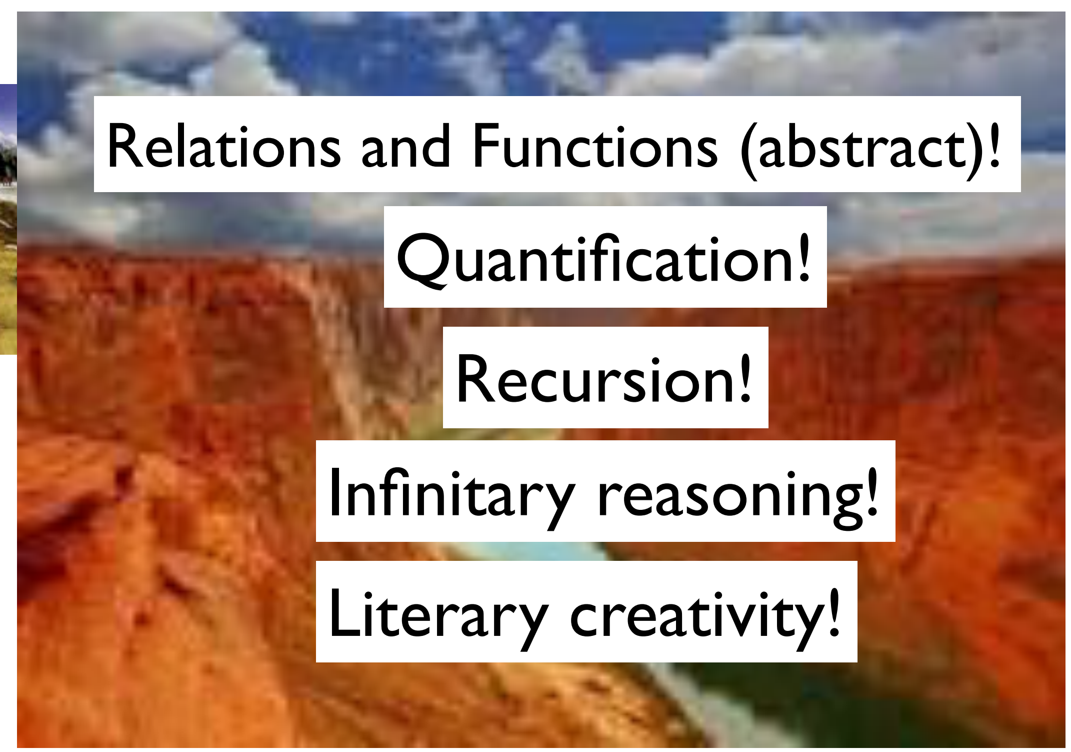
The Canyon of Discontinuity (or Darwin's Dread)

gradual increase in
cognitive powers

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

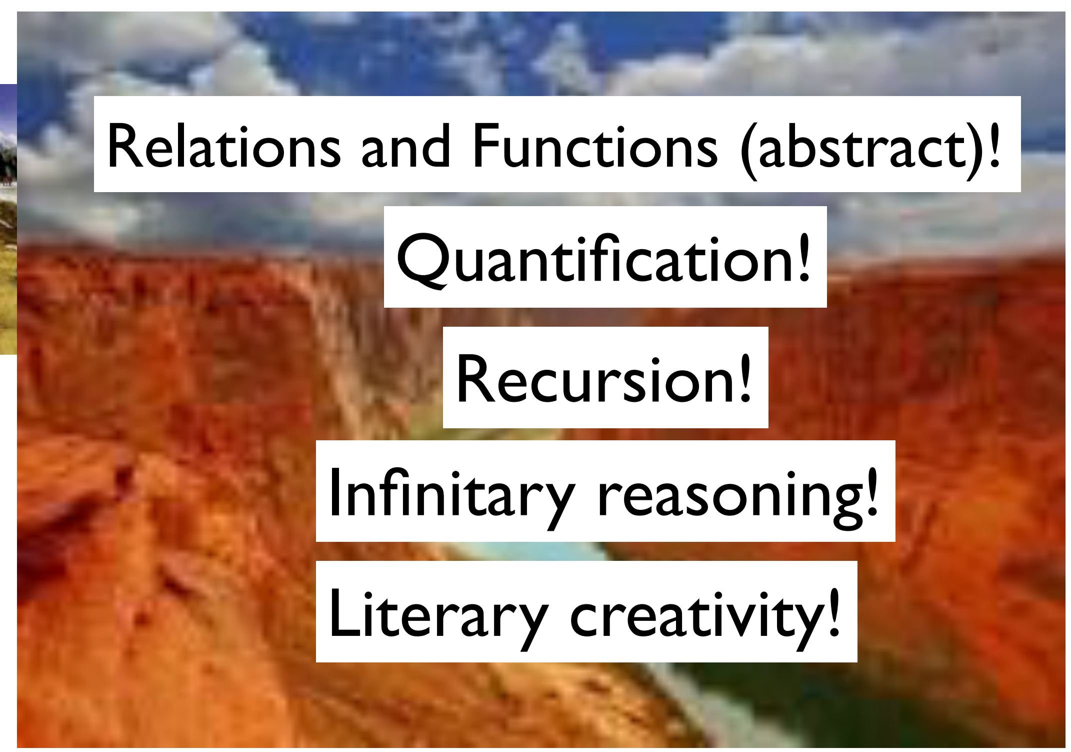



gradual increase in cognitive powers

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin’s Dread)



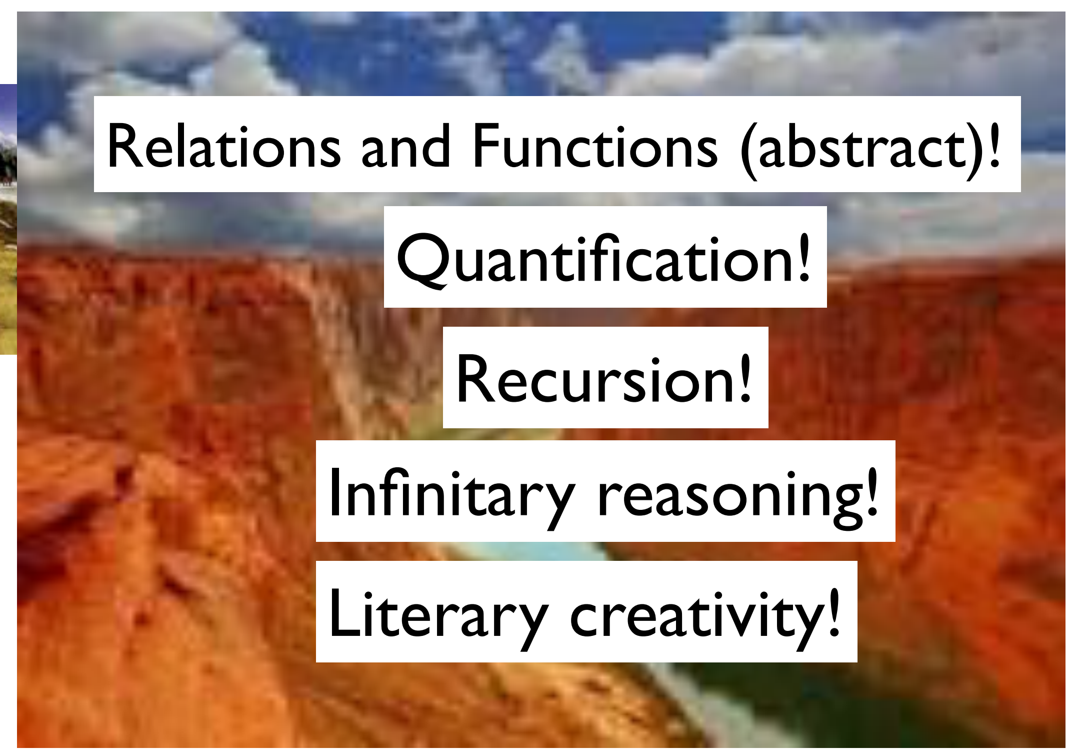
gradual increase in cognitive powers

P^{∞}

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

P^{HG}



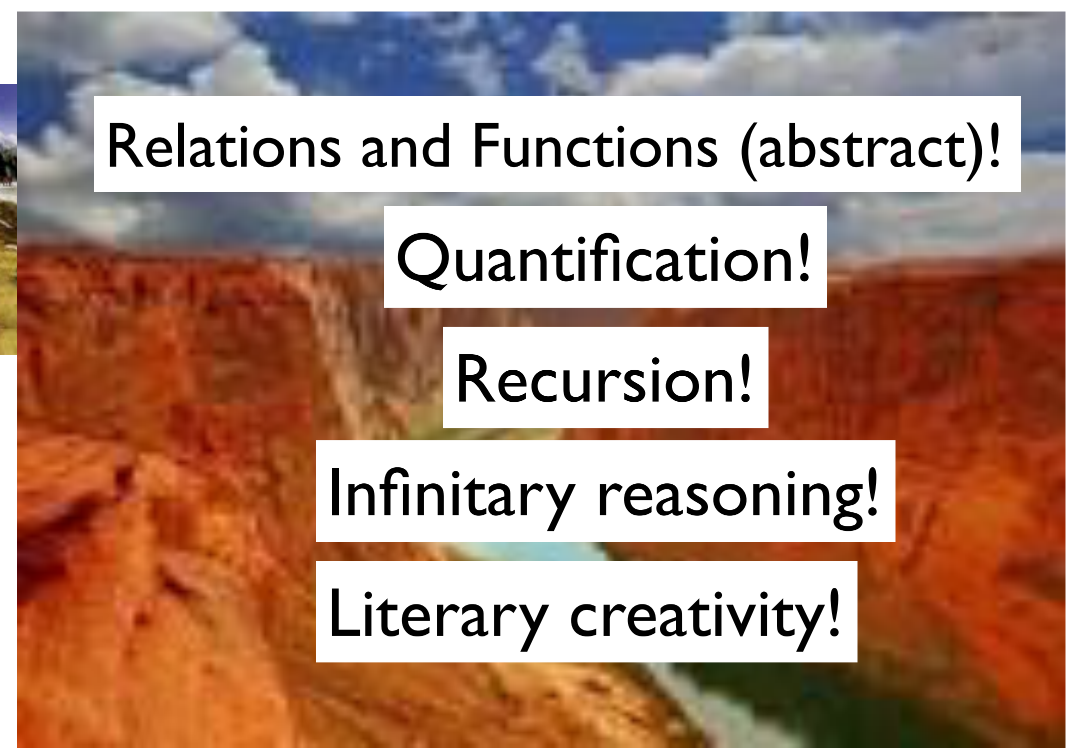
gradual increase in cognitive powers

P^{∞}

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

P^{AI}



P^{HG}

gradual increase in cognitive powers

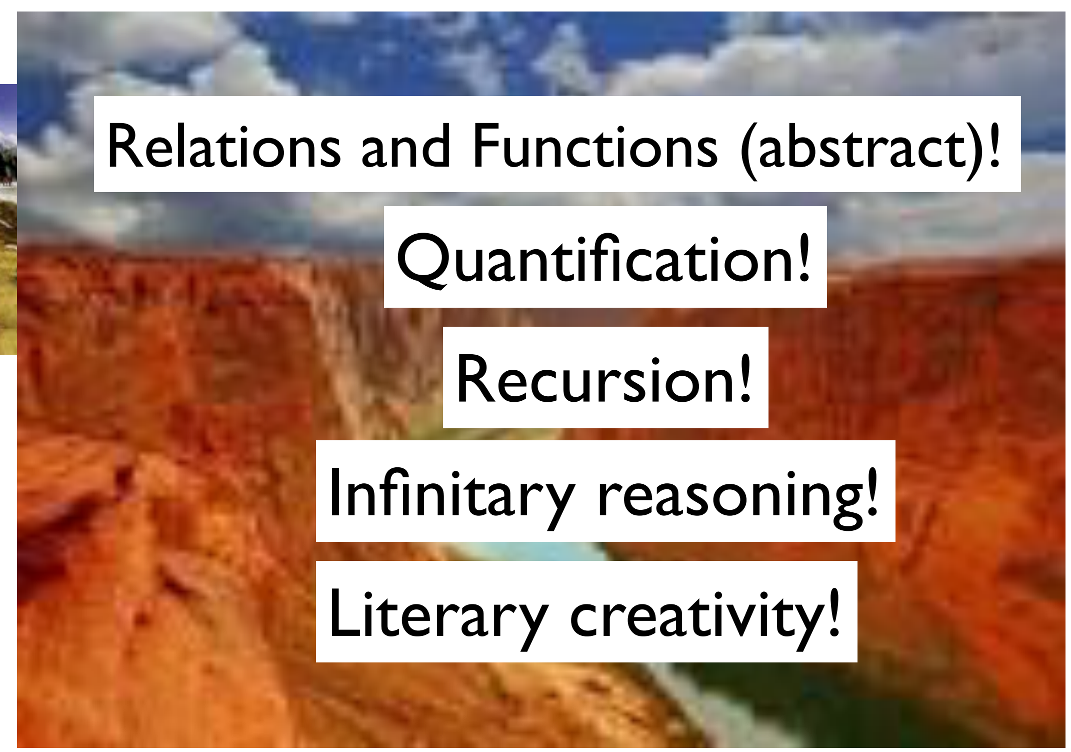
P^{∞}

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

P^{AI}

P^{HG}



P^{∞}

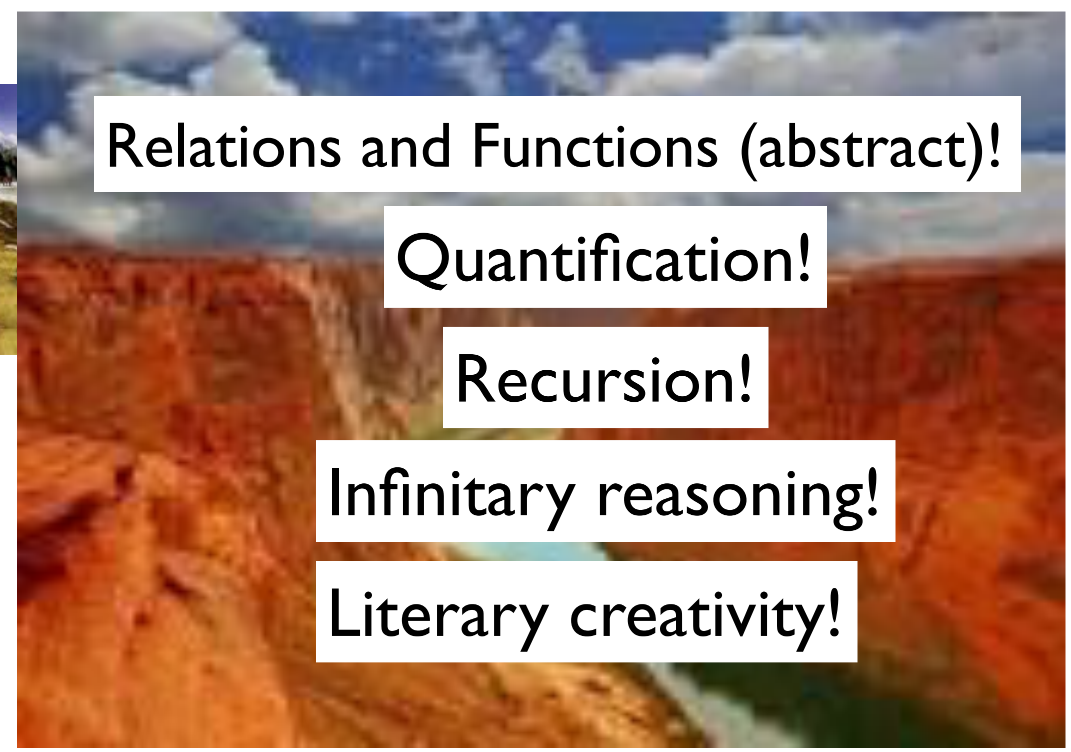
gradual increase in cognitive powers

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

P^{AI}

P^{HG}



gradual increase in cognitive powers

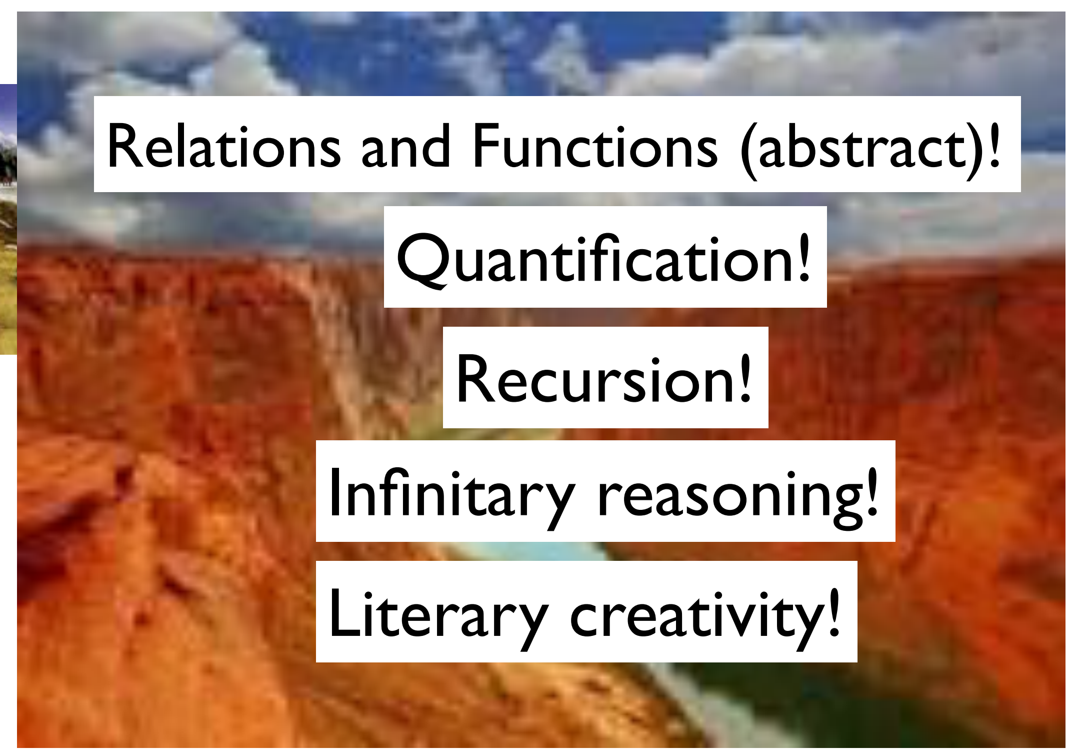
P^{∞}

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

P^{AI}

P^{HG}



P^{∞}

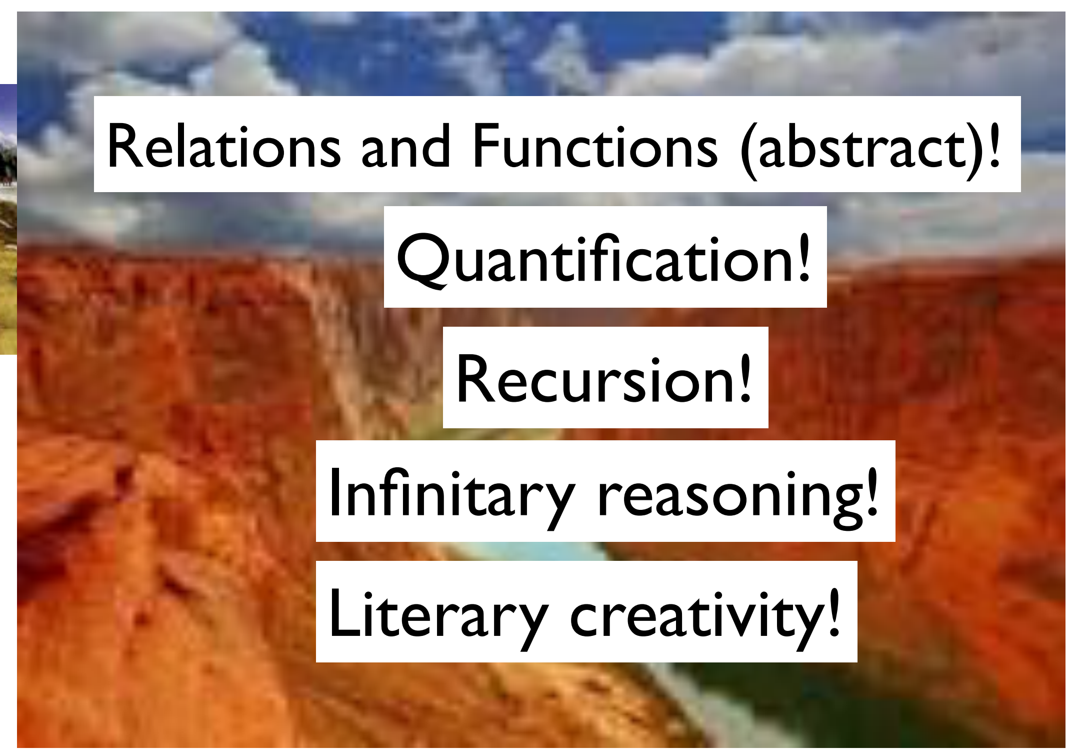
gradual increase in cognitive powers

%% SELMER: Add in e.g. neanderthals.

The Canyon of Discontinuity (or Darwin's Dread)

P^{AI}

P^{HG}



gradual increase in cognitive powers

P^{∞}

The Argument

	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/. . . .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < \ P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ (where P^{AI} is a limit on the cognitive power of AI), and AIs can hunt and gather.	see AI today
∴	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P^{\text{HG}})$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
∴	(8)	God exists.	*modus ponens* (5), (6), (7)

The Argument

	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/. . . .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ (where P^{AI} is a limit on the cognitive power of AI), and AIs can hunt and gather.	see AI today
∴	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P^{\text{HG}})$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
∴	(8)	God exists.	*modus ponens* (5), (6), (7)

The Argument

	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/. . . .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ (where P^{AI} is a limit on the cognitive power of AI), and AIs can hunt and gather.	see AI today
∴	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P^{\text{HG}})$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
∴	(8)	God exists.	*modus ponens* (5), (6), (7)

Set "subtraction": The extreme cognitive powers, but none of the routine ones. So, that which we share with hunter-gatherer activity & the lower animals is irrelevant.

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) “Recursive Syntactic Pattern Learning by Songbirds” *Nature* **440**: 1204–1207.

Humans

Songbirds

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

Songbirds

the birds sing differentially based on what song they hear

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

$\Delta_h^B := (ab)^n$ and $a^n b^n$ are (trivially) both human-decidable

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

$\Delta_h^B := (ab)^n$ and $a^n b^n$ are (trivially) both human-decidable

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

$\Delta_h^B := (ab)^n$ and $a^n b^n$ are (trivially) both human-decidable

---

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

$\Delta_h^B := (ab)^n$ and $a^n b^n$ are (trivially) both human-decidable

---

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Some elements of some formalized nonhuman-animal behavior overlap some elements of some formalized human behavior.

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

$\Delta_h^B := (ab)^n$ and $a^n b^n$ are (trivially) both human-decidable

---

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Some elements of some formalized nonhuman-animal behavior overlap some elements of some formalized human behavior. $\Delta_a^B \cap \Delta_h^B \neq \emptyset!$

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

$\Delta_h^B := (ab)^n$ and $a^n b^n$ are (trivially) both human-decidable

---

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Some elements of some formalized nonhuman-animal behavior overlap some elements of some formalized human behavior. $\Delta_a^B \cap \Delta_h^B \neq \emptyset!$

After all, birds can see; humans can too; but but this is not relevant to my present purposes, which is to focus on:

Gentner, T., Fenn, K., Margoliash, D. & Nusbaum, H. (2006) "Recursive Syntactic Pattern Learning by Songbirds" *Nature* **440**: 1204–1207.

Humans

humans can easily enough decide these languages

$\Delta_h^B := (ab)^n$ and $a^n b^n$ are (trivially) both human-decidable

- -

Songbirds

the birds sing differentially based on what song they hear

a "rattle" coded as *a*; a "warble" as *b*

$\Delta_a^B := (ab)^n$ and $a^n b^n$ are *both* starling-decidable

Some elements of some formalized nonhuman-animal behavior overlap some elements of some formalized human behavior. $\Delta_a^B \cap \Delta_h^B \neq \emptyset!$

After all, birds can see; humans can too; but but this is not relevant to my present purposes, which is to focus on:

$$P^{\infty} - P^{\mathrm{HG}} / P^{\infty} - P^{\mathrm{AI}}$$

The Argument

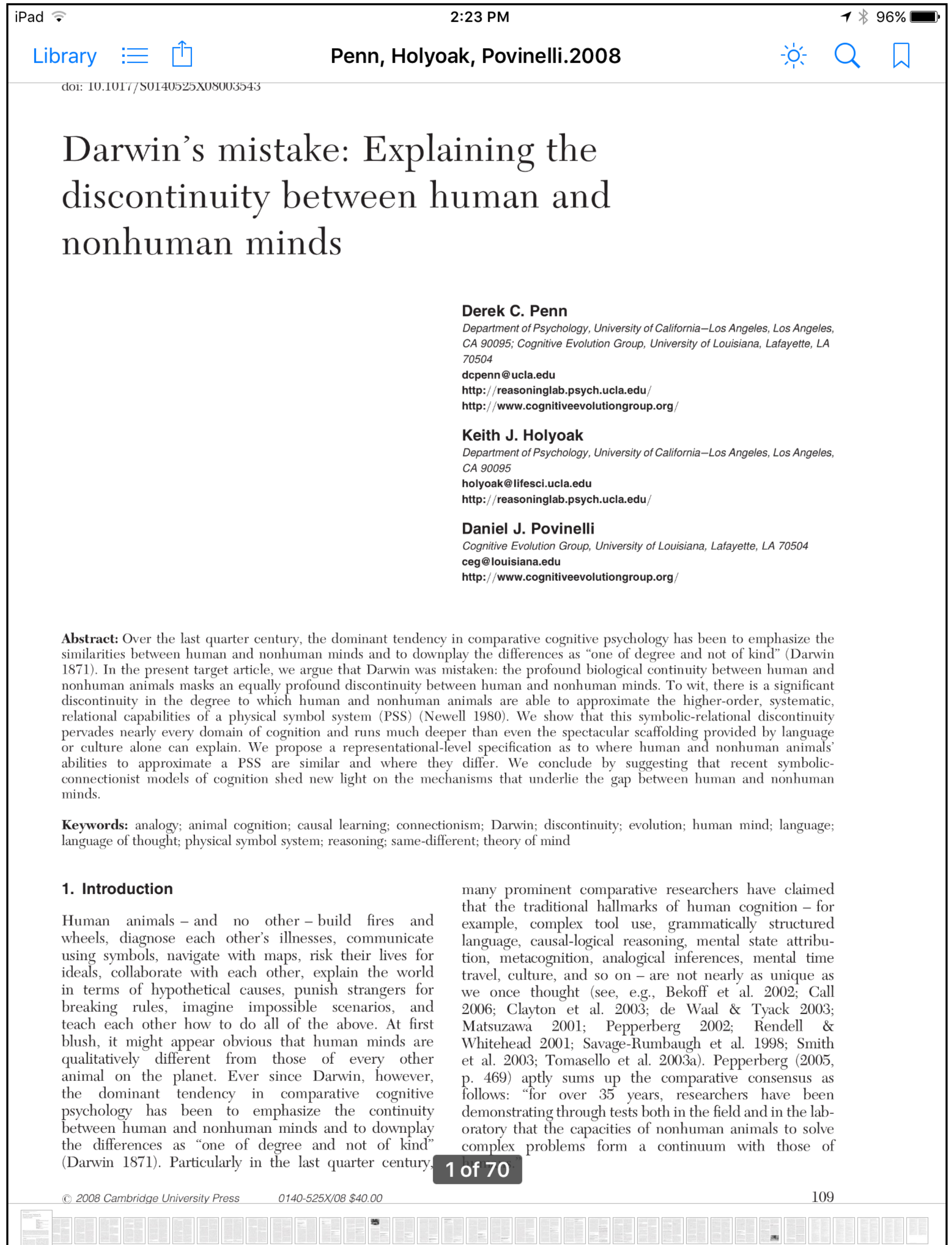
	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/… .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ and, where P^{AI} is a limit on the cognitive power of AI, AIs can hunt and gather.	see AI today
∴	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P)$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
∴	(8)	God exists.	*modus ponens* (5), (6), (7)

The Argument

	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/… .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ and, where P^{AI} is a limit on the cognitive power of AI, AIs can hunt and gather.	see AI today
$\therefore$	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P)$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
$\therefore$	(8)	God exists.	*modus ponens* (5), (6), (7)

See …

See …


iPad 2:23 PM 96%

Library Penn, Holyoak, Povinelli.2008

doi: 10.1017/S0140525X08003543

Darwin's mistake: Explaining the discontinuity between human and nonhuman minds

Derek C. Penn
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095; Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
dcpenn@ucla.edu
http://reasoninglab.psych.ucla.edu/
http://www.cognitiveevolutiongroup.org/

Keith J. Holyoak
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095
holyoak@lifesci.ucla.edu
http://reasoninglab.psych.ucla.edu/

Daniel J. Povinelli
Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
ceg@louisiana.edu
http://www.cognitiveevolutiongroup.org/

Abstract: Over the last quarter century, the dominant tendency in comparative cognitive psychology has been to emphasize the similarities between human and nonhuman minds and to downplay the differences as "one of degree and not of kind" (Darwin 1871). In the present target article, we argue that Darwin was mistaken: the profound biological continuity between human and nonhuman animals masks an equally profound discontinuity between human and nonhuman minds. To wit, there is a significant discontinuity in the degree to which human and nonhuman animals are able to approximate the higher-order, systematic, relational capabilities of a physical symbol system (PSS) (Newell 1980). We show that this symbolic-relational discontinuity pervades nearly every domain of cognition and runs much deeper than even the spectacular scaffolding provided by language or culture alone can explain. We propose a representational-level specification as to where human and nonhuman animals' abilities to approximate a PSS are similar and where they differ. We conclude by suggesting that recent symbolic-connectionist models of cognition shed new light on the mechanisms that underlie the gap between human and nonhuman minds.

Keywords: analogy; animal cognition; causal learning; connectionism; Darwin; discontinuity; evolution; human mind; language; language of thought; physical symbol system; reasoning; same-different; theory of mind

1. Introduction

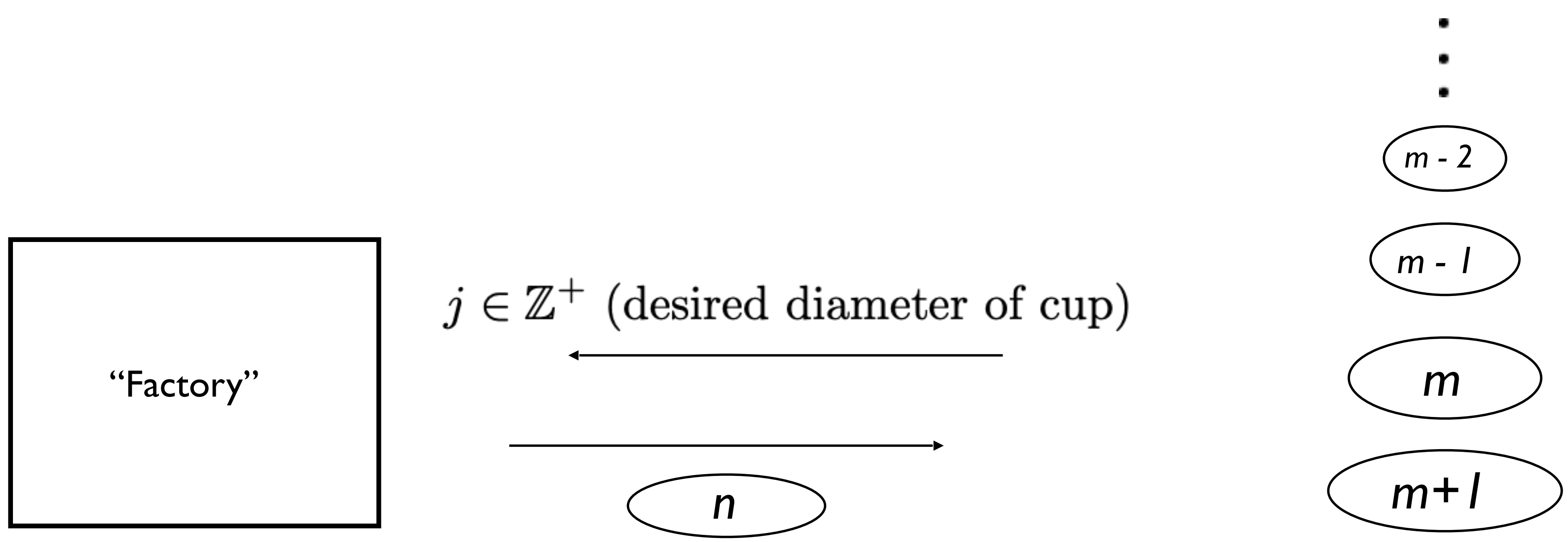
Human animals – and no other – build fires and wheels, diagnose each other's illnesses, communicate using symbols, navigate with maps, risk their lives for ideals, collaborate with each other, explain the world in terms of hypothetical causes, punish strangers for breaking rules, imagine impossible scenarios, and teach each other how to do all of the above. At first blush, it might appear obvious that human minds are qualitatively different from those of every other animal on the planet. Ever since Darwin, however, the dominant tendency in comparative cognitive psychology has been to emphasize the continuity between human and nonhuman minds and to downplay the differences as "one of degree and not of kind" (Darwin 1871). Particularly in the last quarter century, many prominent comparative researchers have claimed that the traditional hallmarks of human cognition – for example, complex tool use, grammatically structured language, causal-logical reasoning, mental state attribution, metacognition, analogical inferences, mental time travel, culture, and so on – are not nearly as unique as we once thought (see, e.g., Bekoff et al. 2002; Call 2006; Clayton et al. 2003; de Waal & Tyack 2003; Matsuzawa 2001; Pepperberg 2002; Rendell & Whitehead 2001; Savage-Rumbaugh et al. 1998; Smith et al. 2003; Tomasello et al. 2003a). Pepperberg (2005, p. 469) aptly sums up the comparative consensus as follows: "for over 35 years, researchers have been demonstrating through tests both in the field and in the laboratory that the capacities of nonhuman animals to solve complex problems form a continuum with those of

1 of 70

 0140-525X/08 $40.00

Selmer's Seriated Cup Challenge #1

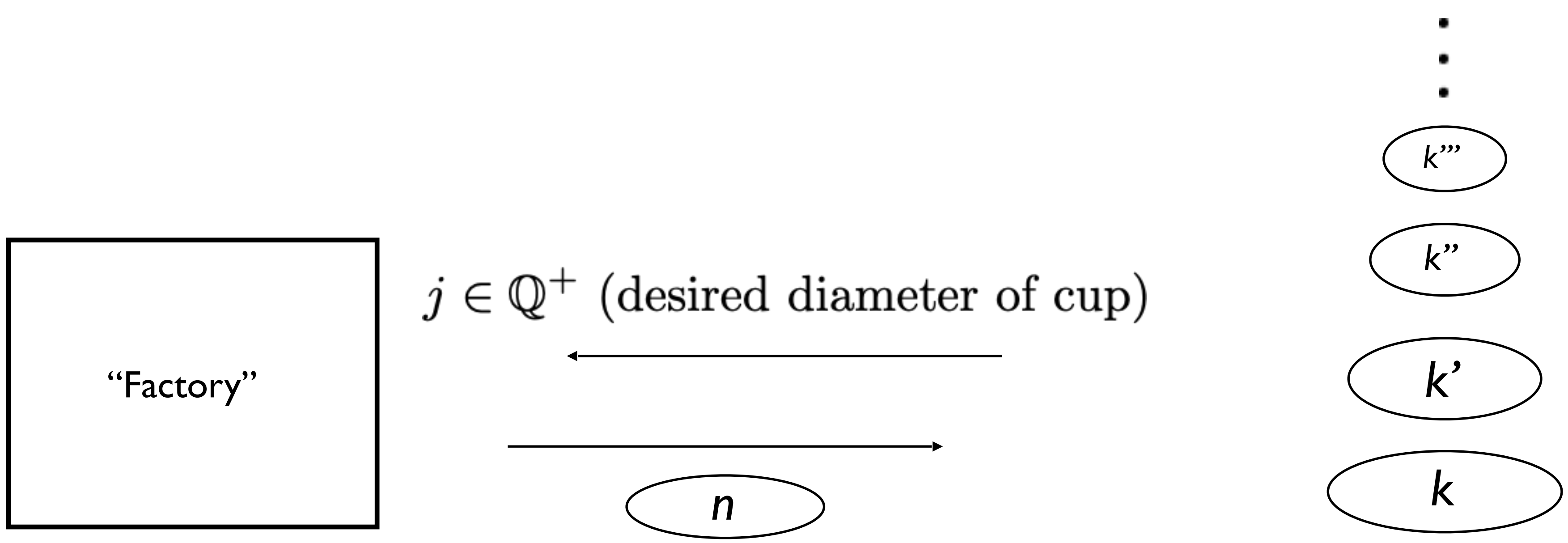
Suppose you have at your disposal a "factory" that, upon hearing you announce a number *j*, can quickly output a cup having a diameter of precisely *j* units. Can you insert a new cup between two of the seriated, stacked cups in the tower shown here? — where the *j* you send in *must* be a positive integer, *m* is likewise a positive integer, and every cup in every tower must be more in diameter than the one immediately above it, and less in diameter than the one immediately below it?** Prove that your answer is correct.



**E.g., if *m* = 3, the tower in that case will have a base cup 4 units in diameter, immediately above that a cup 3 units in diameter, then a cup 2 units in diameter, and then finally a top cup of 1 unit in diameter.

Selmer's Seriated Cup Challenge #2

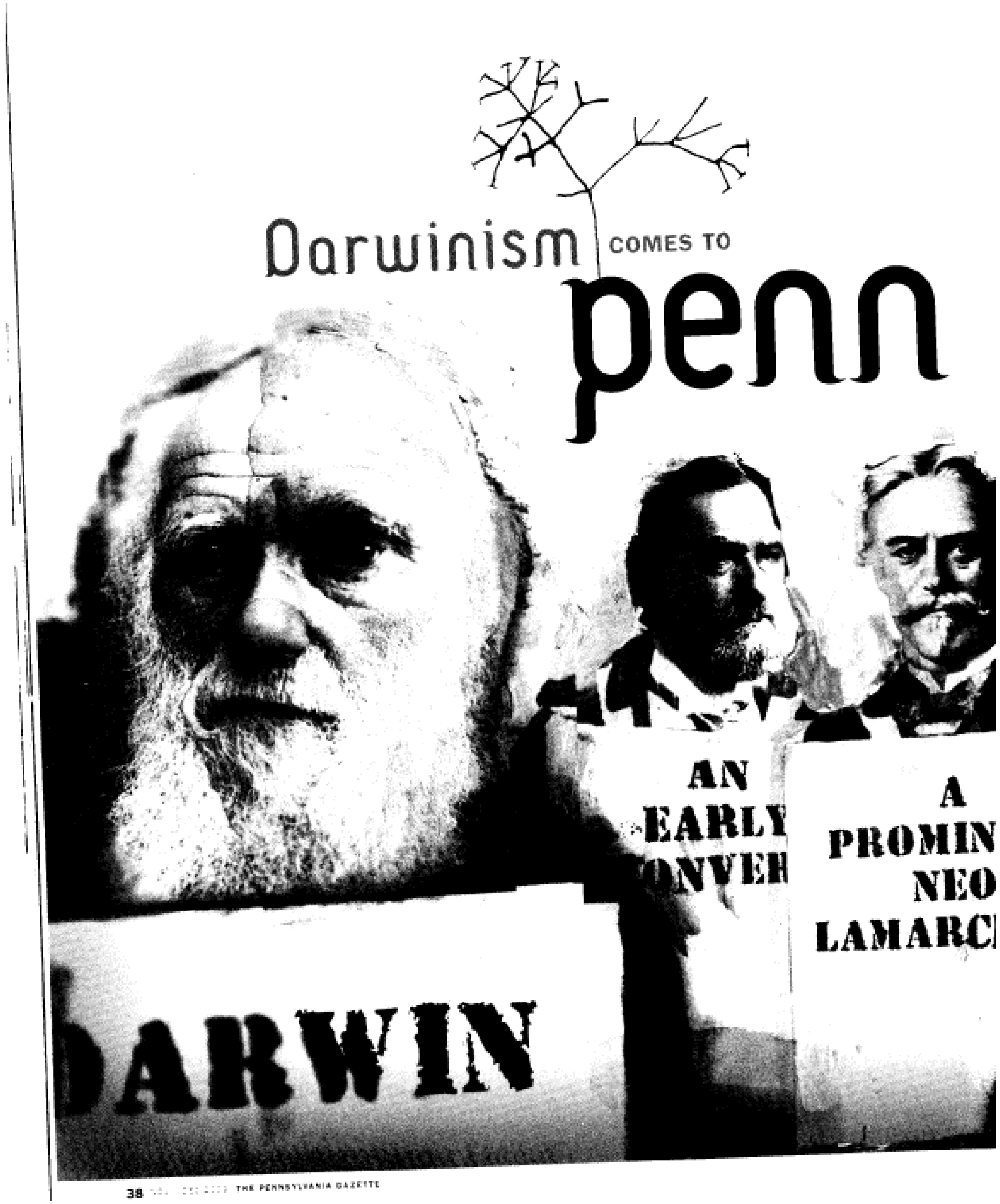
Suppose you have at your disposal a "factory" that, upon hearing you announce a number *j*, can quickly output a cup having a diameter of precisely *j* units. Can you insert a new cup between two of the seriated, stacked cups in the tower shown here? — where the *j* you send in *must* be a positive rational number; *k*, *k'*, *k''*, *k'''* … are likewise positive rational numbers, and every cup in every tower must be more in diameter than the one immediately above it, and less in diameter than the one immediately below it?** Prove that your answer is correct.



**E.g., if $k = \frac{1}{2}$, the tower in that case will have a base cup $\frac{1}{2}$ units in diameter, immediately above that there could be a cup $\frac{1}{3}$ units in diameter, then perhaps a cup $\frac{1}{4}$ units in diameter, and then perhaps finally a top cup of $\frac{1}{32}$ units in diameter.

Check your history books ...

From my alma mater: *Pennsylvania Gazette* Nov/Dec 2009



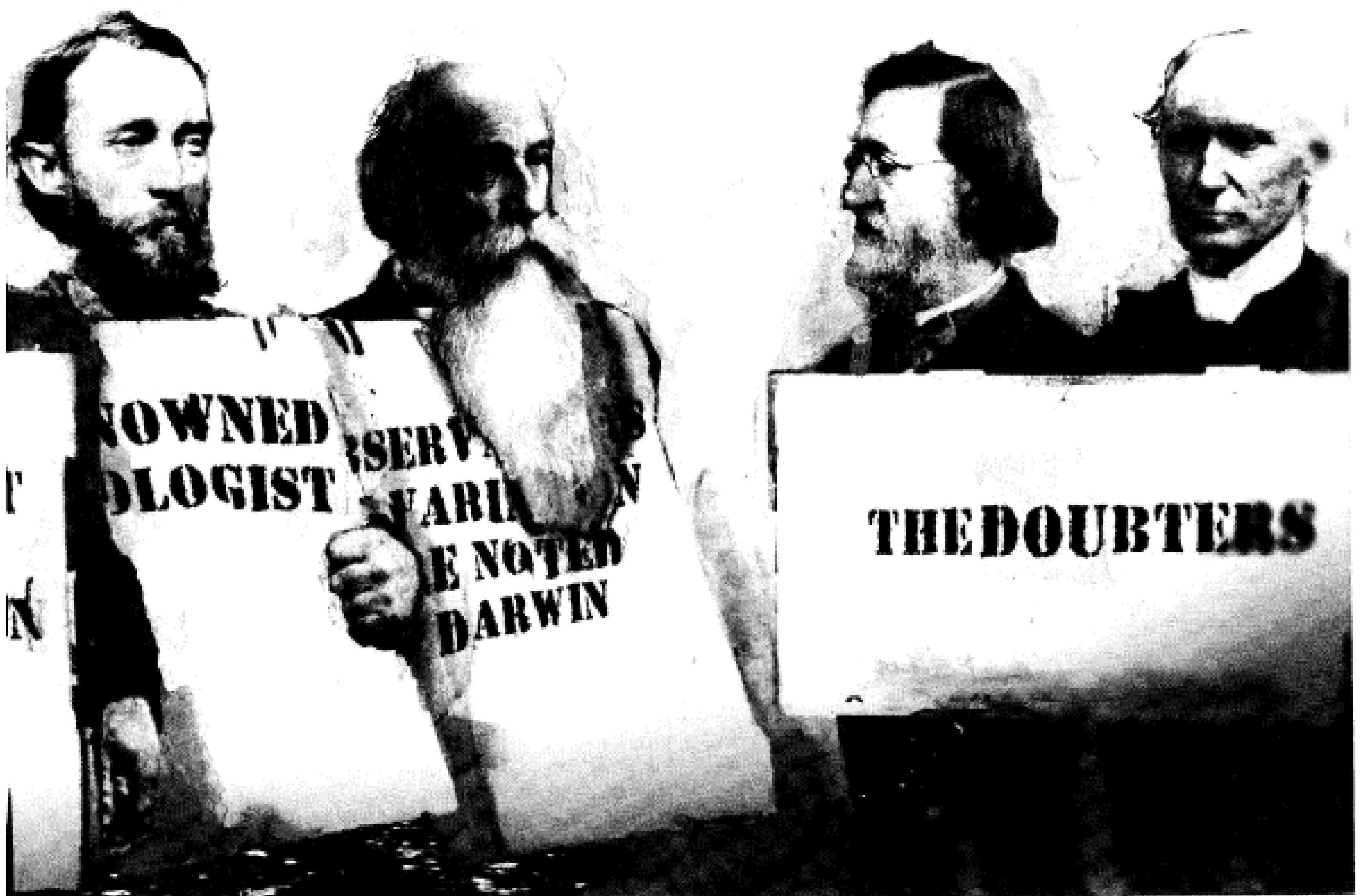
From my alma mater: *Pennsylvania Gazette* Nov/Dec 2009

From my alma mater: *Pennsylvania Gazette* Nov/Dec 2009

A century-and-a-half after the November 1859 publication of *On the Origin of Species*, a Penn microbiologist looks back at how Darwin's ideas were received by some of the University's leading thinkers. **BY HOWARD GOLDFINE**

ON June 18, 1858, Charles Darwin received a manuscript from Alfred Russel Wallace, which outlined a theory of evolution based on natural selection. Wallace's letter came from an island in the Malay Archipelago, where he was collecting field specimens and studying the distribution of species. Wallace, like Darwin, invoked the Malthusian concept that a struggle for existence within rapidly expanding populations would be the driving force for selection of natural variants within a species. Darwin's immediate reaction was one of dismay. He had been working on his "big book on species" since his five-year voyage on the Beagle (1831-36) and a relatively unknown naturalist had forestalled him. Darwin wrote to Charles Lyell, "If Wallace had my [manuscript] sketch written out in 1842, he could not have written out a better short abstract!"

Fortunately, Darwin had previously outlined his theory to his friends, the distinguished geologist Lyell and the botanist Joseph D. Hooker, and in a brief, unpublished draft to Asa Gray, a botanist at Harvard. Lyell and Hooker immediately arranged for Wallace's paper and a brief summary of Darwin's theory to be read simultaneously at the Linnaean Society in London on July 1, 1858. These were received with little comment. The president of the society later noted that nothing of great interest had happened that year.



A century-and-a-half after the November 1859 publication of *On the Origin of Species*, a Penn microbiologist looks back at how Darwin's ideas were received by some of the University's leading thinkers. **BY HOWARD GOLDFINE**

ON June 18, 1858, Charles Darwin received a manuscript from Alfred Russel Wallace, which outlined a theory of evolution based on natural selection. Wallace's letter came from an island in the Malay Archipelago, where he was collecting field specimens and studying the distribution of species. Wallace, like Darwin, invoked the Malthusian concept that a struggle for existence within rapidly expanding populations would be the driving force for selection of natural variants within a species. Darwin's immediate reaction was one of dismay. He had been working on his "big book on species" since his five-year voyage on the Beagle (1831-36) and a relatively unknown naturalist had forestalled him. Darwin wrote to Charles Lyell, "If Wallace had my [manuscript] sketch written out in 1842, he could not have written out a better short abstract!"

Fortunately, Darwin had previously outlined his theory to his friends, the distinguished geologist Lyell and the botanist Joseph D. Hooker, and in a brief, unpublished draft to Asa Gray, a botanist at Harvard. Lyell and Hooker immediately arranged for Wallace's paper and a brief summary of Darwin's theory to be read simultaneously at the Linnaean Society in London on July 1, 1858. These were received with little comment. The president of the society later noted that nothing of great interest had happened that year.

ILLUSTRATION BY DAVID HOLLENBACH

"On June 18, 1858, Charles Darwin received a manuscript from Alfred Russel Wallace, which outlined a theory of evolution based on natural selection. ... Darwin's immediate reaction was one of dismay. ... [That year] Wallace's paper, and a brief summary of Darwin's theory [were] read simultaneously (sic) at the Linnaean Society in London on July 1, 1858. ..."

Wallace rejected the claim that the human mind, with its capacity for abstract, rational thought, is the product of evolution by mutation and natural selection, on the basis of reasoned argument (Wallace's Paradox).

Wallace rejected the claim that the human mind, with its capacity for abstract, rational thought, is the product of evolution by mutation and natural selection, on the basis of reasoned argument (Wallace's Paradox).

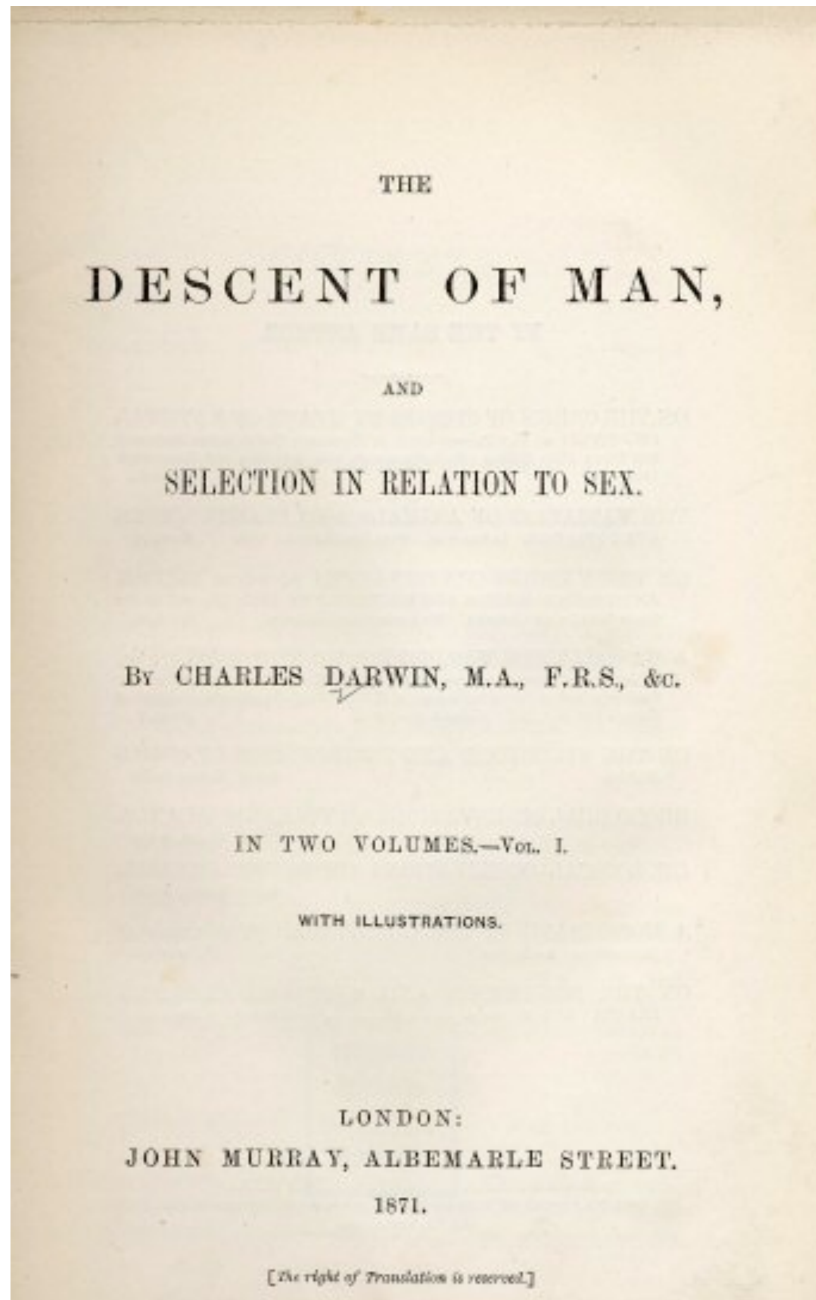
Darwin did not.
And he defended his position in a book:
Descent of Man.

Wallace rejected the claim that the human mind, with its capacity for abstract, rational thought, is the product of evolution by mutation and natural selection, on the basis of reasoned argument (Wallace's Paradox).

Darwin did not.
And he defended his position in a book:
Descent of Man.

Wallace seems to me to be right; Darwin to be wrong…

The book that shook the world, and supposedly obliterated the stupid notion that human persons are made in (in Milton's unpacked version of the phrase) God's image.



THE

DESCENT OF MAN,

AND

SELECTION IN RELATION TO SEX.

BY CHARLES DARWIN, M.A., F.R.S., &c.

IN TWO VOLUMES.—VOL. I.

WITH ILLUSTRATIONS.

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1871.

[The right of Translation is reserved.]

Praise for Darwin & *DoM*

Back cover of my Amazon.com version of *DoM*:
"Darwin's engaging literary style, charming modesty, brilliant argument, and discursive method of proof makes the book an exhilarating romp through Earth's natural history and Man's history ..."

Praise for Darwin & *DoM*

Back cover of my Amazon.com version of *DoM*:
"Darwin's engaging literary style, charming modesty, brilliant argument, and discursive method of proof makes the book an exhilarating romp through Earth's natural history and Man's history ..."

Really?

I found no brilliant arguments, and not a single proof.

A Key Proposition

A Key Proposition

$$\bar{A}$$

There is at least one mental power possessed by human persons, but not by any mere animal; and the mental powers of human persons are of a wholly different nature than those of mere animals.

Efficient Refutation of Darwin's *DoM*

	(1)	If human persons are the product of evolution, then it's not the case that $\bar{A}$ holds.	
	(2)	$\bar{A}$ does hold.	
$\therefore$	(3)	Human persons are not the product of evolution.	from (1), (2) by *modus tollens*
		QED	

Efficient Refutation of Darwin's *DoM*

	(1)	If human persons are the product of evolution, then it's not the case that $\bar{A}$ holds.	
	(2)	$\bar{A}$ does hold.	
$\therefore$	(3)	Human persons are not the product of evolution.	from (1), (2) by *modus tollens*
		QED	

Note: (3) doesn't deductively entail that *no* parts of human personhood are the product of evolution. In other words, (3) can be rephrased as: "Human persons are not solely and completely the product of evolution." As seen shortly, the power of human persons to carry out abstract, infinitary reasoning (as in the case of developing the tensor calculus) would be — according to Wallace & Bringsjord — something that evolution didn't produce.

Whence comes the first premise in this argument?

From Darwin Himself

From Darwin Himself

“If no organic being excepting man had possessed any mental power, or if his powers had been of a wholly different nature from those of the lower animals, then we should never have been able to convince ourselves that our high faculties had been gradually developed.”

(*Descent of Man*, Part One, Chapter Two)

The Argument

	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/. . . .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ and, where P^{AI} is a limit on the cognitive power of AI, AIs can hunt and gather.	see AI today
∴	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P)$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
∴	(8)	God exists.	*modus ponens* (5), (6), (7)

E.g., from S.J. Gould: A big mutation happened but lay dormant for ~ 250,000 years.

The Argument

	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/... .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ and, where P^{AI} is a limit on the cognitive power of AI, AIs can hunt and gather.	see AI today
$\therefore$	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P)$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
$\therefore$	(8)	God exists.	*modus ponens* (5), (6), (7)

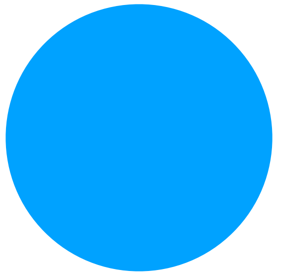
?

E.g., from S.J. Gould: A big mutation happened but lay dormant for ~ 250,000 years.

The Argument

	(1)	Hunter-gatherers possessed the cognitive power P^{∞} to e.g. invent the calculus and create literary art of the caliber of Blecher/Proust/Ibsen/. . . .	undisputed
	(2)	AI shows us that these early versions of us, to hunt and gather, needed only humble cognitive power P^{HG}, where $P^{\text{HG}} < P^{\infty}$, because $P^{\text{AI}} \approx P^{\text{HG}}$ and, where P^{AI} is a limit on the cognitive power of AI, AIs can hunt and gather.	see AI today
$\therefore$	(3)	We have $(P^{\infty} - P^{\text{HG}})$.	abstraction (1), (2)
	(4)	Our having $(P^{\infty} - P^{\text{HG}})$, *contra* Darwin, is inexplicable by gradual mutation and natural selection (i.e. P^{∞} is **discontinuous** from P^{HG}).	see critique of *DoM* see theorem/proof
	(5)	If our having $(P^{\infty} - P^{\text{HG}})$ is explicable, then $E_1 \vee E_2 \vee$ God exists.	sub-arg
	(6)	Our having $(P^{\infty} - P)$ is explicable.	undeniable
	(7)	$\neg E_1 \wedge \neg E_2$	sub-argument
$\therefore$	(8)	God exists.	*modus ponens* (5), (6), (7)

?



The Springboard Paper

The Springboard Paper

doi: 10.1017/S0140525X08003543

Darwin's mistake: Explaining the discontinuity between human and nonhuman minds

Derek C. Penn
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095; Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
dcpenn@ucla.edu
http://reasoninglab.psych.ucla.edu/
http://www.cognitiveevolutiongroup.org/

Keith J. Holyoak
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095
holyoak@lifesci.ucla.edu
http://reasoninglab.psych.ucla.edu/

Daniel J. Povinelli
Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
ceg@louisiana.edu
http://www.cognitiveevolutiongroup.org/

Abstract: Over the last quarter century, the dominant tendency in comparative cognitive psychology has been to emphasize the similarities between human and nonhuman minds and to downplay the differences as "one of degree and not of kind" (Darwin 1871). In the present target article, we argue that Darwin was mistaken: the profound biological continuity between human and nonhuman animals masks an equally profound discontinuity between human and nonhuman minds. To wit, there is a significant discontinuity in the degree to which human and nonhuman animals are able to approximate the higher-order, systematic, relational capabilities of a physical symbol system (PSS) (Newell 1980). We show that this symbolic-relational discontinuity pervades nearly every domain of cognition and runs much deeper than even the spectacular scaffolding provided by language or culture alone can explain. We propose a representational-level specification as to where human and nonhuman animals' abilities to approximate a PSS are similar and where they differ. We conclude by suggesting that recent symbolic-connectionist models of cognition shed new light on the mechanisms that underlie the gap between human and nonhuman minds.

Keywords: analogy; animal cognition; causal learning; connectionism; Darwin; discontinuity; evolution; human mind; language; language of thought; physical symbol system; reasoning; same-different; theory of mind

1. Introduction

Human animals – and no other – build fires and wheels, diagnose each other's illnesses, communicate using symbols, navigate with maps, risk their lives for ideals, collaborate with each other, explain the world in terms of hypothetical causes, punish strangers for breaking rules, imagine impossible scenarios, and teach each other how to do all of the above. At first blush, it might appear obvious that human minds are qualitatively different from those of every other animal on the planet. Ever since Darwin, however, the dominant tendency in comparative cognitive psychology has been to emphasize the continuity between human and nonhuman minds and to downplay the differences as "one of degree and not of kind" (Darwin 1871). Particularly in the last quarter century, many prominent comparative researchers have claimed that the traditional hallmarks of human cognition – for example, complex tool use, grammatically structured language, causal-logical reasoning, mental state attribution, metacognition, analogical inferences, mental time travel, culture, and so on – are not nearly as unique as we once thought (see, e.g., Bekoff et al. 2002; Call 2006; Clayton et al. 2003; de Waal & Tyack 2003; Matsuzawa 2001; Pepperberg 2002; Rendell & Whitehead 2001; Savage-Rumbaugh et al. 1998; Smith et al. 2003; Tomasello et al. 2003a). Pepperberg (2005, p. 469) aptly sums up the comparative consensus as follows: "for over 35 years, researchers have been demonstrating through tests both in the field and in the laboratory that the capacities of nonhuman animals to solve complex problems form a continuum with those of

 0140-525X/08 $40.00

doi: 10.1017/S0140525X08003543

Darwin's mistake: Explaining the discontinuity between human and nonhuman minds

Derek C. Penn
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095; Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
dcpenn@ucla.edu
http://reasoninglab.psych.ucla.edu/
http://www.cognitiveevolutiongroup.org/

Keith J. Holyoak
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095
holyoak@lifesci.ucla.edu
http://reasoninglab.psych.ucla.edu/

Daniel J. Povinelli
Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
ceg@louisiana.edu
http://www.cognitiveevolutiongroup.org/

Abstract: Over the last quarter century, the dominant tendency in comparative cognitive psychology has been to emphasize the similarities between human and nonhuman minds and to downplay the differences as "one of degree and not of kind" (Darwin 1871). In the present target article, we argue that Darwin was mistaken: the profound biological continuity between human and nonhuman animals masks an equally profound discontinuity between human and nonhuman minds. To wit, there is a significant discontinuity in the degree to which human and nonhuman animals are able to approximate the higher-order, systematic, relational capabilities of a physical symbol system (PSS) (Newell 1980). We show that this symbolic-relational discontinuity pervades nearly every domain of cognition and runs much deeper than even the spectacular scaffolding provided by language or culture alone can explain. We propose a representational-level specification as to where human and nonhuman animals' abilities to approximate a PSS are similar and where they differ. We conclude by suggesting that recent symbolic-connectionist models of cognition shed new light on the mechanisms that underlie the gap between human and nonhuman

doi: 10.1017/S0140525X08003543

Darwin's mistake: Explaining the discontinuity between human and nonhuman minds

Derek C. Penn
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095; Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
dcpenn@ucla.edu
http://reasoninglab.psych.ucla.edu/
http://www.cognitiveevolutiongroup.org/

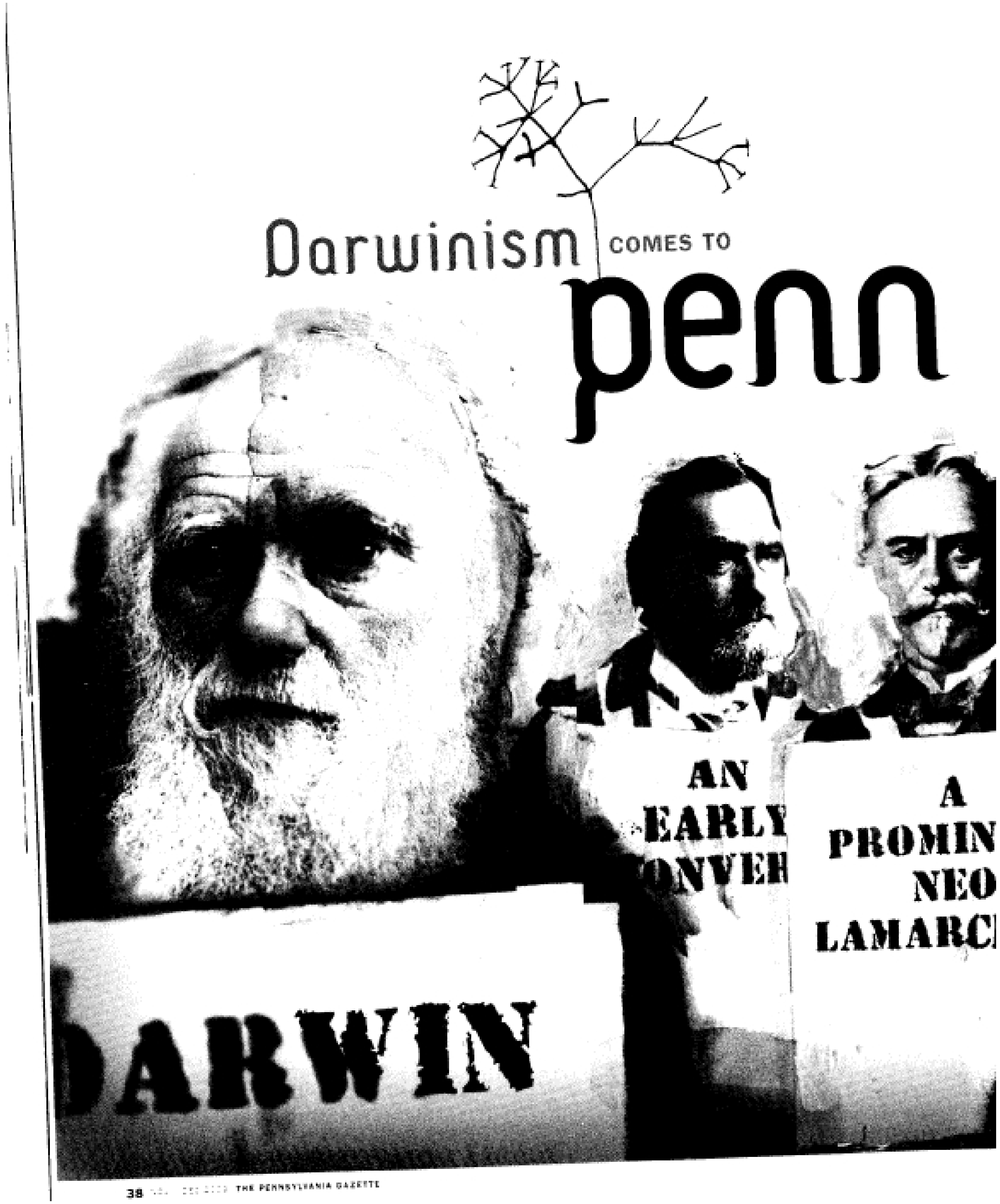
Keith J. Holyoak
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095
holyoak@lifesci.ucla.edu
http://reasoninglab.psych.ucla.edu/

Daniel J. Povinelli
Cognitive Evolution Group, University of Louisiana, Lafayette, LA 70504
ceg@louisiana.edu
http://www.cognitiveevolutiongroup.org/

Abstract: Over the last quarter century, the dominant tendency in comparative cognitive psychology has been to emphasize the similarities between human and nonhuman minds and to downplay the differences as "one of degree and not of kind" (Darwin 1871). In the present target article, we argue that Darwin was mistaken: the profound biological continuity between human and nonhuman animals masks an equally profound discontinuity between human and nonhuman minds. To wit, there is a significant discontinuity in the degree to which human and nonhuman animals are able to approximate the higher-order, systematic, relational capabilities of a physical symbol system (PSS) (Newell 1980). We show that this symbolic-relational discontinuity pervades nearly every domain of cognition and runs much deeper than even the spectacular scaffolding provided by language or culture alone can explain. We propose a representational-level specification as to where human and nonhuman animals' abilities to approximate a PSS are similar and where they differ. We conclude by suggesting that recent symbolic-connectionist models of cognition shed new light on the mechanisms that underlie the gap between human and nonhuman

Check your history books ...

From my alma mater: *Pennsylvania Gazette* Nov/Dec 2009



From my alma mater: *Pennsylvania Gazette* Nov/Dec 2009

From my alma mater: *Pennsylvania Gazette* Nov/Dec 2009

A century-and-a-half after the November 1859 publication of *On the Origin of Species*, a Penn microbiologist looks back at how Darwin's ideas were received by some of the University's leading thinkers. **BY HOWARD GOLDFINE**

ON June 18, 1858, Charles Darwin received a manuscript from Alfred Russel Wallace, which outlined a theory of evolution based on natural selection. Wallace's letter came from an island in the Malay Archipelago, where he was collecting field specimens and studying the distribution of species. Wallace, like Darwin, invoked the Malthusian concept that a struggle for existence within rapidly expanding populations would be the driving force for selection of natural variants within a species. Darwin's immediate reaction was one of dismay. He had been working on his "big book on species" since his five-year voyage on the Beagle (1831-36) and a relatively unknown naturalist had forestalled him. Darwin wrote to Charles Lyell, "If Wallace had my [manuscript] sketch written out in 1842, he could not have written out a better short abstract!"

Fortunately, Darwin had previously outlined his theory to his friends, the distinguished geologist Lyell and the botanist Joseph D. Hooker, and in a brief, unpublished draft to Asa Gray, a botanist at Harvard. Lyell and Hooker immediately arranged for Wallace's paper and a brief summary of Darwin's theory to be read simultaneously at the Linnaean Society in London on July 1, 1858. These were received with little comment. The president of the society later noted that nothing of great interest had happened that year.

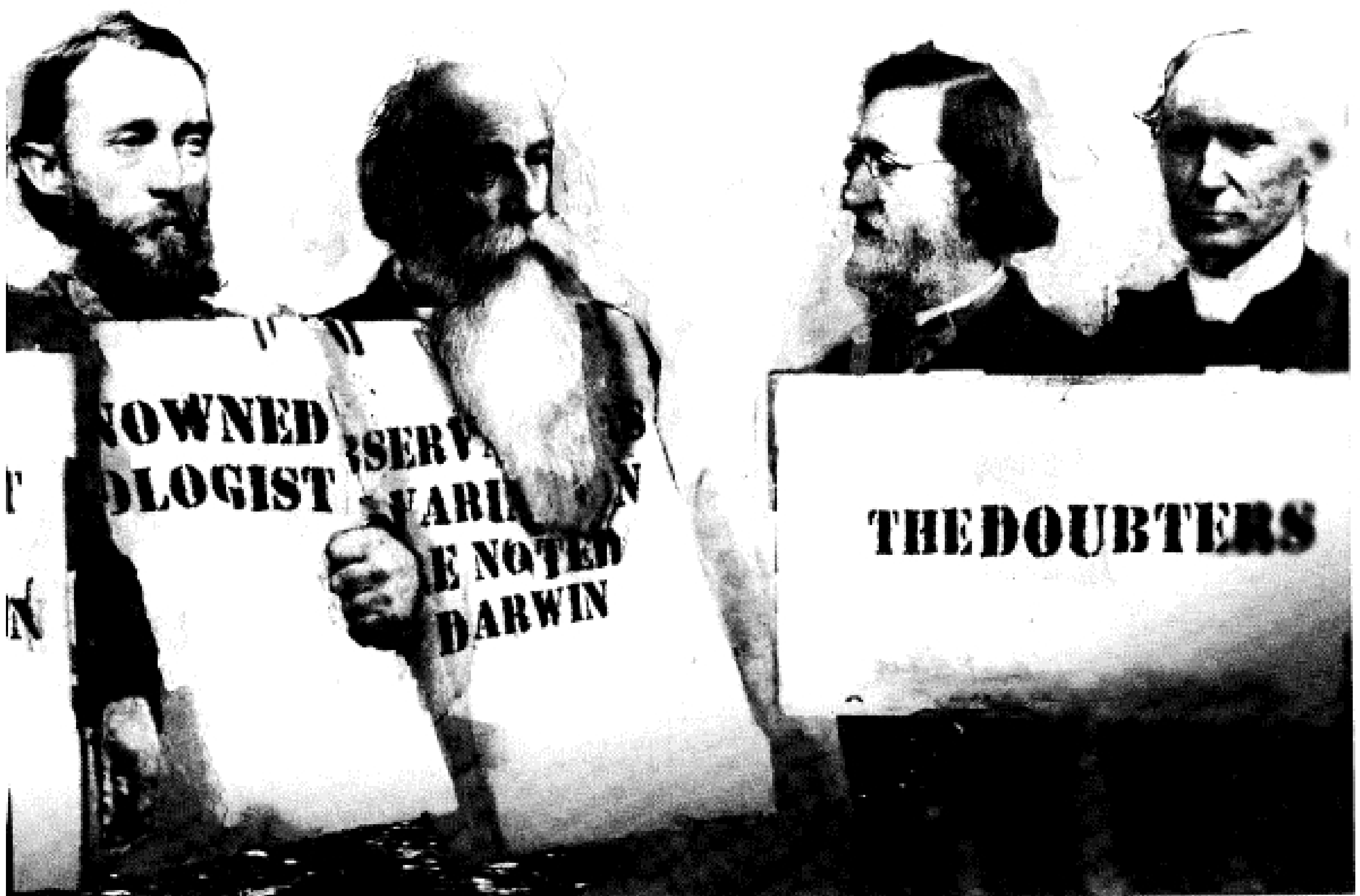


ILLUSTRATION BY DAVID HOLLENBACH

A century-and-a-half after the November 1859 publication of *On the Origin of Species*, a Penn microbiologist looks back at how Darwin's ideas were received by some of the University's leading thinkers. **BY HOWARD GOLDFINE**

ON June 18, 1858, Charles Darwin received a manuscript from Alfred Russel Wallace, which outlined a theory of evolution based on natural selection. Wallace's letter came from an island in the Malay Archipelago, where he was collecting field specimens and studying the distribution of species. Wallace, like Darwin, invoked the Malthusian concept that a struggle for existence within rapidly expanding populations would be the driving force for selection of natural variants within a species. Darwin's immediate reaction was one of dismay. He had been working on his "big book on species" since his five-year voyage on the Beagle (1831-36) and a relatively unknown naturalist had forestalled him. Darwin wrote to Charles Lyell, "If Wallace had my [manuscript] sketch written out in 1842, he could not have written out a better short abstract!"

Fortunately, Darwin had previously outlined his theory to his friends, the distinguished geologist Lyell and the botanist Joseph D. Hooker, and in a brief, unpublished draft to Asa Gray, a botanist at Harvard. Lyell and Hooker immediately arranged for Wallace's paper and a brief summary of Darwin's theory to be read simultaneously at the Linnaean Society in London on July 1, 1858. These were received with little comment. The president of the society later noted that nothing of great interest had happened that year.

"On June 18, 1858, Charles Darwin received a manuscript from Alfred Russel Wallace, which outlined a theory of evolution based on natural selection. ... Darwin's immediate reaction was one of dismay. ... [That year] Wallace's paper, and a brief summary of Darwin's theory [were] read simultaneously (sic) at the Linnaean Society in London on July 1, 1858. ..."

Wallace rejected the claim that the human mind, with its capacity for abstract, rational thought, is the product of evolution by mutation and natural selection, on the basis of reasoned argument (Wallace's Paradox).

Wallace rejected the claim that the human mind, with its capacity for abstract, rational thought, is the product of evolution by mutation and natural selection, on the basis of reasoned argument (Wallace's Paradox).

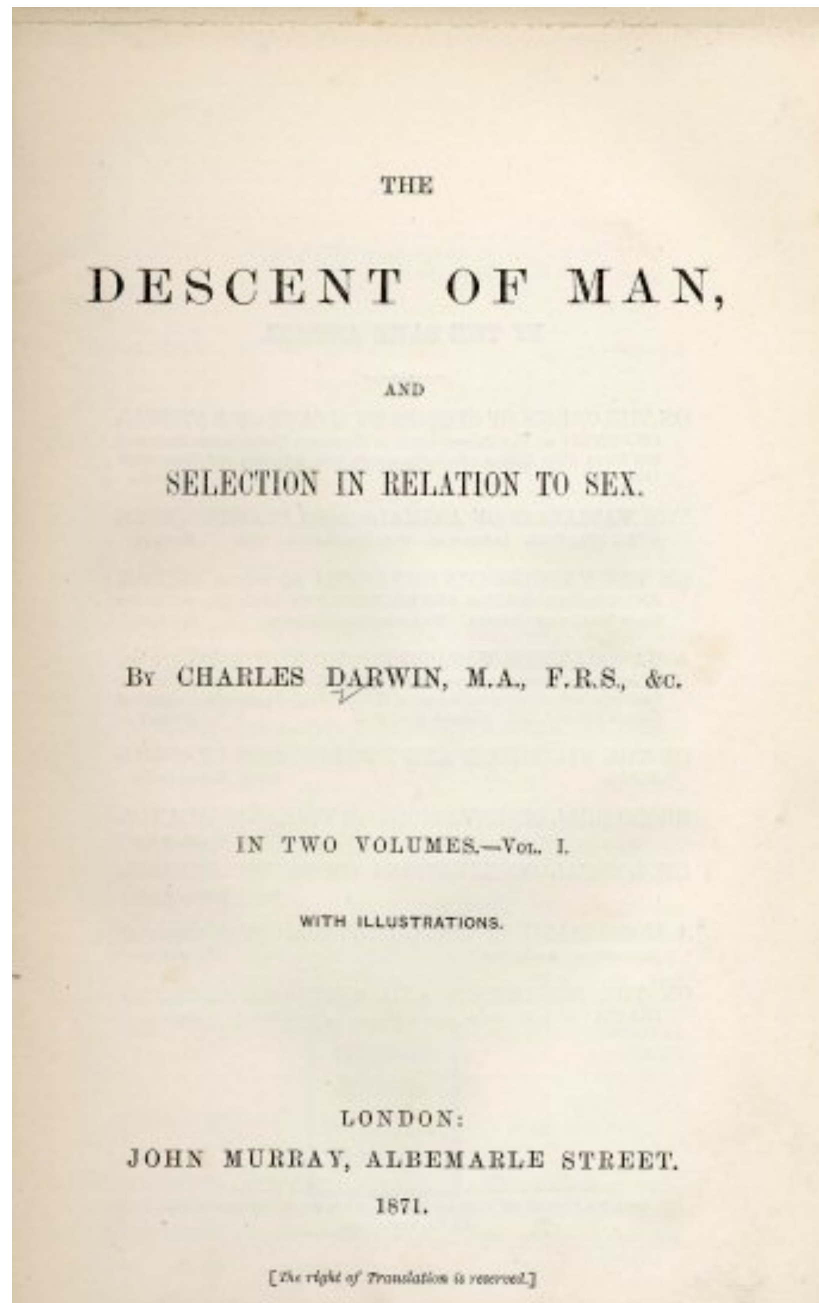
Darwin did not.
And he defended his position in a book:
Descent of Man.

Wallace rejected the claim that the human mind, with its capacity for abstract, rational thought, is the product of evolution by mutation and natural selection, on the basis of reasoned argument (Wallace's Paradox).

Darwin did not.
And he defended his position in a book:
Descent of Man.

Wallace seems to me to be right; Darwin to be wrong…

The book that shook the world, and supposedly obliterated the stupid notion that human persons are made in (in Milton's unpacked version of the phrase) God's image.


THE

DESCENT OF MAN,

AND

SELECTION IN RELATION TO SEX.

BY CHARLES DARWIN, M.A., F.R.S., &c.

IN TWO VOLUMES.—VOL. I.

WITH ILLUSTRATIONS.

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1871.

[*The right of Translation is reserved.*]

Praise for Darwin & *DoM*

Back cover of my Amazon.com version of *DoM*:
“Darwin’s engaging literary style, charming modesty, brilliant argument, and discursive method of proof makes the book an exhilarating romp through Earth’s natural history and Man’s history ...”

Praise for Darwin & *DoM*

Back cover of my Amazon.com version of *DoM*:
"Darwin's engaging literary style, charming modesty, brilliant argument, and discursive method of proof makes the book an exhilarating romp through Earth's natural history and Man's history ..."

Really?

I found no brilliant arguments, and not a single proof.

Perhaps the emperors have no clothes.

A Key Proposition

A Key Proposition

$$\bar{A}$$

There is at least one mental power possessed by human persons, but not by any mere animal; and the mental powers of human persons are of a wholly different nature than those of mere animals.

Efficient Refutation of Darwin's *DoM*

	(1)	If human persons are the product of evolution, then it's not the case that $\bar{A}$ holds.	
	(2)	$\bar{A}$ does hold.	
$\therefore$	(3)	Human persons are not the product of evolution.	from (1), (2) by *modus tollens*
		QED	

Efficient Refutation of Darwin's *DoM*

	(1)	If human persons are the product of evolution, then it's not the case that $\bar{A}$ holds.	
	(2)	$\bar{A}$ does hold.	
$\therefore$	(3)	Human persons are not the product of evolution.	from (1), (2) by *modus tollens*
		QED	

Note: (3) doesn't deductively entail that *no* parts of human personhood are the product of evolution. In other words, (3) can be rephrased as: "Human persons are not solely and completely the product of evolution." As seen shortly, the power of human persons to carry out abstract, infinitary reasoning (as in the case of developing the tensor calculus) would be — according to Wallace & Bringsjord — something that evolution didn't produce.

Whence comes the first premise in this argument?

From Darwin Himself

From Darwin Himself

"If no organic being excepting man had possessed any mental power, or if his powers had been of a wholly different nature from those of the lower animals, then we should never have been able to convince ourselves that our high faculties had been gradually developed."

(*Descent of Man*, Part One, Chapter Two)

So, Darwin devotes himself to trying to overthrow

$$\bar{A}.$$

So, Darwin devotes himself to trying to overthrow

$$\bar{A}.$$

How?

Darwin's Defense

	(1)	Story or anecdote S.	
$\therefore$	(2)	There exist animals manifesting behavior B.	from (1)
	(3)	Anything behaving as in B has purportedly differentiating mental powers $M_1, \ldots, M_k$.	
$\therefore$	(4)	There exist animals having purportedly differentiating mental powers $M_1, \ldots, M_k$.	from (2), (3), (4)
$\therefore$	(5)	$\neg \bar{A}$	(4), def of $\bar{A}$
$\therefore$	(6)	Bringsjord's intended refutation fails.	(5), def of refutation

Darwin's Defense wrt Reasoning

	(1)	Story or anecdote S.	
$\therefore$	(2)	There exist animals manifesting behavior B.	from (1)
	(3)	Anything behaving as in B has the purportedly differentiating mental power of reasoning.	
$\therefore$	(4)	There exist animals the having purportedly differentiating mental power of reasoning.	from (2), (3), (4)

What *is* reasoning?

What *is* reasoning?

- Well, deductive, inductive/probabilistic, abductive, analogical?

What *is* reasoning?

- Well, deductive, inductive/probabilistic, abductive, analogical?
- All varieties, if even marginally rigorous, presuppose *deductive* reasoning.

What *is* reasoning?

- Well, deductive, inductive/probabilistic, abductive, analogical?
- All varieties, if even marginally rigorous, presuppose *deductive* reasoning.
- Examples:

What *is* reasoning?

- Well, deductive, inductive/probabilistic, abductive, analogical?
- All varieties, if even marginally rigorous, presuppose *deductive* reasoning.
- Examples:
  - Wason Selection Task cracked, & others seen …

What *is* reasoning?

- Well, deductive, inductive/probabilistic, abductive, analogical?
- All varieties, if even marginally rigorous, presuppose *deductive* reasoning.
- Examples:
  - Wason Selection Task cracked, & others seen …
  - “Intergalactic Diplomacy” ... (see end of slide deck)

What *is* reasoning?

- Well, deductive, inductive/probabilistic, abductive, analogical?
- All varieties, if even marginally rigorous, presuppose *deductive* reasoning.
- Examples:
  - Wason Selection Task cracked, & others seen …
  - "Intergalactic Diplomacy" ... (see end of slide deck)
  - Karkooking Problem …

What *is* reasoning?

- Well, deductive, inductive/probabilistic, abductive, analogical?
- All varieties, if even marginally rigorous, presuppose *deductive* reasoning.
- Examples:
  - Wason Selection Task cracked, & others seen …
  - "Intergalactic Diplomacy" ... (see end of slide deck)
  - Karkooking Problem …
  - And *infinitary* deductive reasoning: "Gödel-level" Theorems ... (see Bringsjord, S. *Gödel's Great Theorems*, forthcoming from Oxford Univ Press)

Karkooking Problem …

Everyone karkooks anyone who karkooks someone.

Alvin karkooks Bill.

Can you infer that everyone karkooks Bill?

ANSWER:

JUSTIFICATION:

Larking Problem …

Everyone larks anyone who larks someone.

Quantification!

Alvin larks Bill.

Can you infer that everyone larks Bill?

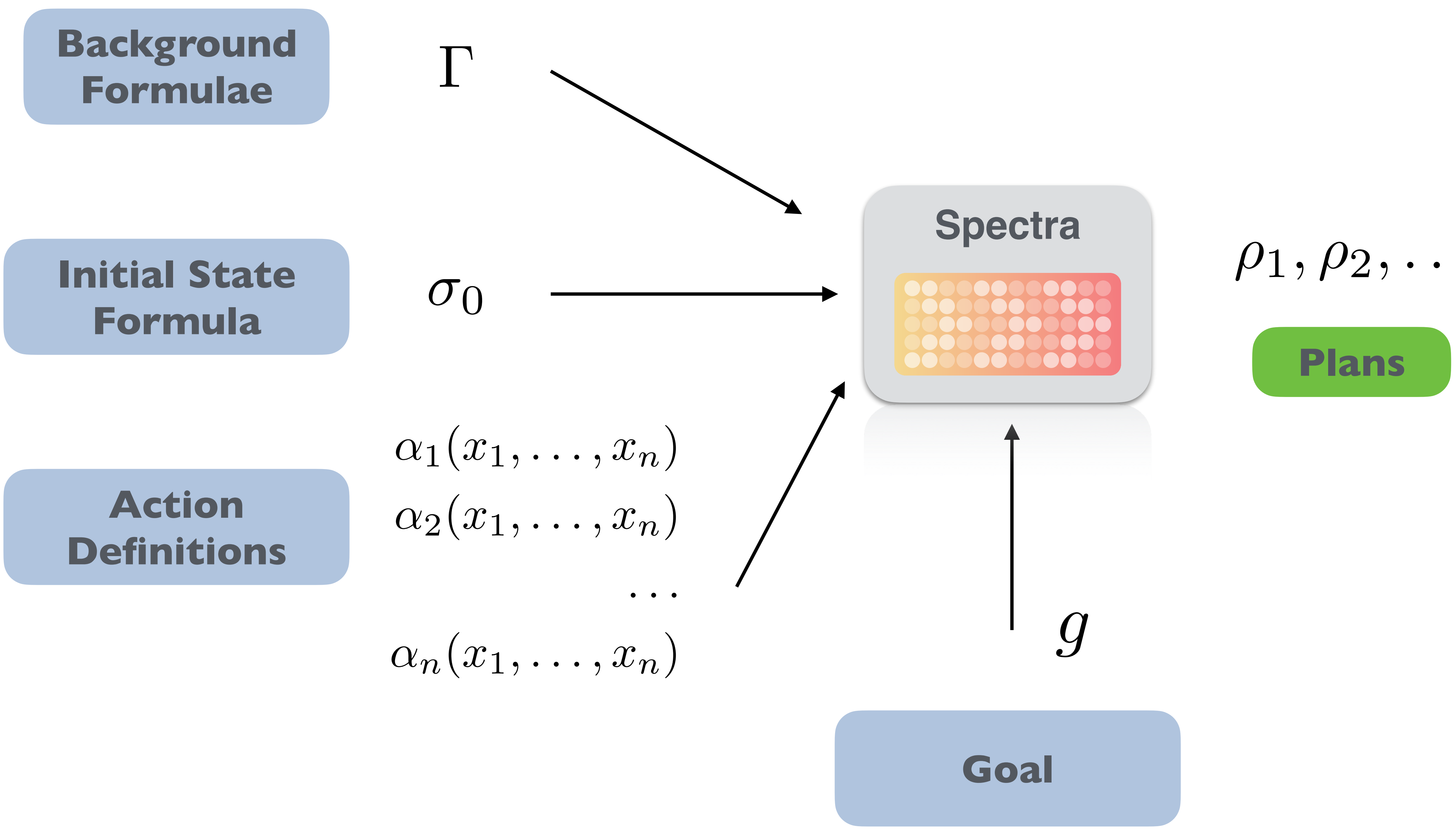
Recursion!

ANSWER:

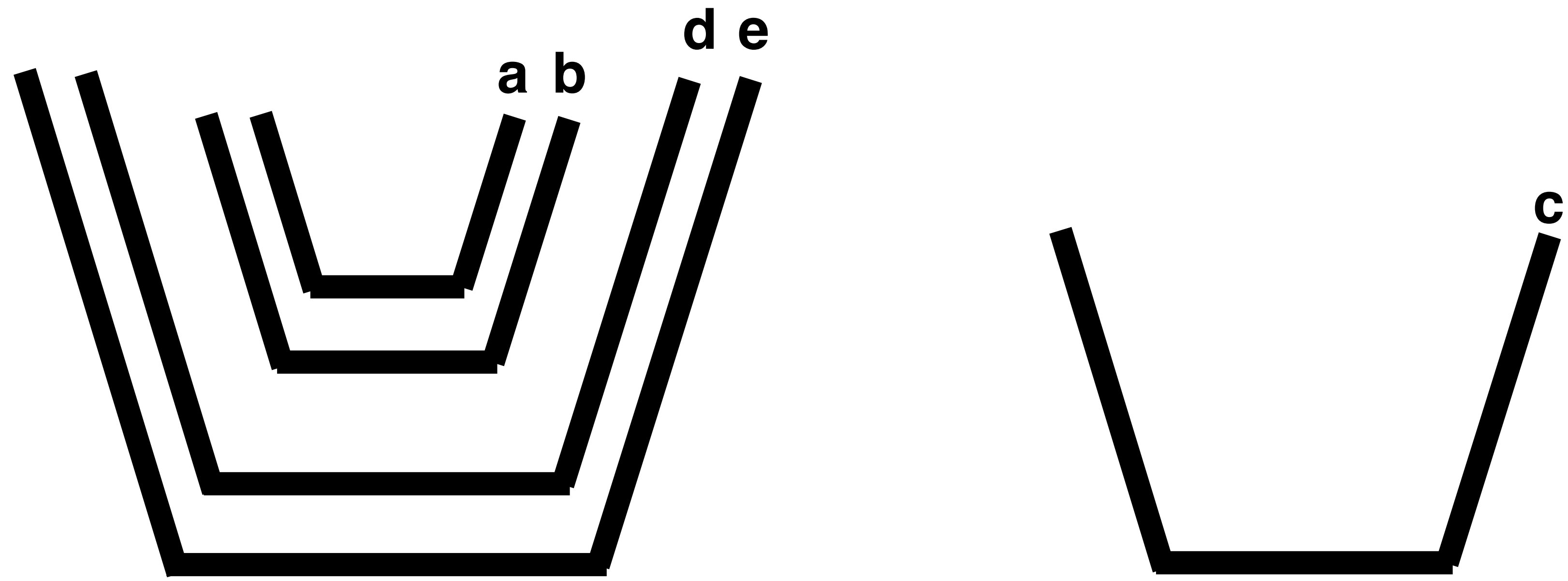
JUSTIFICATION:

SERIATED CUP CHALLENGE

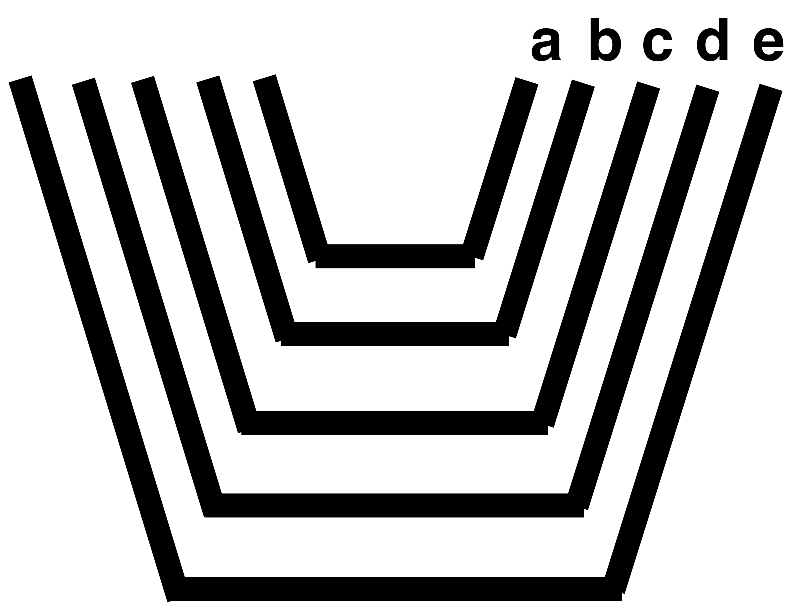
Spectra (planner)



Start



Goal State



```
:background [ ;; Transitivity of <
              (forall [?x ?y ?z]
                      (if (and (< (size ?x) (size ?y))
                               (< (size ?y) (size ?z)))
                          (< (size ?x) (size ?z))))
              ;; Asymmetry of <
              (forall [?x ?y]
                      (iff (< (size ?x) (size ?y))
                           (not (< (size ?y) (size ?x)))))

              ;; If there is something inside a cup, it is not empty.
              (forall [?y]
                      (if (exists [?x] (In ?x ?y))
                          (not (Empty ?y))))

              ;;; Sizes of cups
              (< (size a) (size b))
              (< (size b) (size c))
              (< (size c) (size d))
              (< (size d) (size e))]
```

Background Formulae

Start

```
:start        [(In a b)
               (In b d)
               (In d e)
               (Empty c)]
```

Action Definitions

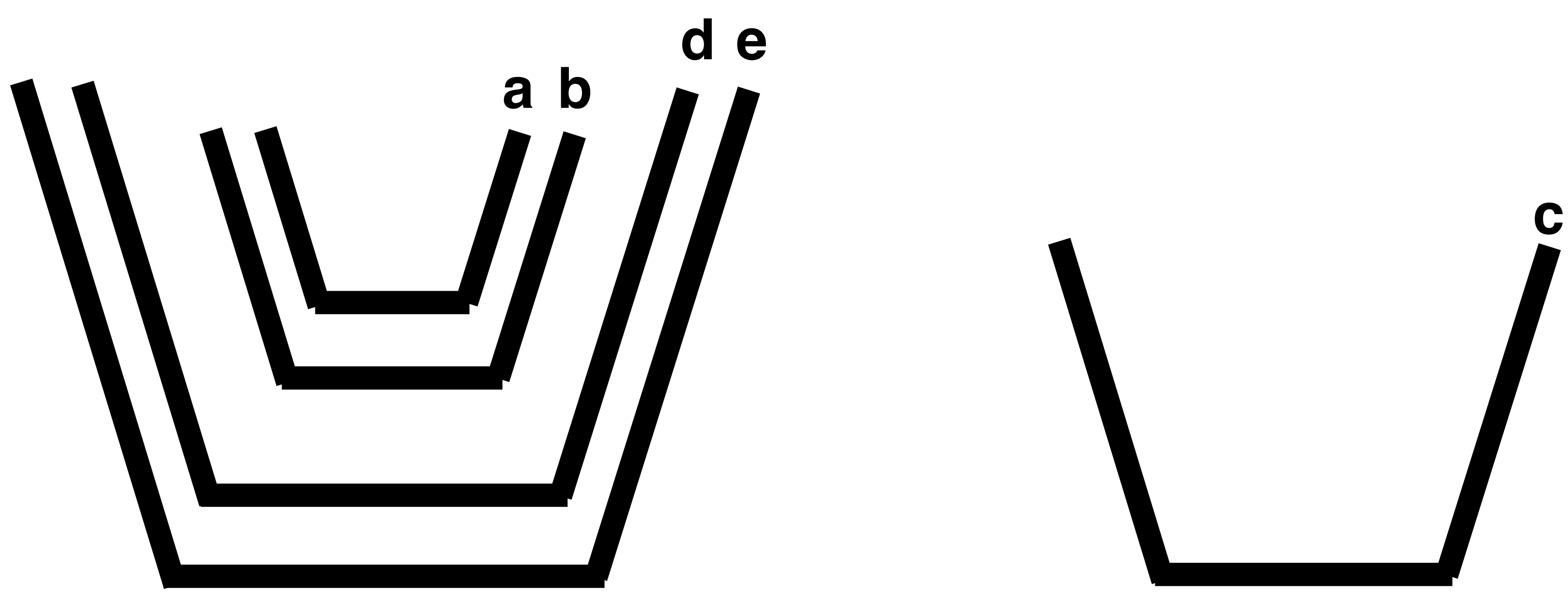
```
:actions  [(define-action placeInside [?x ?y]
                          {:preconditions [(< (size ?x) (size ?y))
                                           (Empty ?y)]
                           :additions     [(In ?x ?y)]
                           :deletions     [(Empty ?y)]})

           (define-action removeFrom [?x ?y]
                          {:preconditions [(In ?x ?y)]
                           :additions     [(Empty ?y)]
                           :deletions     [(In ?x ?y)]} )]
```

Goal

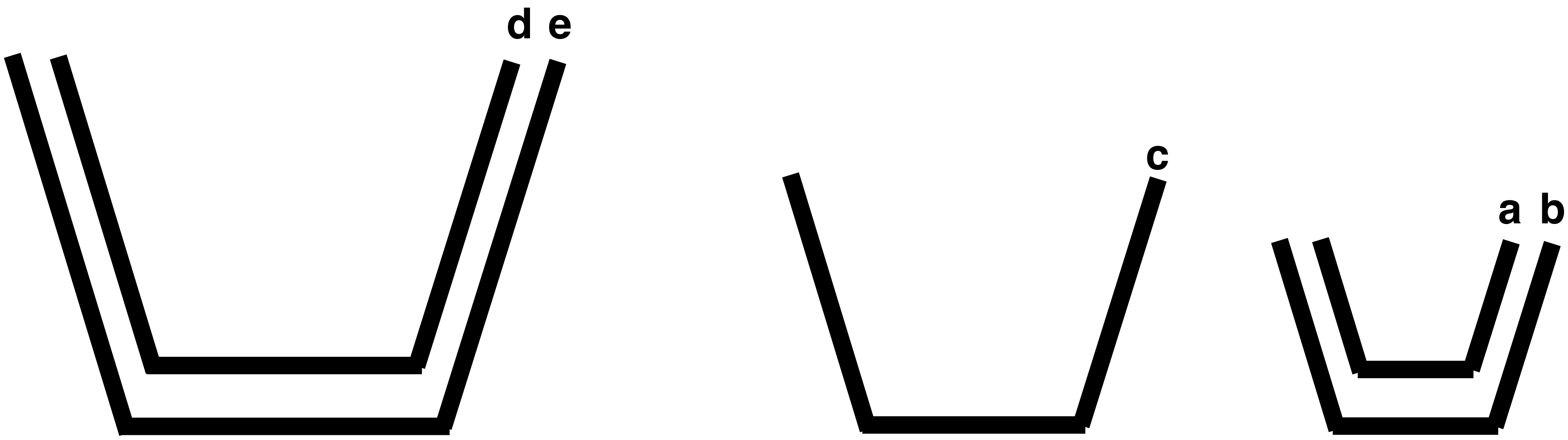
```
:goals {G1 {:priority 1.0
            :state    [(In a b)
                       (In b c)
                       (In c d)
                       (In d e)]}}
```

```
========================
Trying to Add Goal: G1
Successfully added: G1
Current Goals: [G1]
Plan: [(removeFrom  b d), (placeInside  b c), (placeInside  c d)]
--------------------------
Time Taken: 22.99s
```

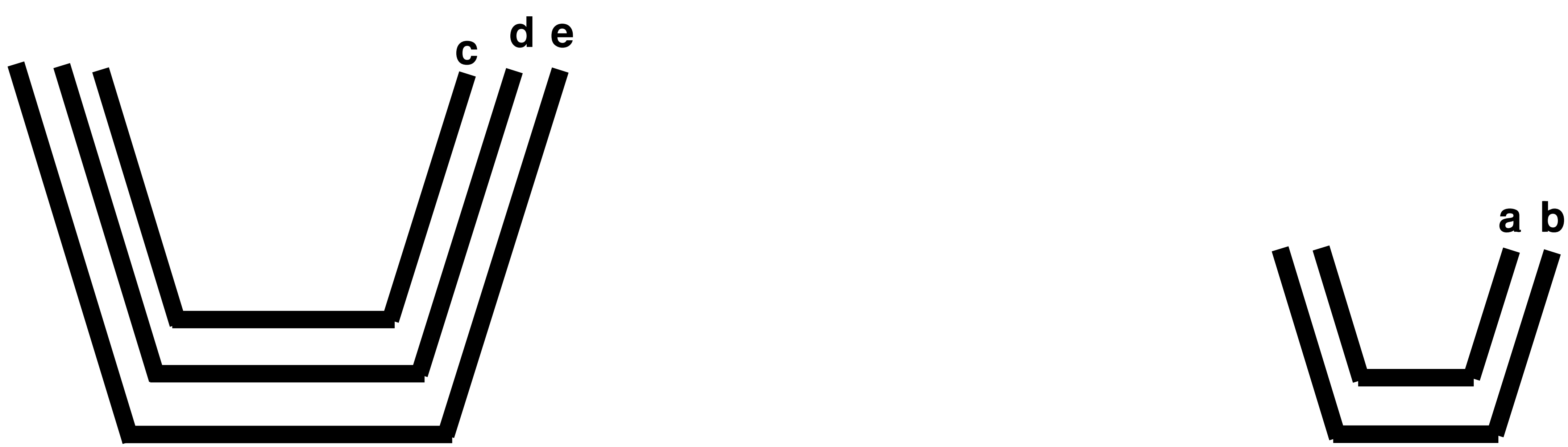
start



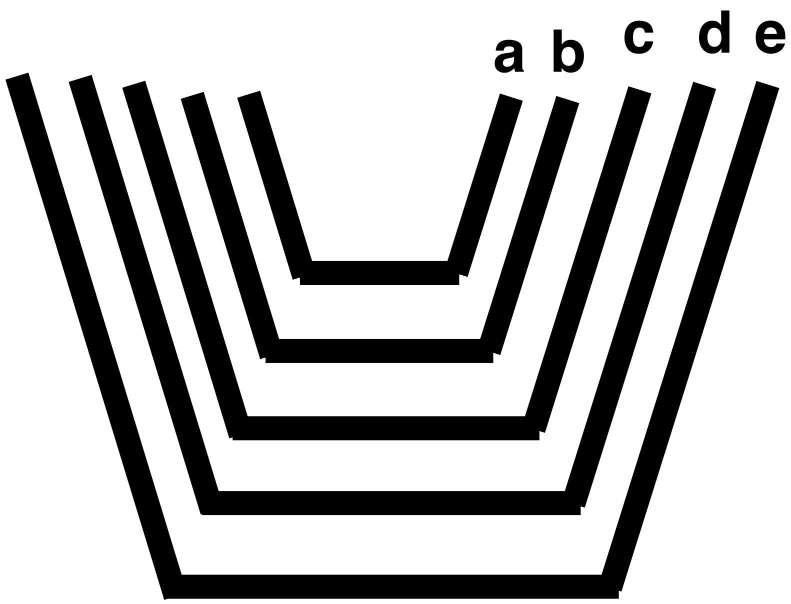
(removeFrom b d)



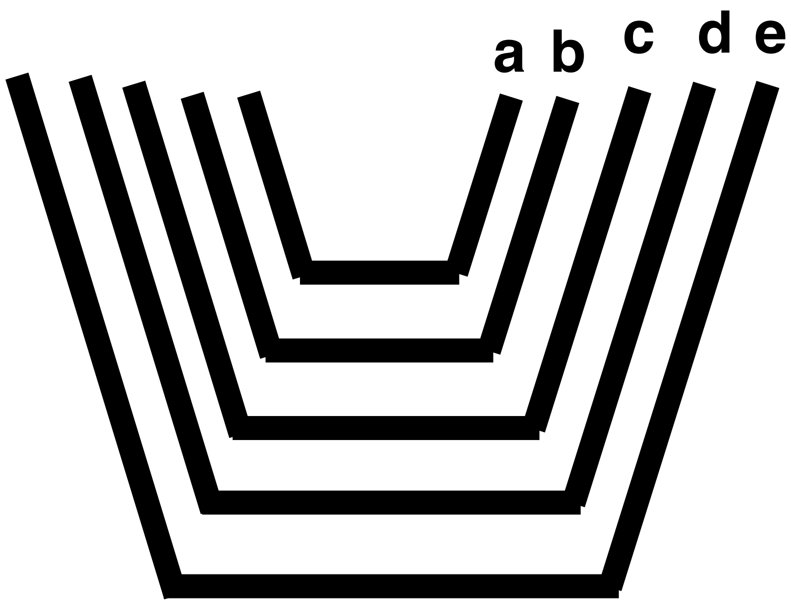
(placeInside b c)



(placeInside b c)



(placeInside b c)



Two Rudimentary Facts re Planning

Two Rudimentary Facts re Planning

Theorem: Planning in cups world 1 (cw1) is **NP**-hard.

Two Rudimentary Facts re Planning

Theorem: Planning in cups world 1 (cw1) is **NP**-hard.

Theorem: Planning in quantified blocks world 1 (qbw1) is Σ_1.

Two Rudimentary Facts re Planning

Theorem: Planning in cups world 1 (cw1) is **NP**-hard.

Theorem: Planning in quantified blocks world 1 (qbw1) is Σ_1.

(And how difficult is it to generate Spectra?)

Humans

Animals

Humans

Humans can on an ongoing basis solve qbw1 problems.

Animals

Humans

Humans can on an ongoing basis solve qbw1 problems.

Animals

Nonhuman animals can’t do anything of the sort; if lucky, they *can* solve a restricted version of the seriated cup challenge — & for the sake or argument we can readily grant that nonhuman animals can solve cw1.

Humans

Humans can on an ongoing basis solve qbw1 problems.

Animals

Nonhuman animals can't do anything of the sort; if lucky, they *can* solve a restricted version of the seriated cup challenge — & for the sake or argument we can readily grant that nonhuman animals can solve cw1.

Some elements of some formalized human-animal planning have zero overlap with any elements of some formalized animal planning!

Humans	Humans can on an ongoing basis solve qbw1 problems.

Animals	Nonhuman animals can't do anything of the sort; if lucky, they *can* solve a restricted version of the seriated cup challenge — & for the sake or argument we can readily grant that nonhuman animals can solve cw1.

Some elements of some formalized human-animal planning have zero overlap with any elements of some formalized animal planning!

Sorry, Darwin.

Humans Humans can on an ongoing basis solve qbw1 problems.

Animals Nonhuman animals can't do anything of the sort; if lucky, they *can* solve a restricted version of the seriated cup challenge — & for the sake or argument we can readily grant that nonhuman animals can solve cw1.

Some elements of some formalized human-animal planning have zero overlap with any elements of some formalized animal planning!

Sorry, Darwin.

Humans Humans can on an ongoing basis solve qbw1 problems.

Animals Nonhuman animals can't do anything of the sort; if lucky, they *can* solve a restricted version of the seriated cup challenge — & for the sake or argument we can readily grant that nonhuman animals can solve cw1.

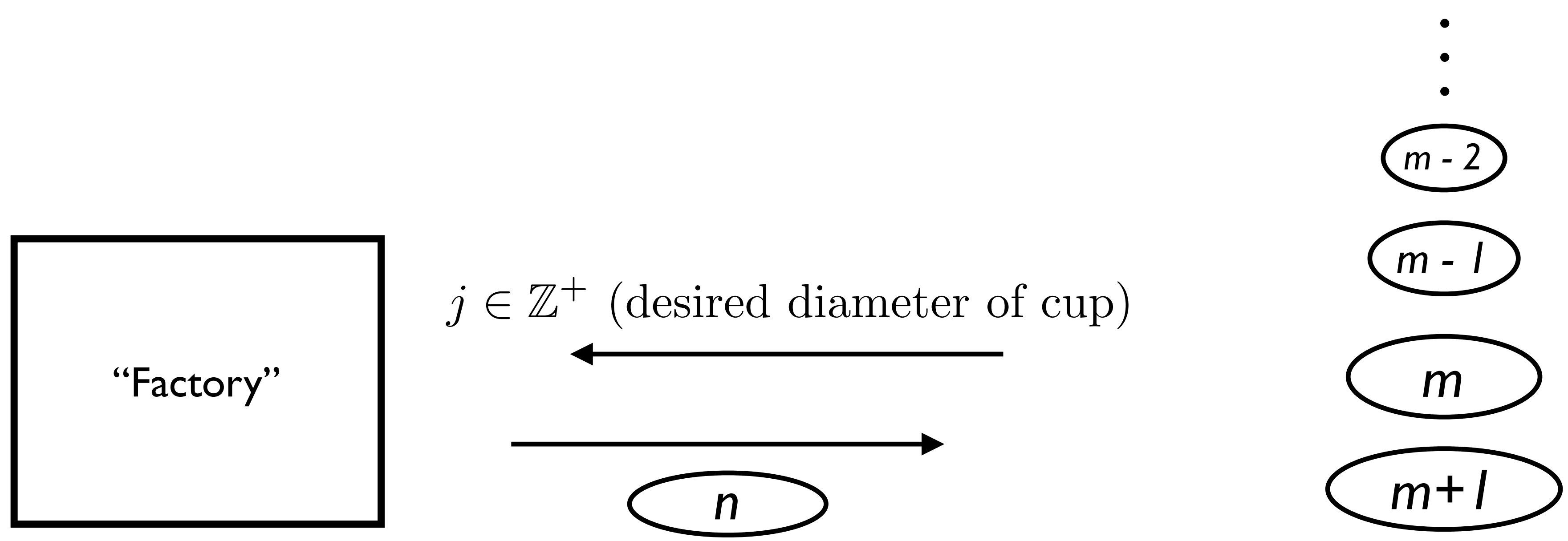
Some elements of some formalized human-animal planning have zero overlap with any elements of some formalized animal planning!

Sorry, Darwin.

And now for the infinitary challenge …

Selmer's Seriated Cup Challenge #1

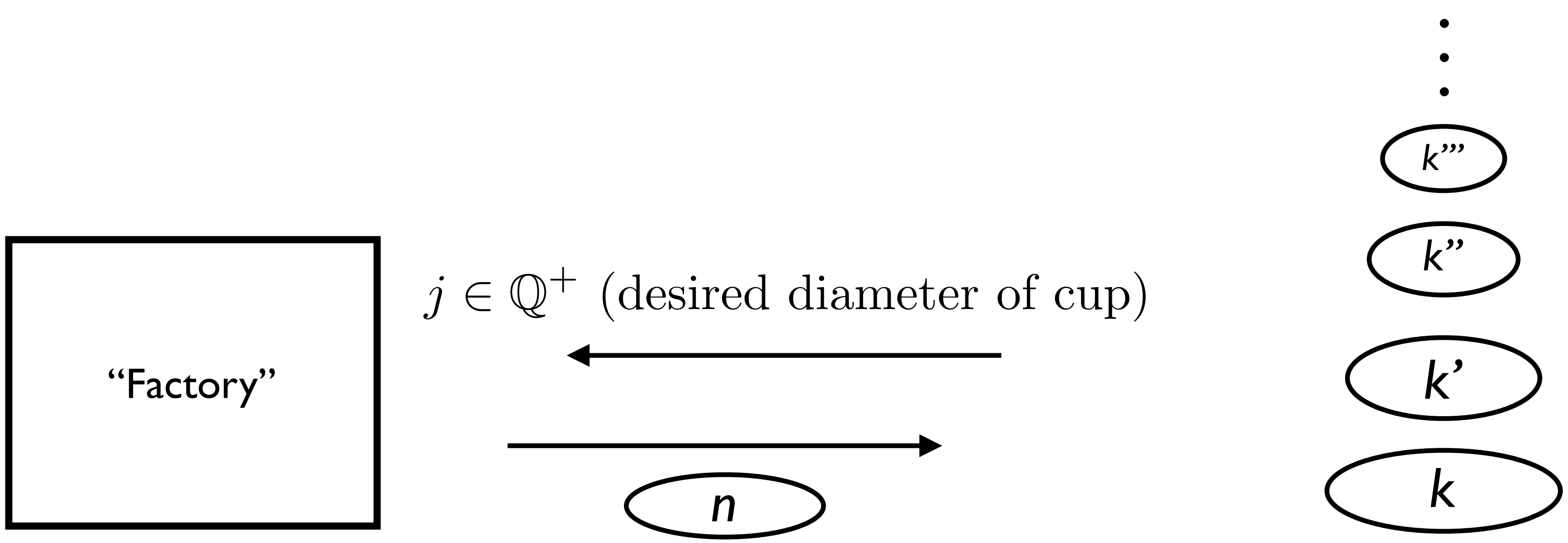
Suppose you have at your disposal a "factory" that, upon hearing you announce a number *j*, can quickly output a cup having a diameter of precisely *j* units. Can you insert a new cup between two of the seriated, stacked cups in the tower shown here? — where the *j* you send in *must* be a positive integer, *m* is likewise a positive integer, and every cup in every tower must be more in diameter than the one immediately above it, and less in diameter than the one immediately below it? ** Prove that your answer is correct.



**E.g., if *m* = 3, the tower in that case will have a base cup 4 units in diameter, immediately above that a cup 3 units in diameter, then a cup 2 units in diameter, and then finally a top cup of 1 unit in diameter.

Selmer's Seriated Cup Challenge #2

Suppose you have at your disposal a "factory" that, upon hearing you announce a number *j*, can quickly output a cup having a diameter of precisely *j* units. Can you insert a new cup between two of the seriated, stacked cups in the tower shown here? — where the *j* you send in *must* be a positive rational number; *k, k', k", k'"* … are likewise positive rational numbers, and every cup in every tower must be more in diameter than the one immediately above it, and less in diameter than the one immediately below it?** Prove that your answer is correct.



**E.g., if *k* = $\frac{1}{2}$, the tower in that case will have a base cup $\frac{1}{2}$ units in diameter, immediately above that there could be a cup $\frac{1}{3}$ units in diameter, then perhaps a cup $\frac{1}{4}$ units in diameter, and then perhaps finally a top cup of $\frac{1}{32}$ units in diameter.

Humans

Animals

Humans

Humans can discover answers and corresponding proofs at the level (minimally) of elementary *infinitary* number and set theory.

Animals

Humans

Humans can discover answers and corresponding proofs at the level (minimally) of elementary *infinitary* number and set theory.

Animals

Nonhuman animals can't do anything of the sort.

Humans

Humans can discover answers and corresponding proofs at the level (minimally) of elementary *infinitary* number and set theory.

Animals

Nonhuman animals can't do anything of the sort.

Some elements of some formalized human-animal behavior have zero overlap with any elements of some formalized animal behavior!

Humans

Humans can discover answers and corresponding proofs at the level (minimally) of elementary *infinitary* number and set theory.

Animals

Nonhuman animals can't do anything of the sort.

Some elements of some formalized human-animal behavior have zero overlap with any elements of some formalized animal behavior!

Humans

Humans can discover answers and corresponding proofs at the level (minimally) of elementary *infinitary* number and set theory.

Animals

Nonhuman animals can't do anything of the sort.

Some elements of some formalized human-animal behavior have zero overlap with any elements of some formalized animal behavior!

Double sorry, Darwin.

So, ...

minimally, deductive reasoning is valid, and grasped as such, when the content-independent *form* of the progression from premise(s) to conclusion accords with certain unassailable, abstract structures that ensure that if the premises are true, the conclusion *must* be true as well. And the production of worthwhile deductive reasoning is based on the search for interesting progressions that accord with such structures.

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn’t the sense of ‘inference’ I’m talking about. This is a highly limited sense of ‘inference’ that can be applied to nearly any organism. Yrs, //Selmer

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn't the sense of 'inference' I'm talking about. This is a highly limited sense of 'inference' that can be applied to nearly any organism. Yrs, //Selmer

Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
doi: 10.1111/phpr.12455

Rational Inference: The Lowest Bounds

CAMERON BUCKNER
University of Houston

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn't the sense of 'inference' I'm talking about. This is a highly limited sense of 'inference' that can be applied to nearly any organism. Yrs, //Selmer

In this paper, I aim to establish the lowest bounds of rational inference, arguing that many (though not all) of these opaque judgments in nonlinguistic animals should be counted as inferential. To clarify, there is a trivial sense of 'inference', common in cognitive psychology, on which any information processing counts as inferential; this is not the sense of inference explored here. This more philosophical notion begins instead with the idea that, at minimum, rational inference (hereafter, just 'inference') is the mental

process of arriving at a conclusion on the basis of reasons which support it. Theoretical inference involves the process of arriving at one belief on the basis of others; for example: Fluffy is a cat, all cats love liver, therefore Fluffy loves liver. Practical inference, on which I shall focus here, involves instead the selection of an action to perform on the basis of beliefs and desires; for example: Fluffy desires to jump on the counter, Fluffy

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn't the sense of 'inference' I'm talking about. This is a highly limited sense of 'inference' that can be applied to nearly any organism. Yrs, //Selmer

In this paper, I aim to establish the lowest bounds of rational inference, arguing that many (though not all) of these opaque judgments in nonlinguistic animals should be counted as inferential. To clarify, there is a trivial sense of 'inference', common in cognitive psychology, on which any information processing counts as inferential; this is not the sense of inference explored here. This more philosophical notion begins instead with the idea that, at minimum, rational inference (hereafter, just 'inference') is the mental

process of arriving at a conclusion on the basis of reasons which support it. Theoretical inference involves the process of arriving at one belief on the basis of others; for example: Fluffy is a cat, all cats love liver, therefore Fluffy loves liver. Practical inference, on which I shall focus here, involves instead the selection of an action to perform on the basis of beliefs and desires; for example: Fluffy desires to jump on the counter, Fluffy

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn't the sense of 'inference' I'm talking about. This is a highly limited sense of 'inference' that can be applied to nearly any organism. Yrs, //Selmer

In this paper, I aim to establish the lowest bounds of rational inference, arguing that many (though not all) of these opaque judgments in nonlinguistic animals should be counted as inferential. To clarify, there is a trivial sense of 'inference', common in cognitive psychology, on which any information processing counts as inferential; this is not the sense of inference explored here. This more philosophical notion begins instead with the idea that, at minimum, rational inference (hereafter, just 'inference') is the mental

RATIONAL INFERENCE: THE LOWEST BOUNDS 1

process of arriving at a conclusion on the basis of reasons which support it. Theoretical inference involves the process of arriving at one belief on the basis of others; for example: Fluffy is a cat, all cats love liver, therefore Fluffy loves liver. Practical inference, on which I shall focus here, involves instead the selection of an action to perform on the basis of beliefs and desires; for example: Fluffy desires to jump on the counter, Fluffy

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn't the sense of 'inference' I'm talking about. This is a highly limited sense of 'inference' that can be applied to nearly any organism. Yrs, //Selmer

In this paper, I aim to establish the lowest bounds of rational inference, arguing that many (though not all) of these opaque judgments in nonlinguistic animals should be counted as inferential. To clarify, there is a trivial sense of 'inference', common in cognitive psychology, on which any information processing counts as inferential; this is not the sense of inference explored here. This more philosophical notion begins instead with the idea that, at minimum, rational inference (hereafter, just 'inference') is the mental

RATIONAL INFERENCE: THE LOWEST BOUNDS 1

process of arriving at a conclusion on the basis of reasons which support it. Theoretical inference involves the process of arriving at one belief on the basis of others; for example: Fluffy is a cat, all cats love liver, therefore Fluffy loves liver. Practical inference, on which I shall focus here, involves instead the selection of an action to perform on the basis of beliefs and desires; for example: Fluffy desires to jump on the counter, Fluffy

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn't the sense of 'inference' I'm talking about. This is a highly limited sense of 'inference' that can be applied to nearly any organism. Yrs, //Selmer

In this paper, I aim to establish the lowest bounds of rational inference, arguing that many (though not all) of these opaque judgments in nonlinguistic animals should be counted as inferential. To clarify, there is a trivial sense of 'inference', common in cognitive psychology, on which any information processing counts as inferential; this is not the sense of inference explored here. This more philosophical notion begins instead with the idea that, at minimum, rational inference (hereafter, just 'inference') is the mental

RATIONAL INFERENCE: THE LOWEST BOUNDS 1

process of arriving at a conclusion on the basis of reasons which support it. Theoretical inference involves the process of arriving at one belief on the basis of others; for example: Fluffy is a cat, all cats love liver, therefore Fluffy loves liver. Practical inference, on which I shall focus here, involves instead the selection of an action to perform on the basis of beliefs and desires; for example: Fluffy desires to jump on the counter, Fluffy

Hi Dan: Thx for bringing the excellent, recent paper to my attention, but this isn't the sense of 'inference' I'm talking about. This is a highly limited sense of 'inference' that can be applied to nearly any organism. Yrs, //Selmer

In this paper, I aim to establish the lowest bounds of rational inference, arguing that many (though not all) of these opaque judgments in nonlinguistic animals should be counted as inferential. To clarify, there is a trivial sense of 'inference', common in cognitive psychology, on which any information processing counts as inferential; this is not the sense of inference explored here. This more philosophical notion begins instead with the idea that, at minimum, rational inference (hereafter, just 'inference') is the mental

RATIONAL INFERENCE: THE LOWEST BOUNDS 1

process of arriving at a conclusion on the basis of reasons which support it. Theoretical inference involves the process of arriving at one belief on the basis of others; for example: Fluffy is a cat, all cats love liver, therefore Fluffy loves liver. Practical inference, on which I shall focus here, involves instead the selection of an action to perform on the basis of beliefs and desires; for example: Fluffy desires to jump on the counter, Fluffy

So, we return to ...
Darwin's Defense wrt Reasoning

	(1)	Story or anecdote S.	
$\therefore$	(2)	There exist animals manifesting behavior B.	from (1)
	(3)	Anything behaving as in B has the purportedly differentiating mental power of reasoning.	
$\therefore$	(4)	There exist animals the having purportedly differentiating mental power of reasoning.	from (2), (3), (4)

Very well. And the stories?

They embarrass me, and Darwin may well have had a dog fetish, but I convey some to you ...

“Dogs on Thin Ice”

“Dr. Hayes, in his work on *The Open Polar Sea*, repeatedly remarks that his dogs, instead of continuing to draw sledges in a compact body, diverged and separated when they came to thin ice, so that their weight might be more evenly distributed.”

“Thirsty Dogs”

“Houzeau relates that, while crossing a wide and arid plain in Texas, his two dogs suffered greatly from thirst, and that between thirty and forty times they rushed down the hollows to search for water. These hollows were not valleys, and there were no trees in them, or any other difference in the vegetation, and as they were absolutely dry there could have been no smell of damp earth. The dogs behaved as if they knew that a dip in the ground offered them the best chance of finding water.”

“A Smart Killer Dog”

“Mr. Colquhoun winged two wild ducks, which fell on the further side of a stream; his retriever tried to bring over both at once, but could not succeed; she then, though never before known to ruffle a feather, deliberately killed one, brought over the other, and returned for the dead bird.”

“A Murderous Dog”

“Col. Hutchinson relates that two partridges were shot at once, one being killed, the other wounded; the latter ran away, and was caught by the retriever, who on her return came across the dead bird: ‘she stopped, evidently greatly puzzled, and after one or two trials, finding she could not take it up without permitting the escape of the winged bird, she considered a moment, then deliberately murdered it by giving it a severe crunch, and afterward brought away both together. This was the only known instance of her ever having willfully injured any game.’ Here we have reason ... they show how strong their reasoning faculty must have been ...”

Please.

Please.

- This comes nearly 2000 years after Aristotle explained what deductive reasoning is, and gave simple but powerful deductive logics to make this clear ... and these dogs are said by a learned man to *reason*?

Please.

- This comes nearly 2000 years after Aristotle explained what deductive reasoning is, and gave simple but powerful deductive logics to make this clear ... and these dogs are said by a learned man to *reason*?
- We can build *non*-reasoning robots to do much more problem-solving than this.

Please.

- This comes nearly 2000 years after Aristotle explained what deductive reasoning is, and gave simple but powerful deductive logics to make this clear ... and these dogs are said by a learned man to *reason*?
- We can build *non*-reasoning robots to do much more problem-solving than this.
- A dog can't even have third-order beliefs.

Please.

- This comes nearly 2000 years after Aristotle explained what deductive reasoning is, and gave simple but powerful deductive logics to make this clear ... and these dogs are said by a learned man to *reason*?
- We can build *non*-reasoning robots to do much more problem-solving than this.
- A dog can't even have third-order beliefs.
    - Does Fido believe that you believe that your mother believes Fido is a good dog at the moment?

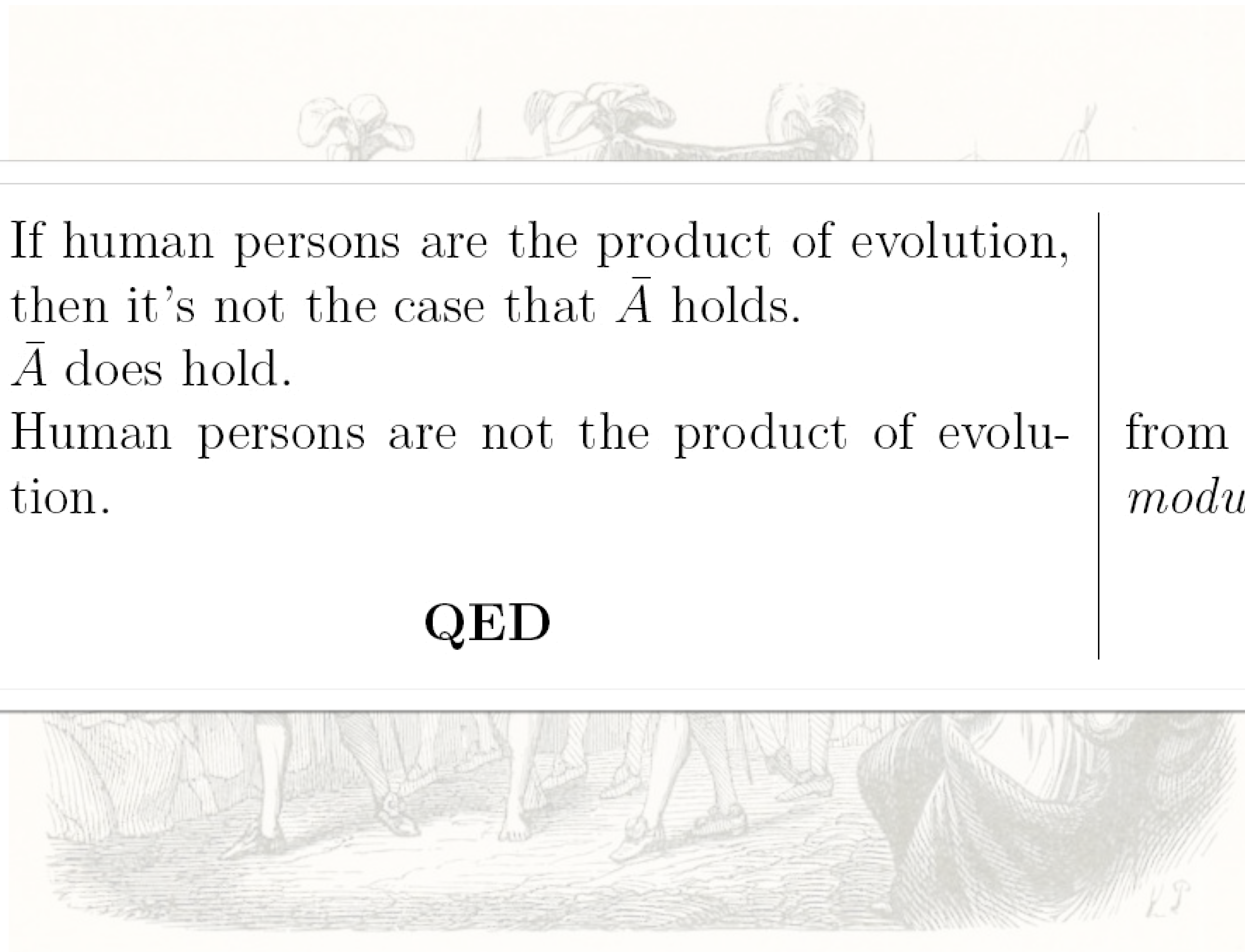
Please.

- This comes nearly 2000 years after Aristotle explained what deductive reasoning is, and gave simple but powerful deductive logics to make this clear ... and these dogs are said by a learned man to *reason*?
- We can build *non*-reasoning robots to do much more problem-solving than this.
- A dog can't even have third-order beliefs.
  - Does Fido believe that you believe that your mother believes Fido is a good dog at the moment?
- Animals can't reason, *certainly* can't reason in infinitary fashion; and so, my friends, I am home free, and part ways with the undressed king and those who follow the groupthink of our age, and hence proclaim with the co-discoverer of evolution, that while my spine may be descended from some brute's in an epoch long past, my mind, and yours alike, is not.

Finis

Finis

Finis

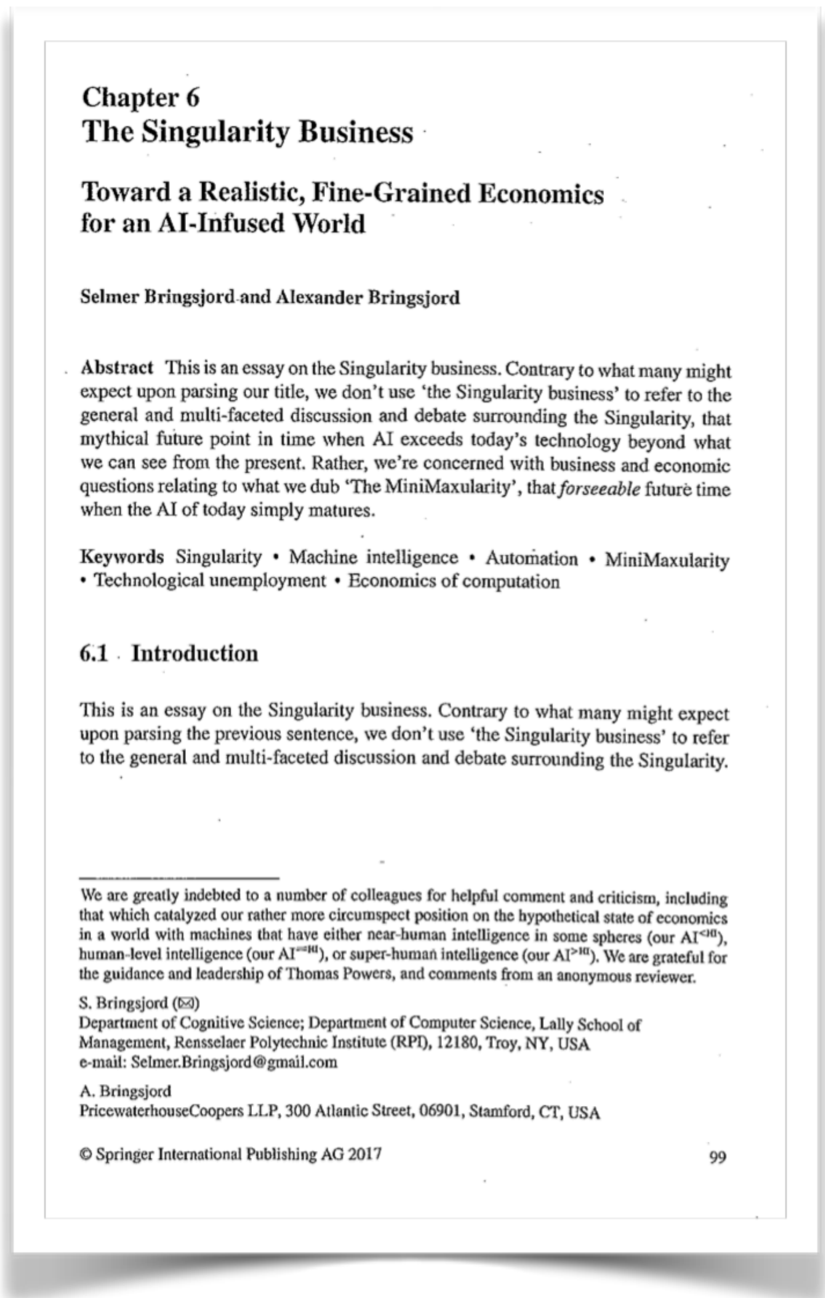


	(1)	If human persons are the product of evolution, then it's not the case that $\bar{A}$ holds.	
	(2)	$\bar{A}$ does hold.	
∴	(3)	Human persons are not the product of evolution.	from (1), (2) by *modus tollens*

QED

Based in large measure on …

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



Chapter 6
The Singularity Business

Toward a Realistic, Fine-Grained Economics for an AI-Infused World

Selmer Bringsjord and Alexander Bringsjord

Abstract This is an essay on the Singularity business. Contrary to what many might expect upon parsing our title, we don't use 'the Singularity business' to refer to the general and multi-faceted discussion and debate surrounding the Singularity, that mythical future point in time when AI exceeds today's technology beyond what we can see from the present. Rather, we're concerned with business and economic questions relating to what we dub 'The MiniMaxularity', that *forseeable* future time when the AI of today simply matures.

Keywords Singularity • Machine intelligence • Automation • MiniMaxularity • Technological unemployment • Economics of computation

6.1 Introduction

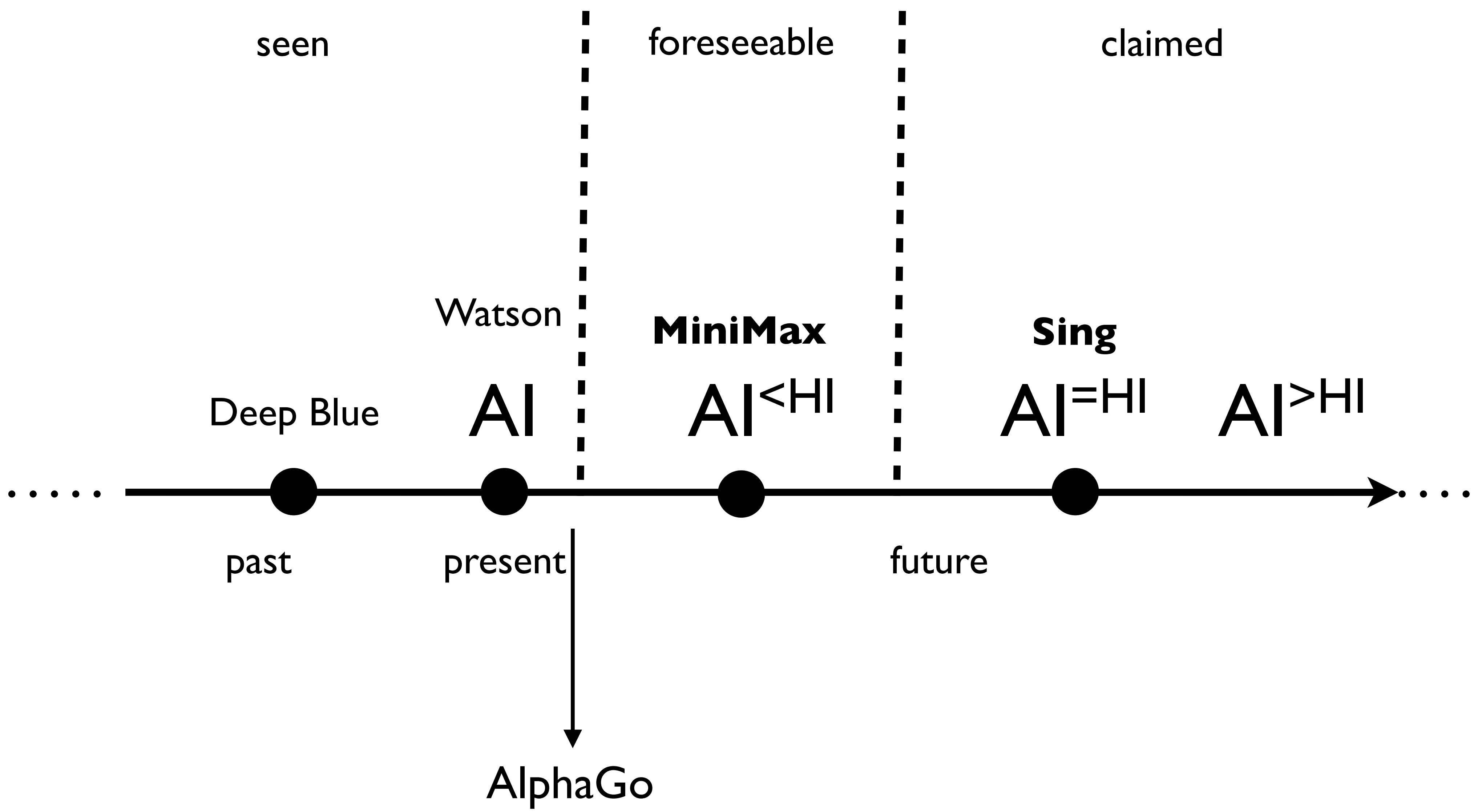
This is an essay on the Singularity business. Contrary to what many might expect upon parsing the previous sentence, we don't use 'the Singularity business' to refer to the general and multi-faceted discussion and debate surrounding the Singularity.

We are greatly indebted to a number of colleagues for helpful comment and criticism, including that which catalyzed our rather more circumspect position on the hypothetical state of economics in a world with machines that have either near-human intelligence in some spheres (our $AI^{<HI}$), human-level intelligence (our $AI^{==HI}$), or super-human intelligence (our $AI^{>HI}$). We are grateful for the guidance and leadership of Thomas Powers, and comments from an anonymous reviewer.

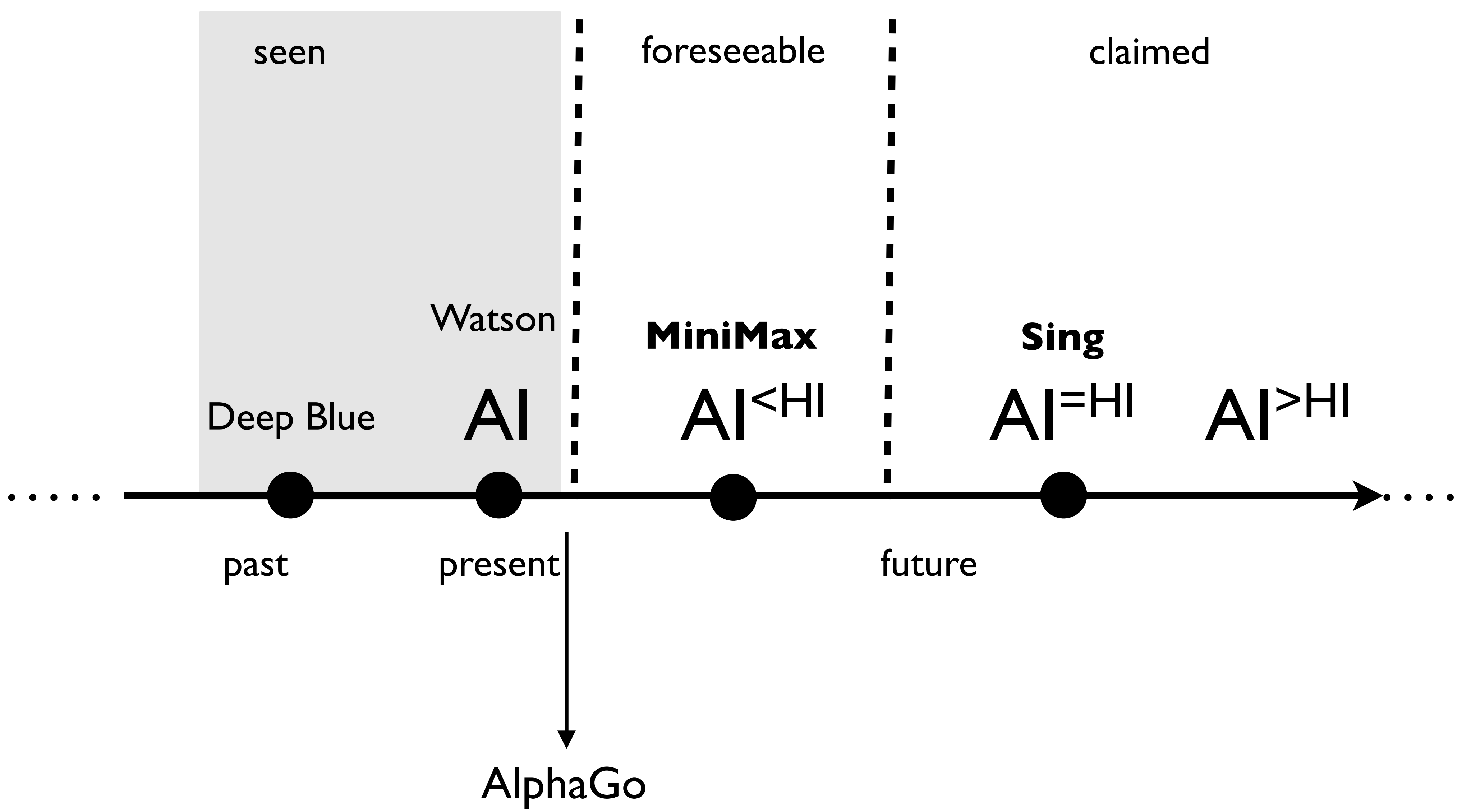
S. Bringsjord (✉)
Department of Cognitive Science; Department of Computer Science, Lally School of Management, Rensselaer Polytechnic Institute (RPI), 12180, Troy, NY, USA
e-mail: Selmer.Bringsjord@gmail.com

A. Bringsjord
PricewaterhouseCoopers LLP, 300 Atlantic Street, 06901, Stamford, CT, USA

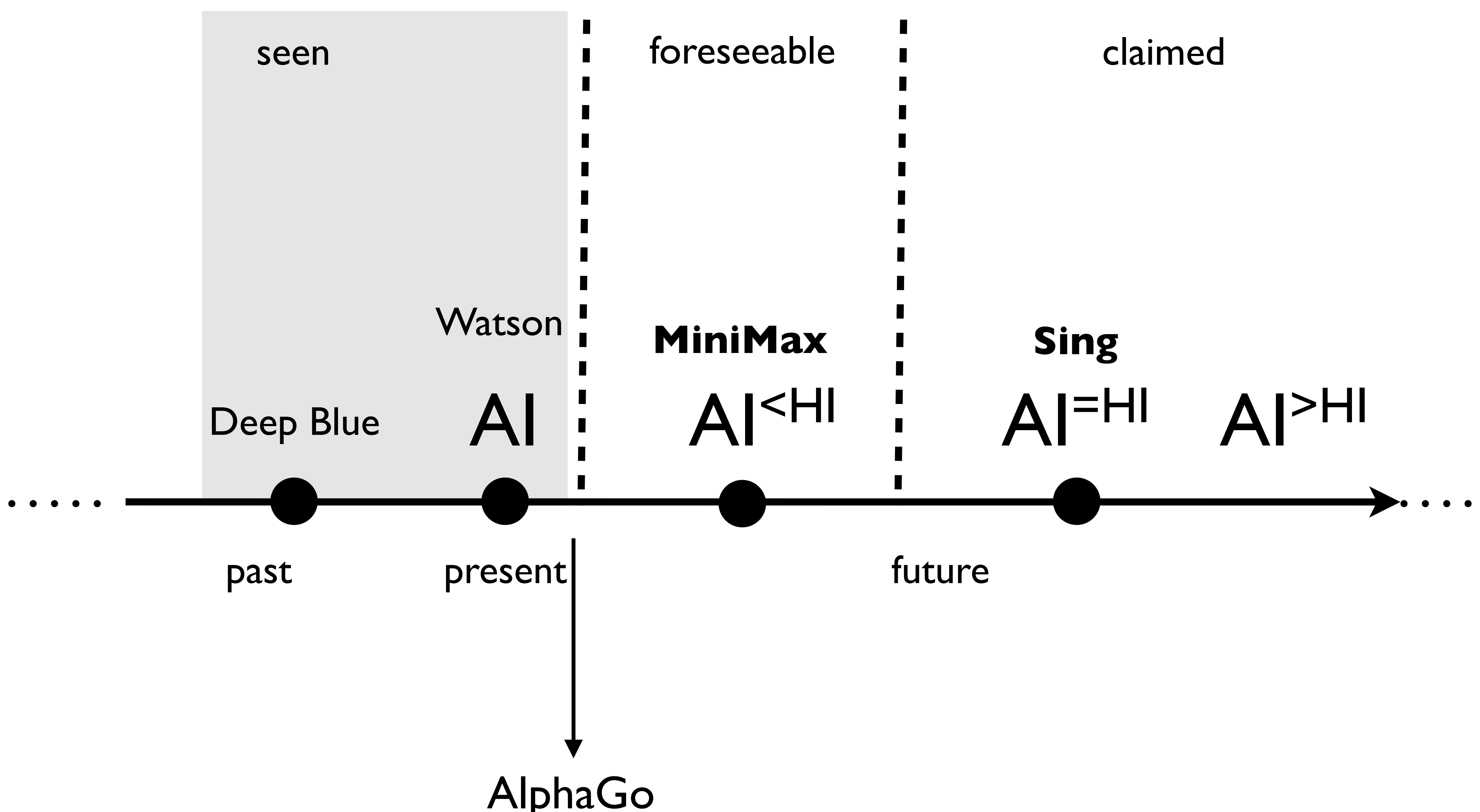
Before: A Realistic Timeline



Before: A Realistic Timeline

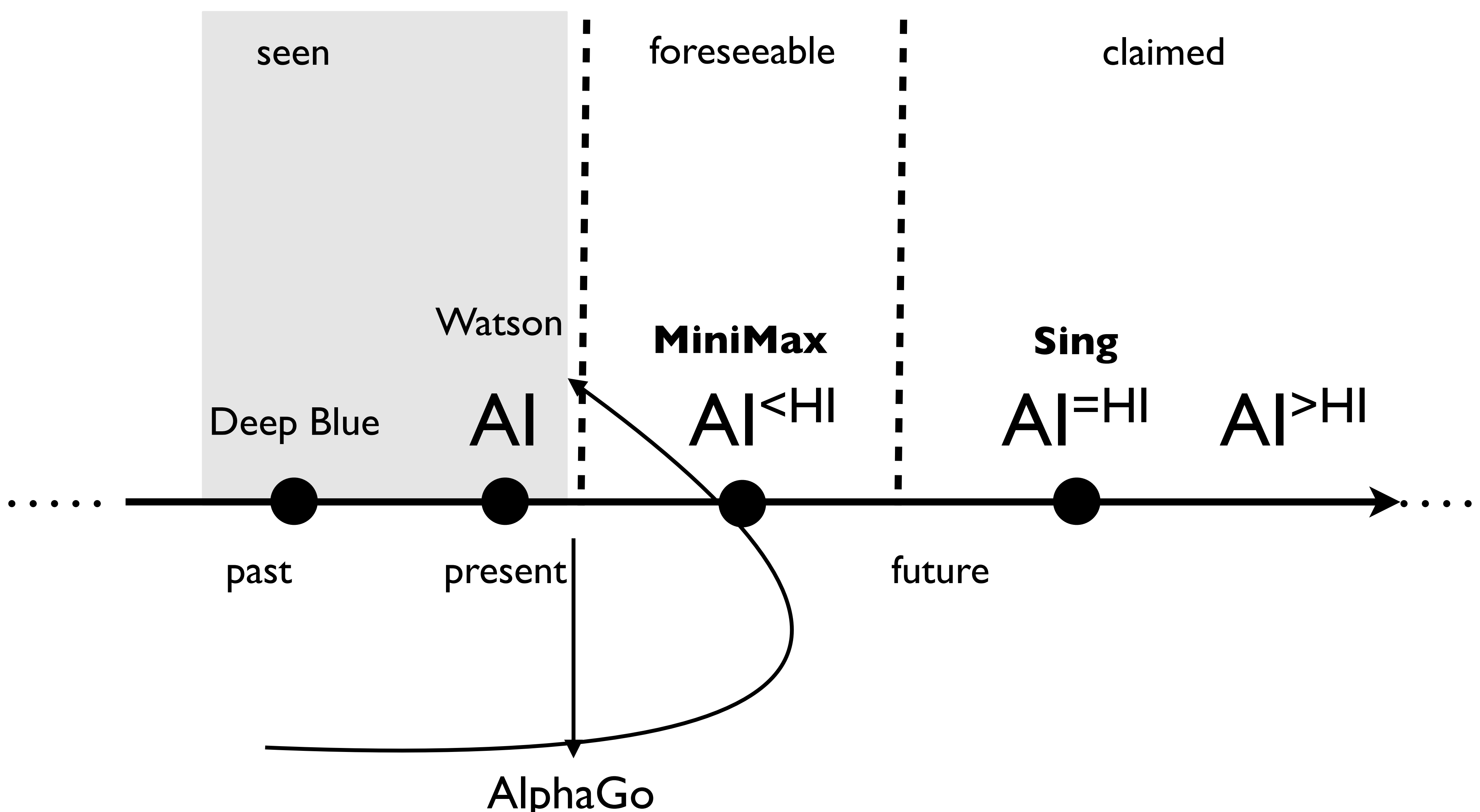


Before: A Realistic Timeline



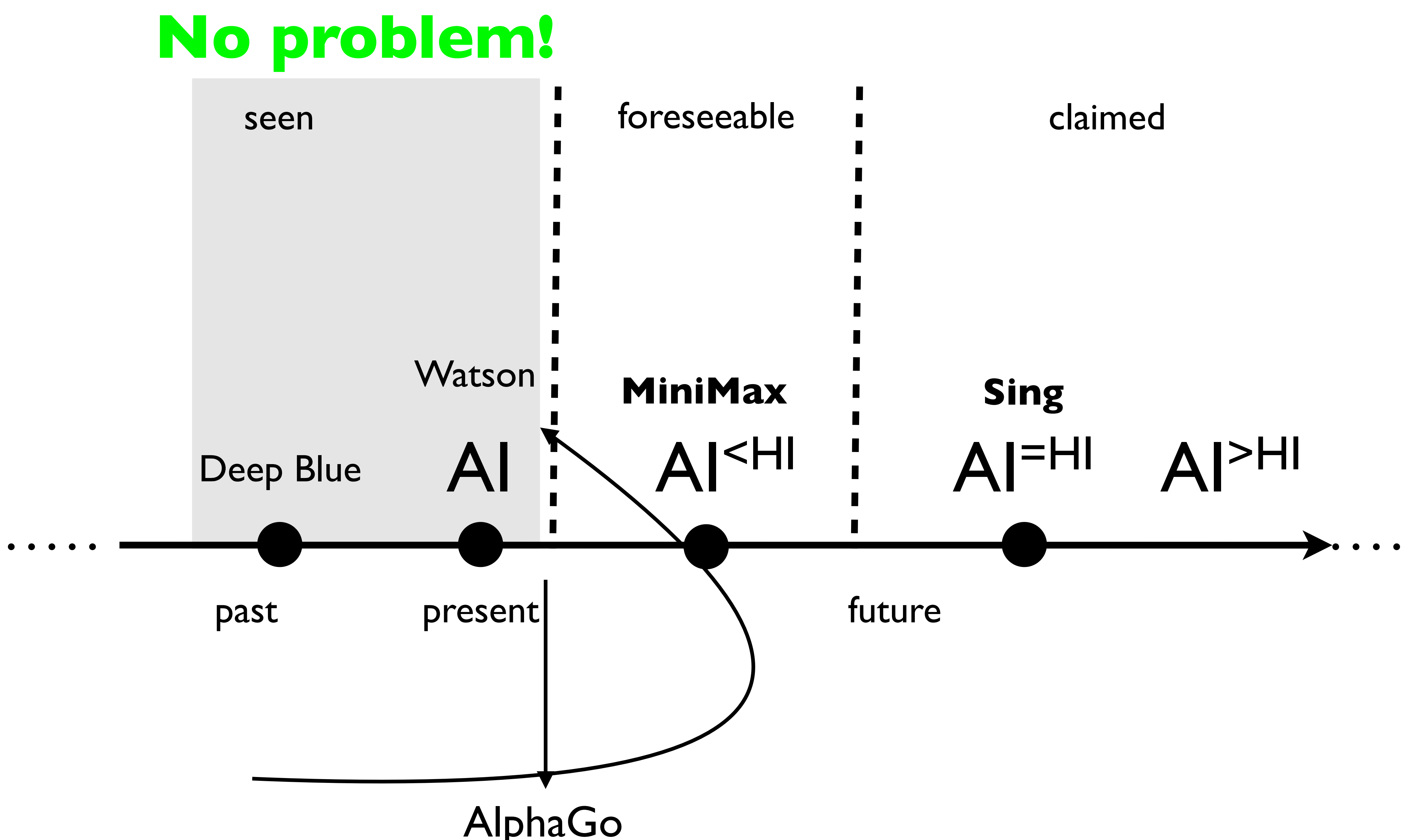
Also, AlphaGo

Before: A Realistic Timeline



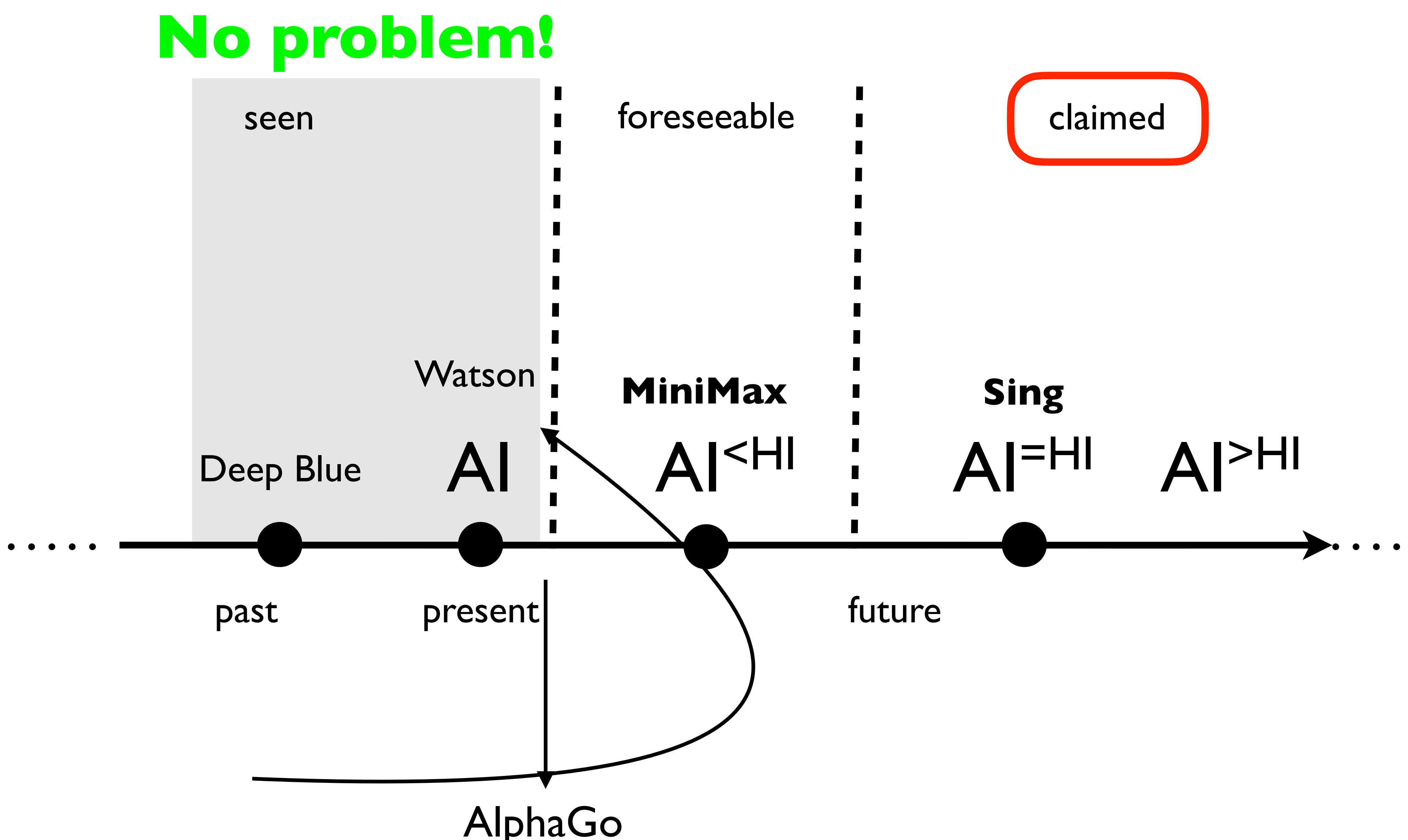
Also, AlphaGo

Before: A Realistic Timeline



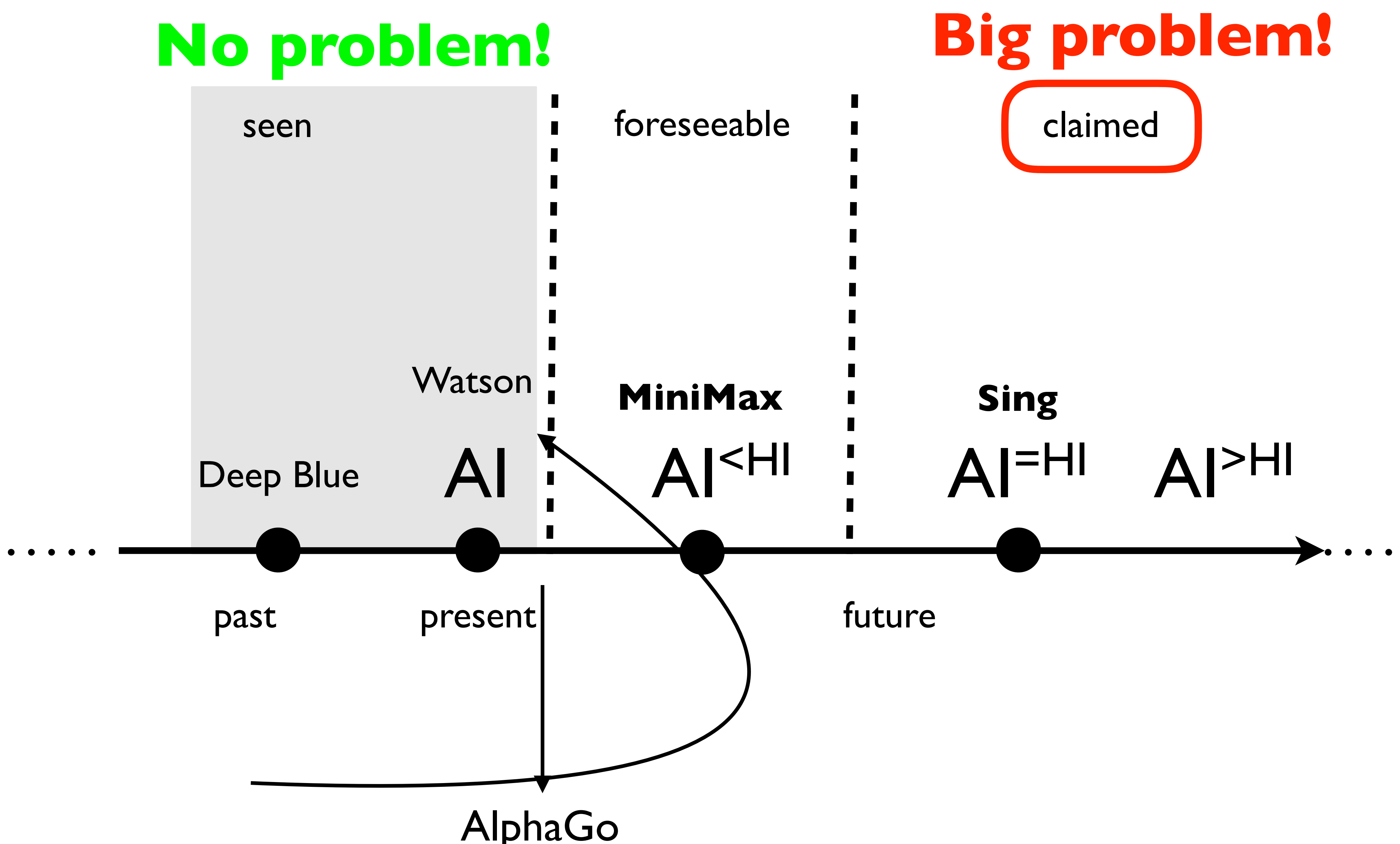
Also, AlphaGo

Before: A Realistic Timeline



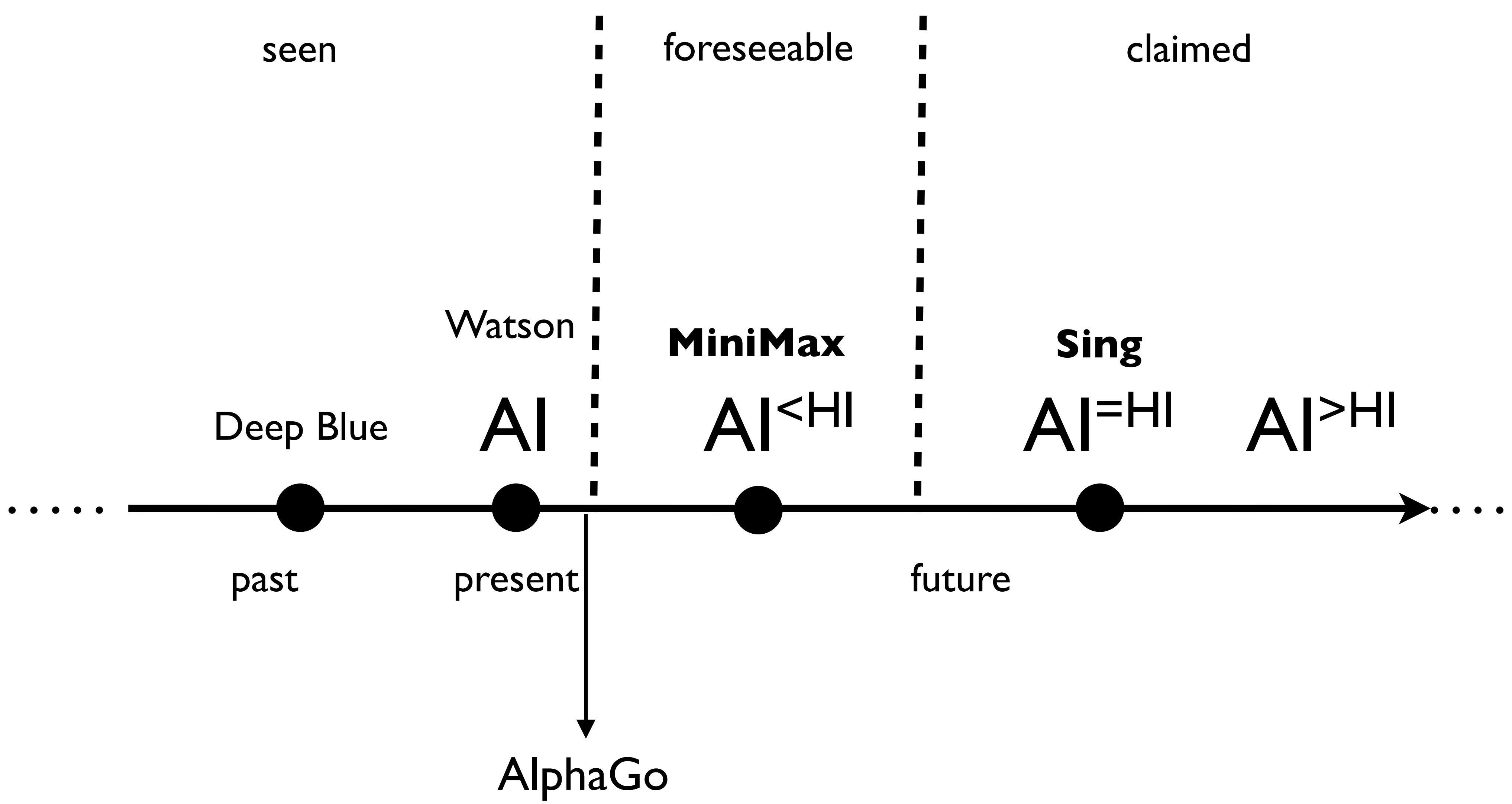
Also, AlphaGo

Before: A Realistic Timeline

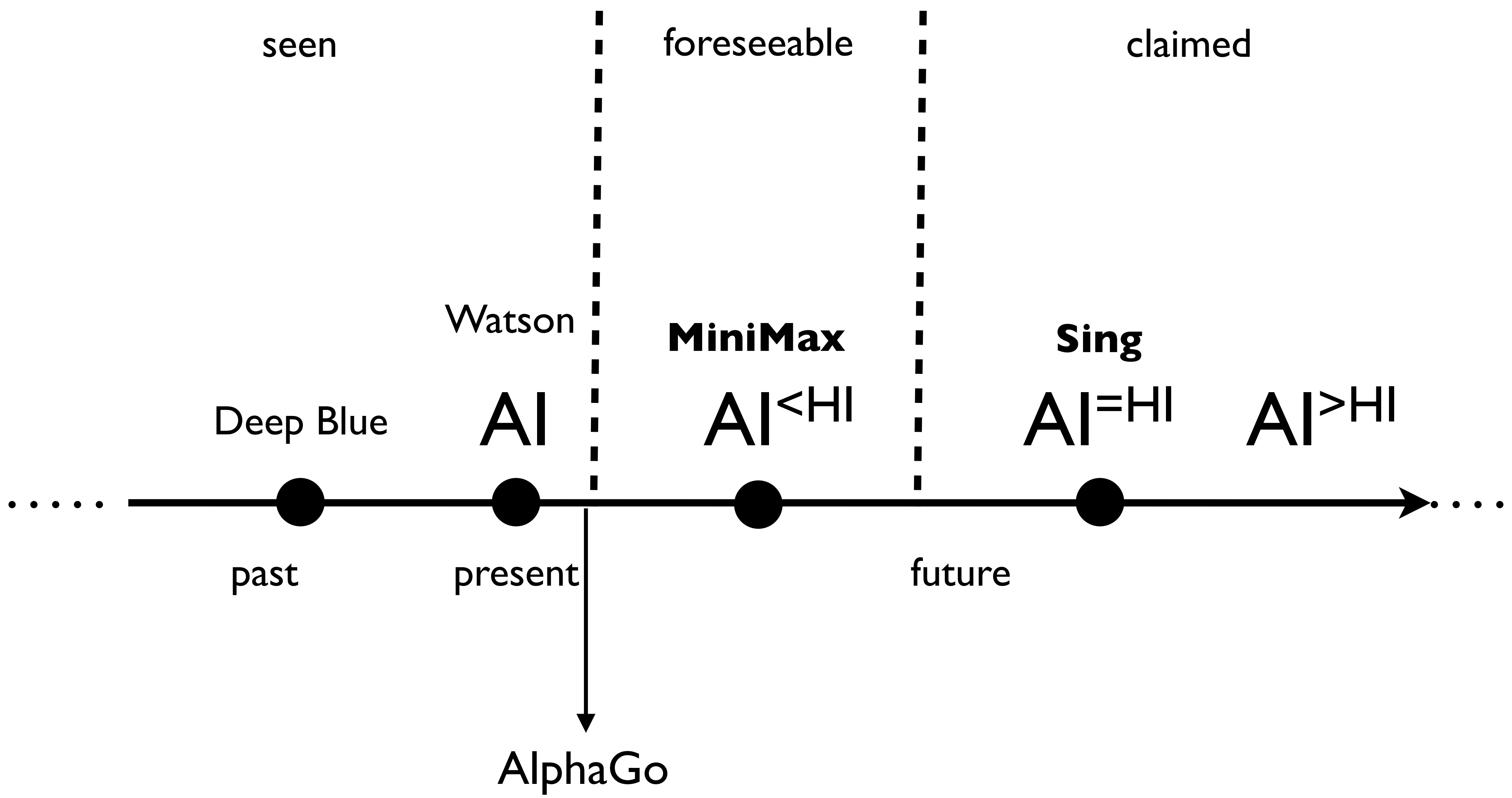


Also, AlphaGo

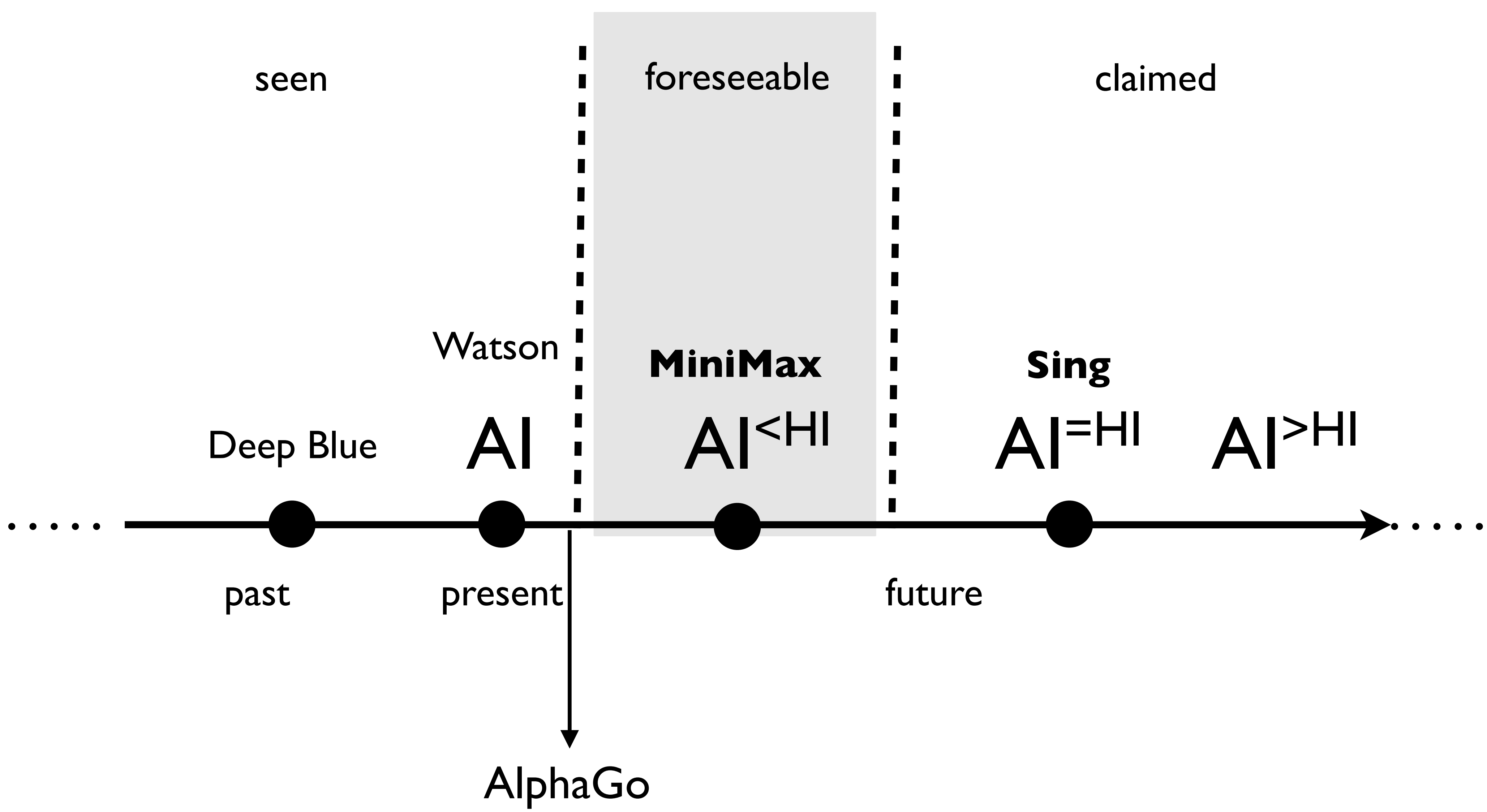
A Realistic Timeline



A Realistic Timeline

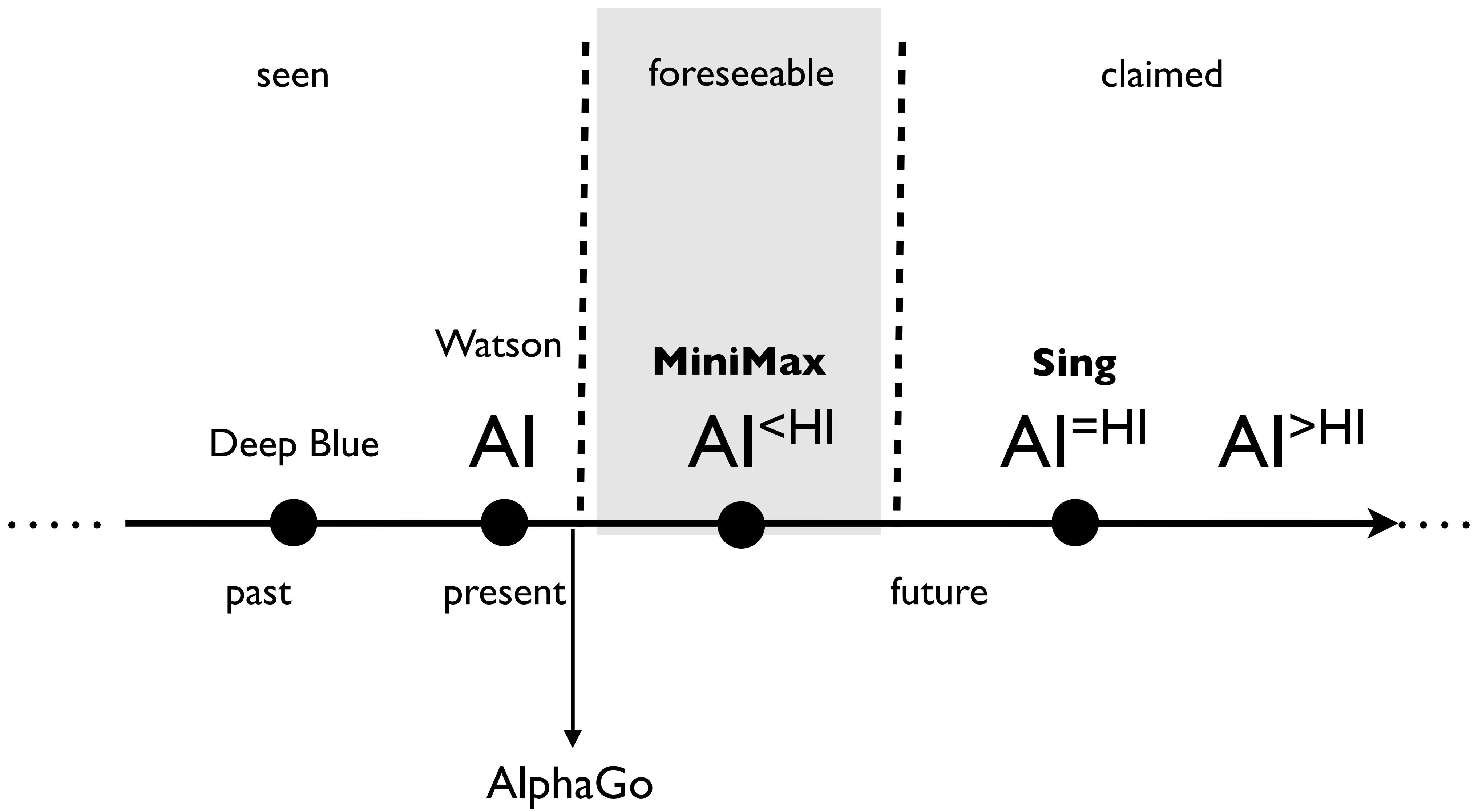


A Realistic Timeline



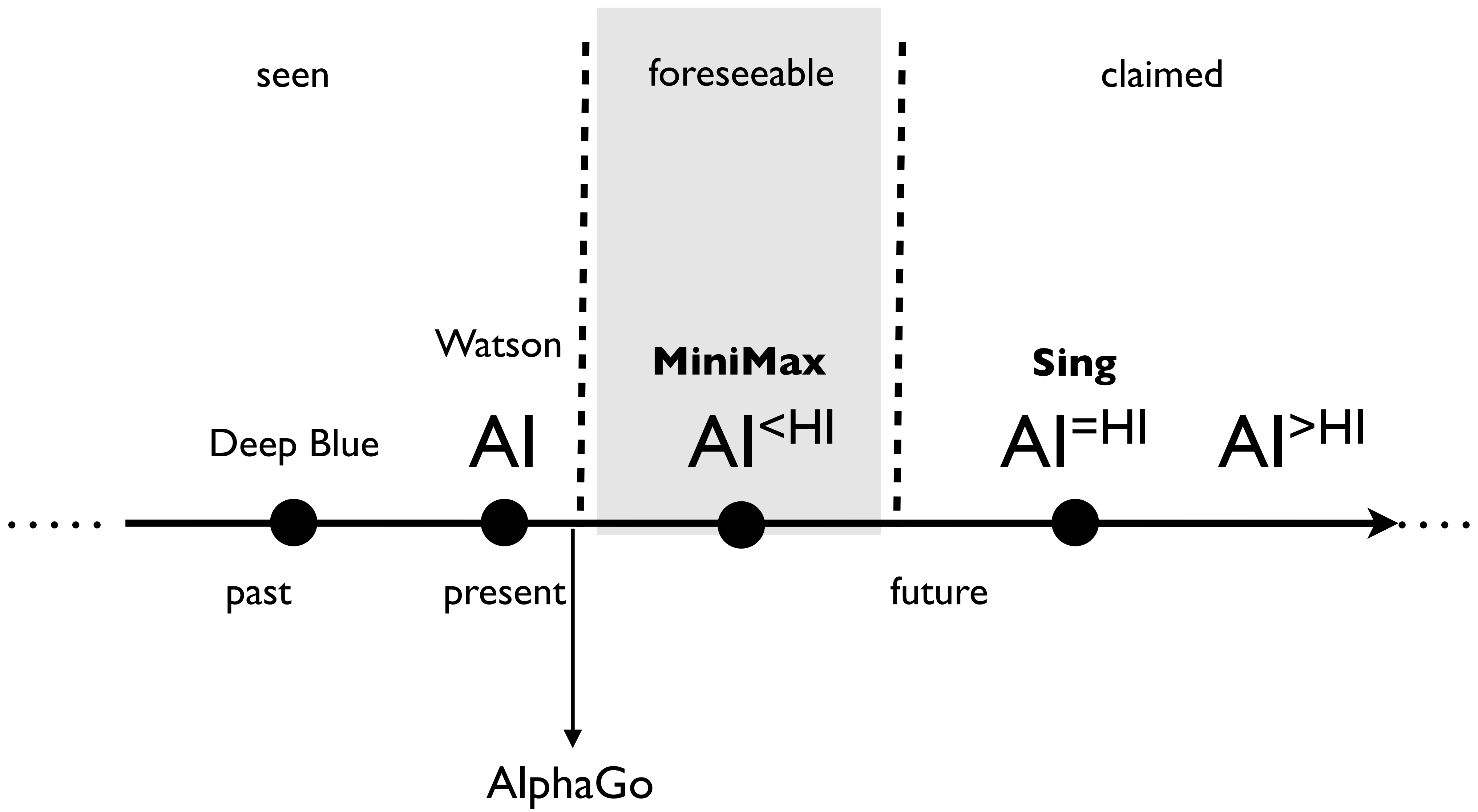
Today

A Realistic Timeline



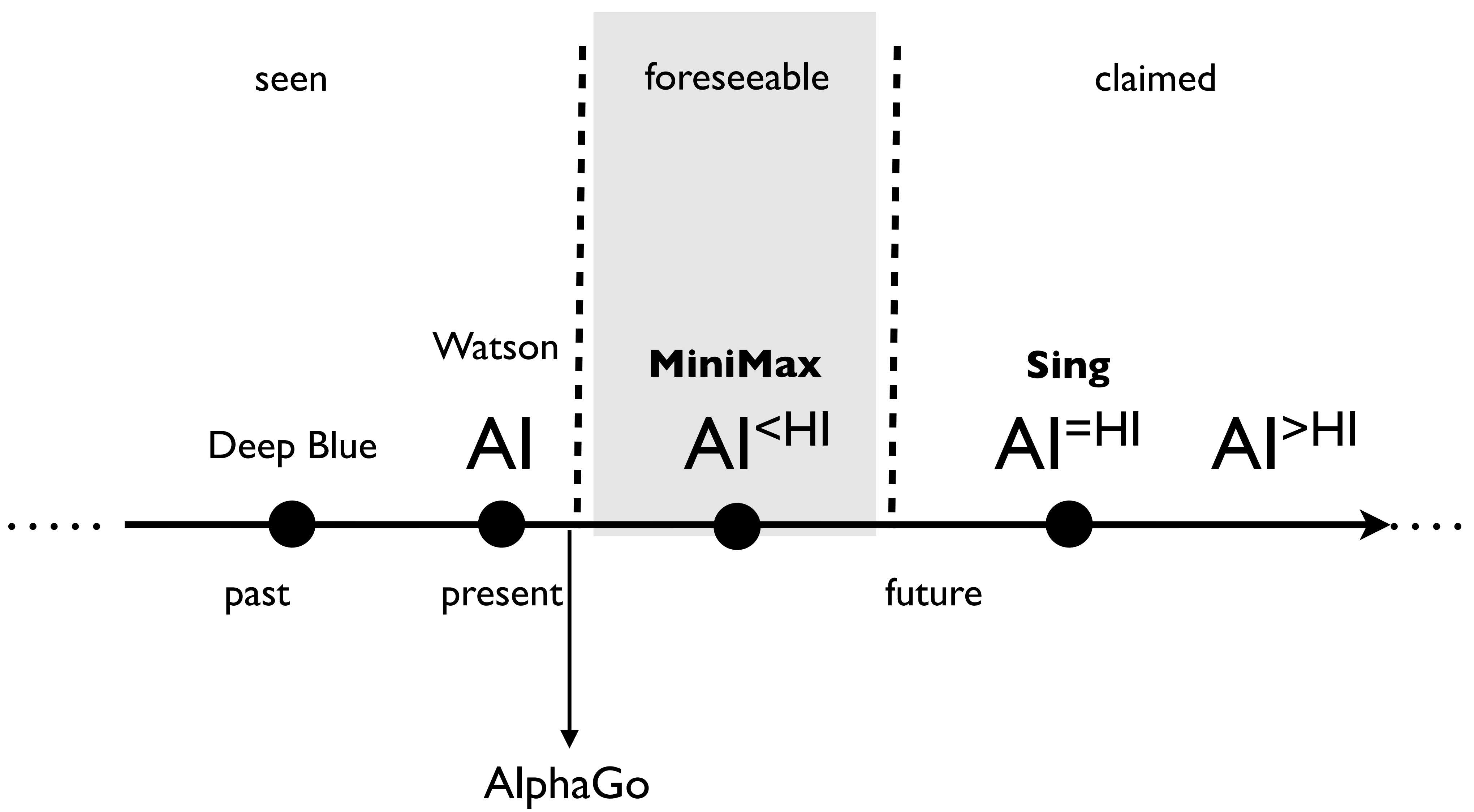
Today

A Realistic Timeline

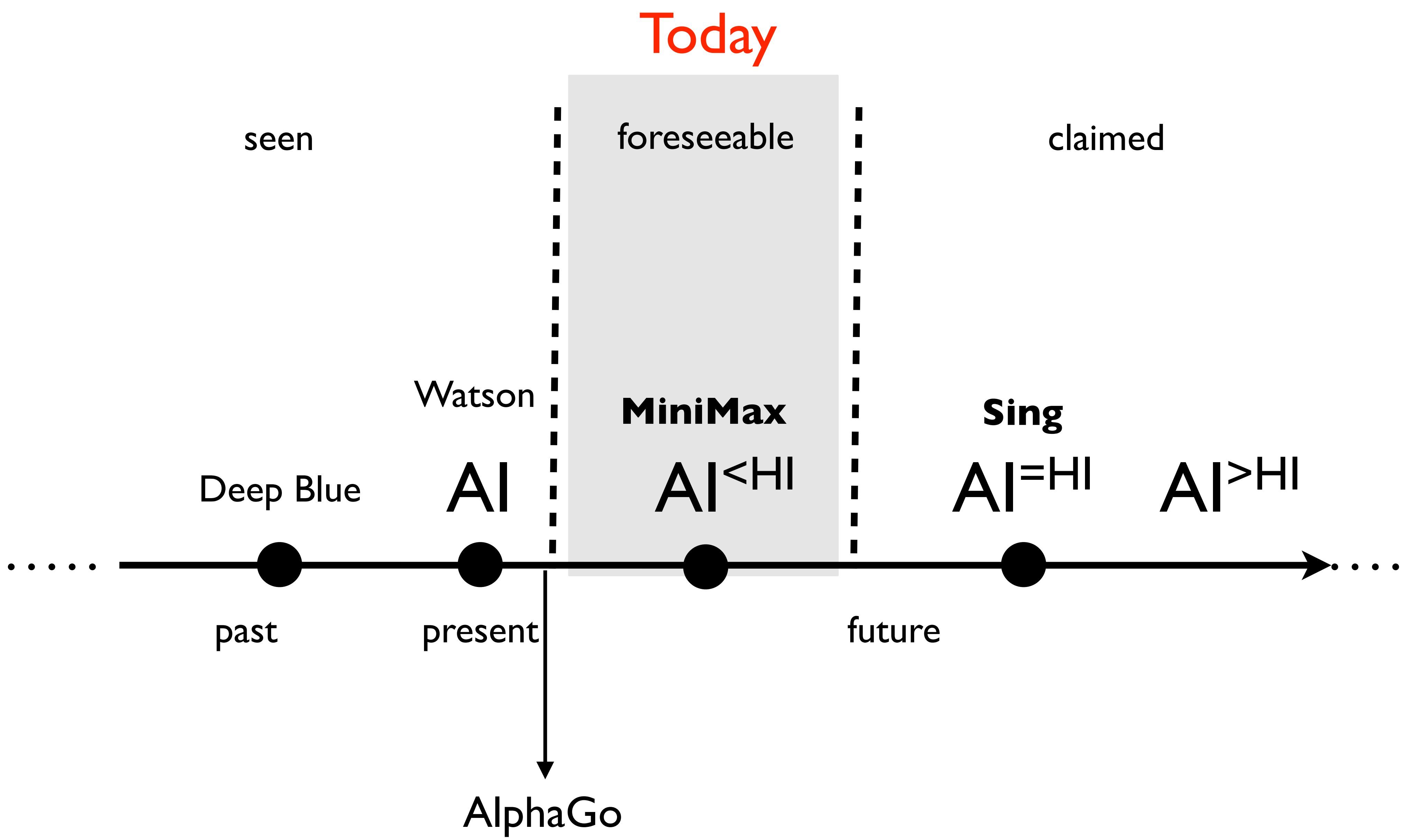


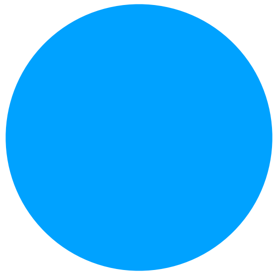
A Realistic Timeline

Today

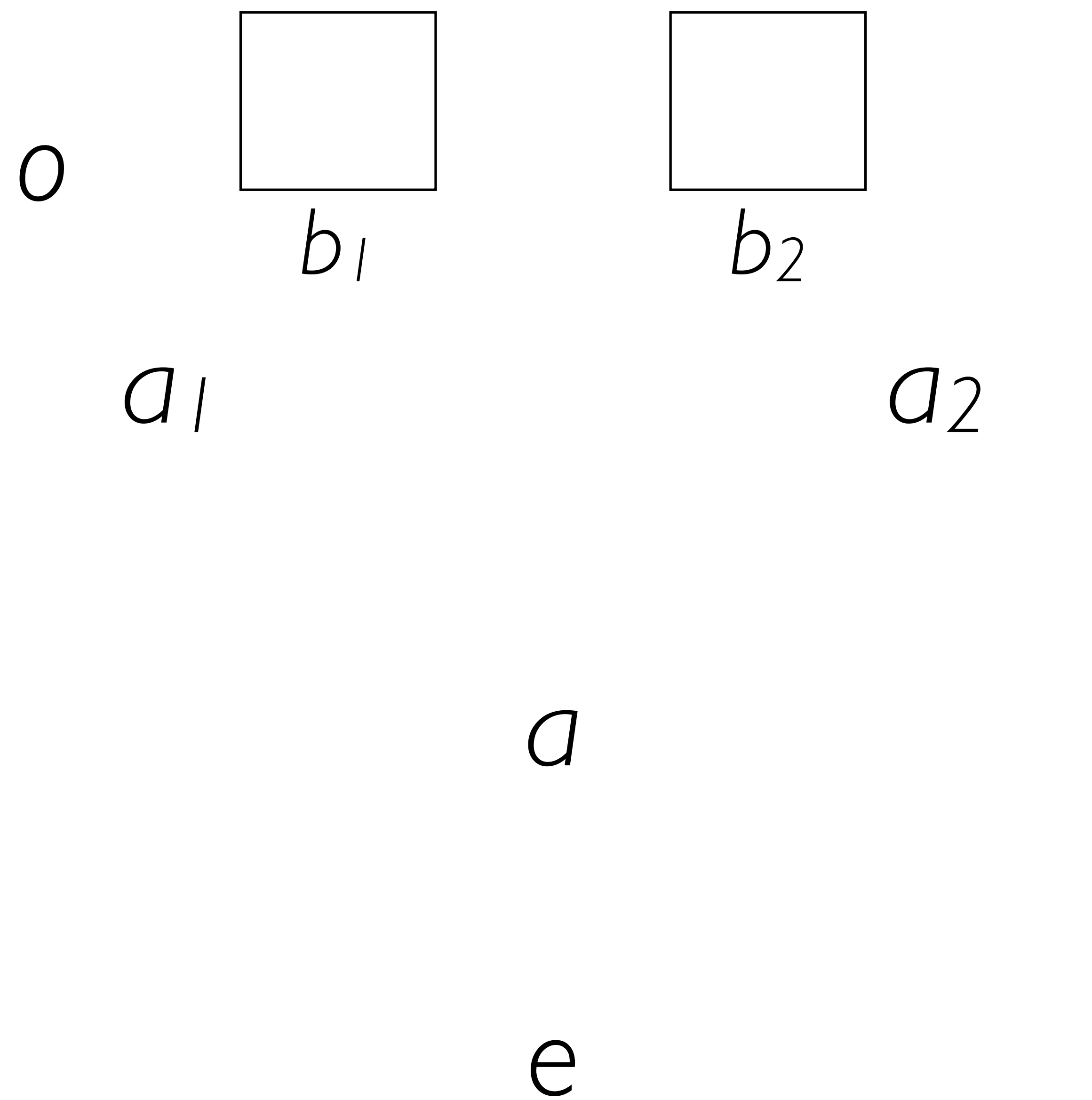


A Realistic Timeline

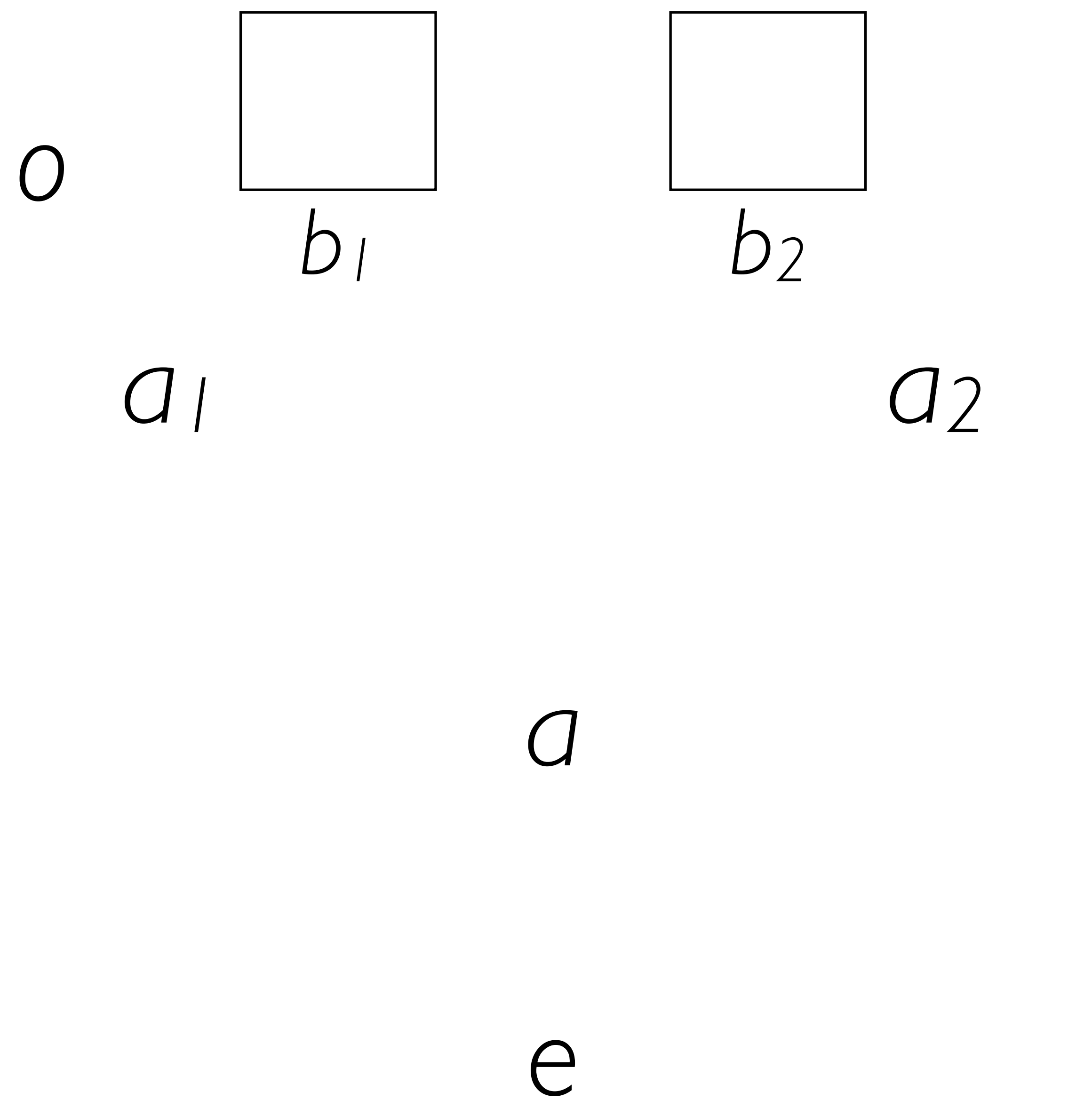




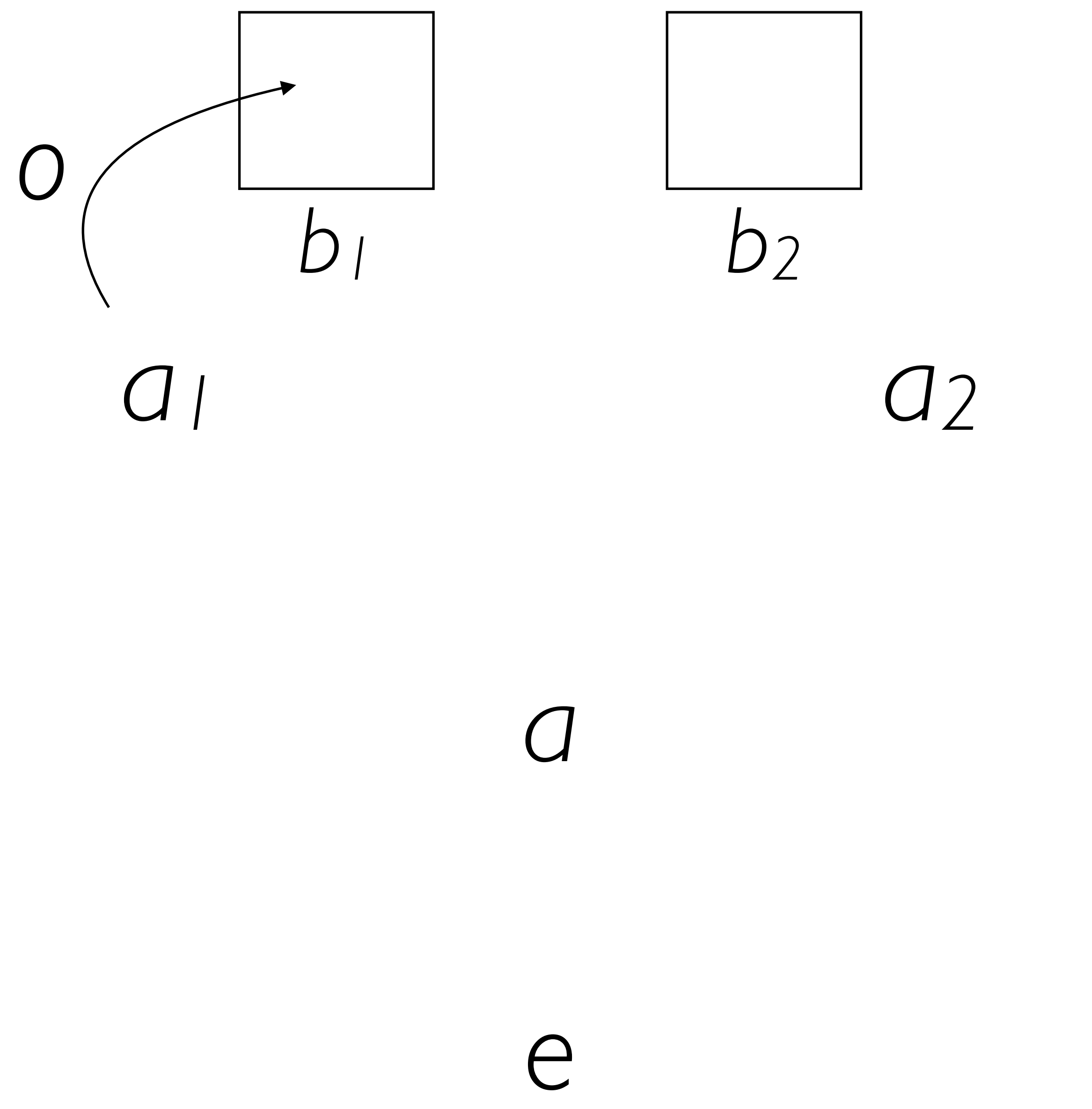
Framework for FBT$^0{}_1$



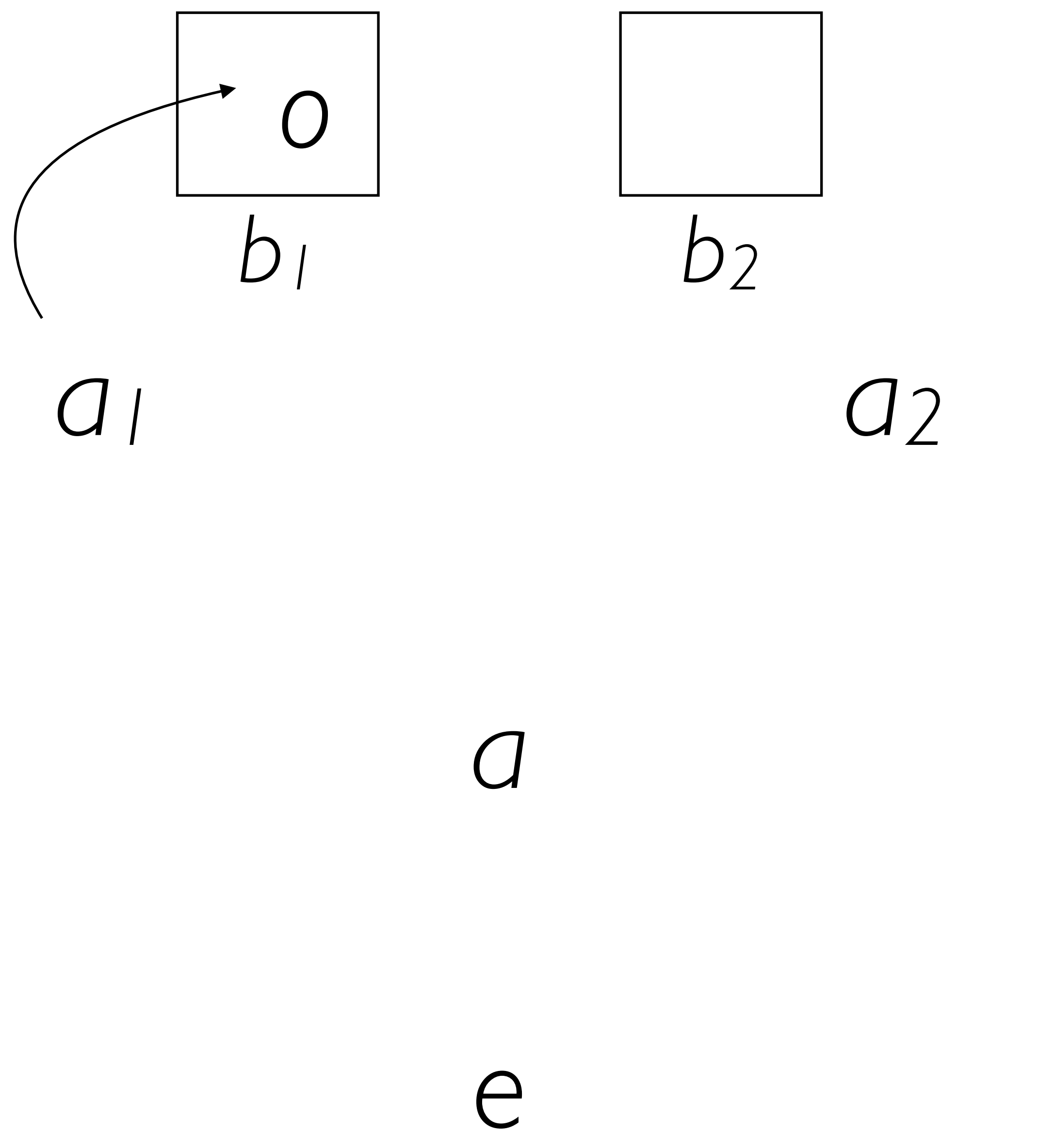
Framework for FBT$^0_1$
(five timepoints)



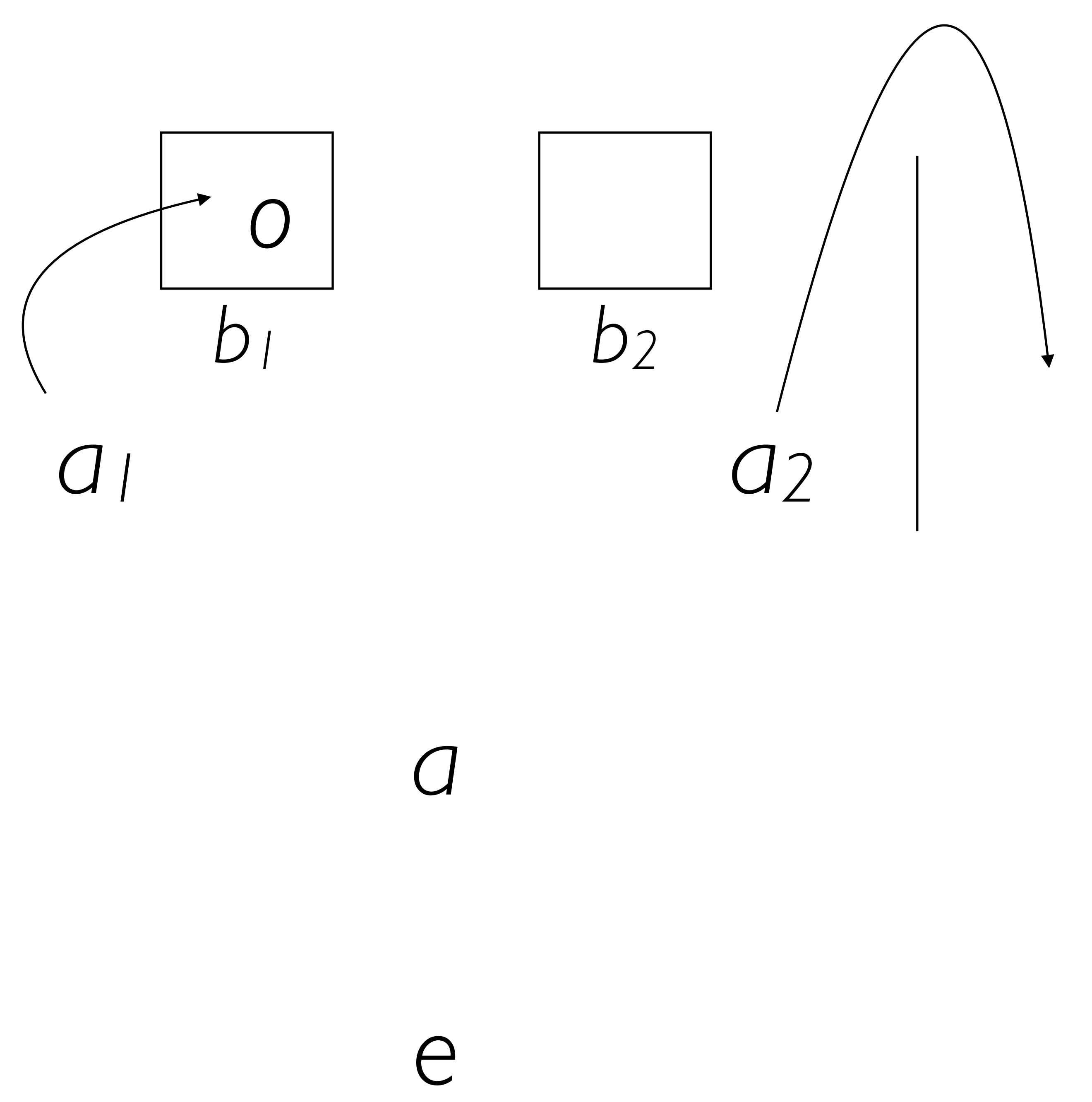
Framework for FBT^0_1
(five timepoints)



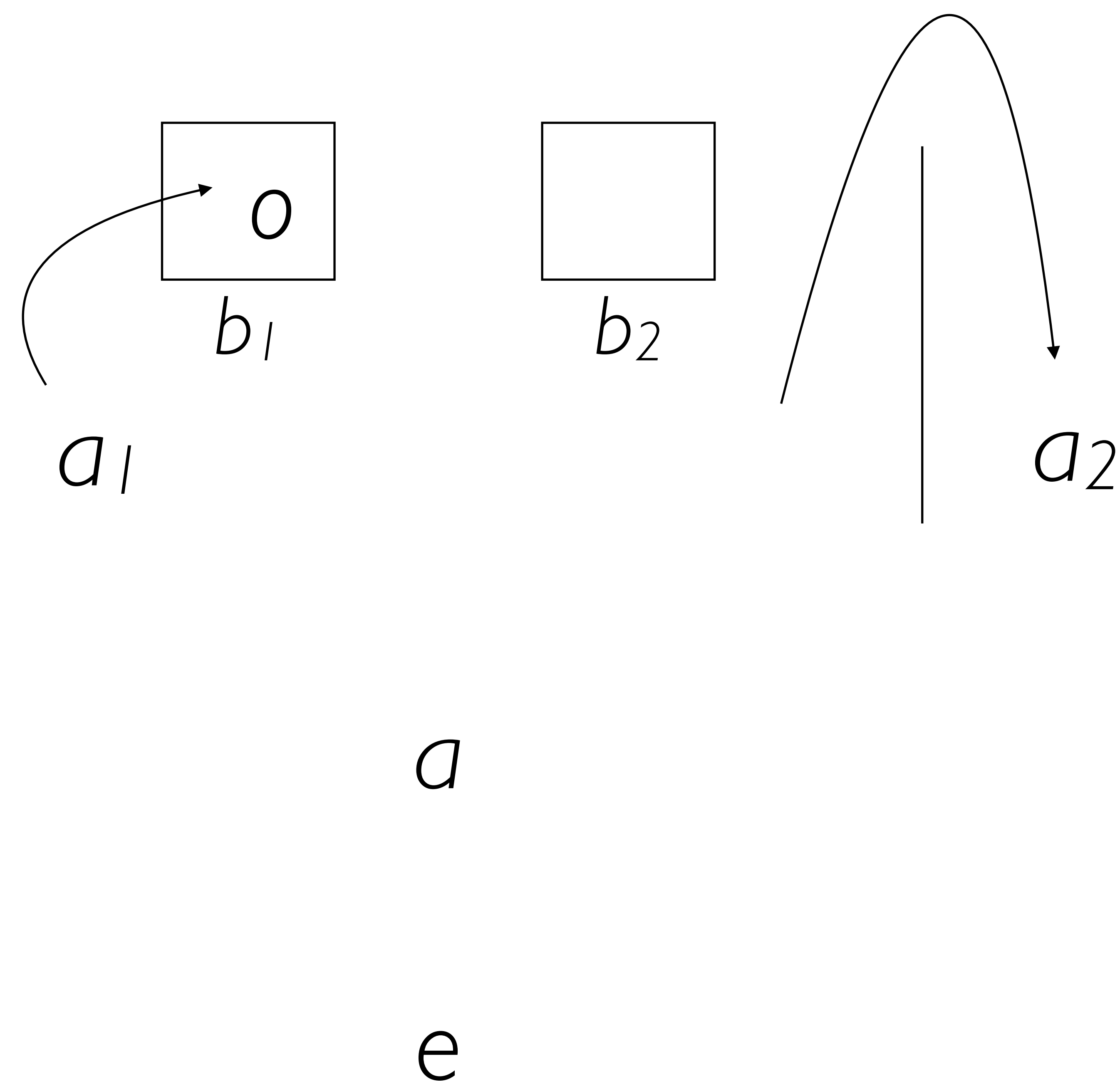
Framework for FBT^0_1
(five timepoints)



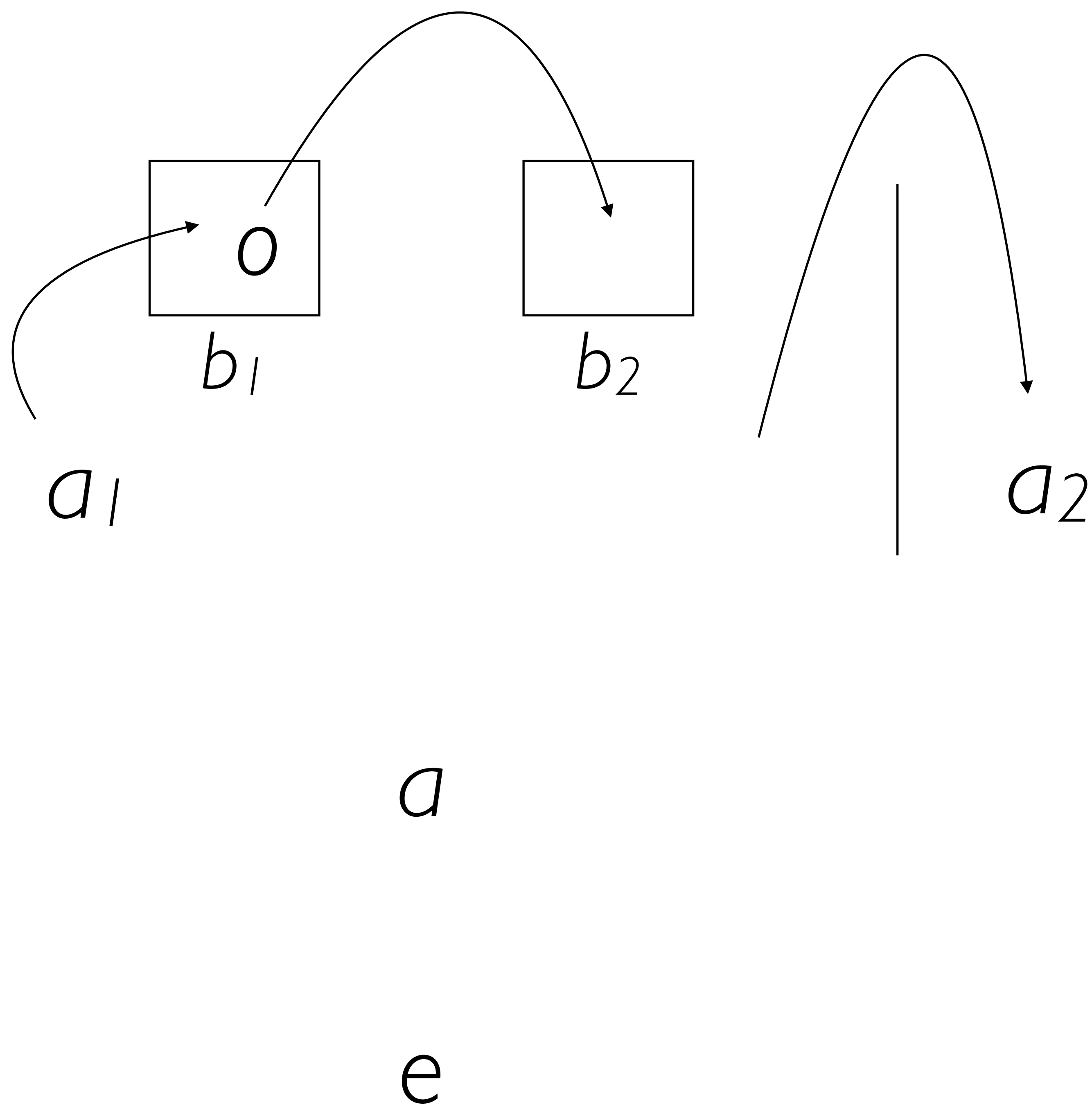
Framework for FBT^0_1
(five timepoints)



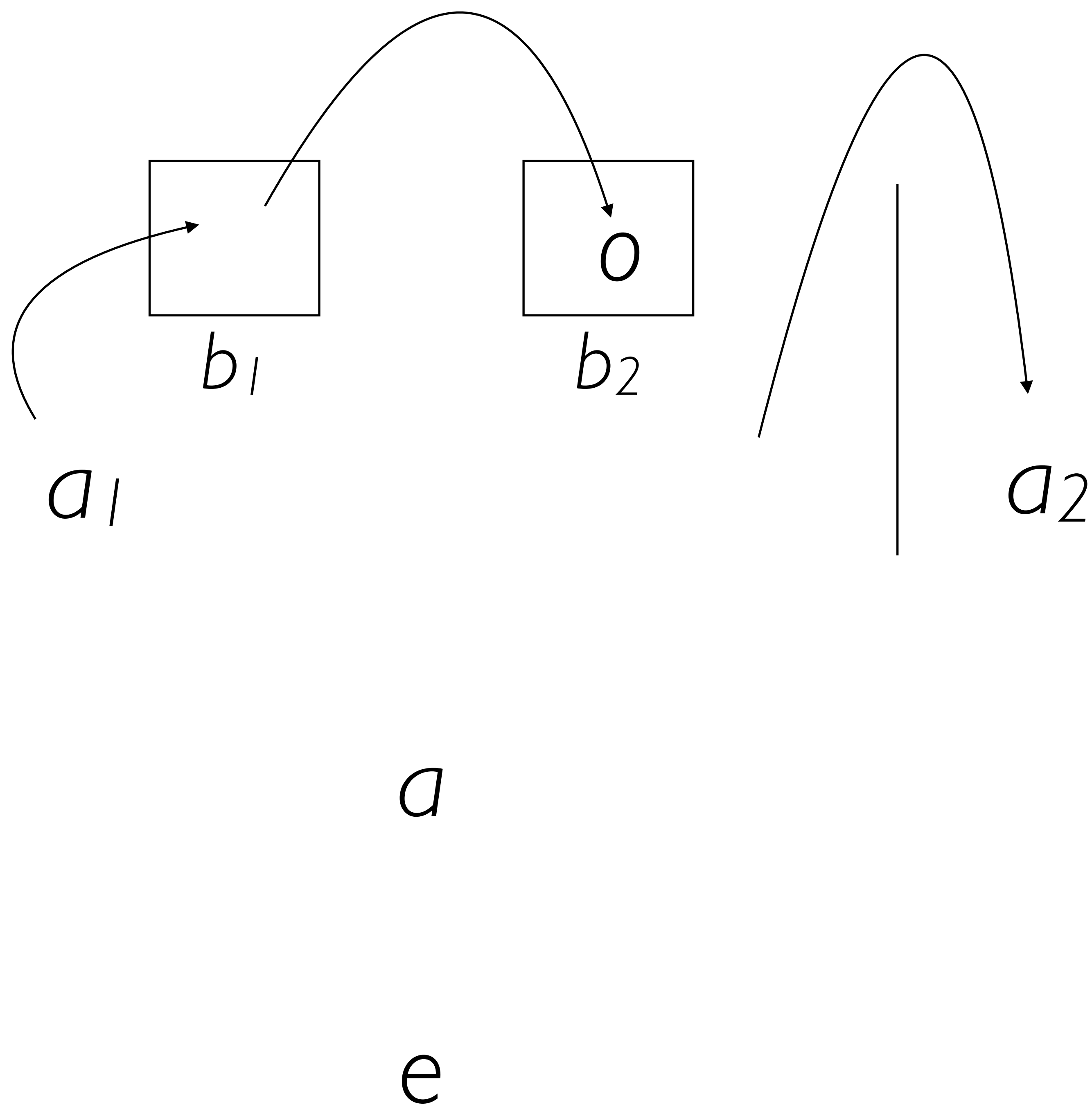
Framework for FBT$^{0}_{1}$
(five timepoints)



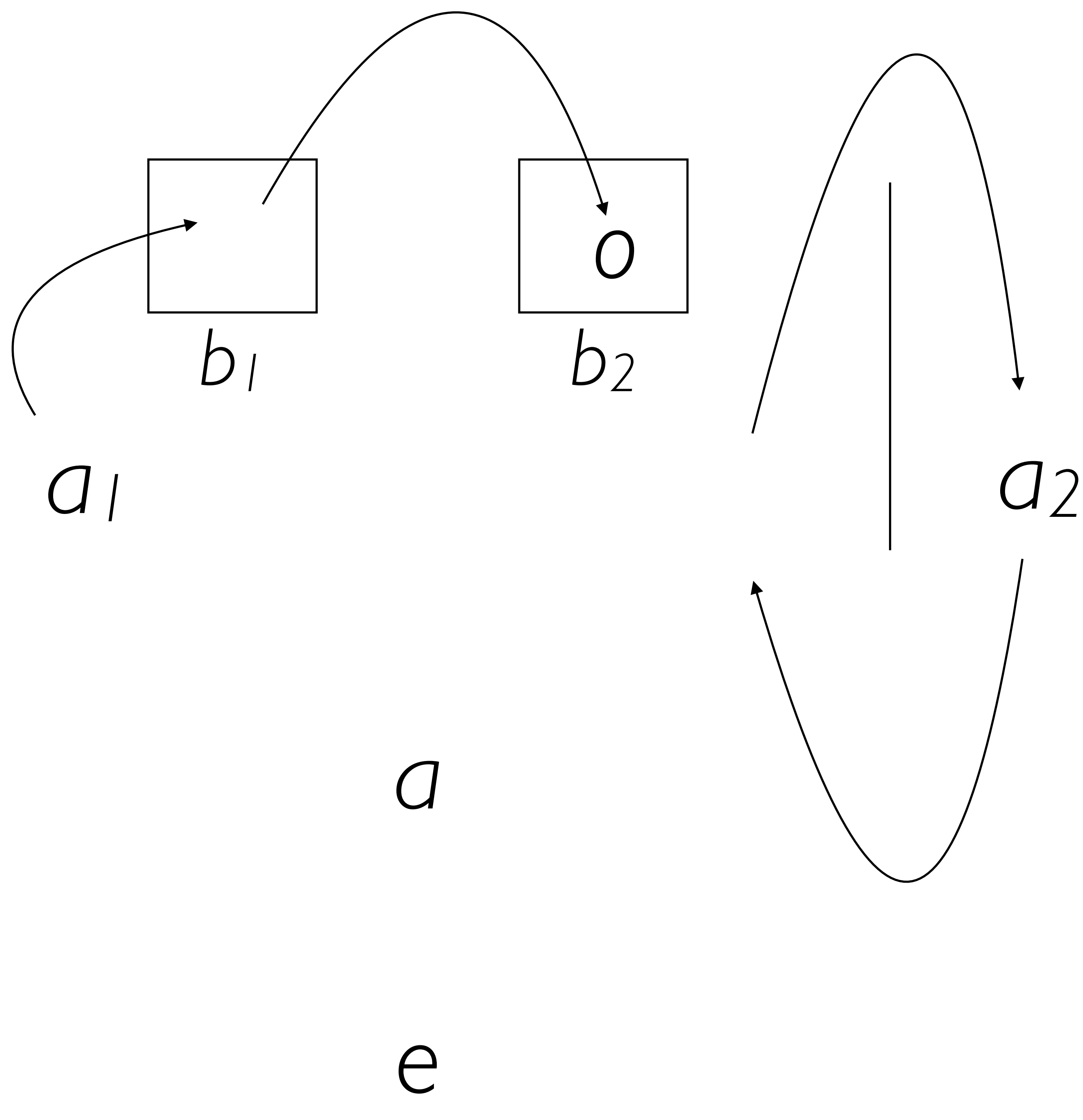
Framework for $FBT^0{}_1$
(five timepoints)



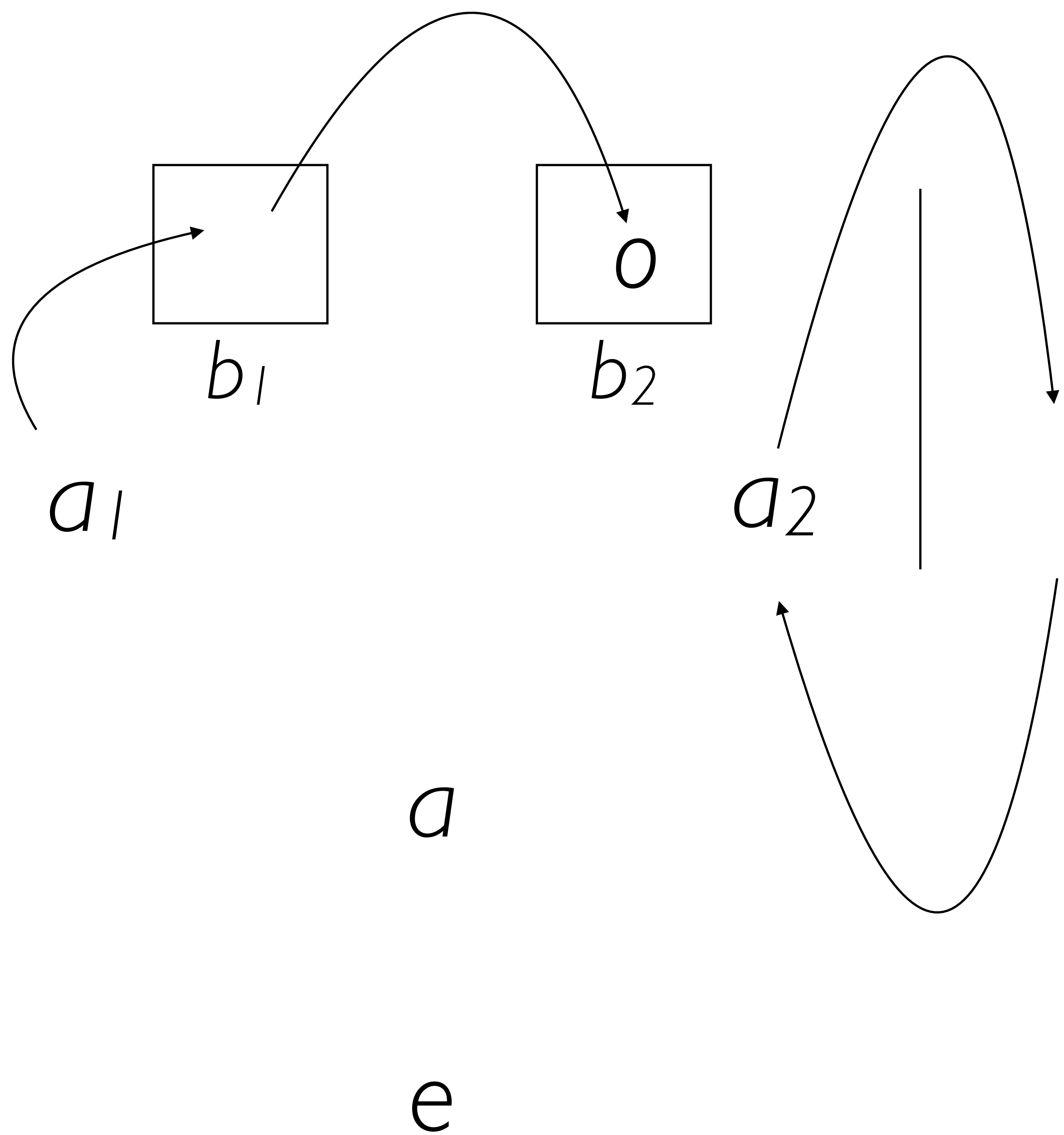
Framework for FBT^0_1
(five timepoints)



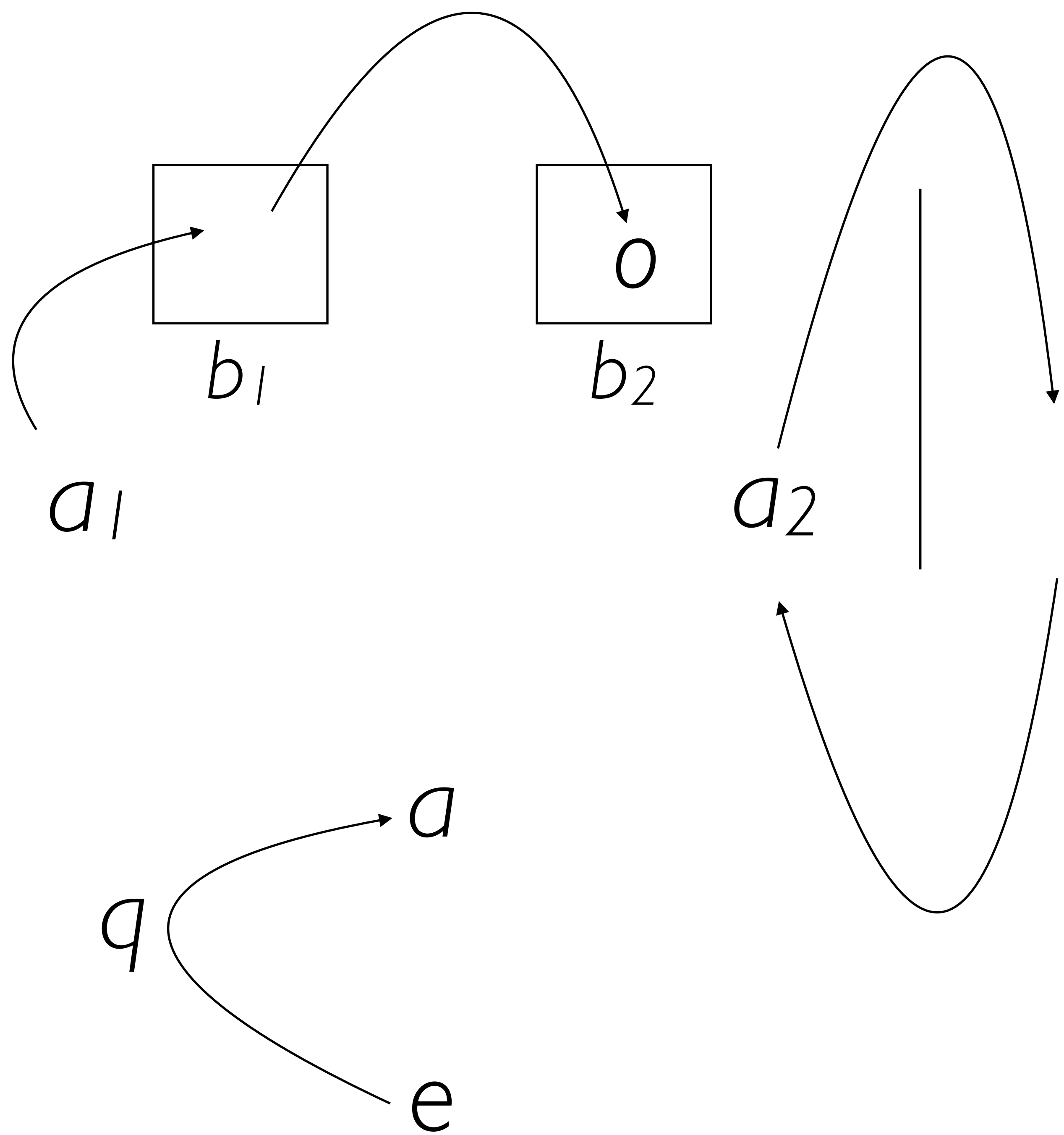
Framework for $FBT^0{}_1$
(five timepoints)



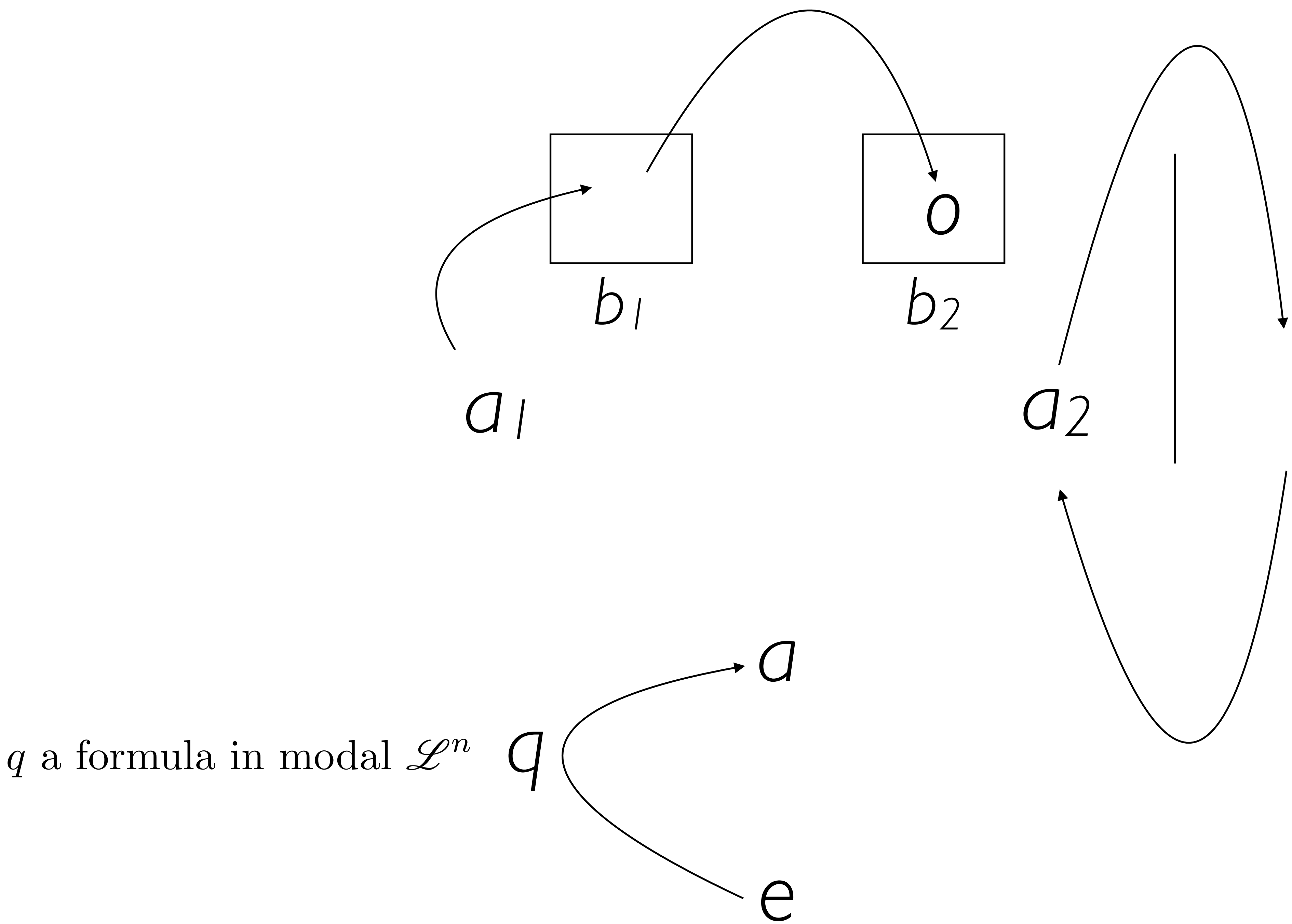
Framework for $FBT^0{}_1$
(five timepoints)



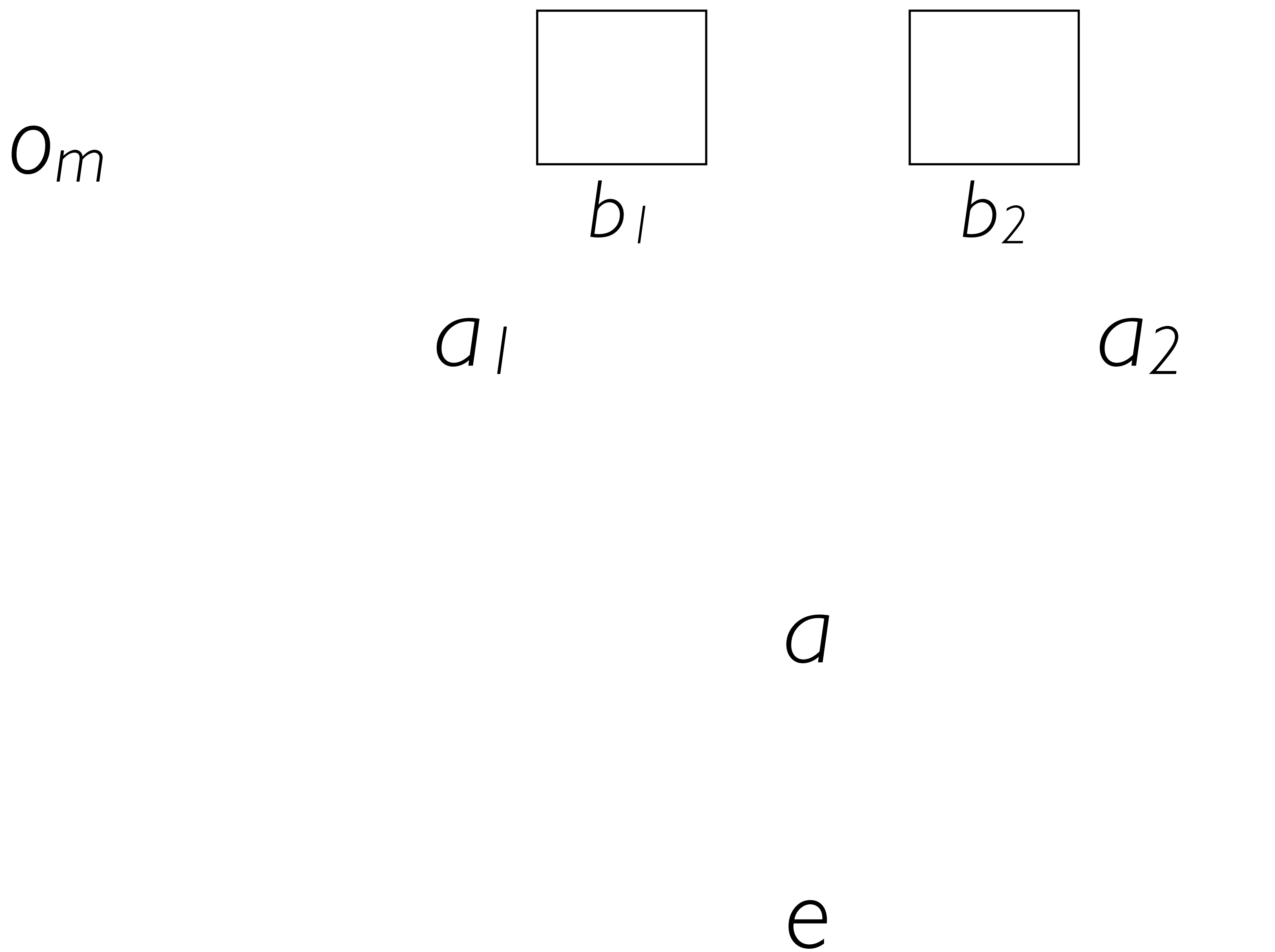
Framework for $FBT^0{}_1$
(five timepoints)



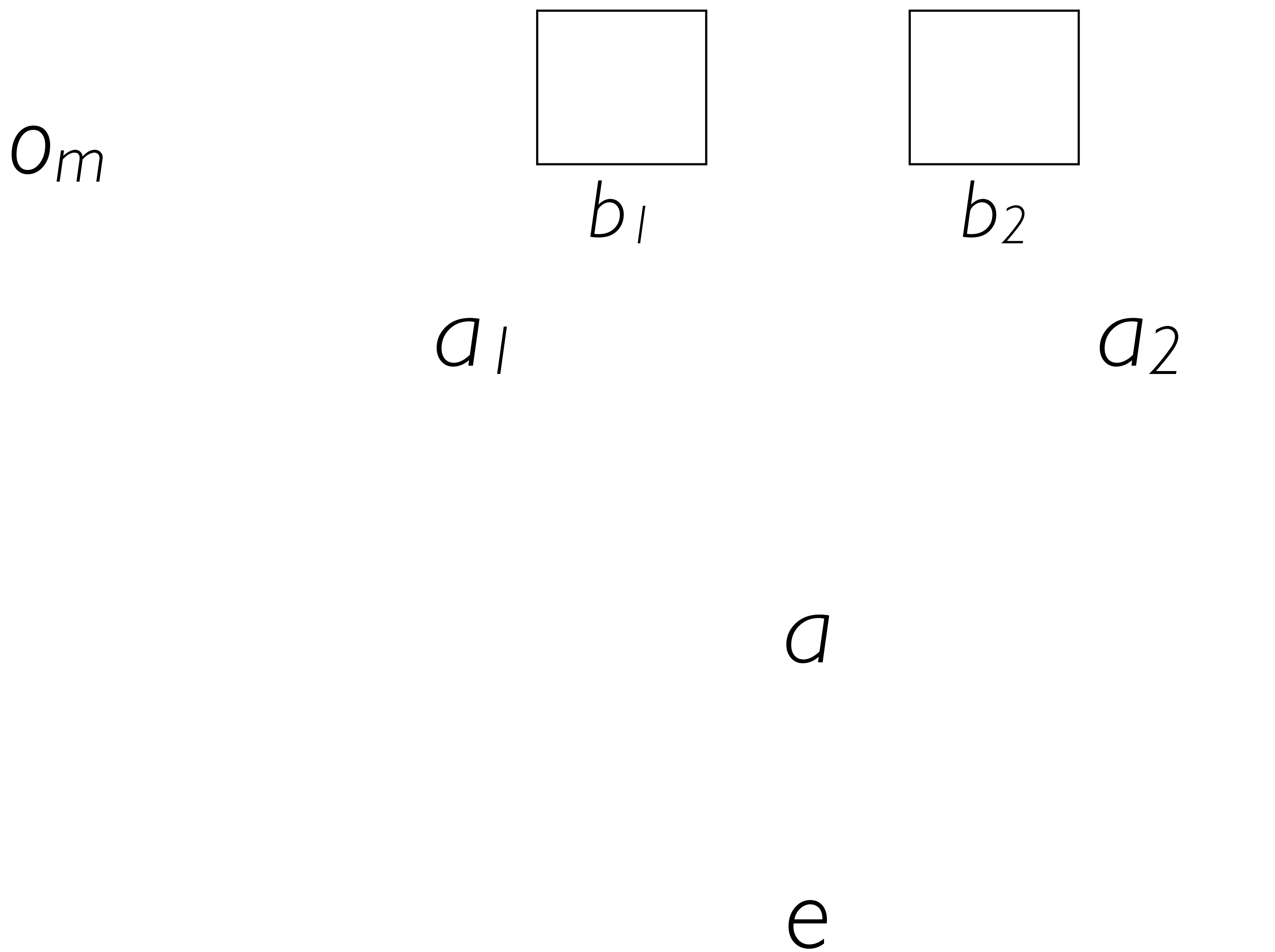
Framework for $FBT^0{}_1$
(five timepoints)



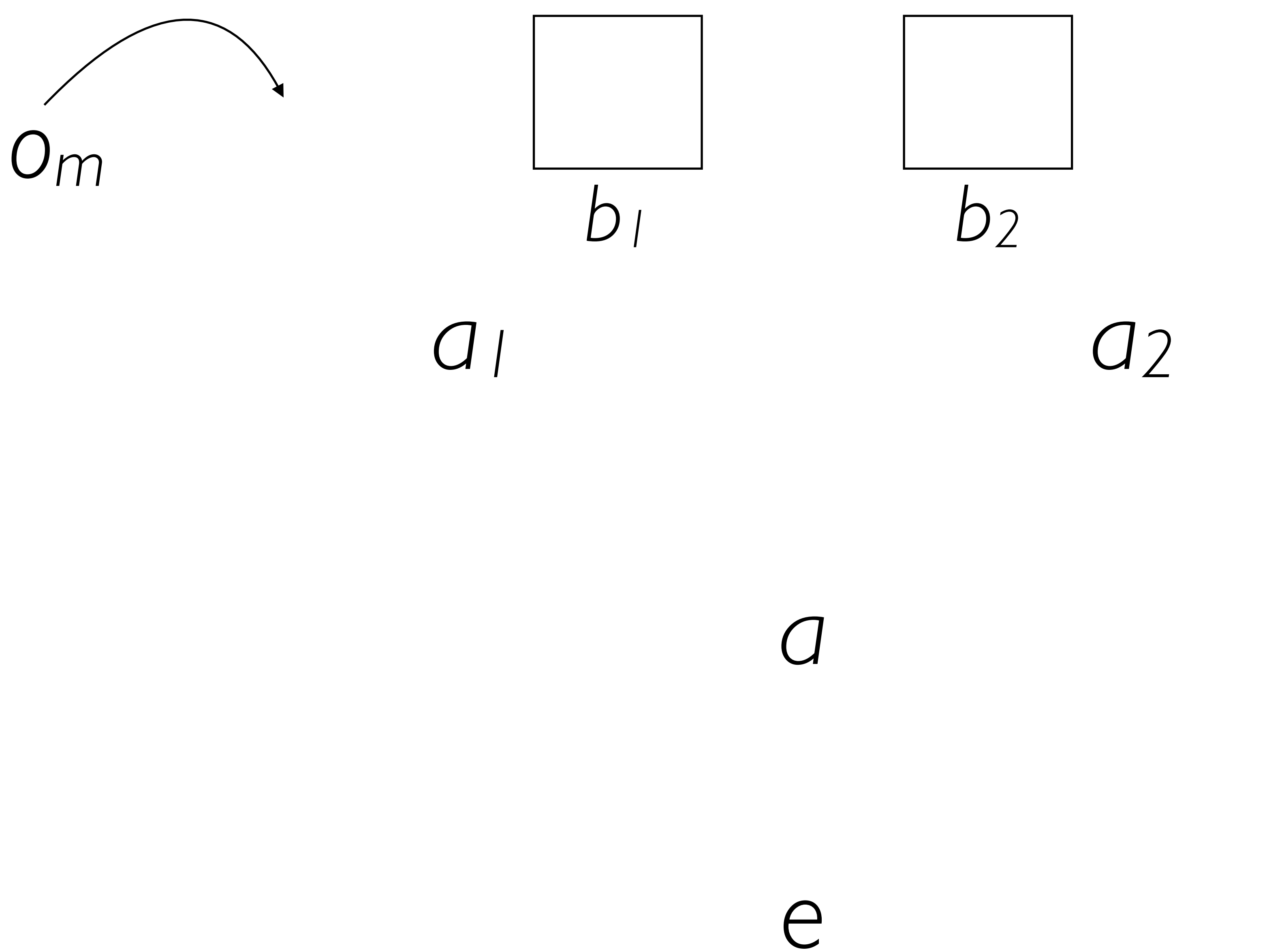
Framework for FBT^1_1



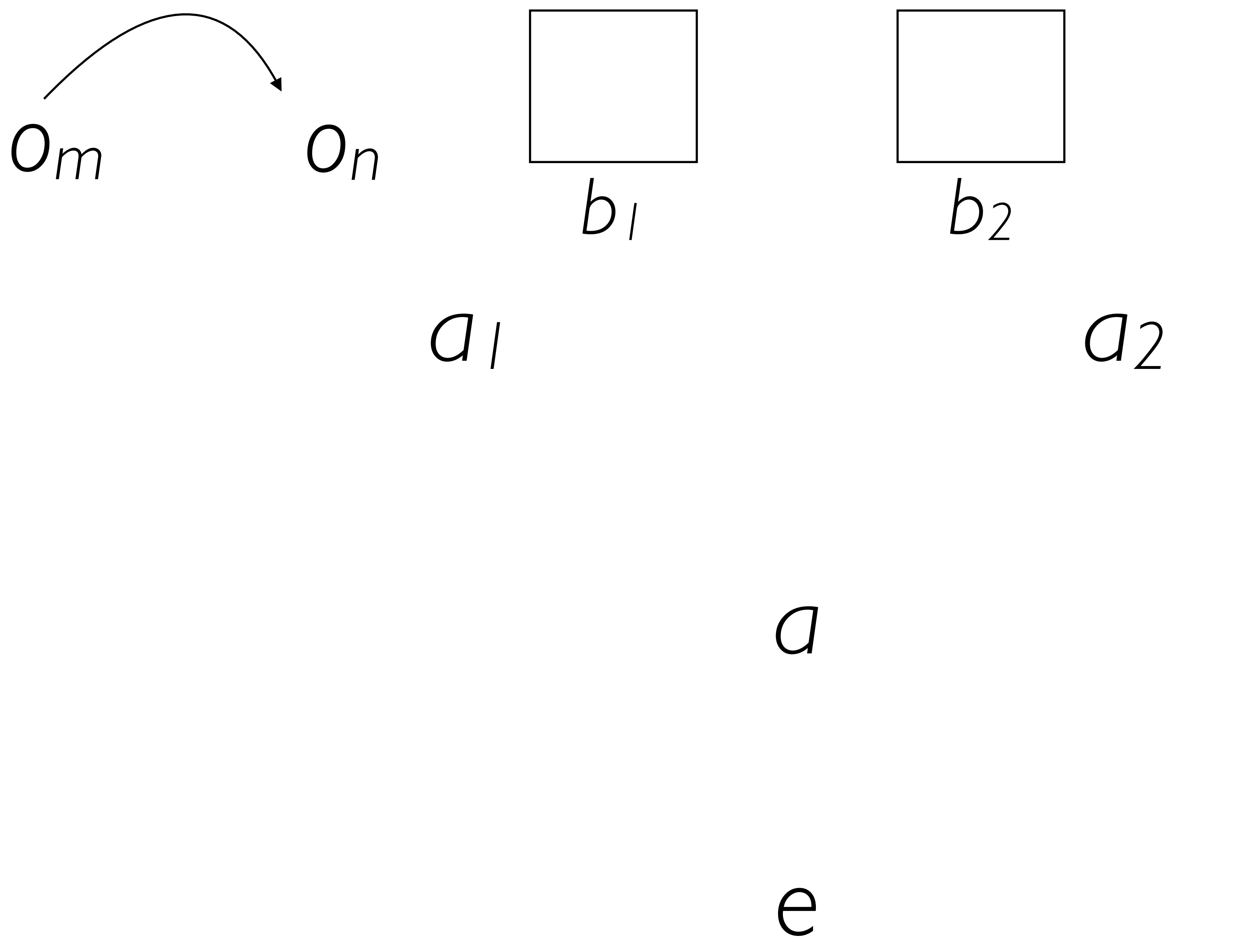
Framework for FBT^1_1
(six timepoints)



Framework for FBT^1_1
(six timepoints)

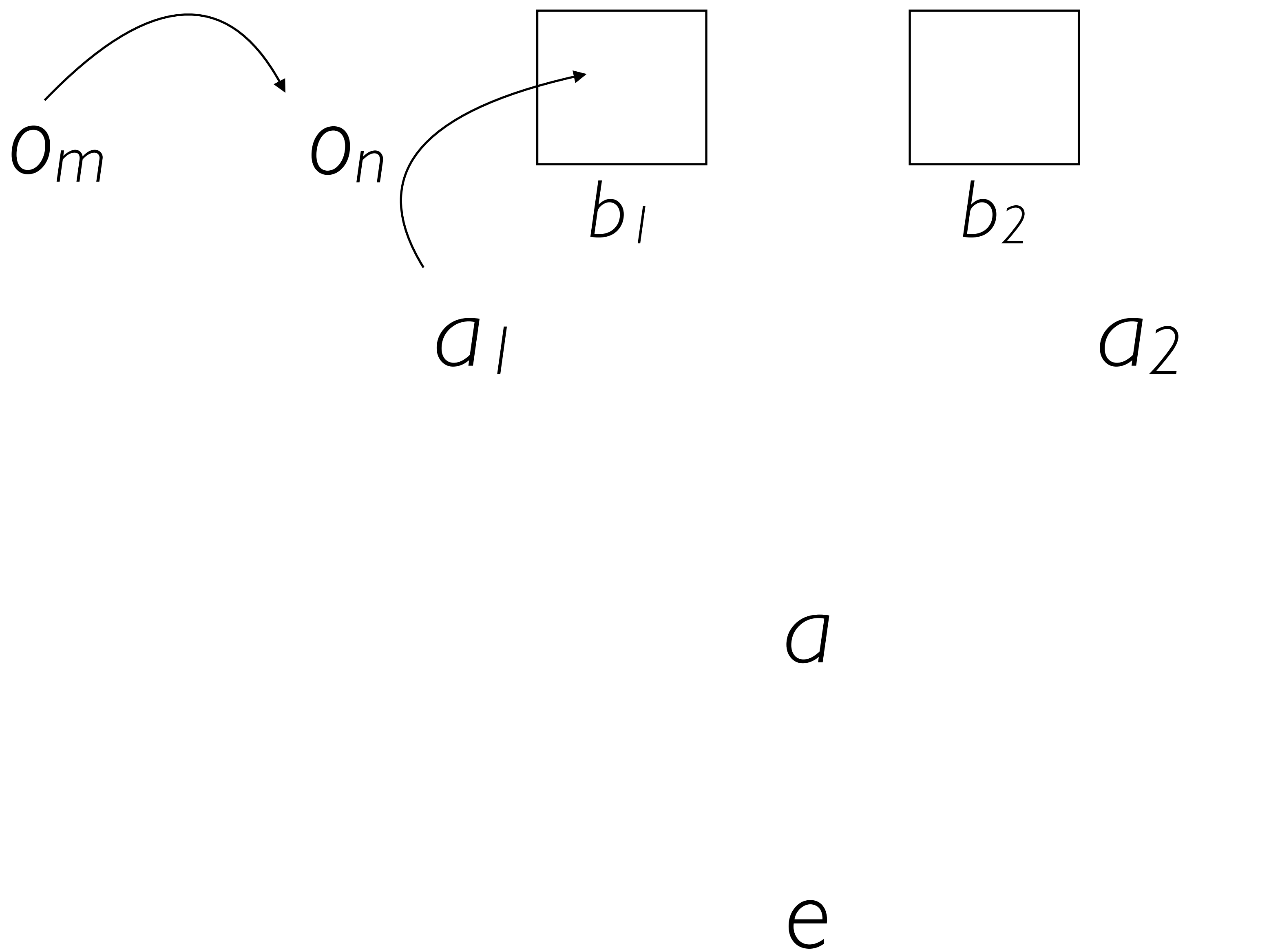


Framework for FBT^1_1
(six timepoints)

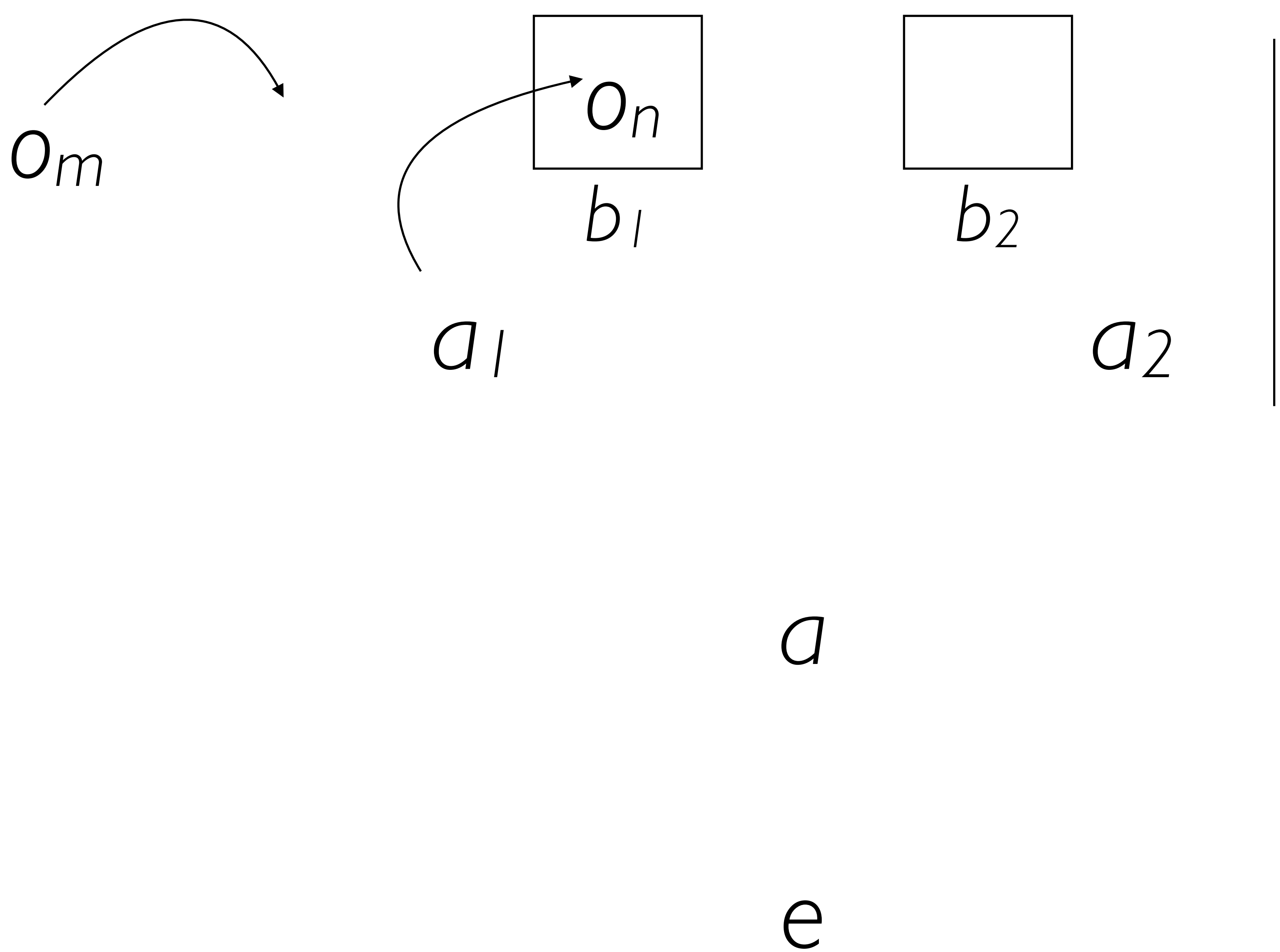


Framework for FBT^{1}_{1}

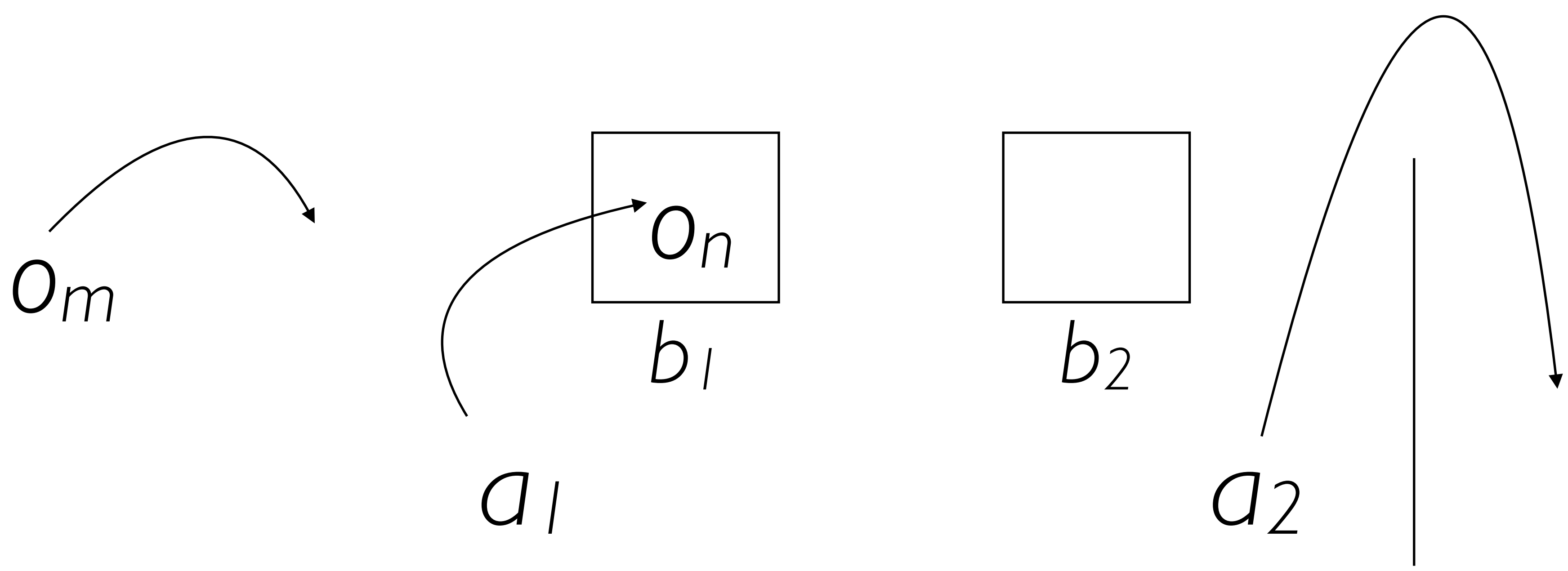
(six timepoints)



Framework for $FBT^1{}_1$
(six timepoints)



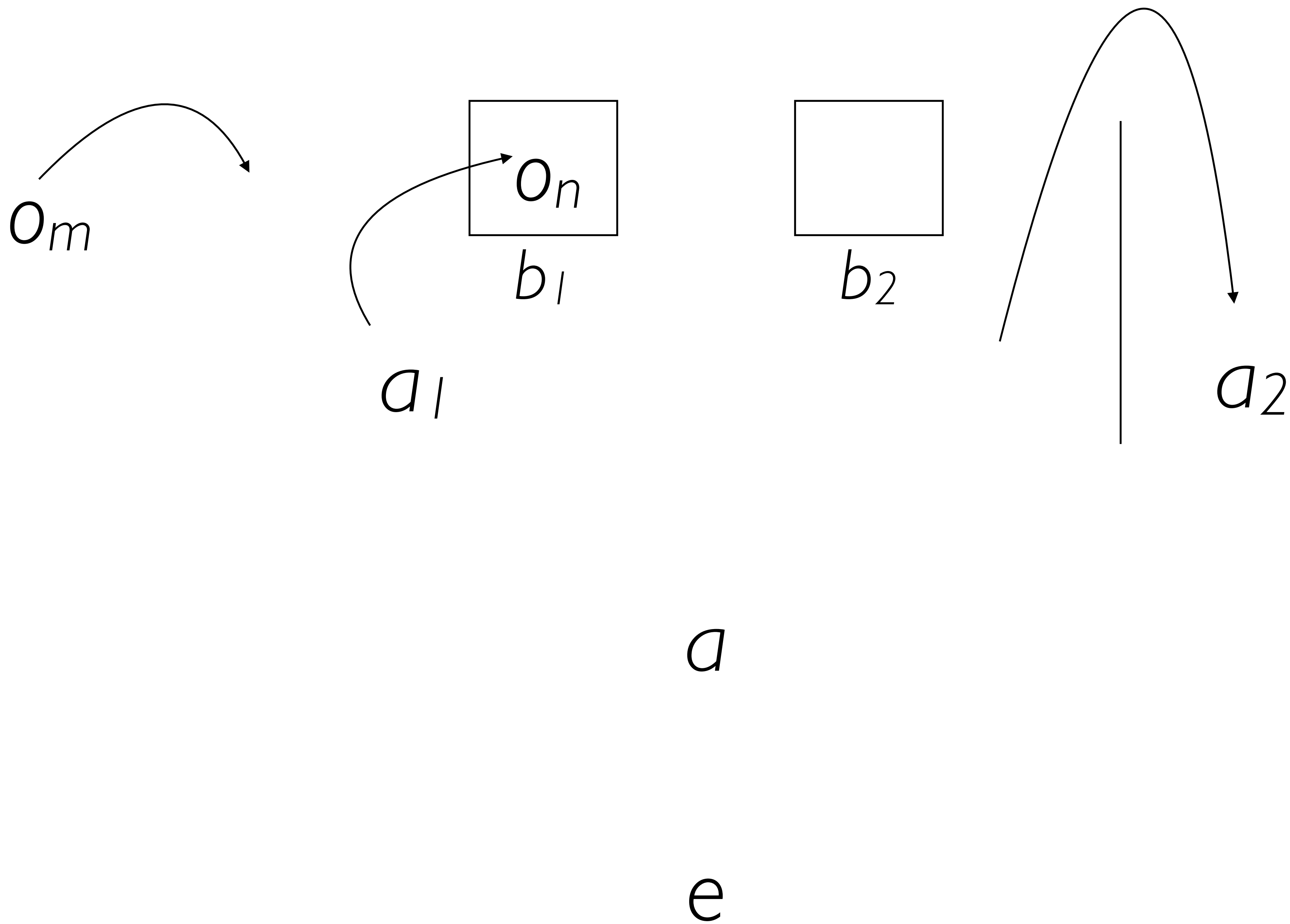
Framework for FBT^{1}_{1}
(six timepoints)



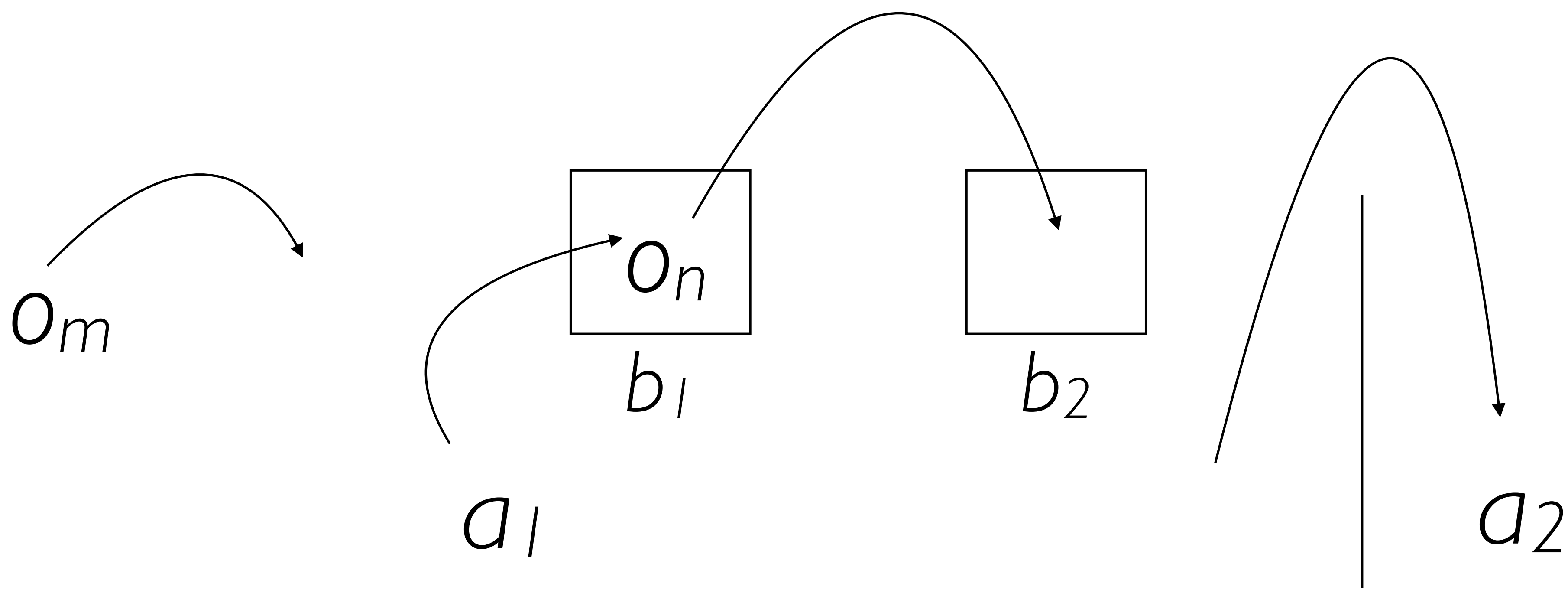
a

e

Framework for FBT^1_1
(six timepoints)



Framework for $FBT^1{}_1$
(six timepoints)

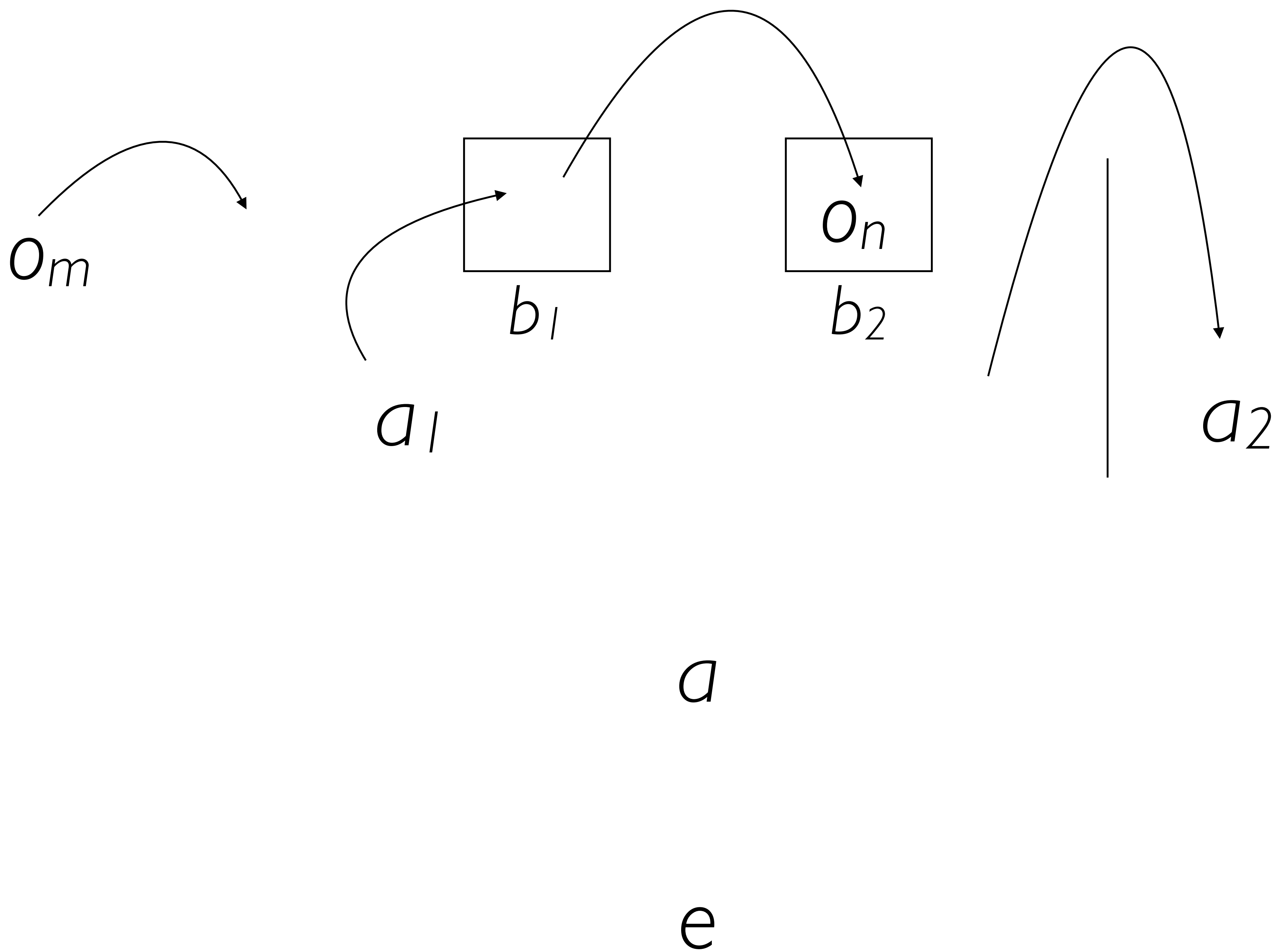


a

e

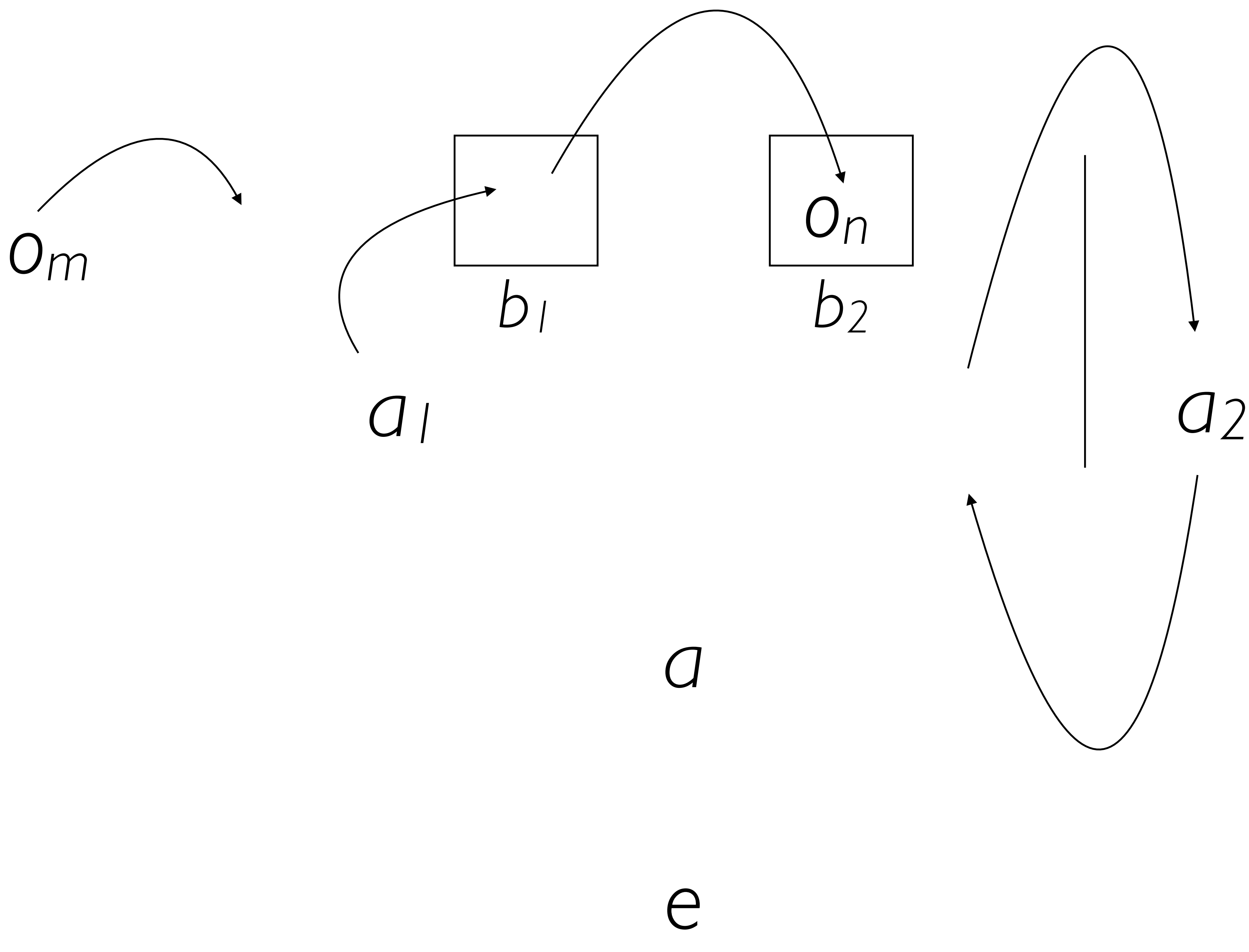
Framework for FBT^1_1

(six timepoints)



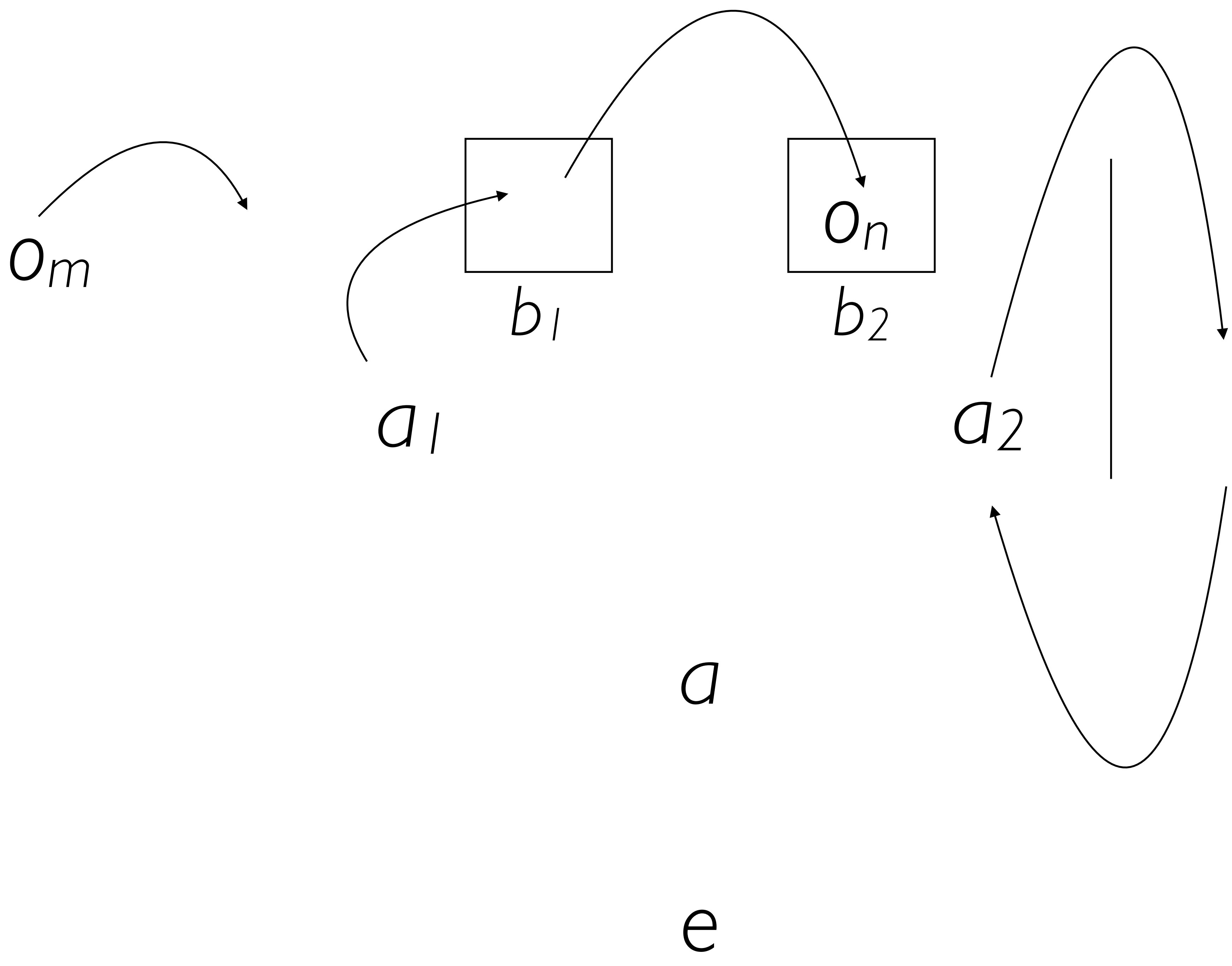
Framework for FBT¹₁

(six timepoints)



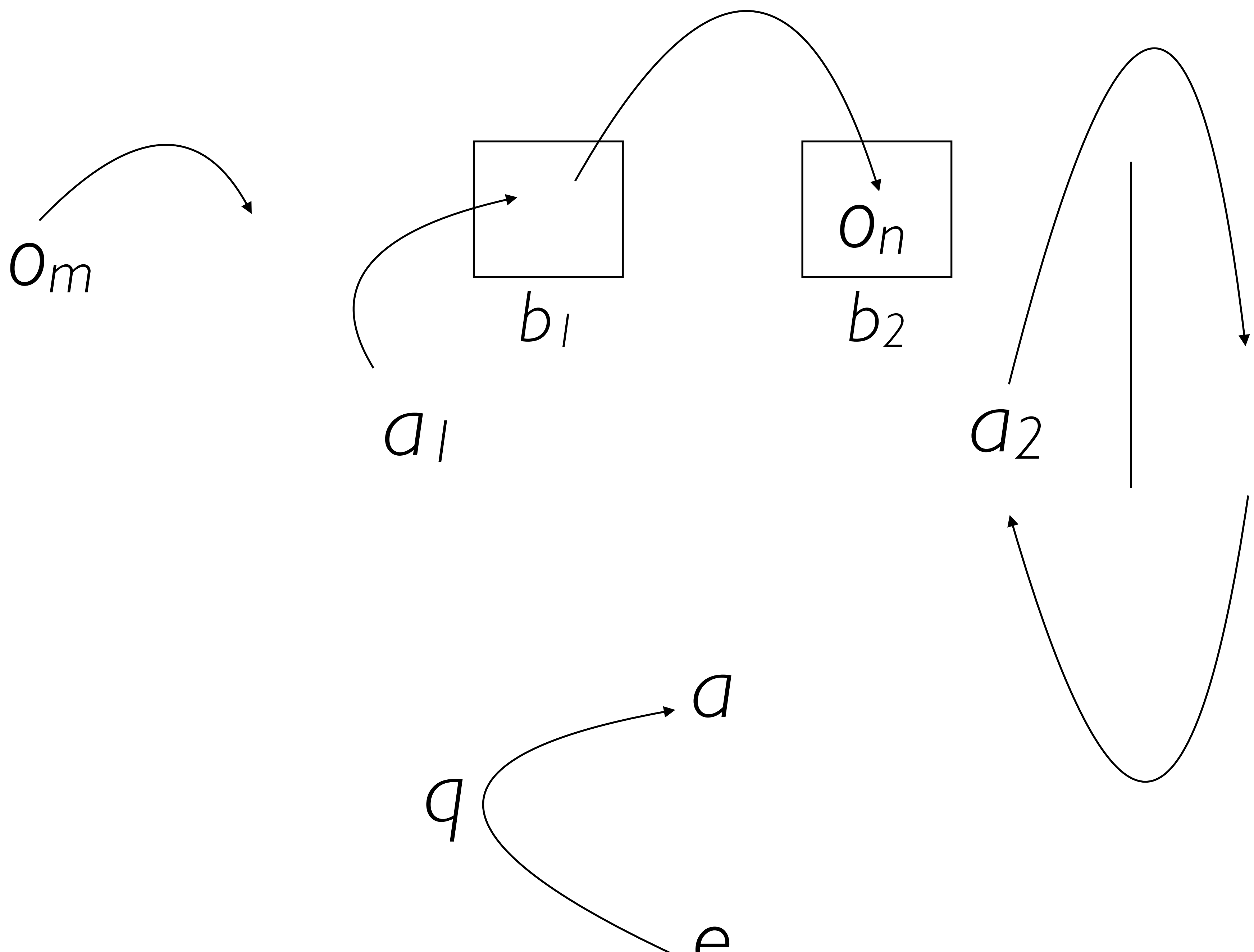
Framework for $FBT^1{}_1$

(six timepoints)



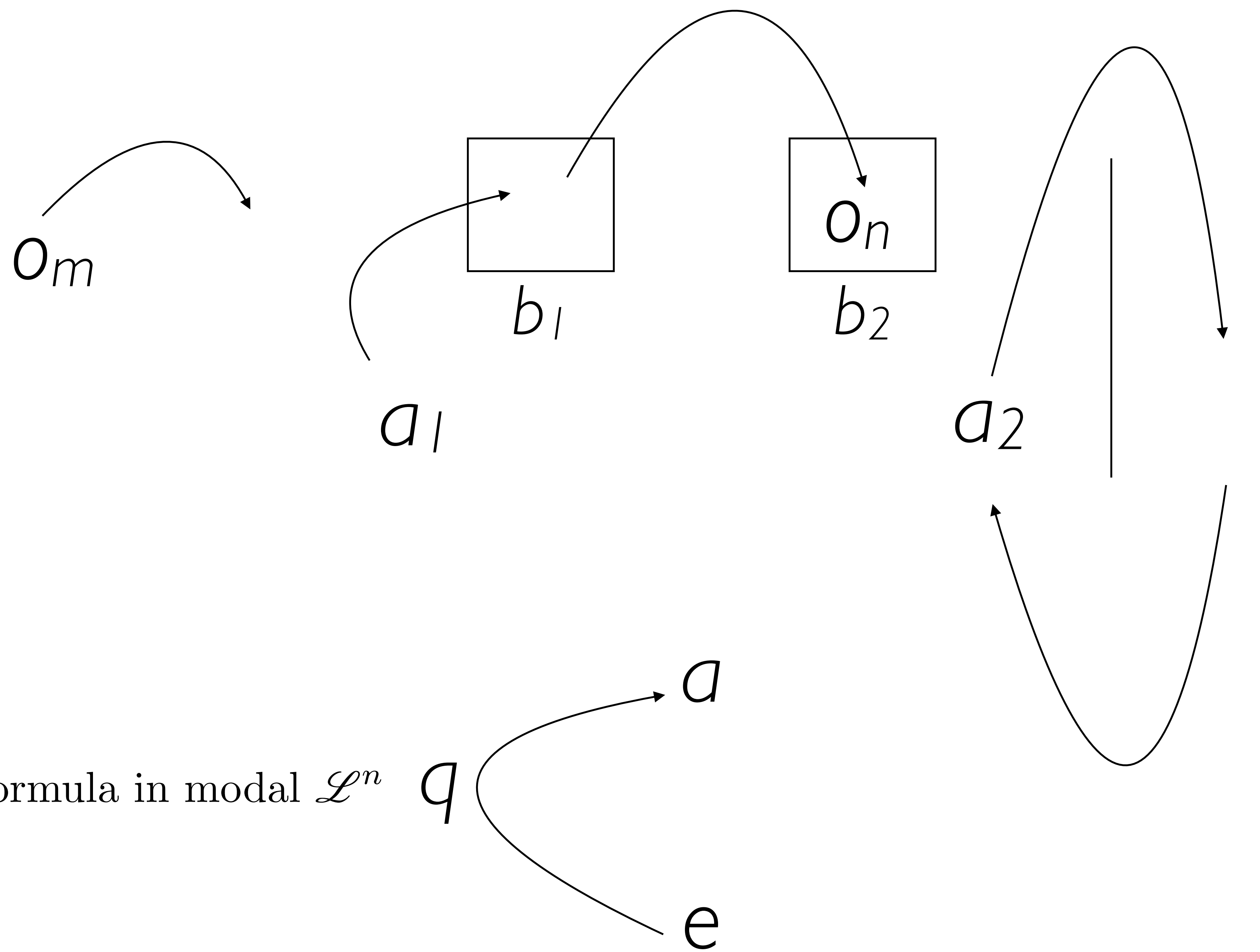
Framework for FBT$^1{}_1$

(six timepoints)



Framework for FBT^1_1

(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.

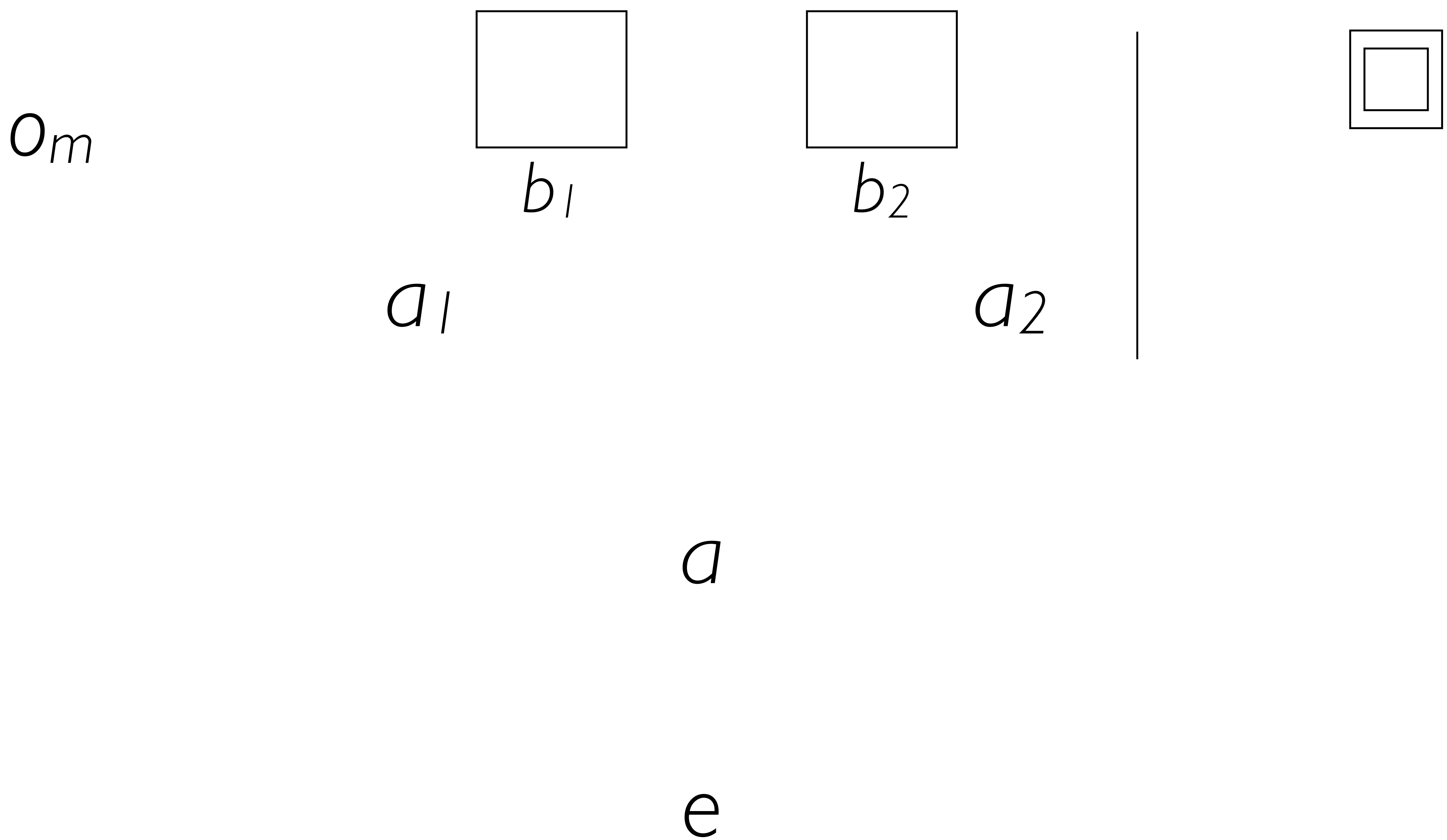
1 Introduction

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

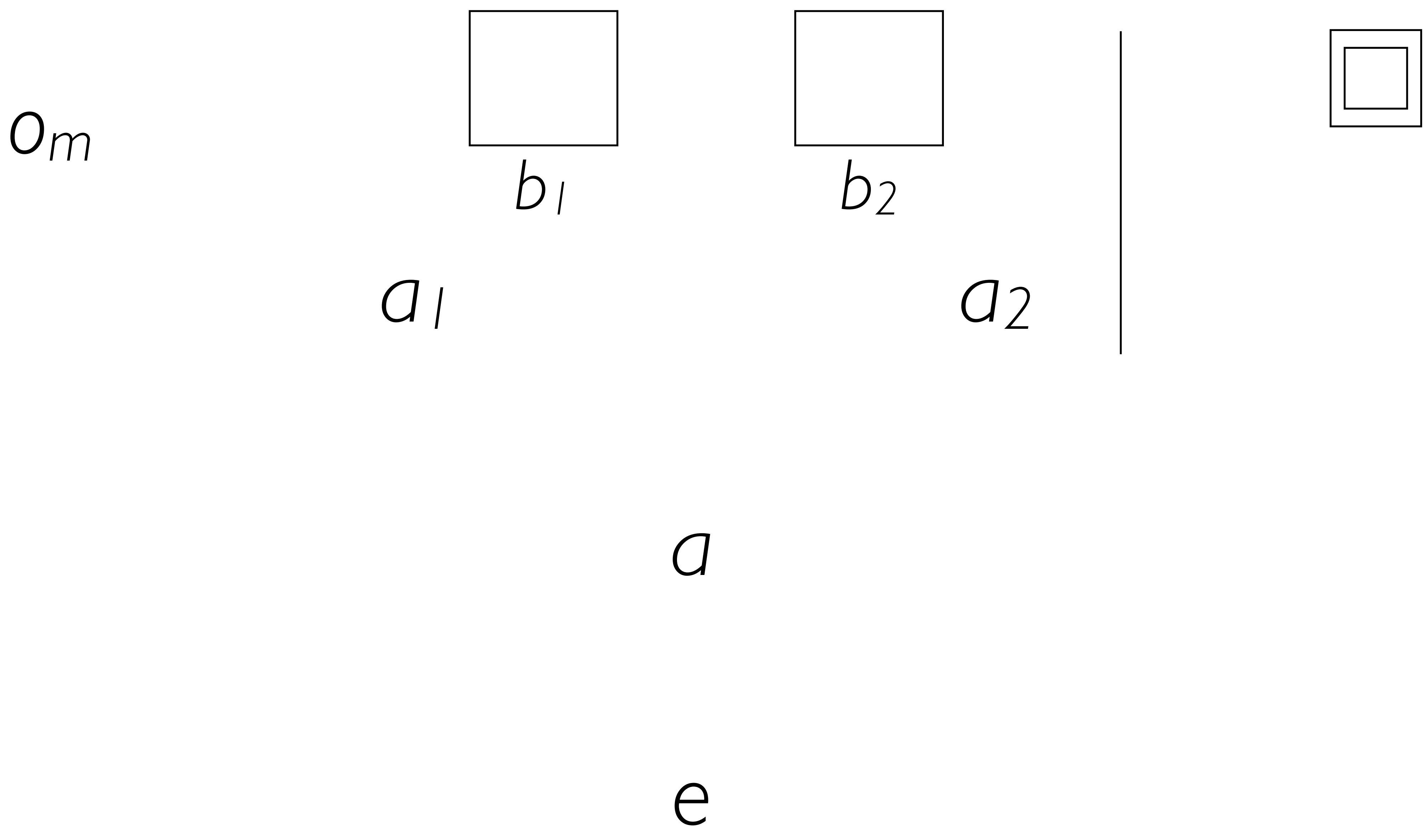
[1] Assuming that a is able to carry out c, that a has no conflicting desires that override his goal that p; and so on.

Framework for FBT^1_2

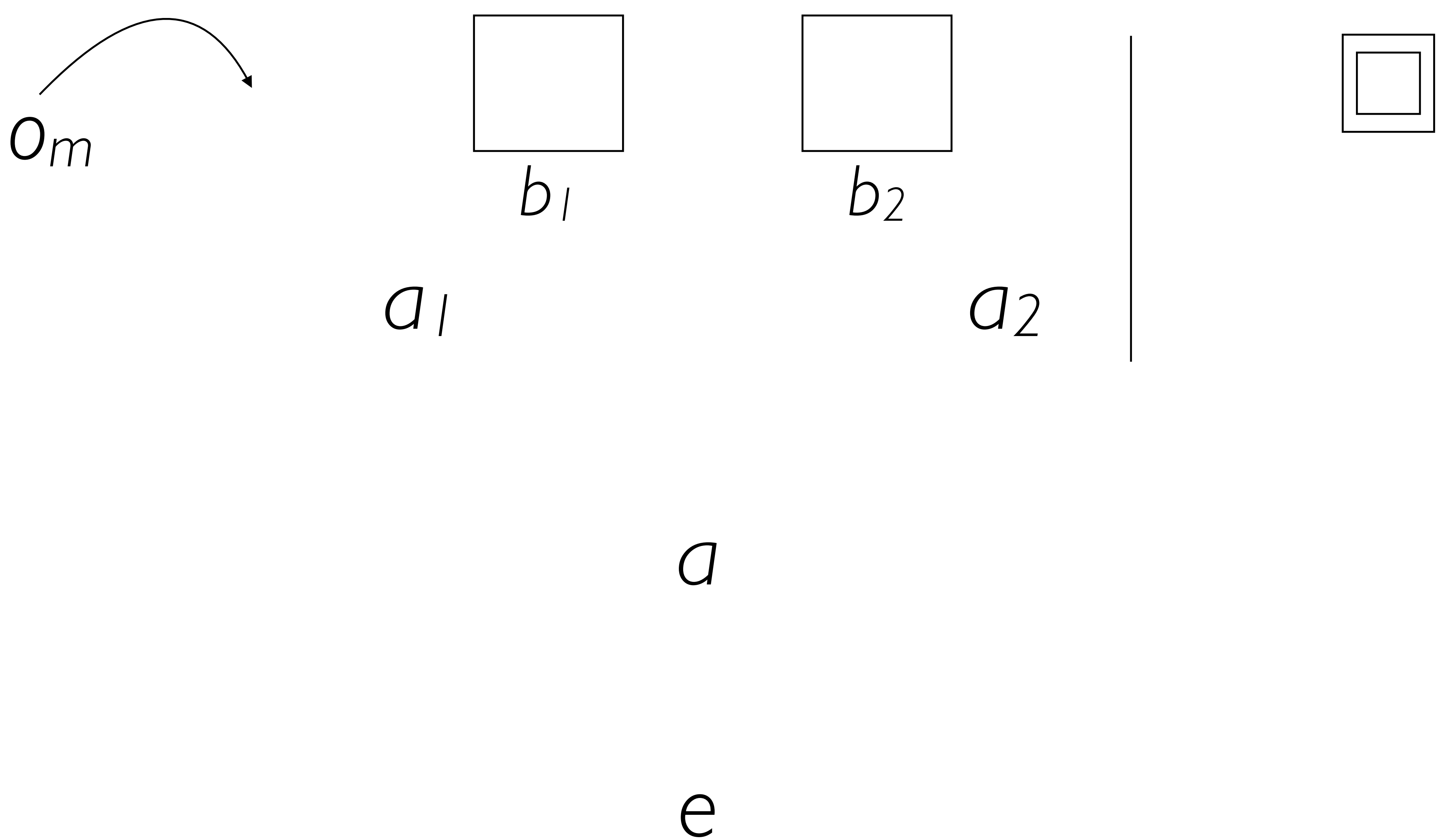


Framework for FBT^1_2

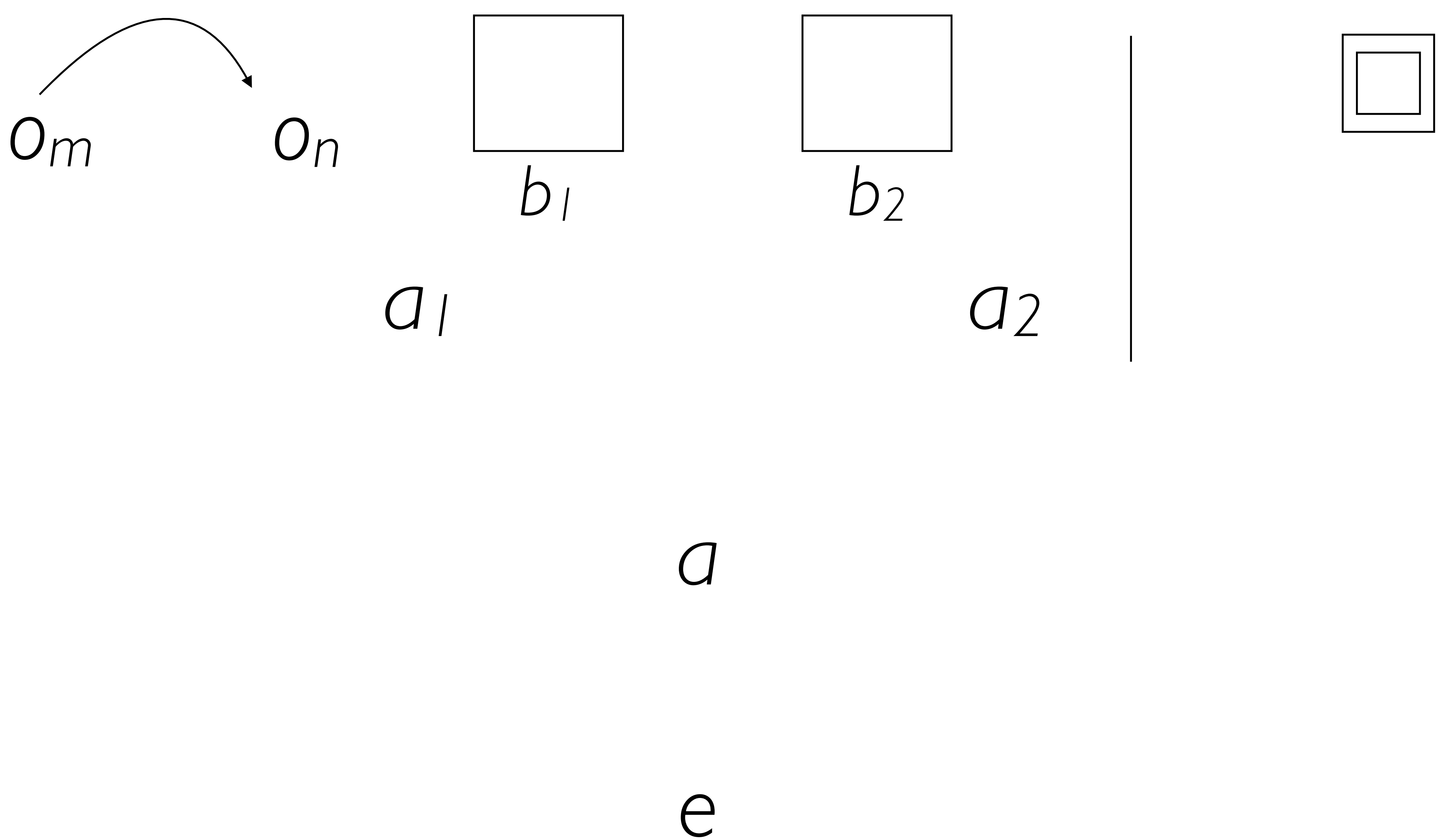
(seven timepoints)



Framework for FBT^1_2
(seven timepoints)

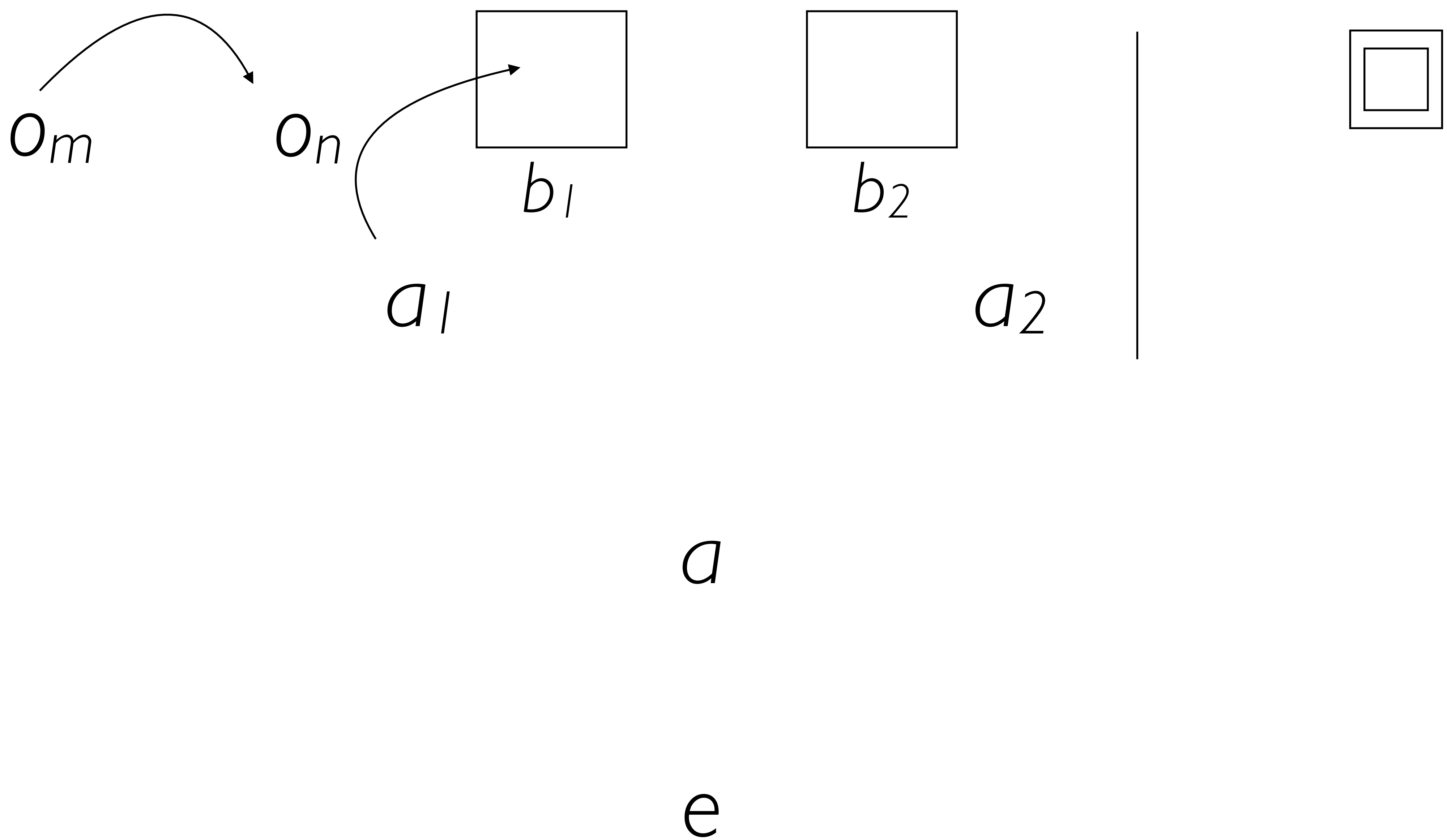


Framework for FBT$^1_2$
(seven timepoints)

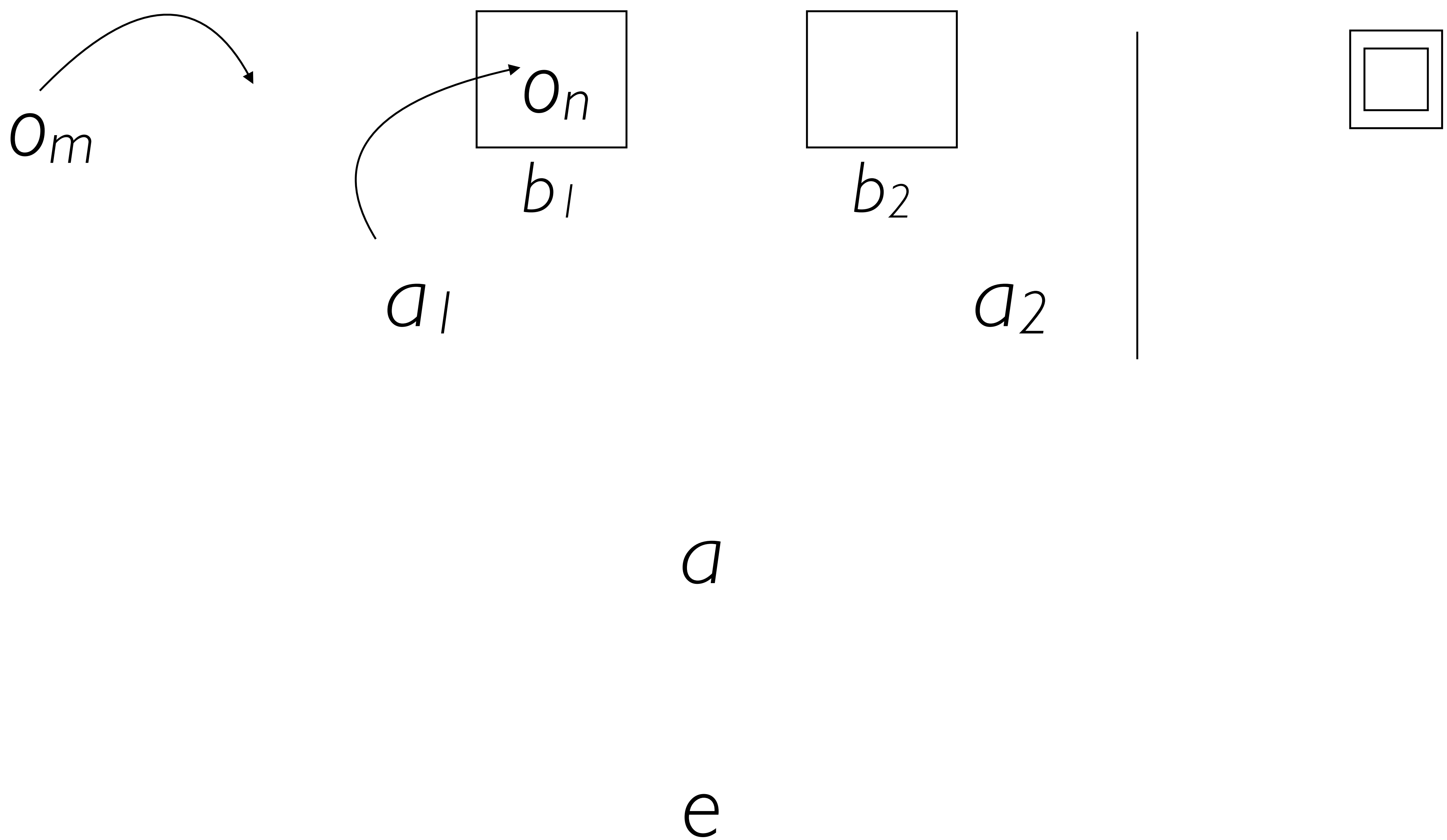


Framework for FBT$^1_2$

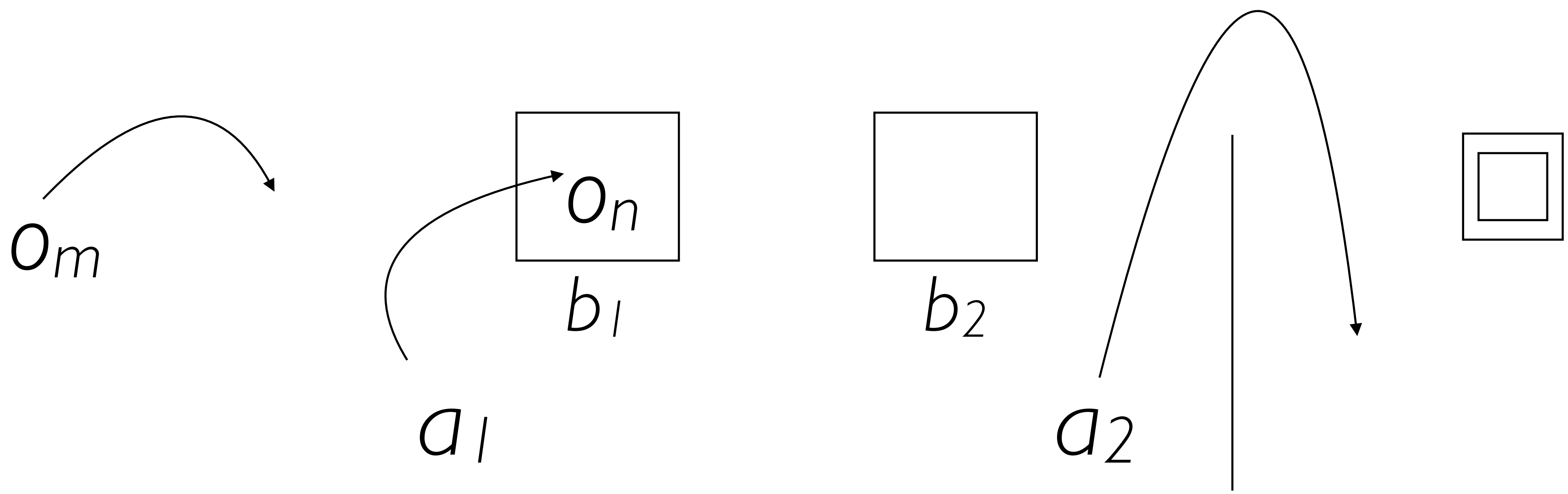
(seven timepoints)



Framework for FBT^1_2
(seven timepoints)



Framework for $FBT^1{}_2$
(seven timepoints)

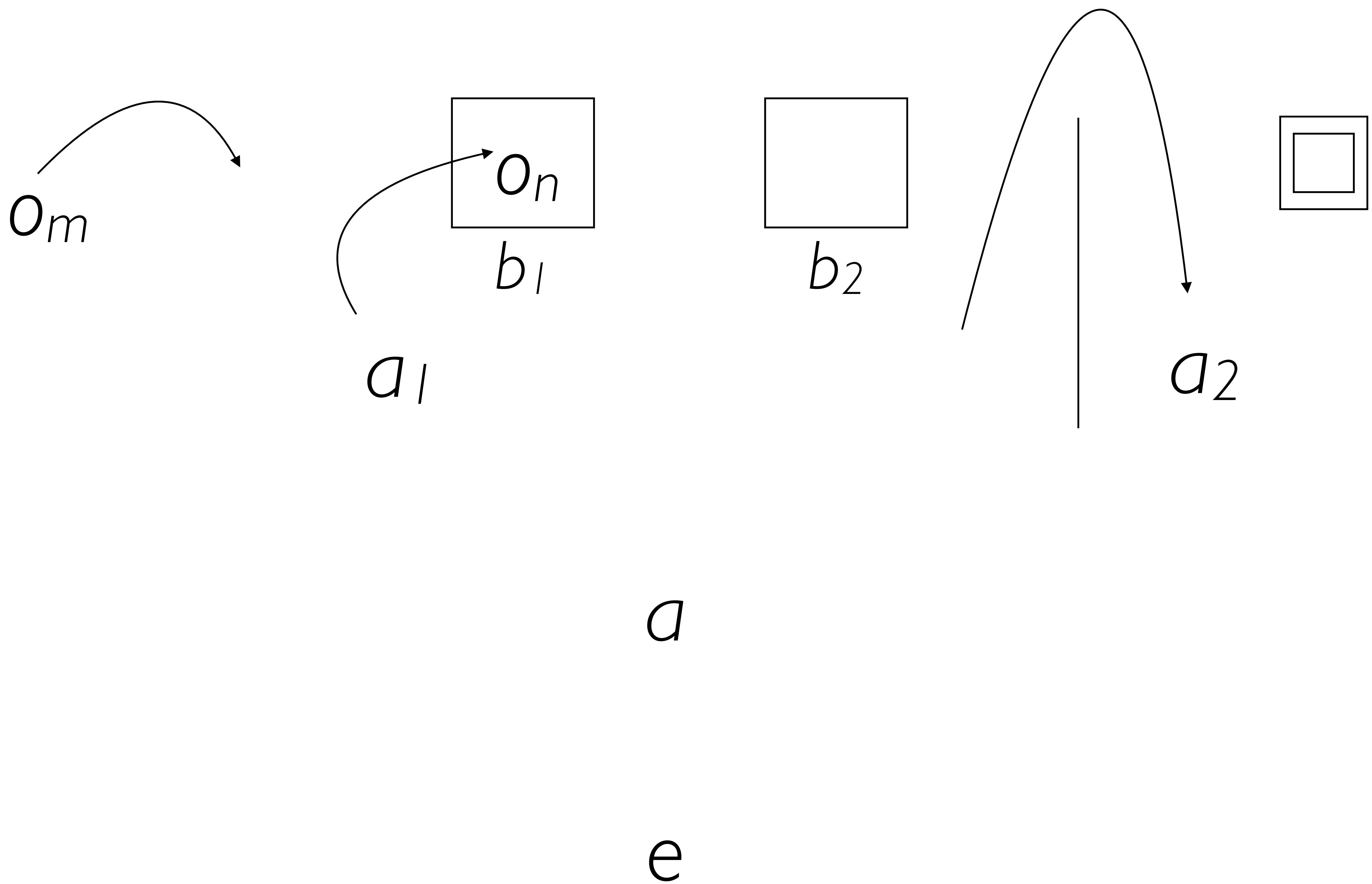


a

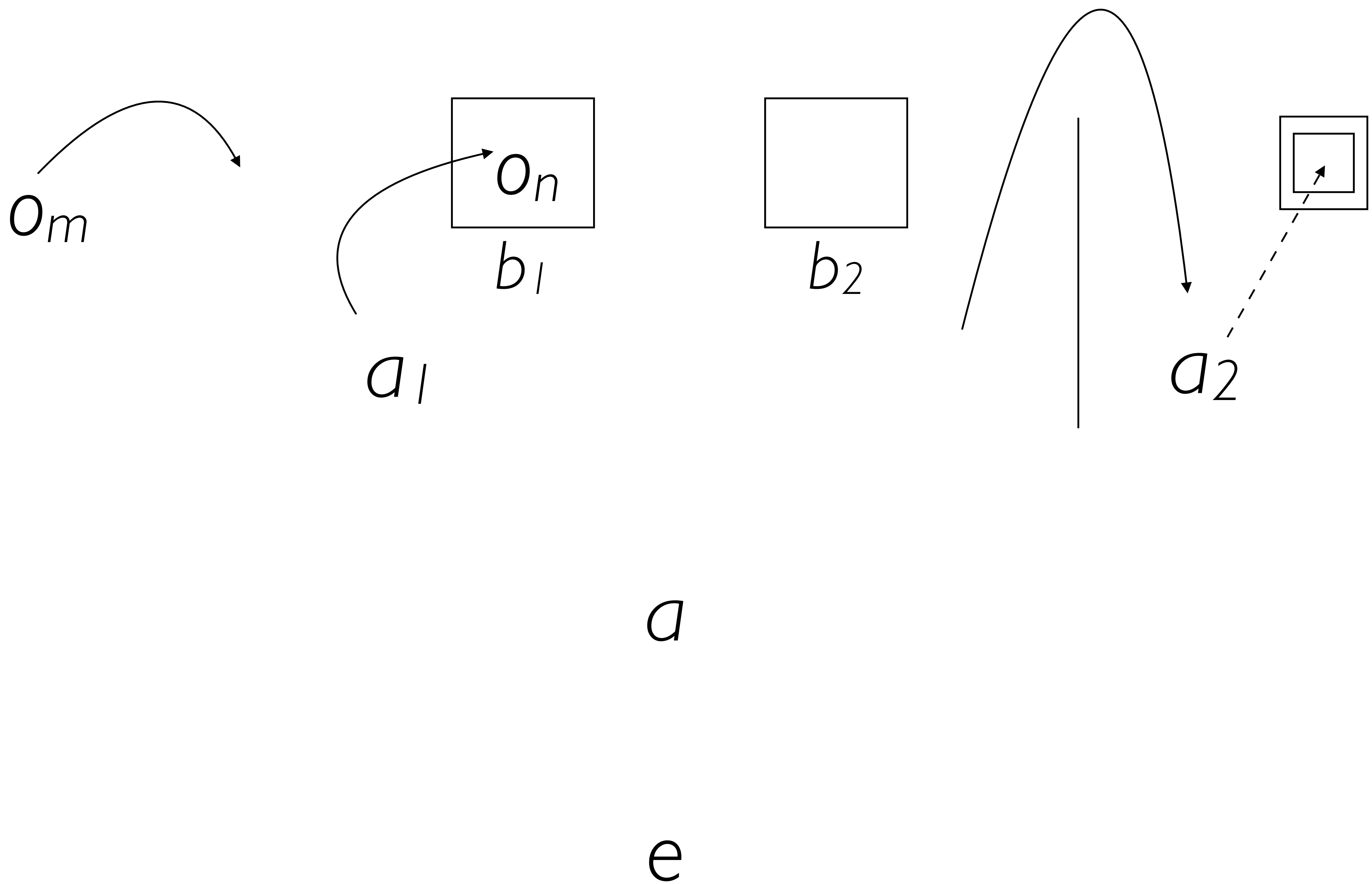
e

Framework for FBT^1_2

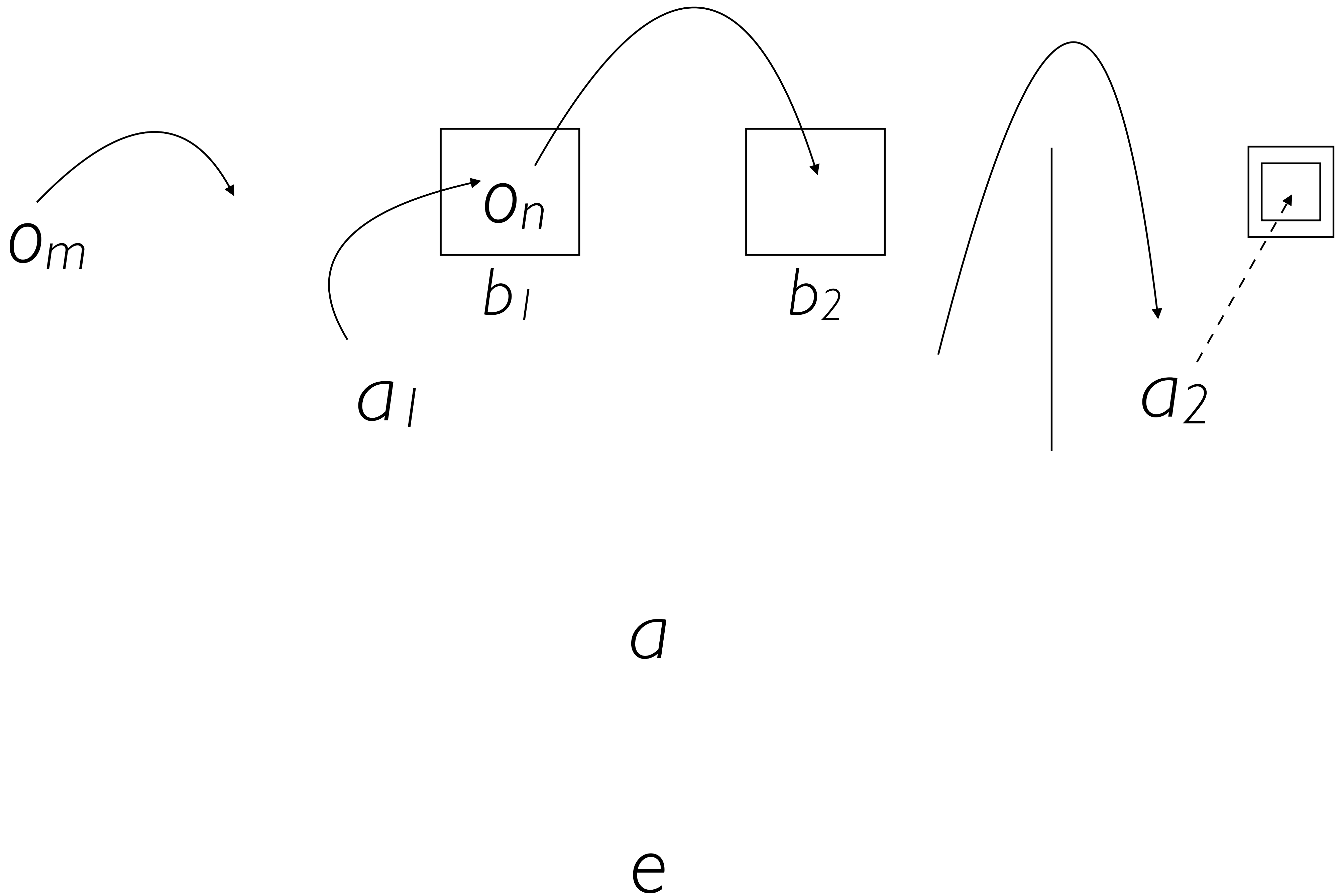
(seven timepoints)



Framework for FBT^1_2
(seven timepoints)

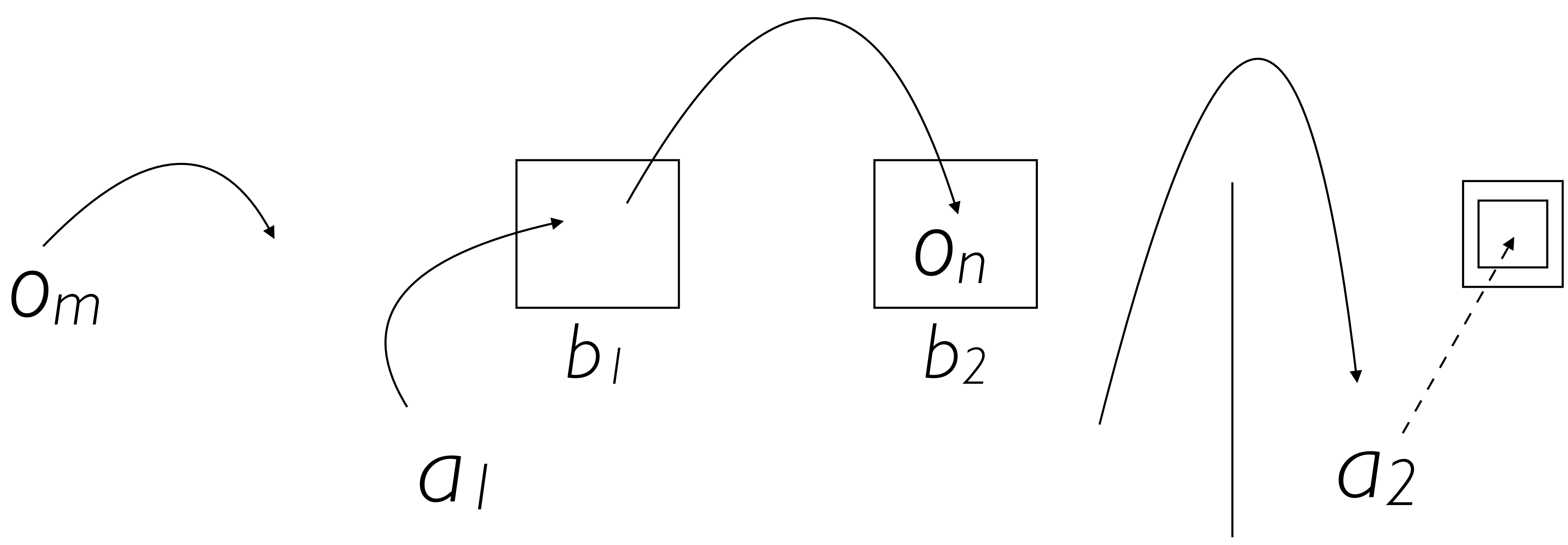


Framework for FBT^{1}_{2}
(seven timepoints)



Framework for $FBT^1{}_2$

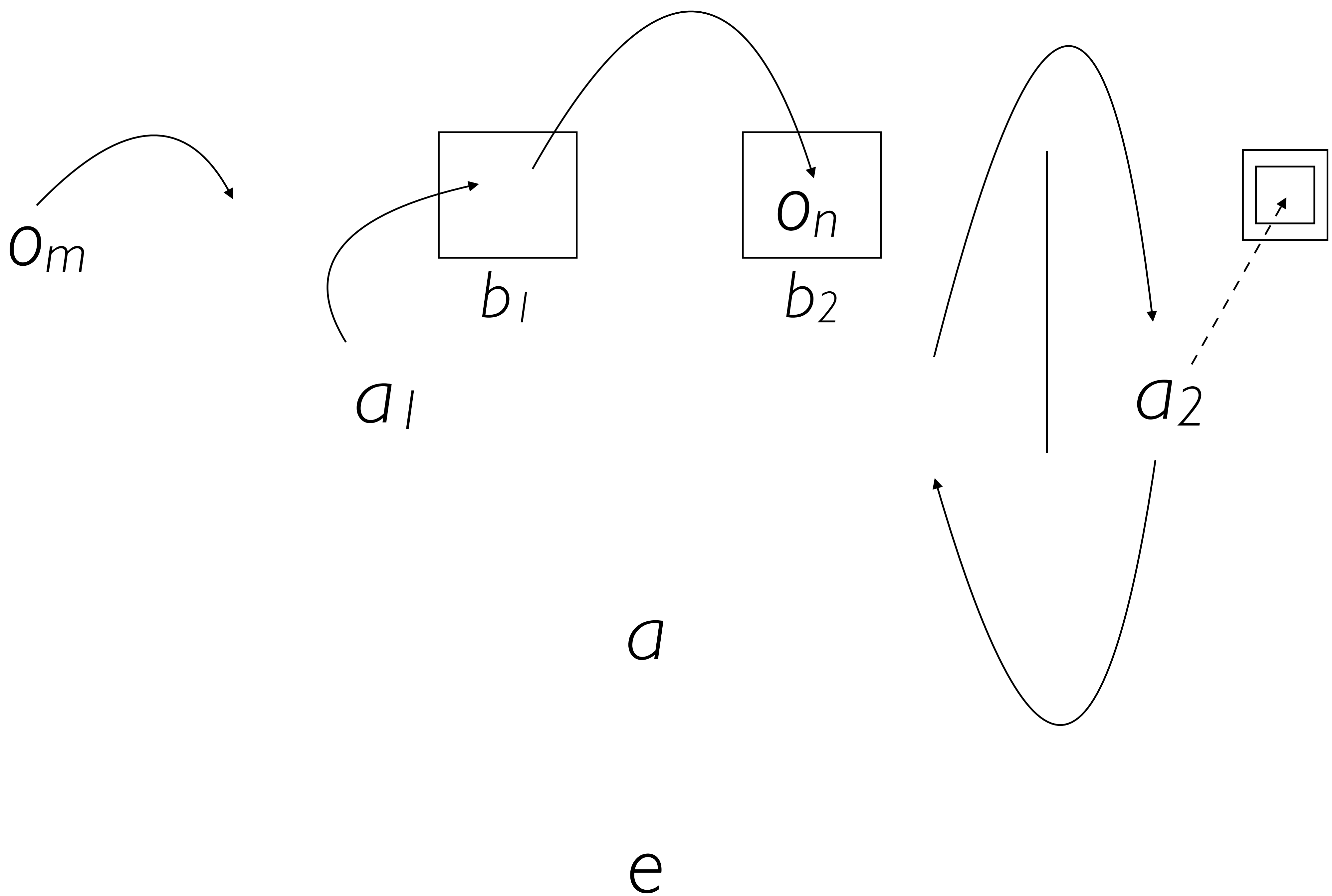
(seven timepoints)



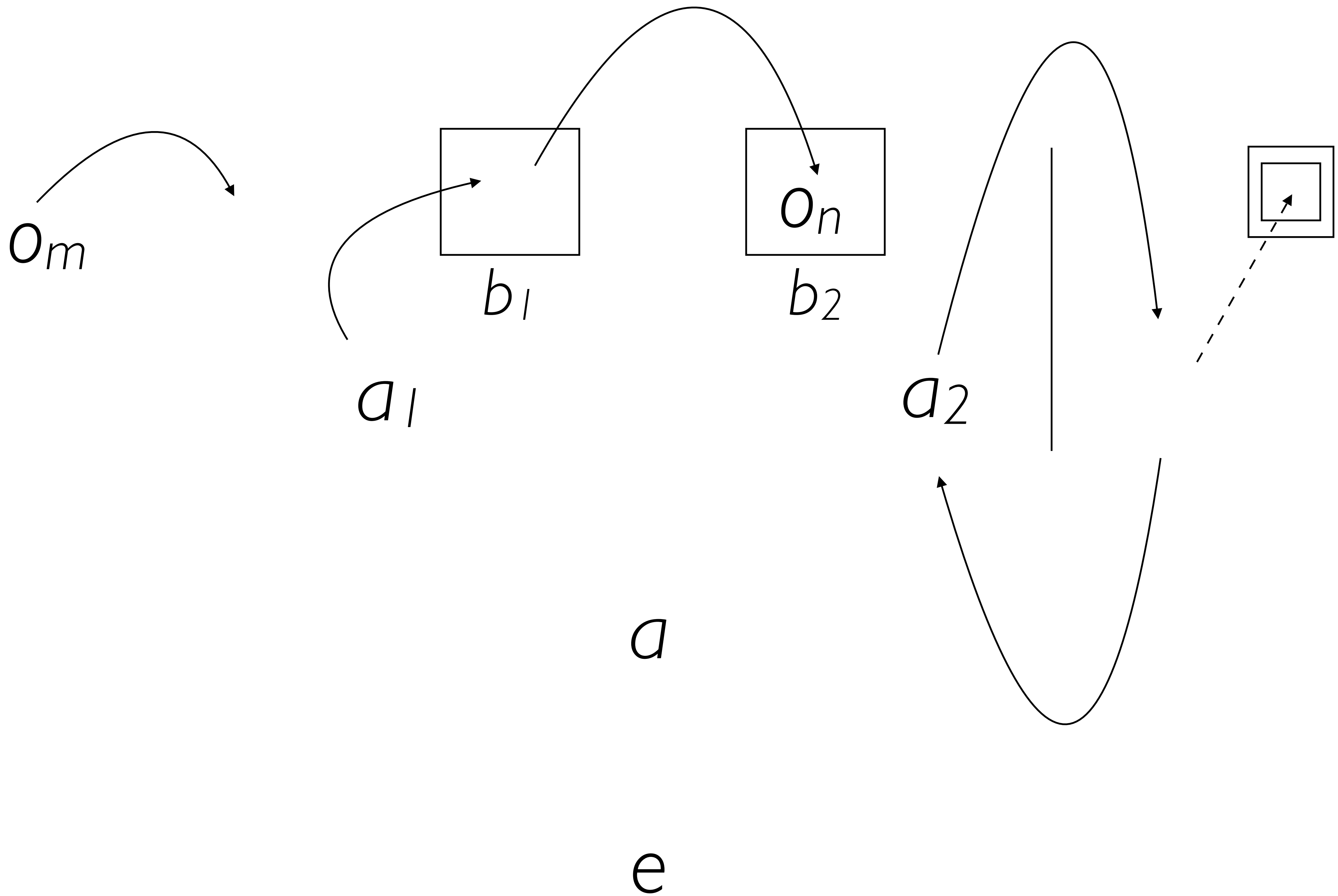
a

e

Framework for FBT$^1{}_2$
(seven timepoints)

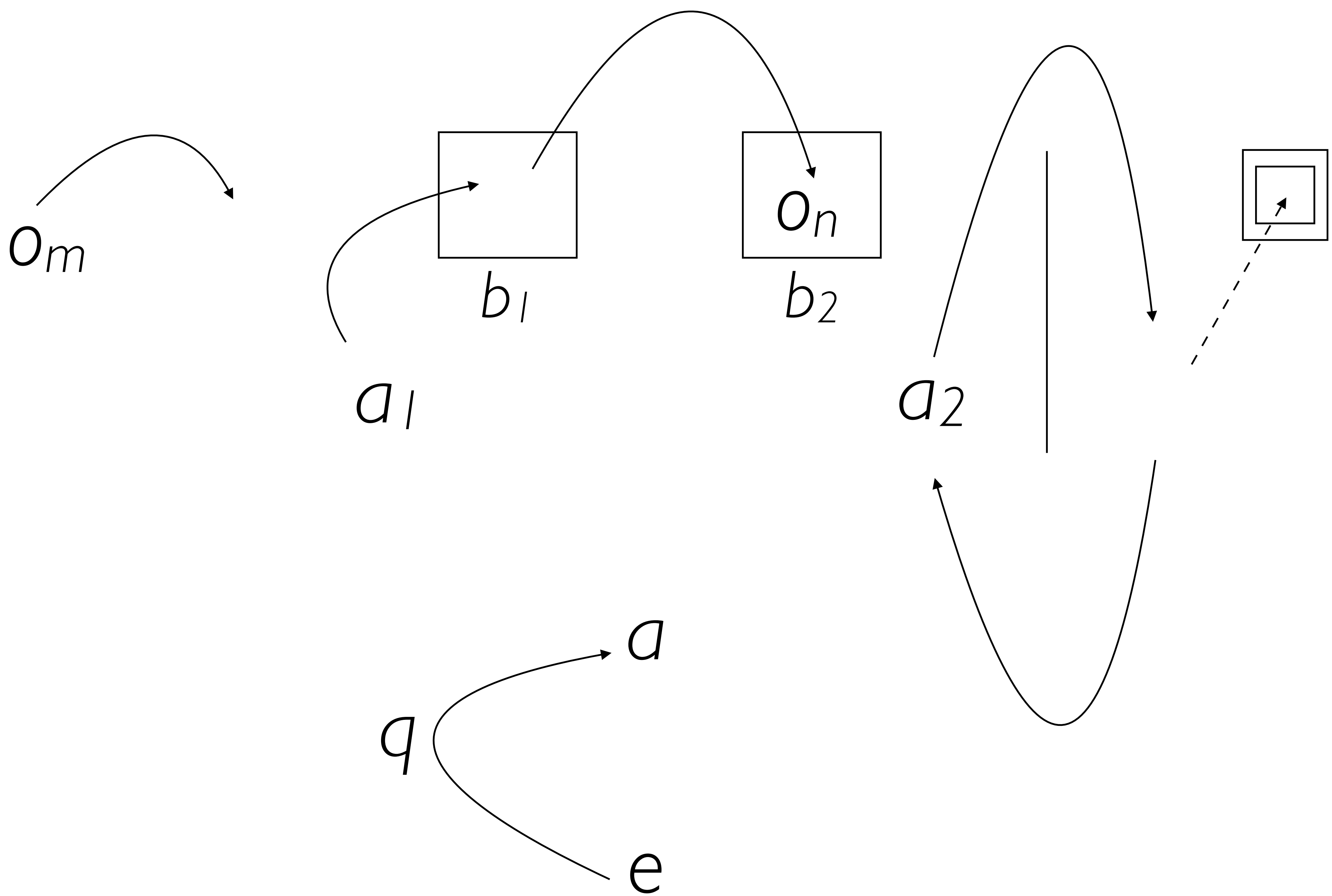


Framework for FBT$^1{}_2$
(seven timepoints)

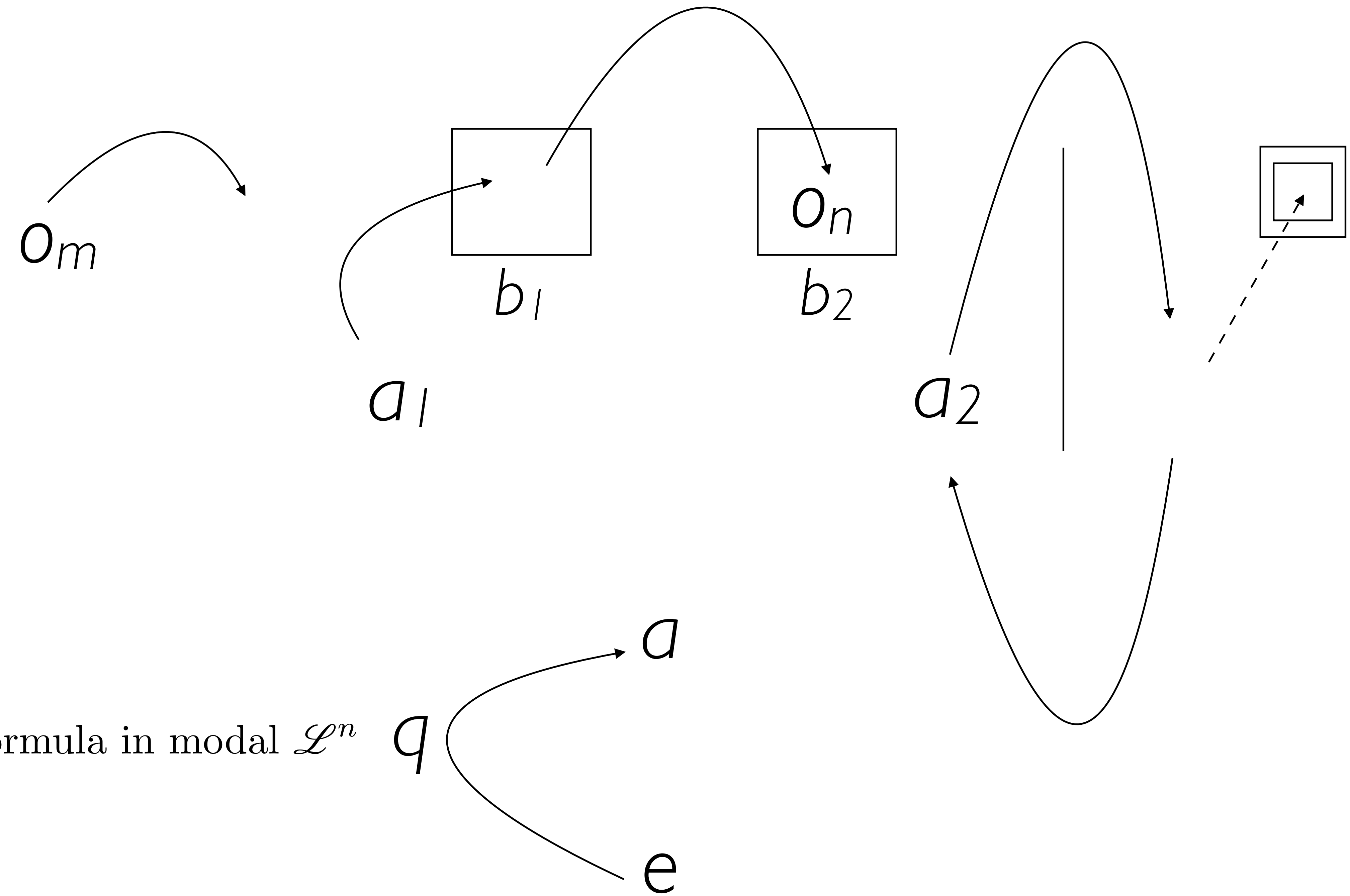


Framework for $FBT^1{}_2$

(seven timepoints)

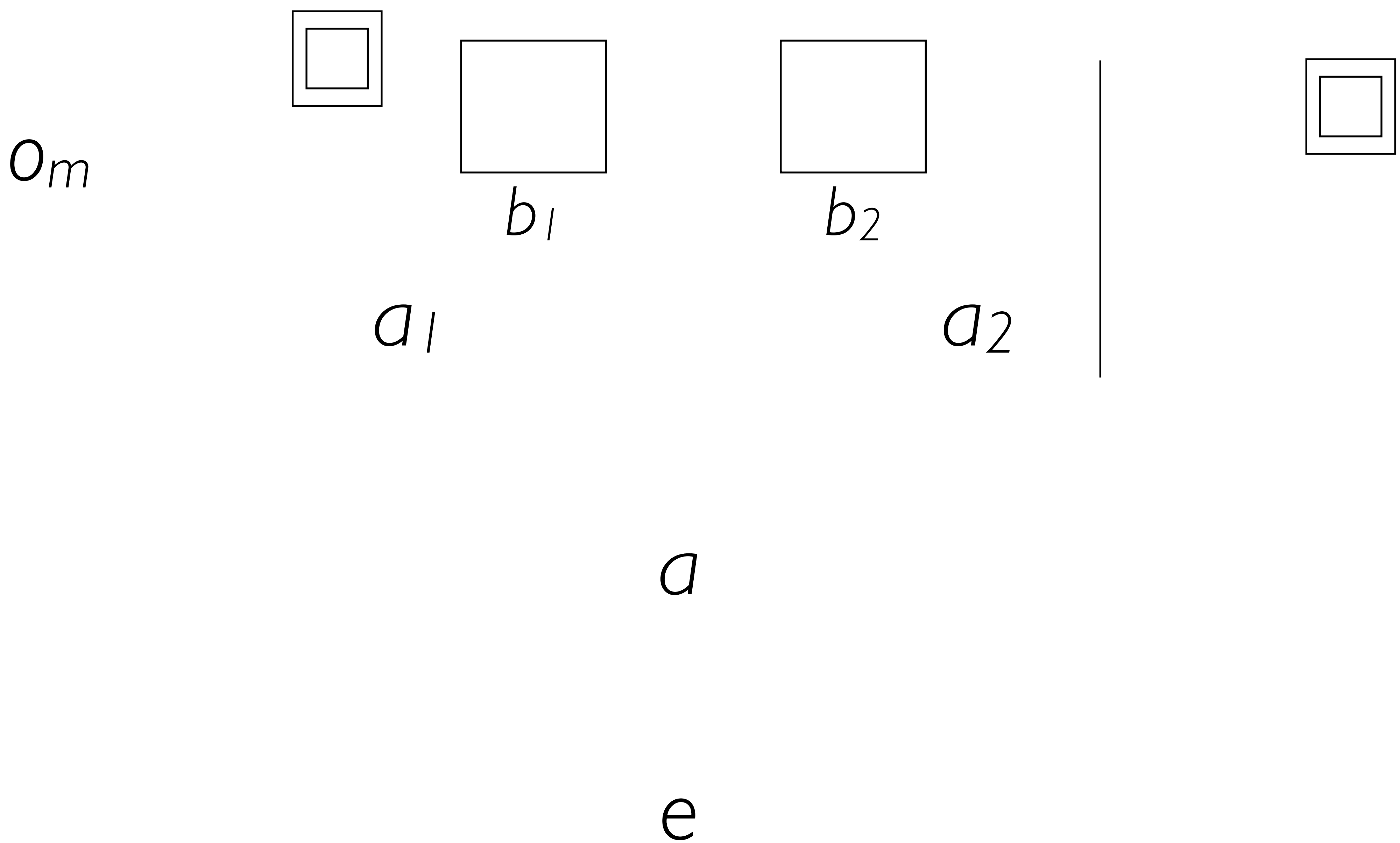


Framework for $FBT^1{}_2$
(seven timepoints)

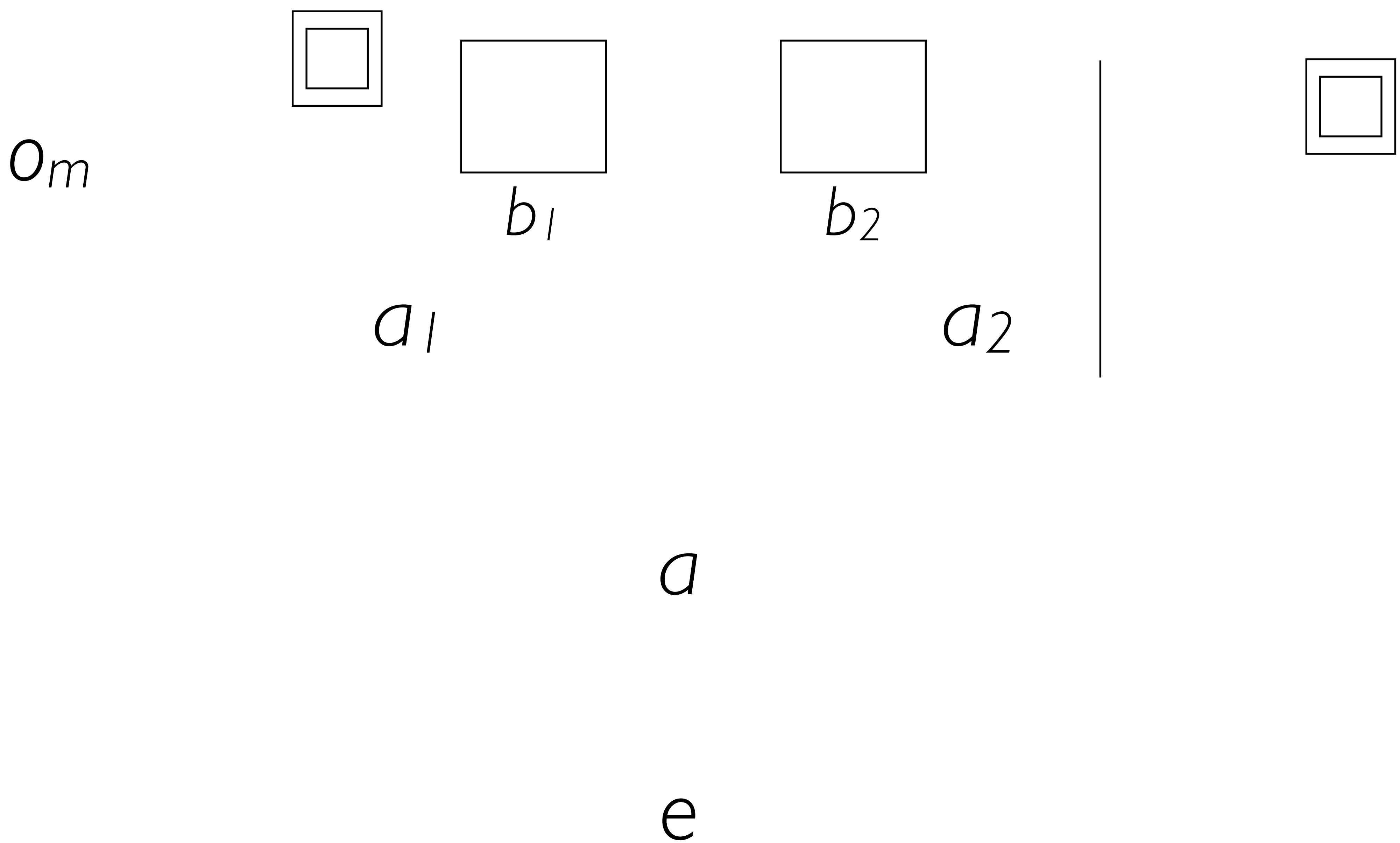


q a formula in modal $\mathscr{L}^n$

Framework for FBT^1_3

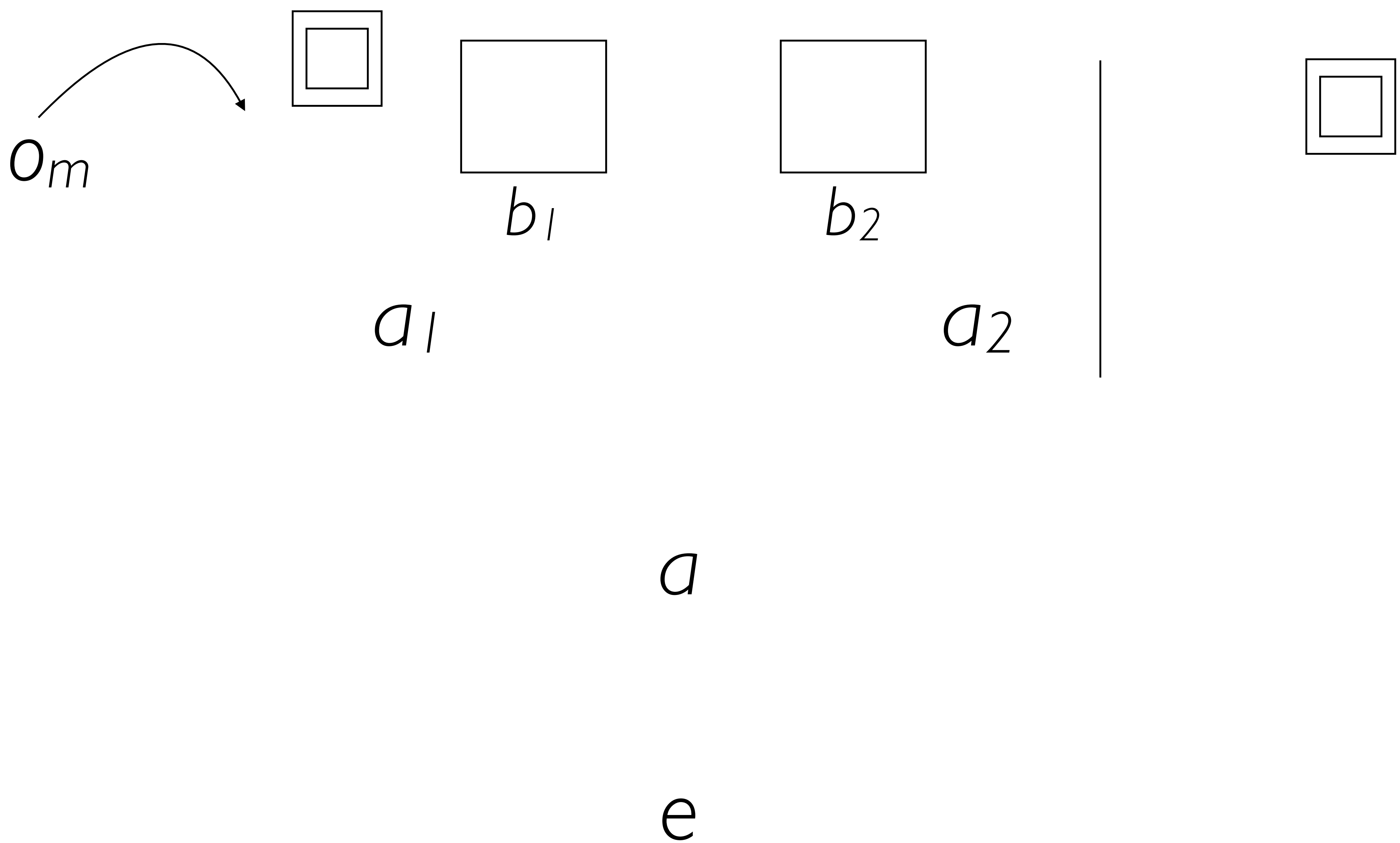


Framework for FBT^1_3
(eight timepoints)

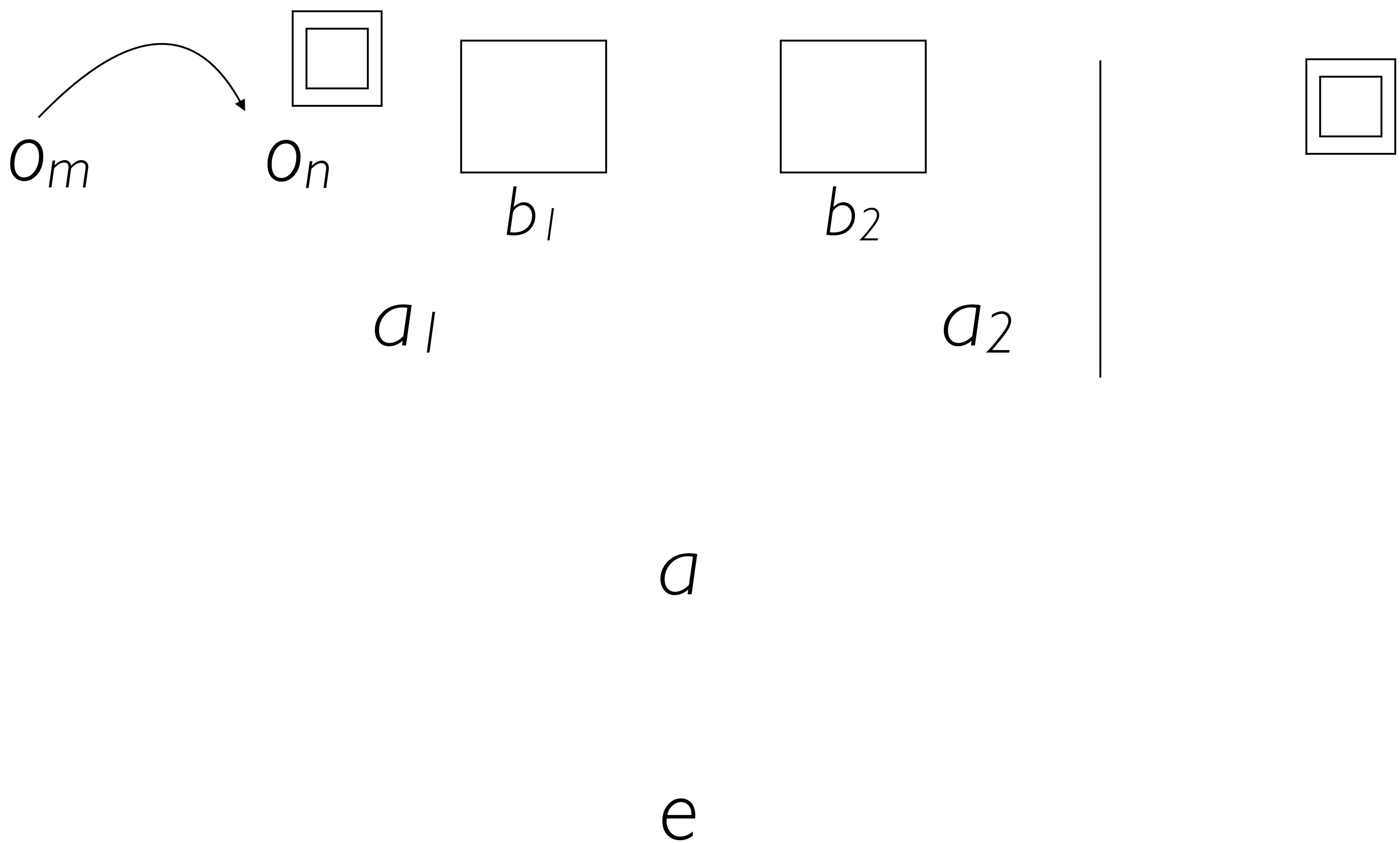


Framework for FBT^1_3

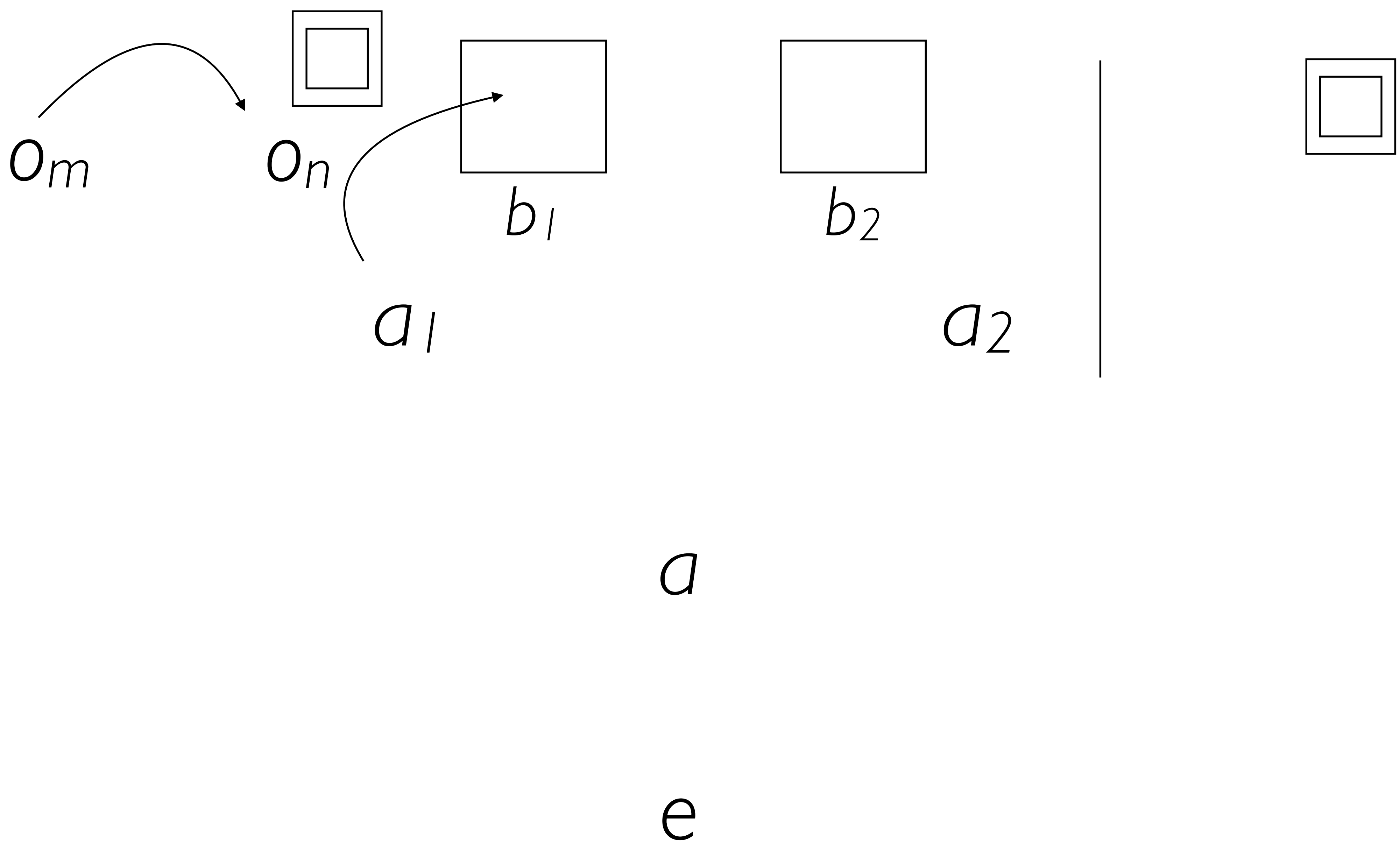
(eight timepoints)



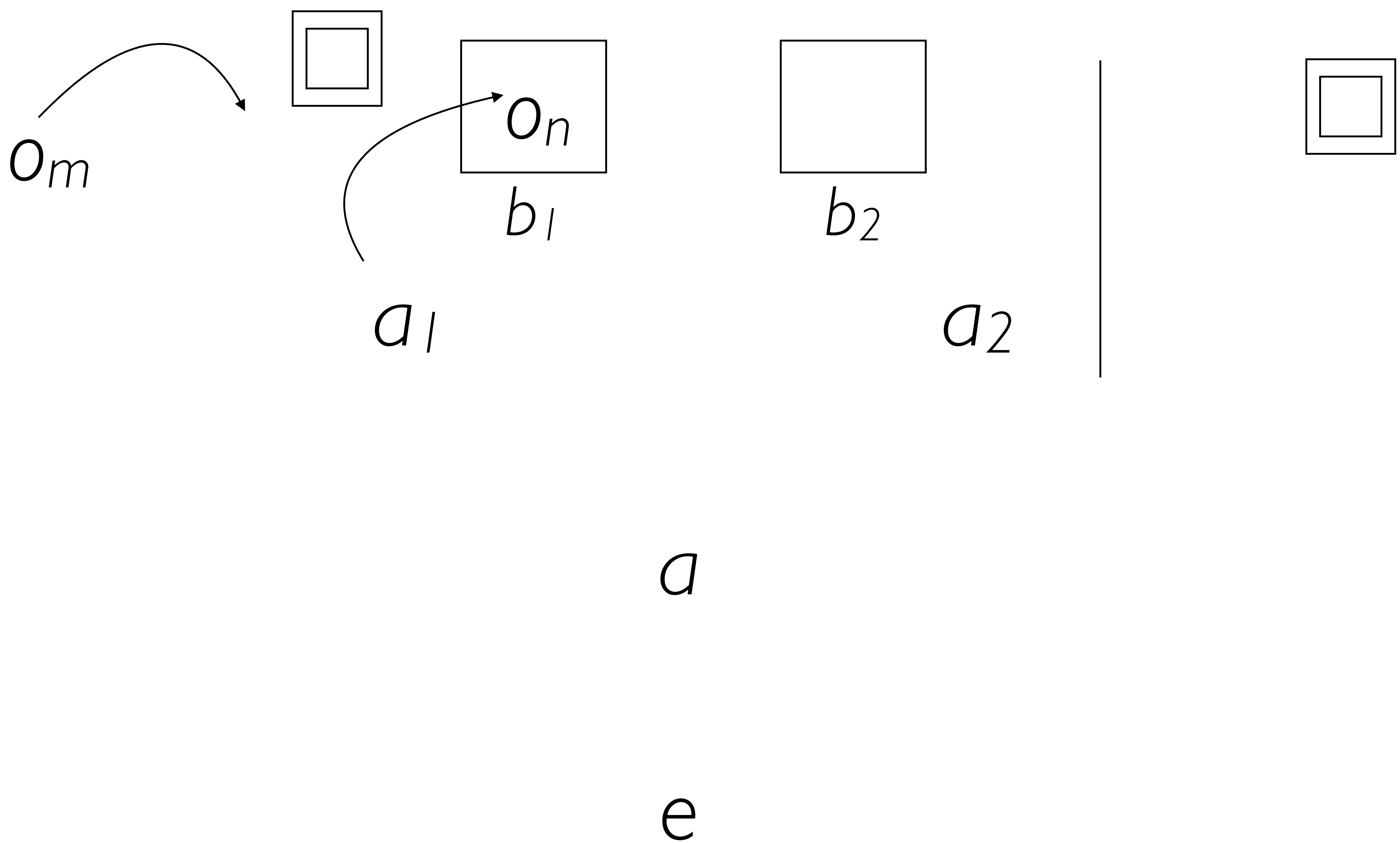
Framework for FBT^1_3
(eight timepoints)



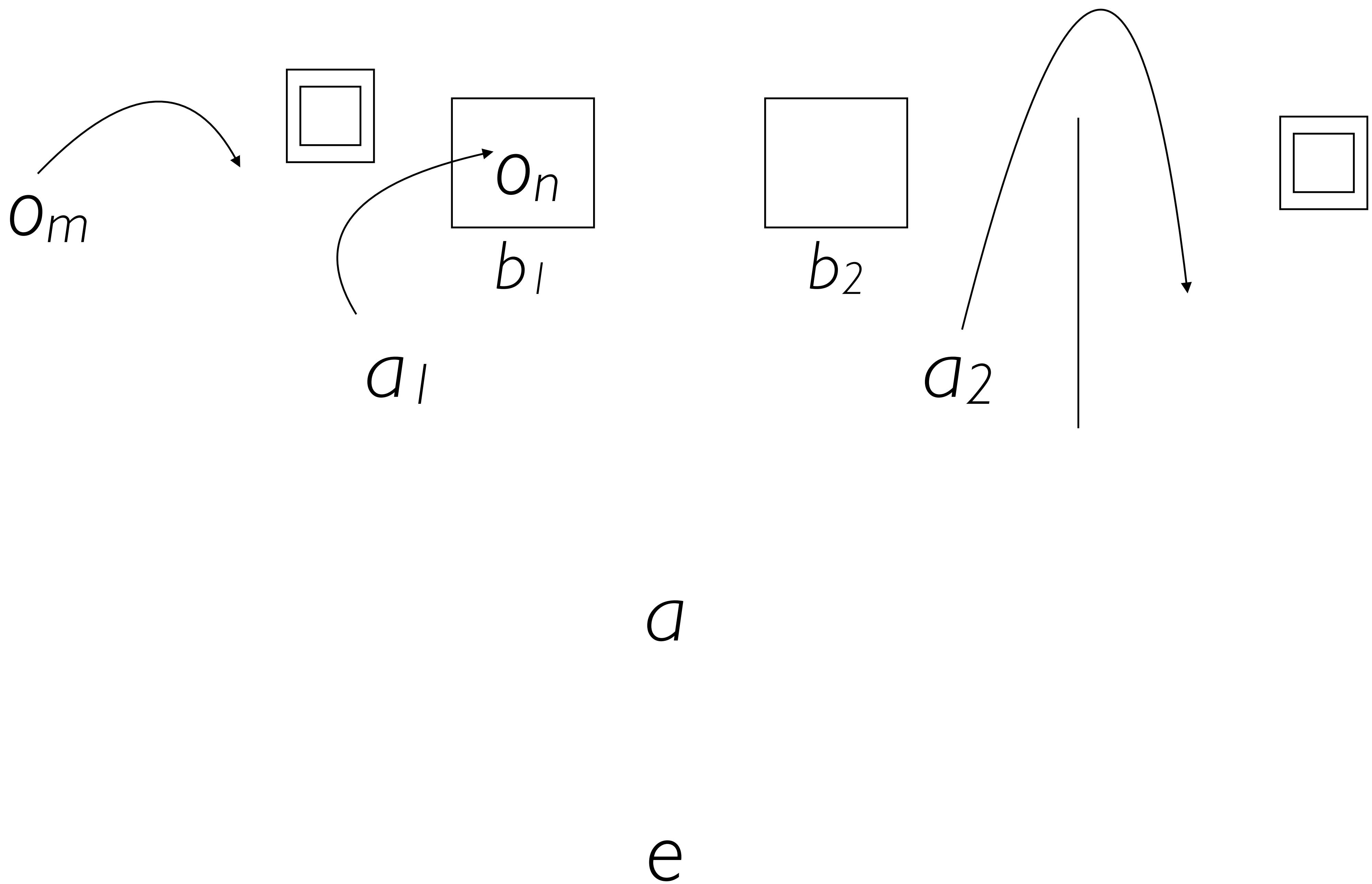
Framework for FBT^1_3
(eight timepoints)



Framework for FBT$^{1}_{3}$
(eight timepoints)

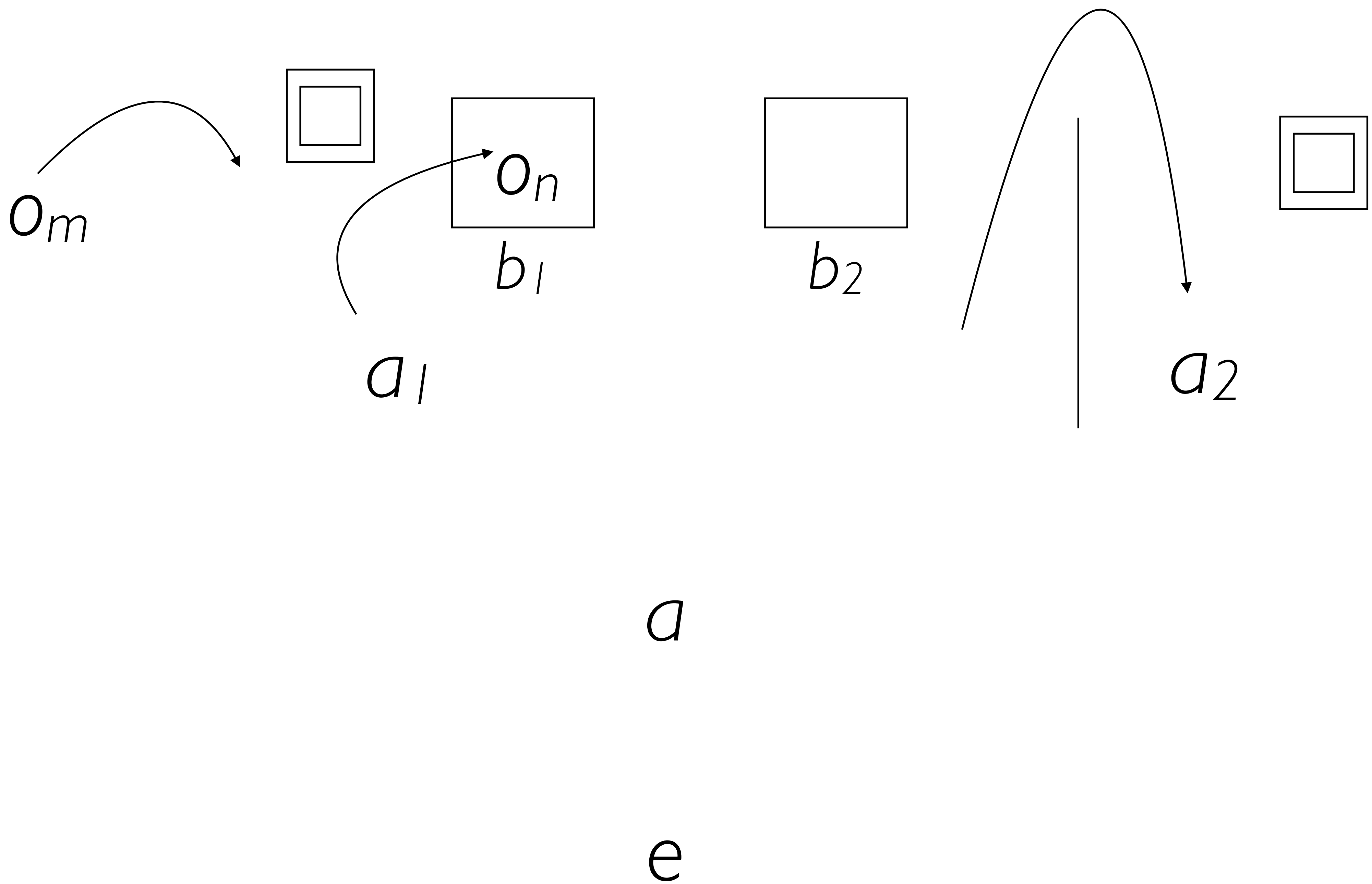


Framework for FBT^1_3
(eight timepoints)

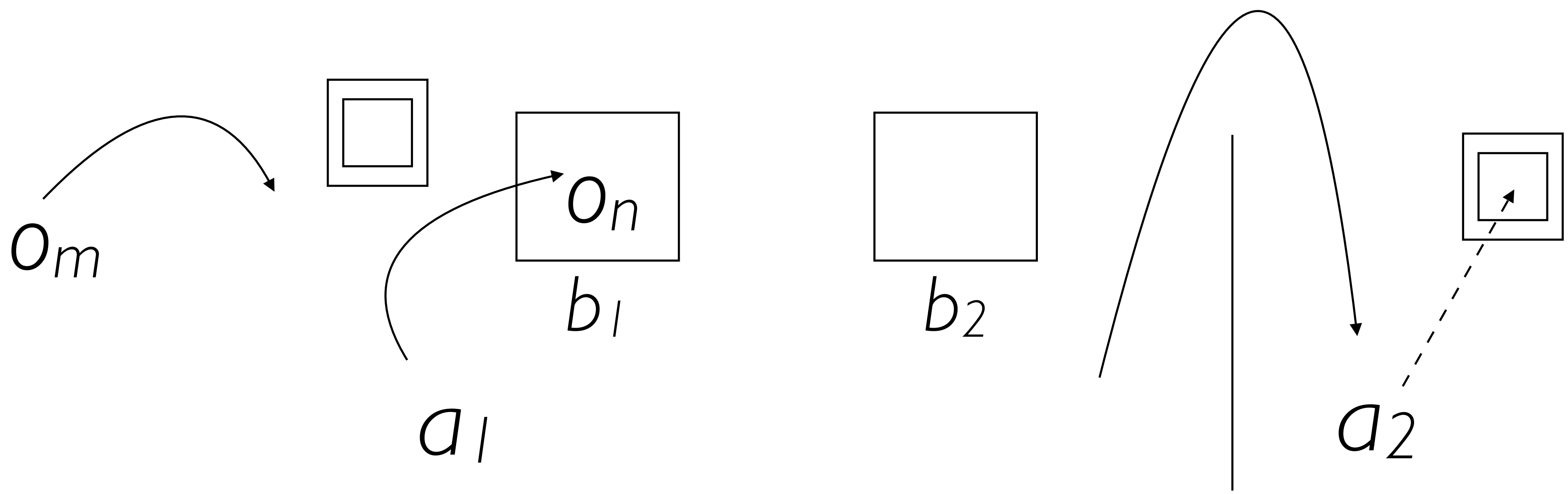


Framework for FBT^{1}_{3}

(eight timepoints)



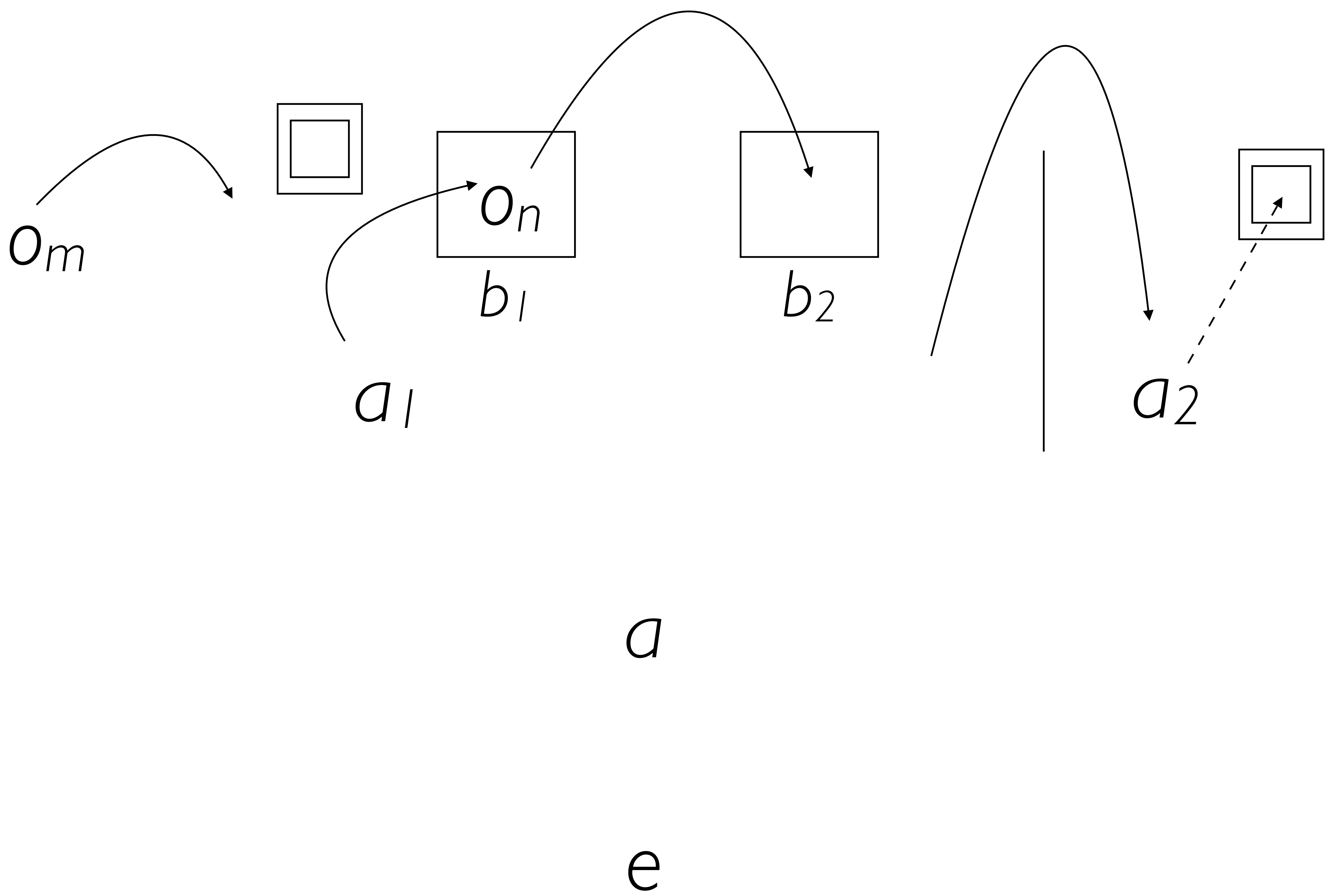
Framework for FBT^1_3
(eight timepoints)



a

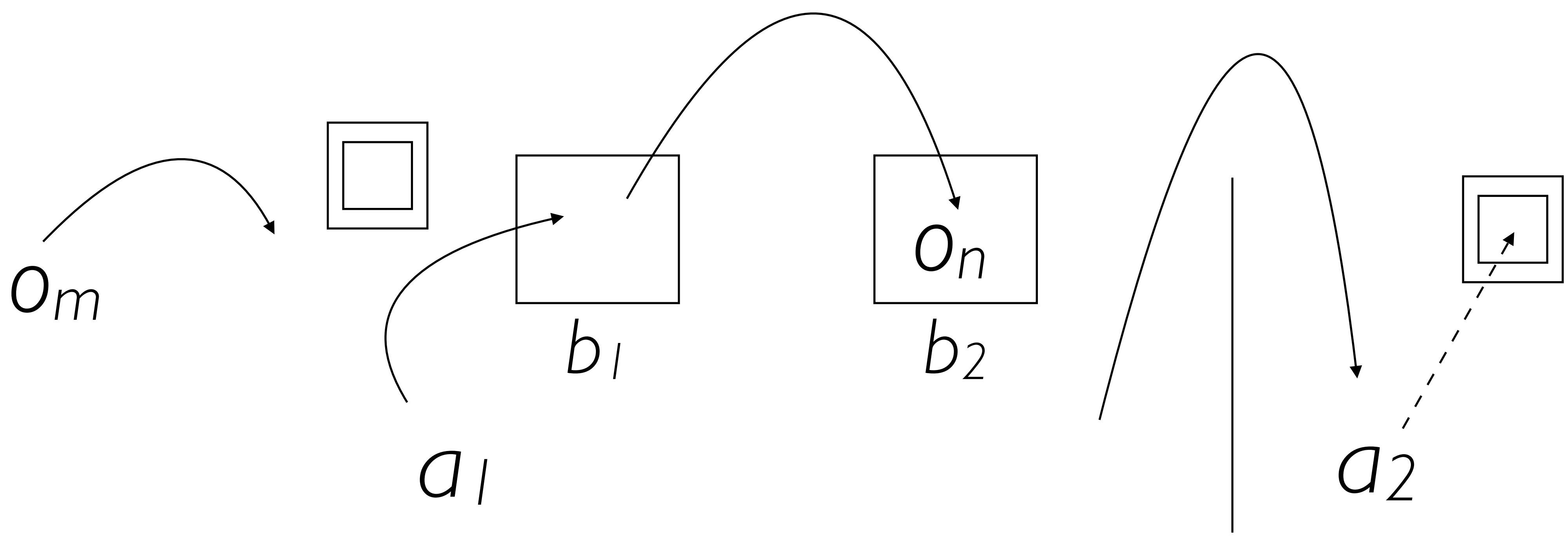
e

Framework for FBT^1_3
(eight timepoints)



Framework for FBT^1_3

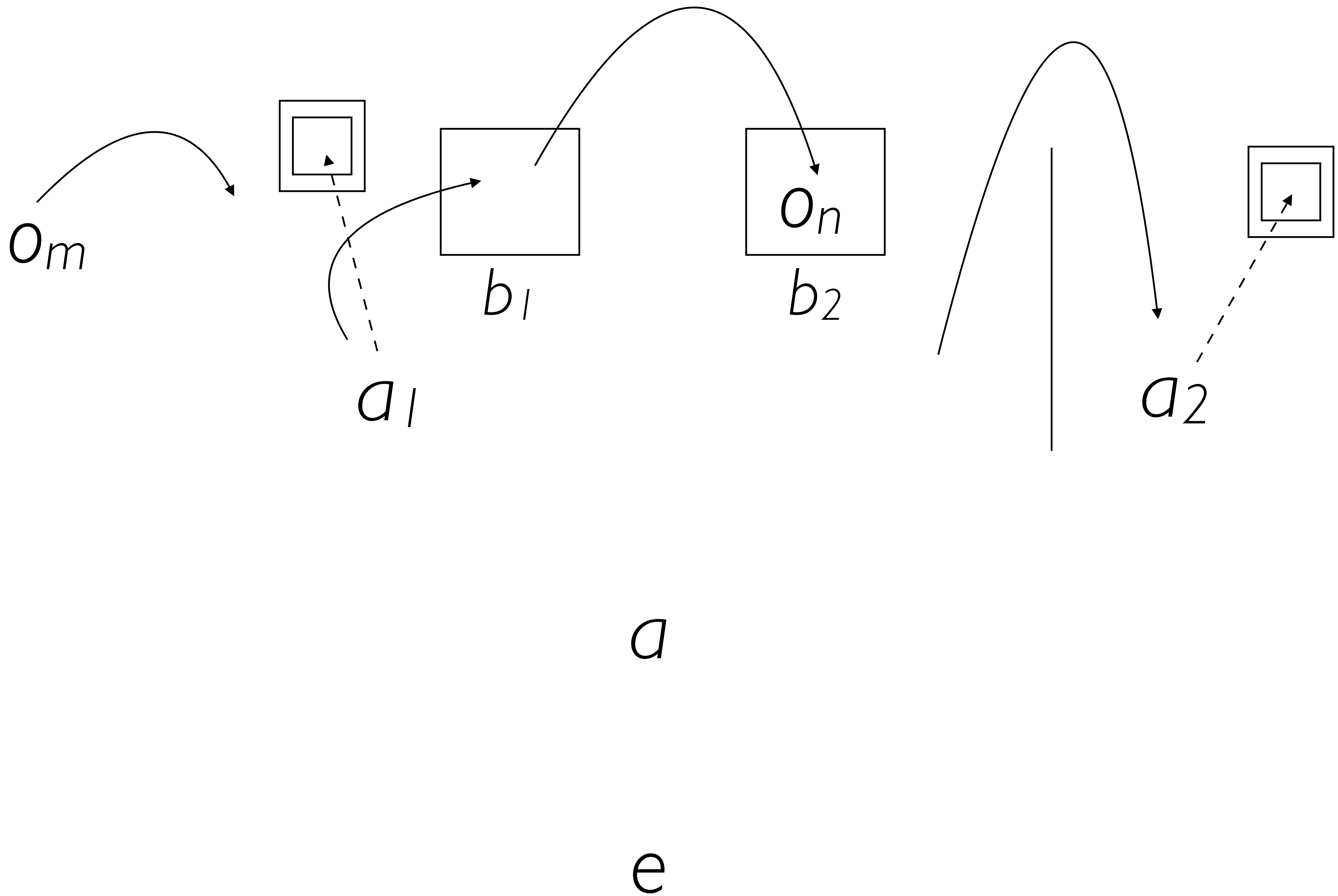
(eight timepoints)



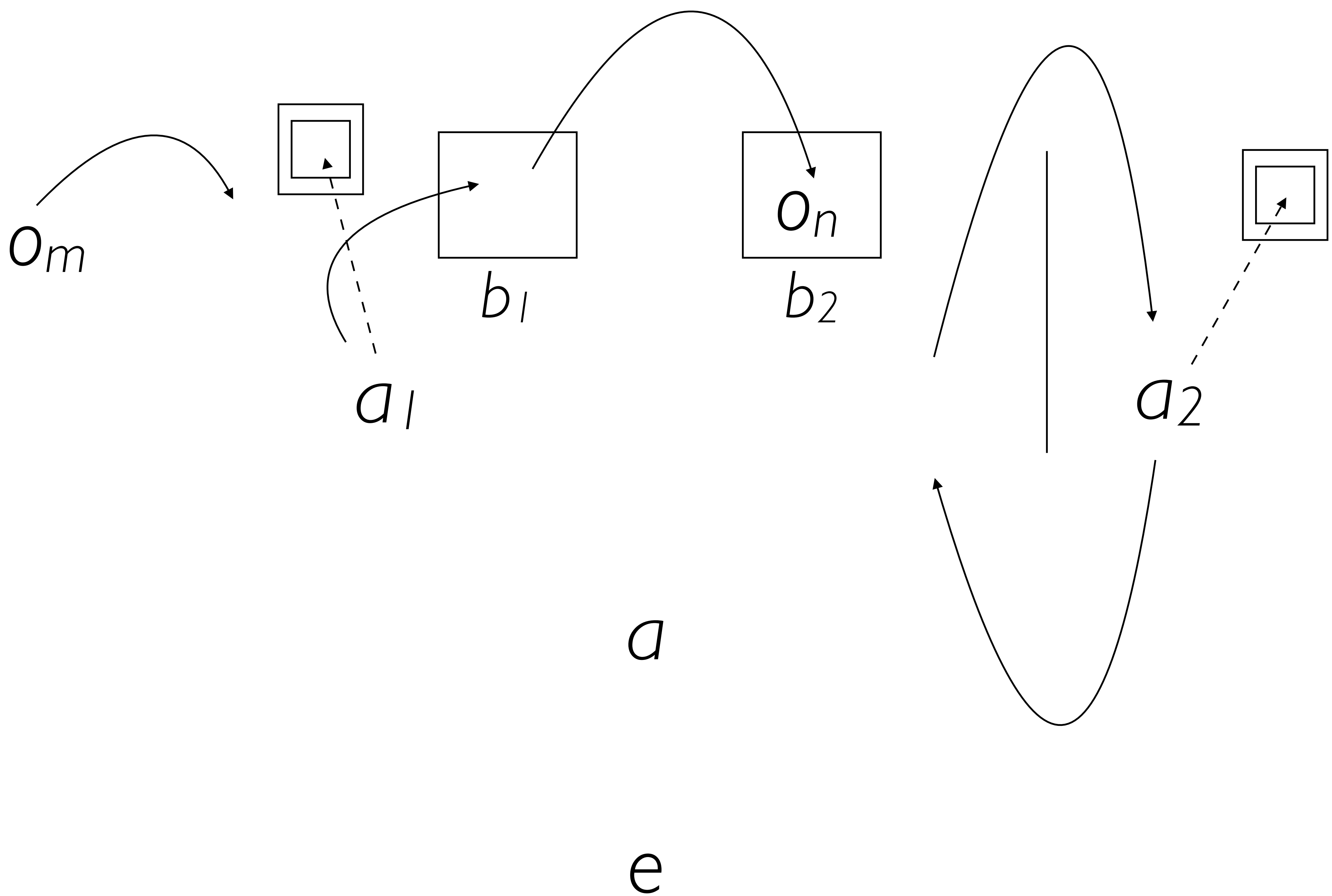
a

e

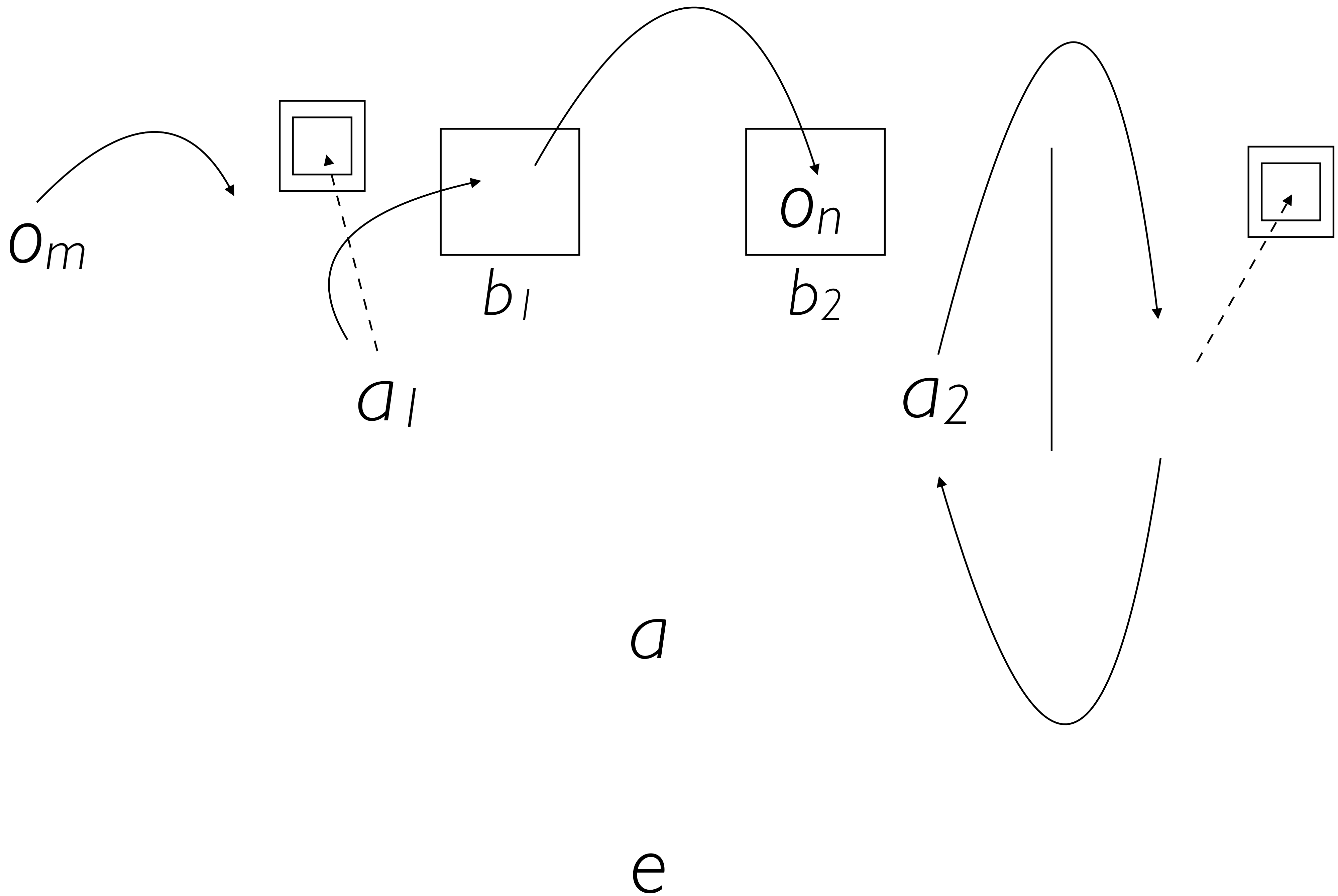
Framework for $FBT^1{}_3$
(eight timepoints)



Framework for FBT^1_3
(eight timepoints)

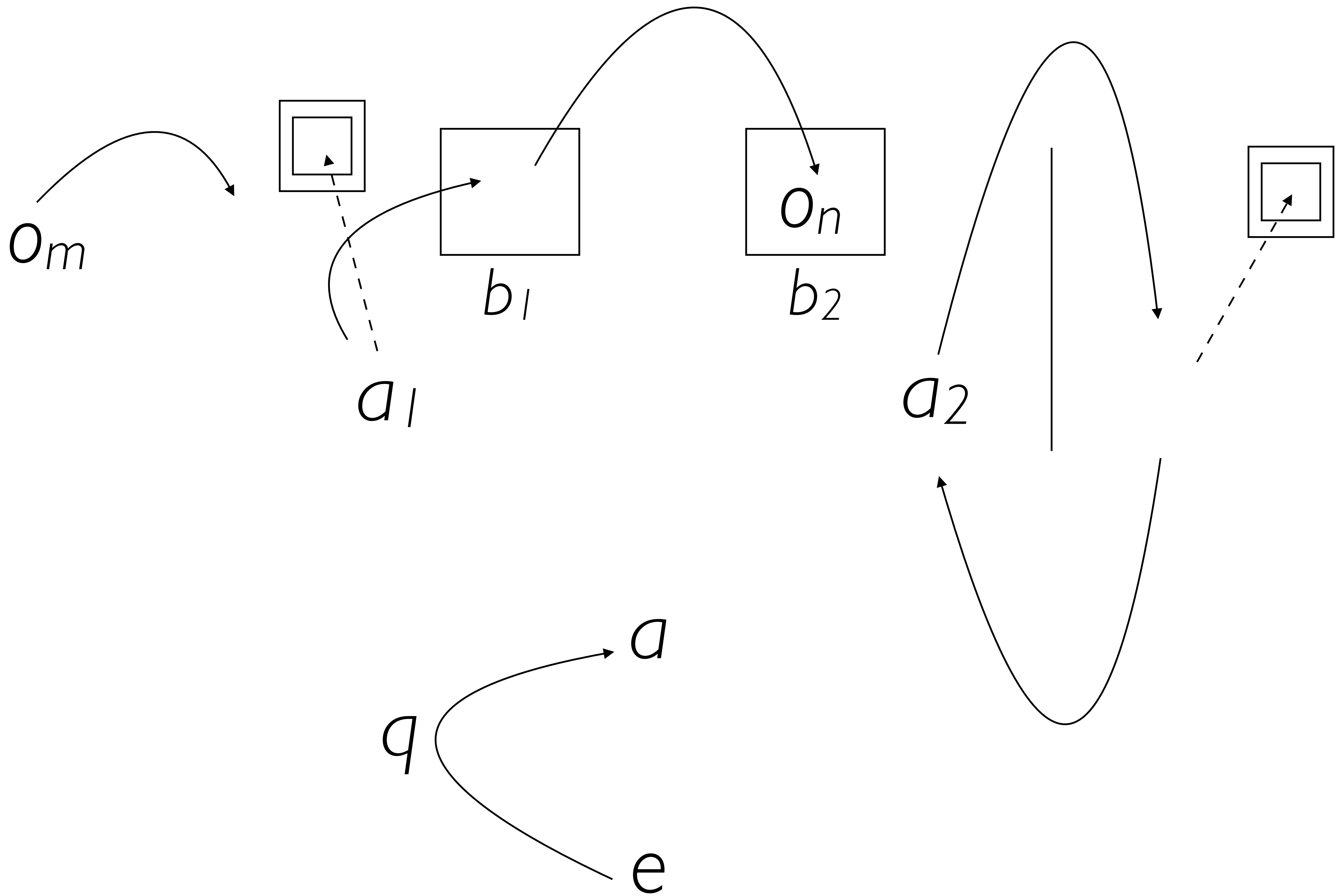


Framework for $FBT^1{}_3$
(eight timepoints)

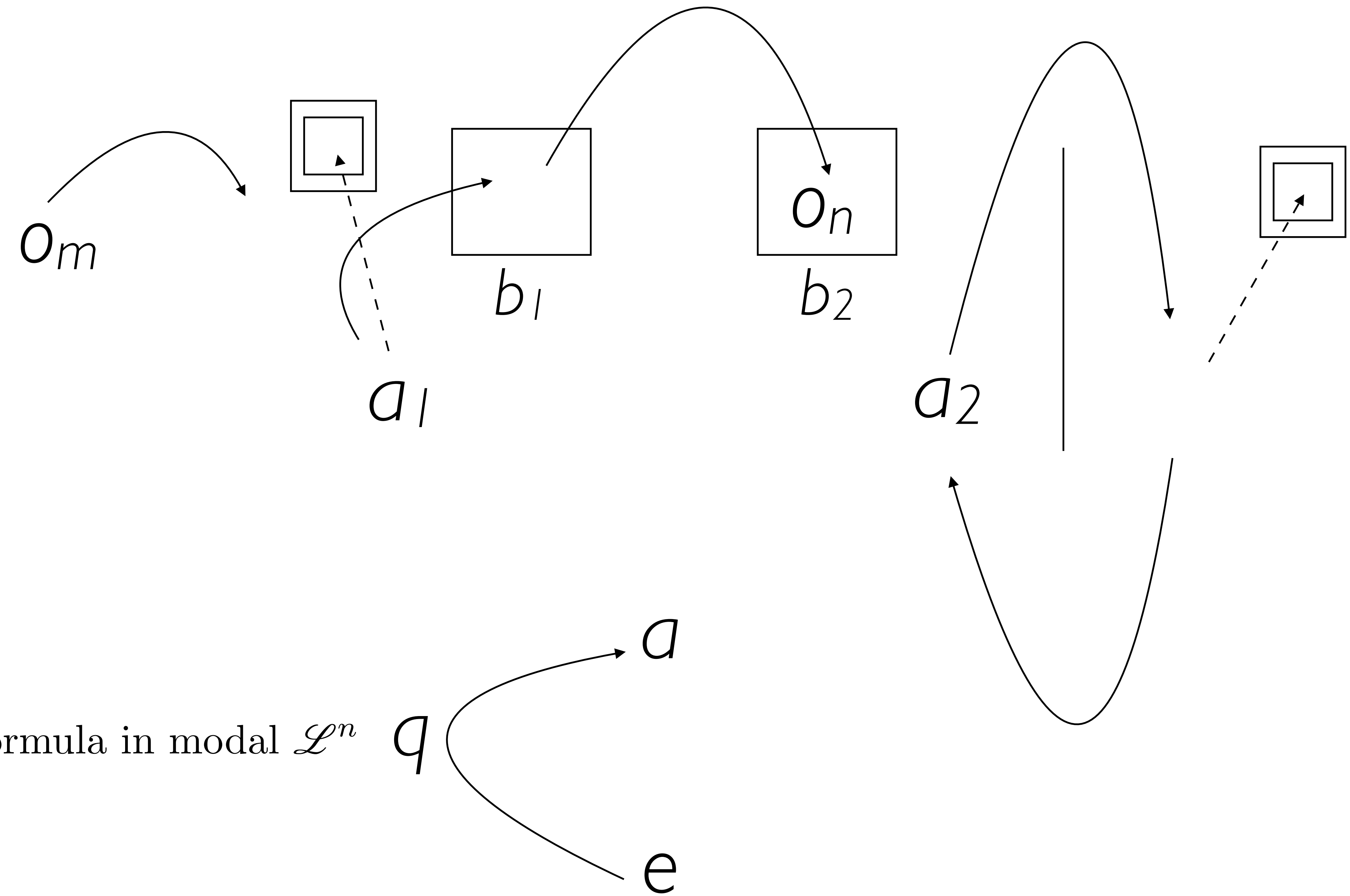


Framework for $FBT^1{}_3$

(eight timepoints)

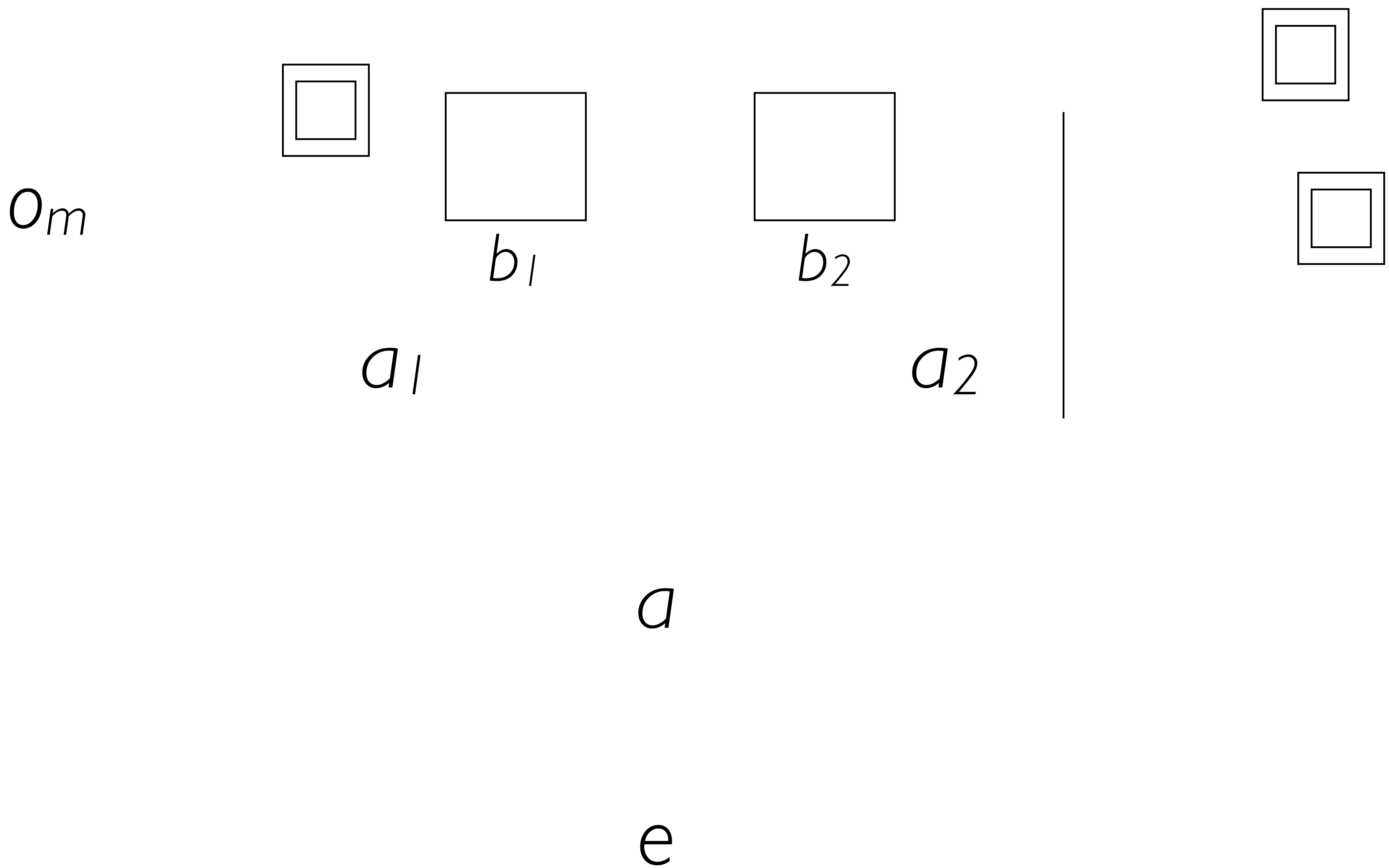


Framework for $FBT^1{}_3$
(eight timepoints)



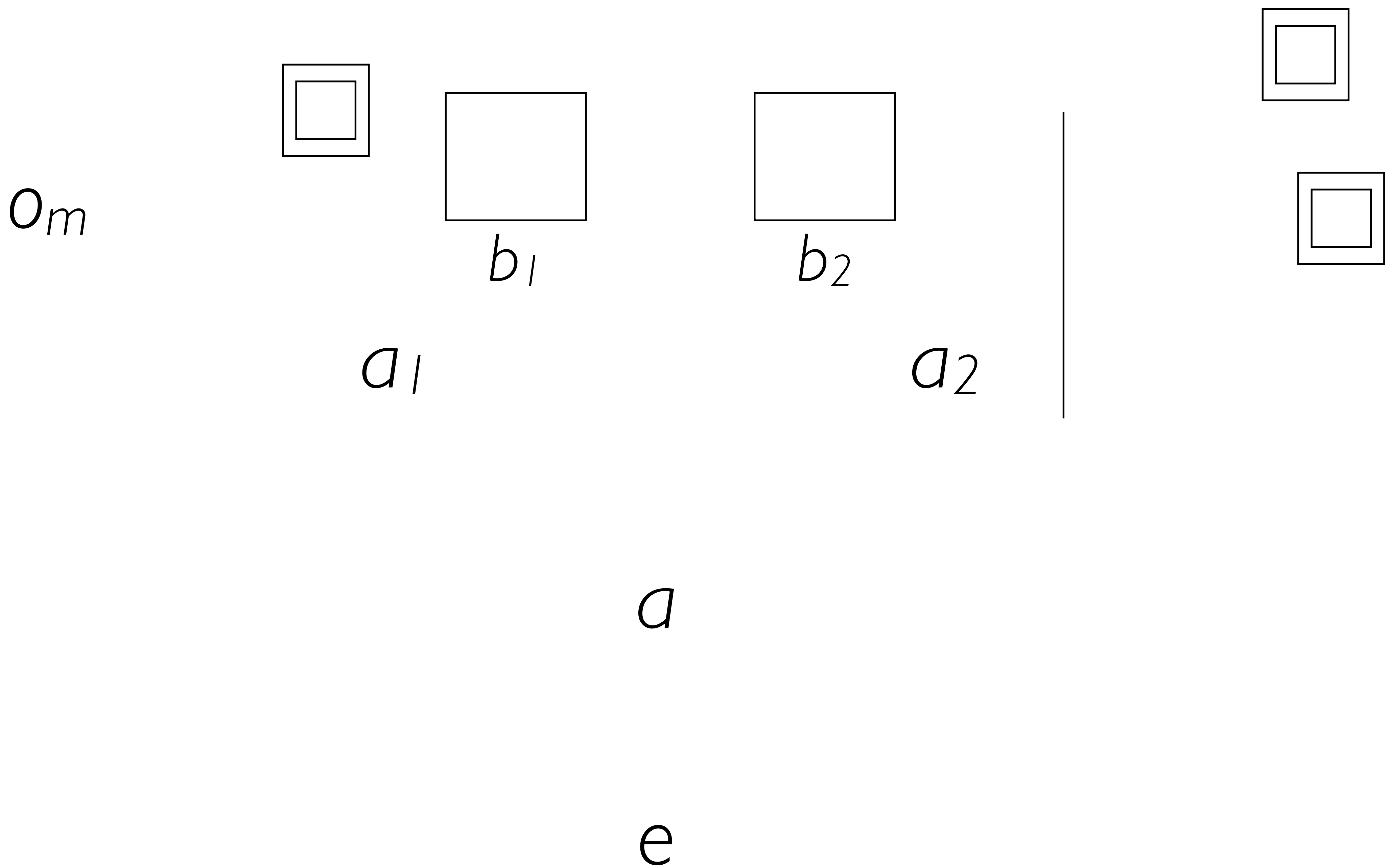
q a formula in modal $\mathscr{L}^n$

Framework for $FBT^1{}_4$

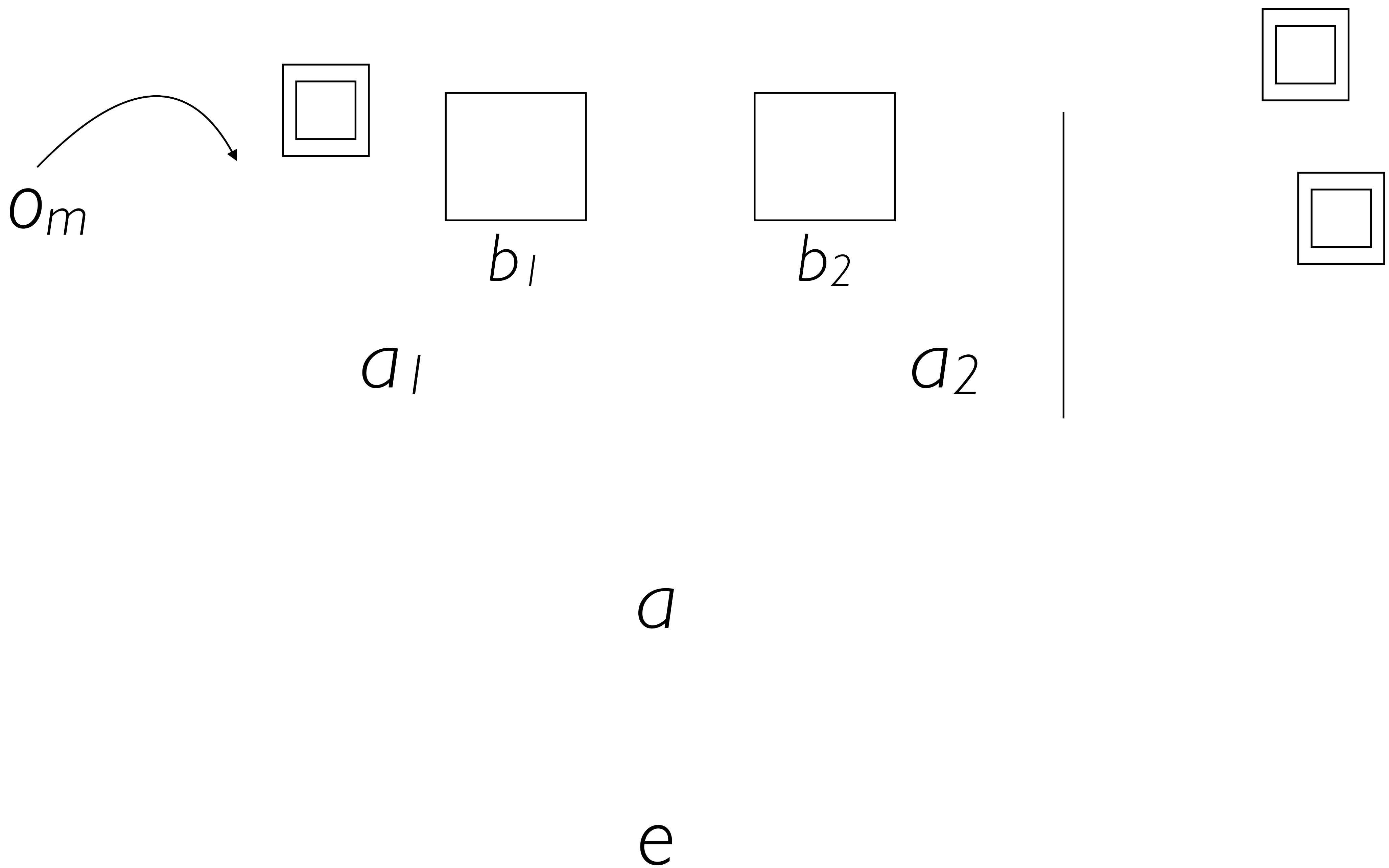


Framework for $FBT^1{}_4$

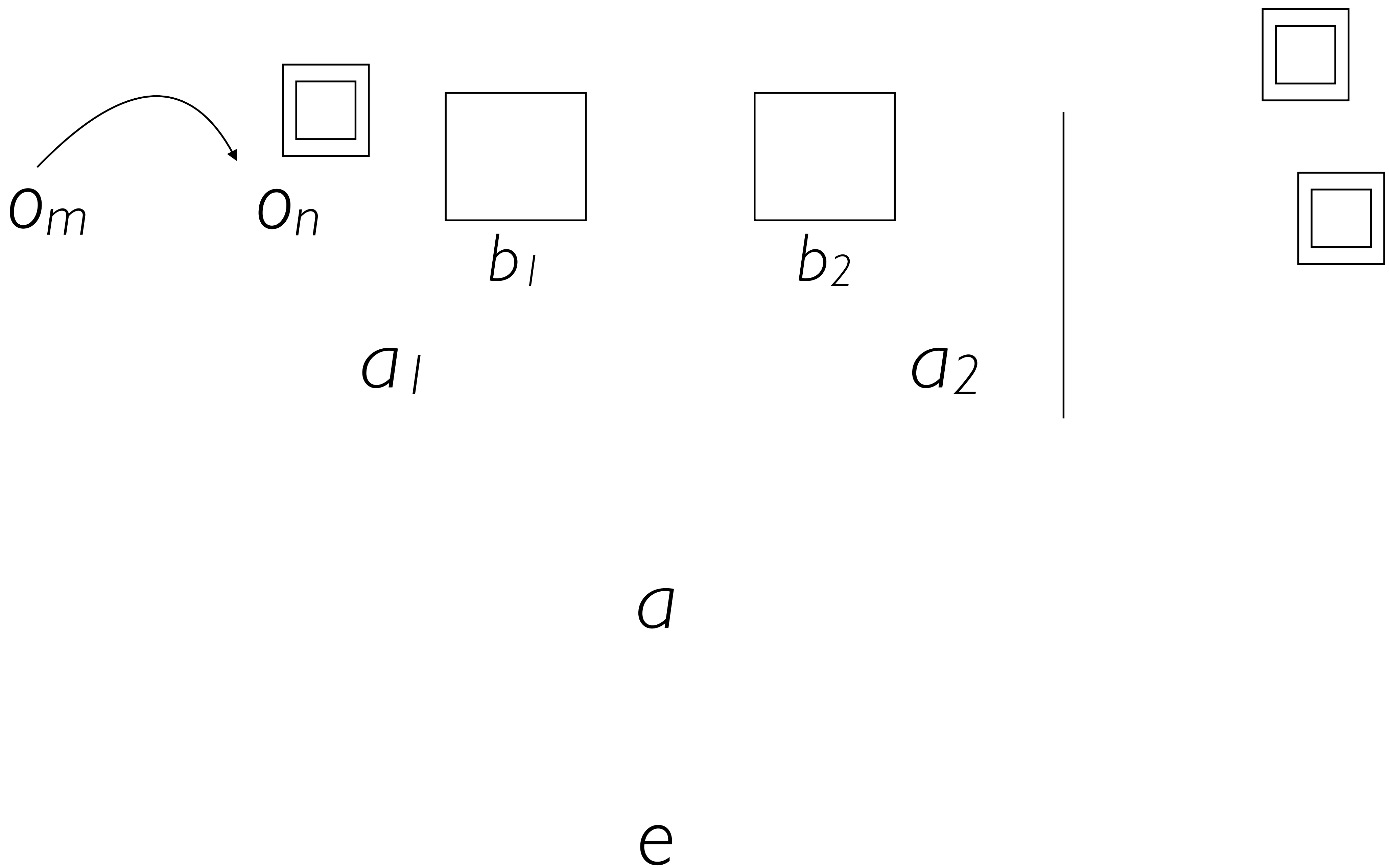
(nine timepoints)



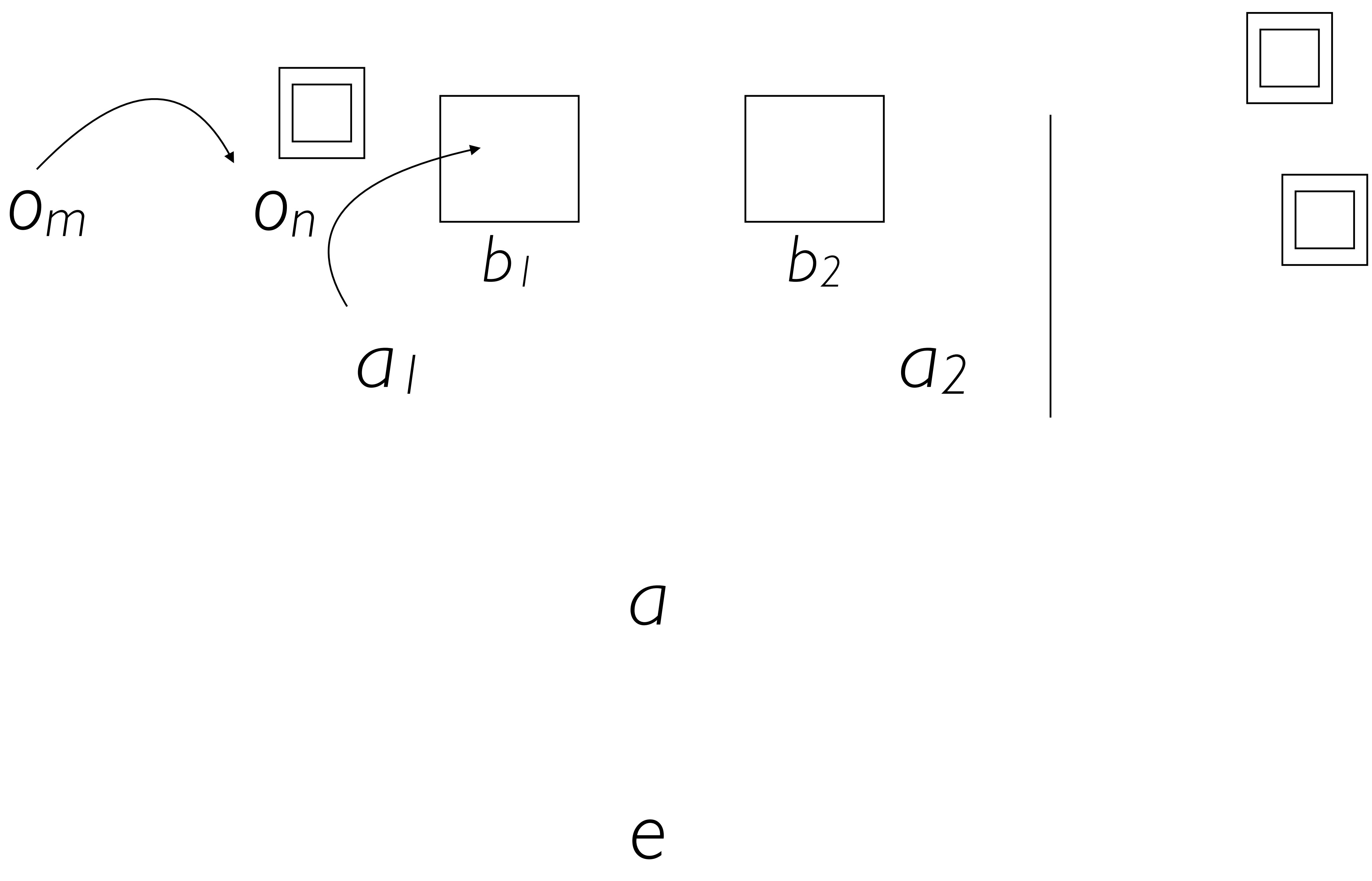
Framework for FBT^{1}_{4}
(nine timepoints)



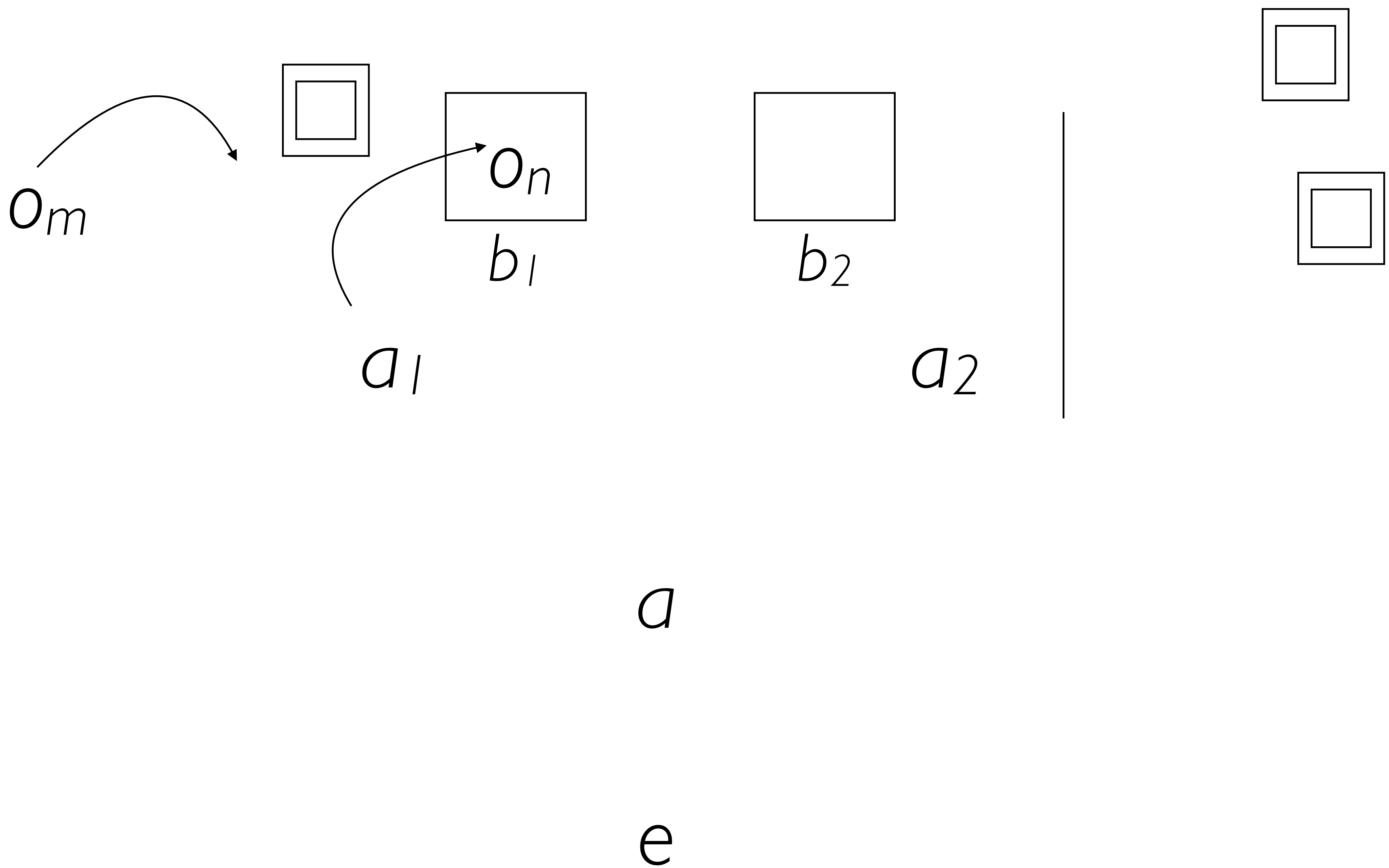
Framework for $FBT^1{}_4$
(nine timepoints)



Framework for $FBT^1{}_4$
(nine timepoints)

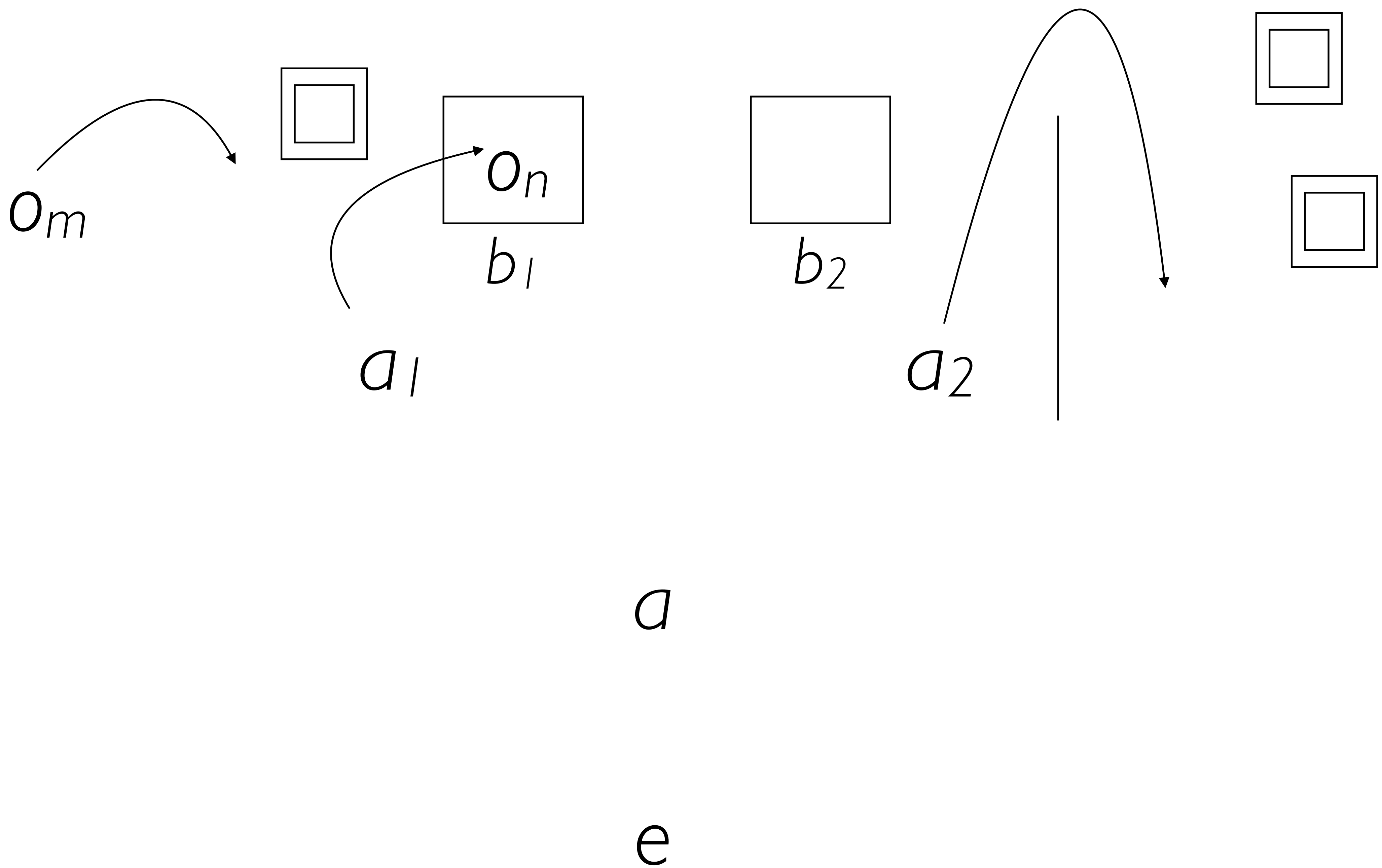


Framework for FBT$^1{}_4$
(nine timepoints)

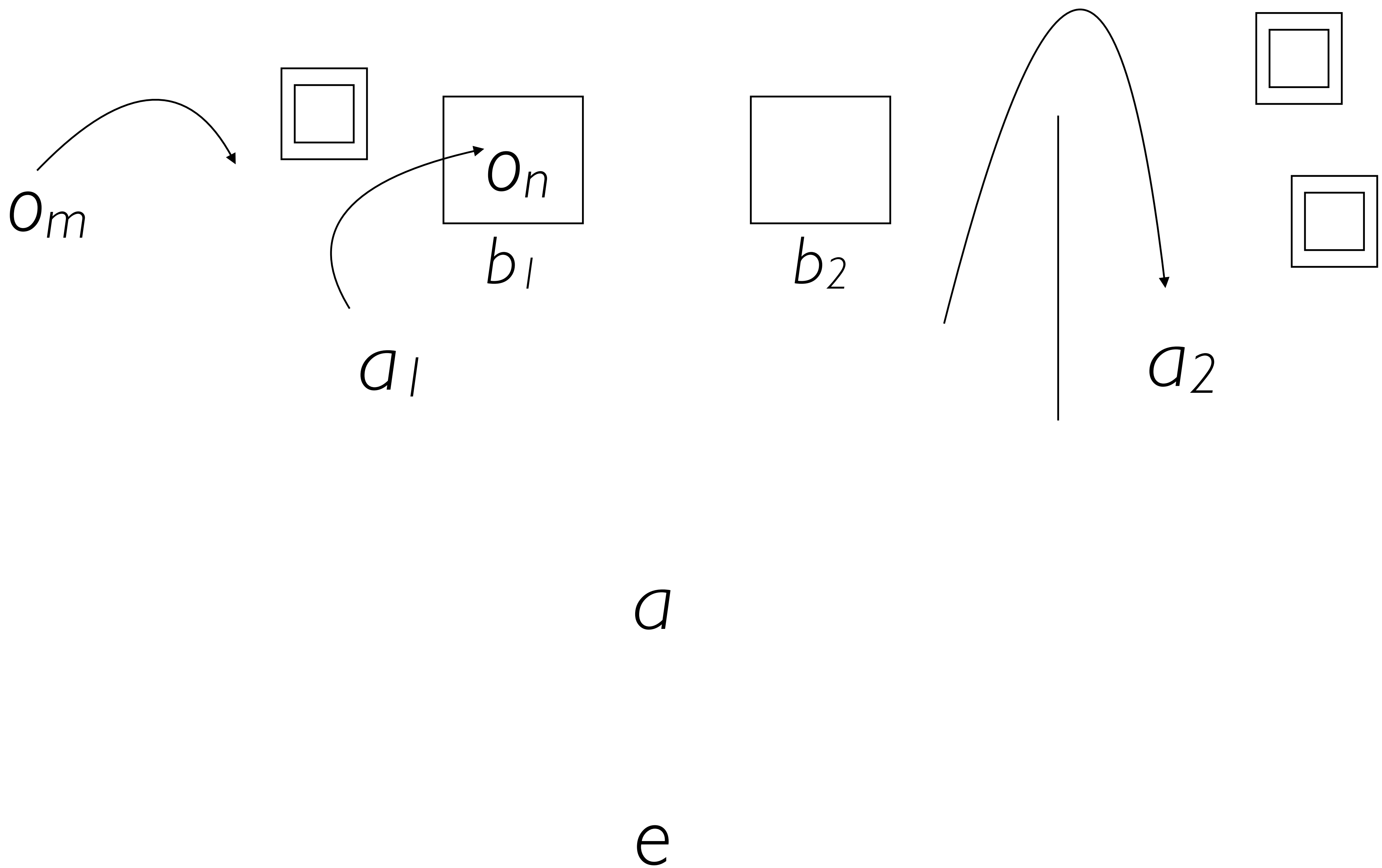


Framework for FBT^1_4

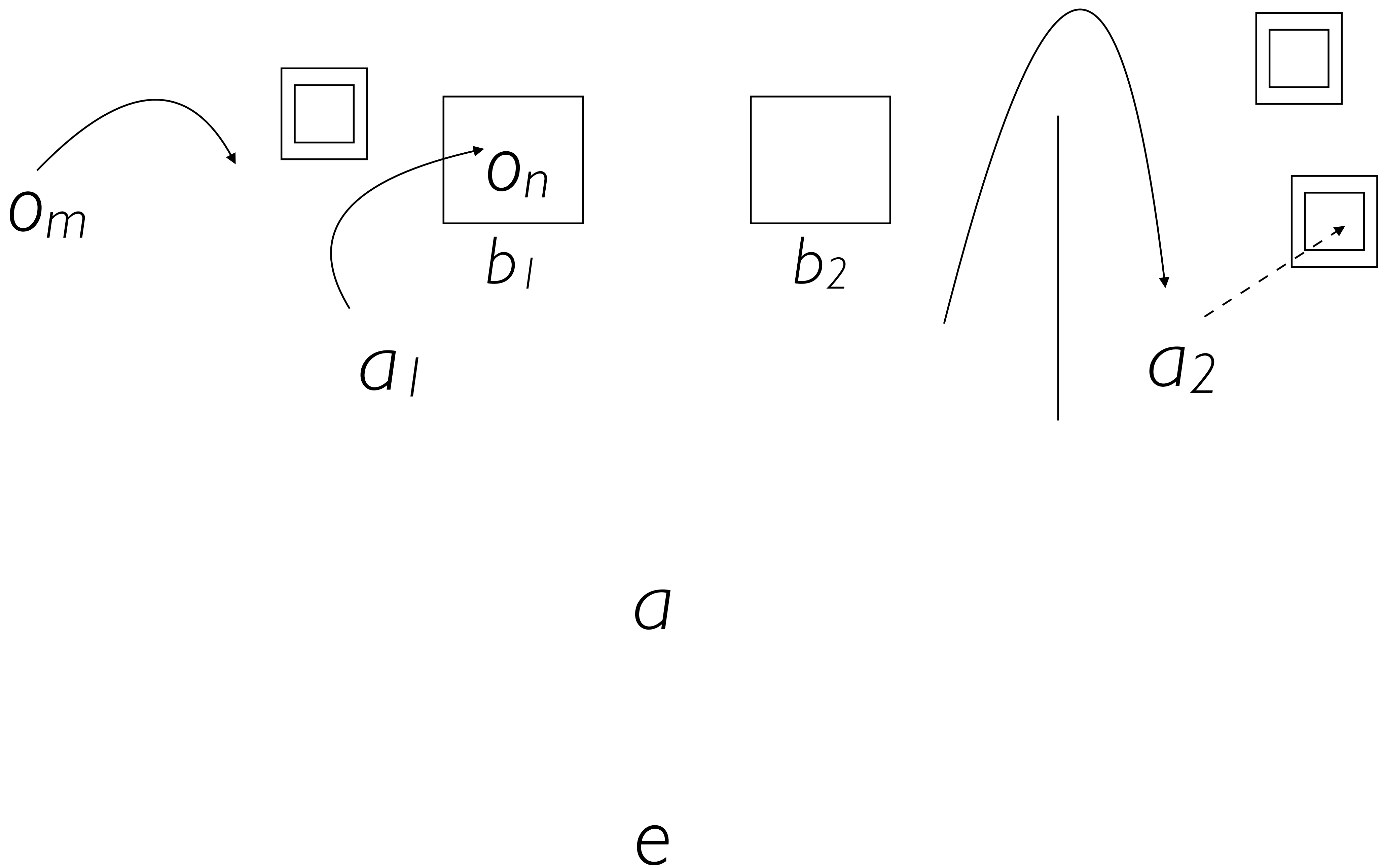
(nine timepoints)



Framework for $FBT^1{}_4$
(nine timepoints)

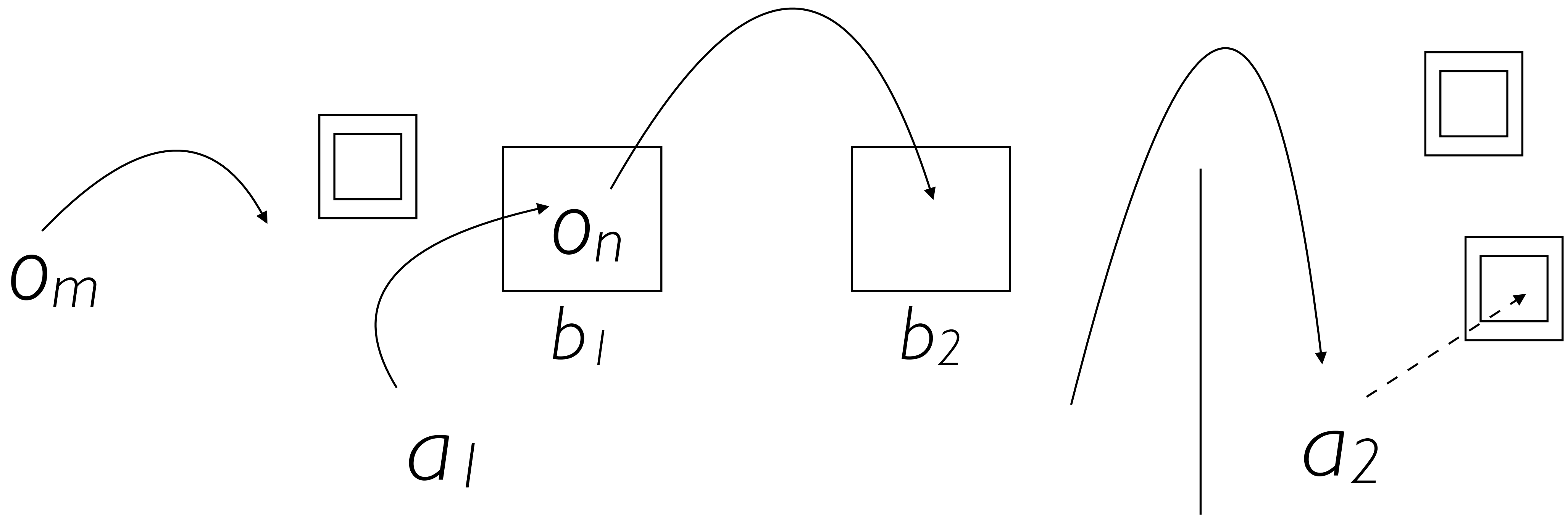


Framework for FBT$^{1}_{4}$
(nine timepoints)



Framework for FBT^1_4

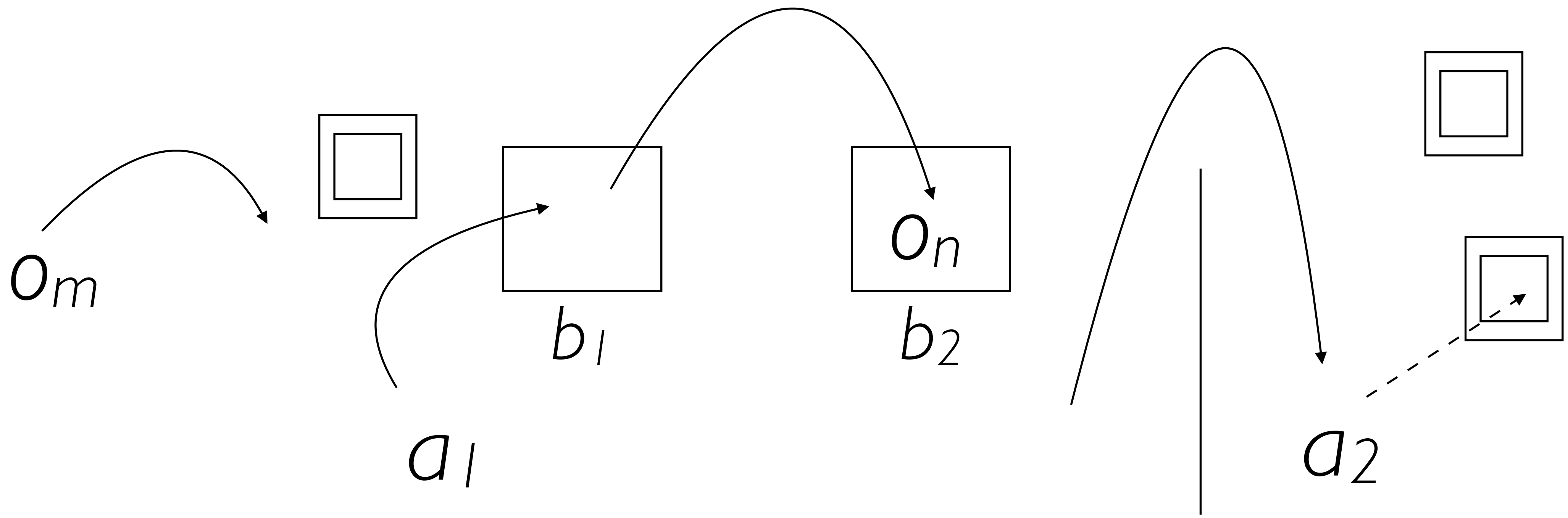
(nine timepoints)



a

e

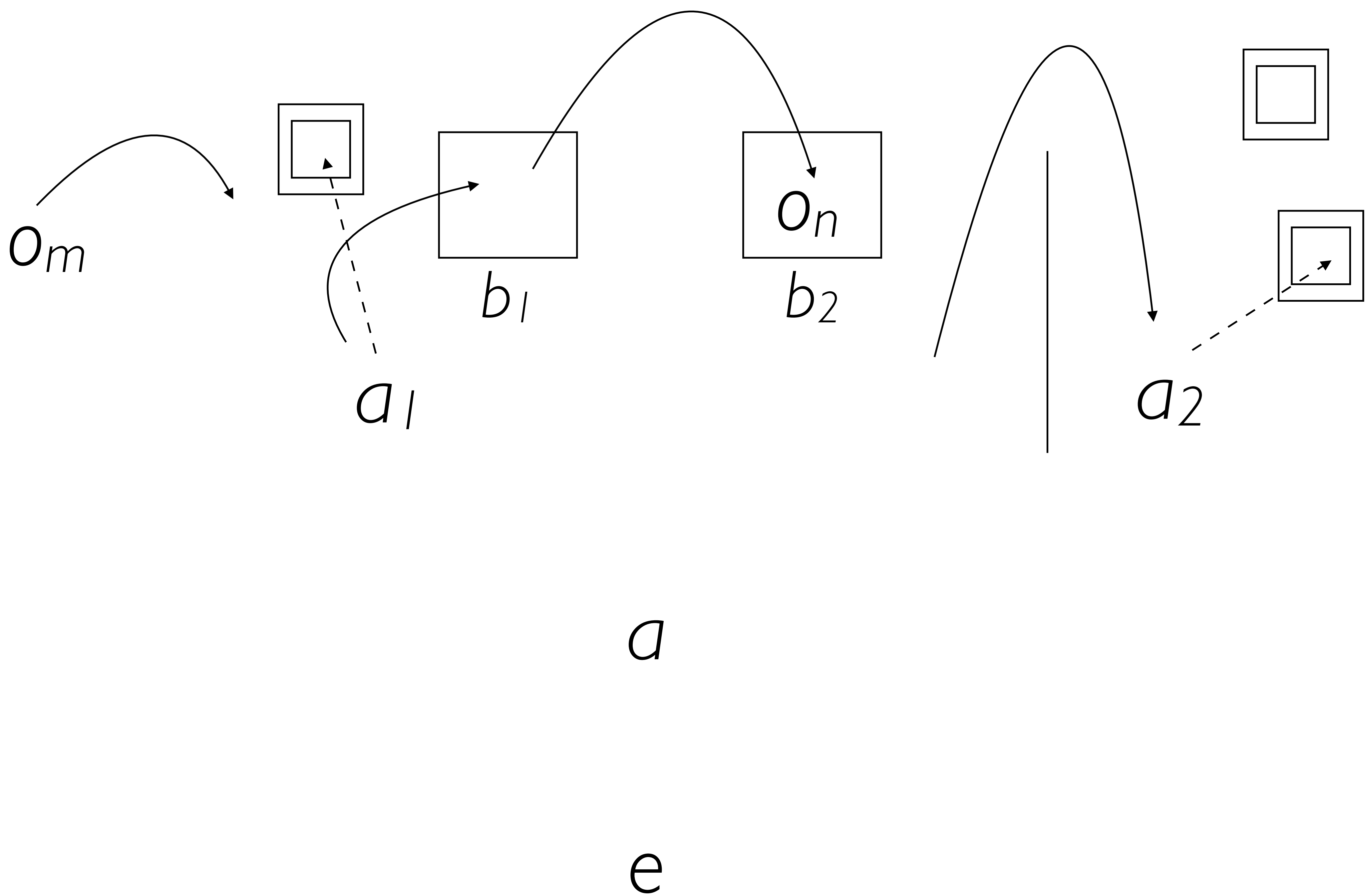
Framework for FBT$^1{}_4$
(nine timepoints)



a

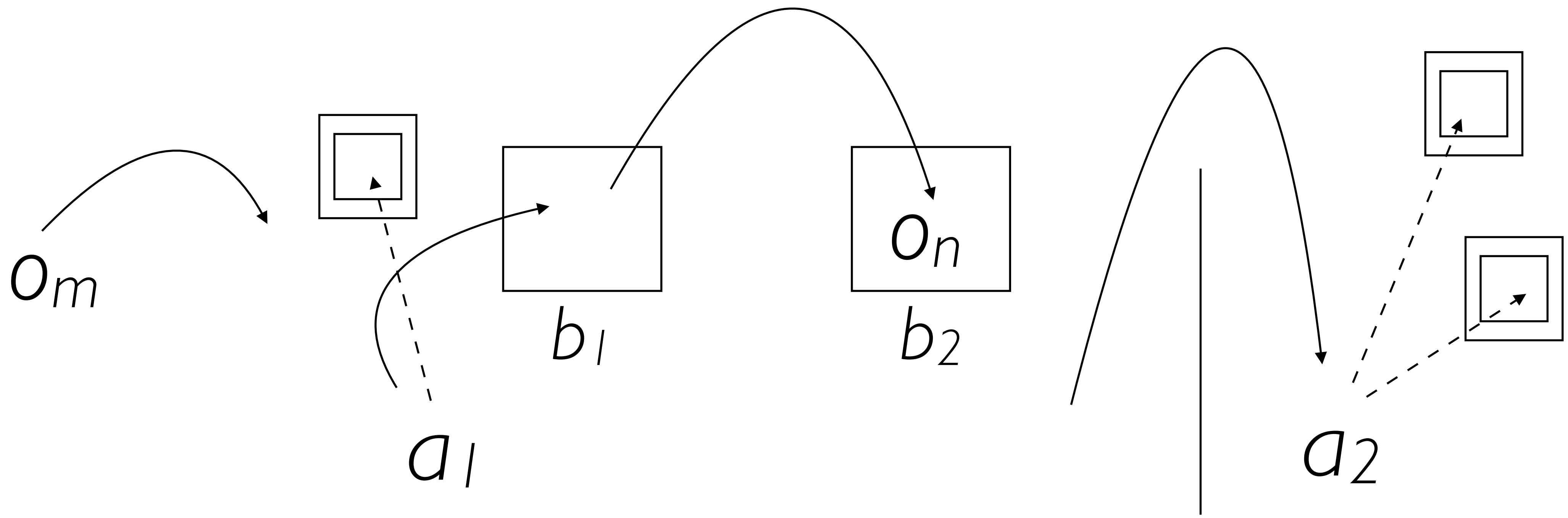
e

Framework for FBT^1_4
(nine timepoints)



Framework for FBT^{1}_{4}

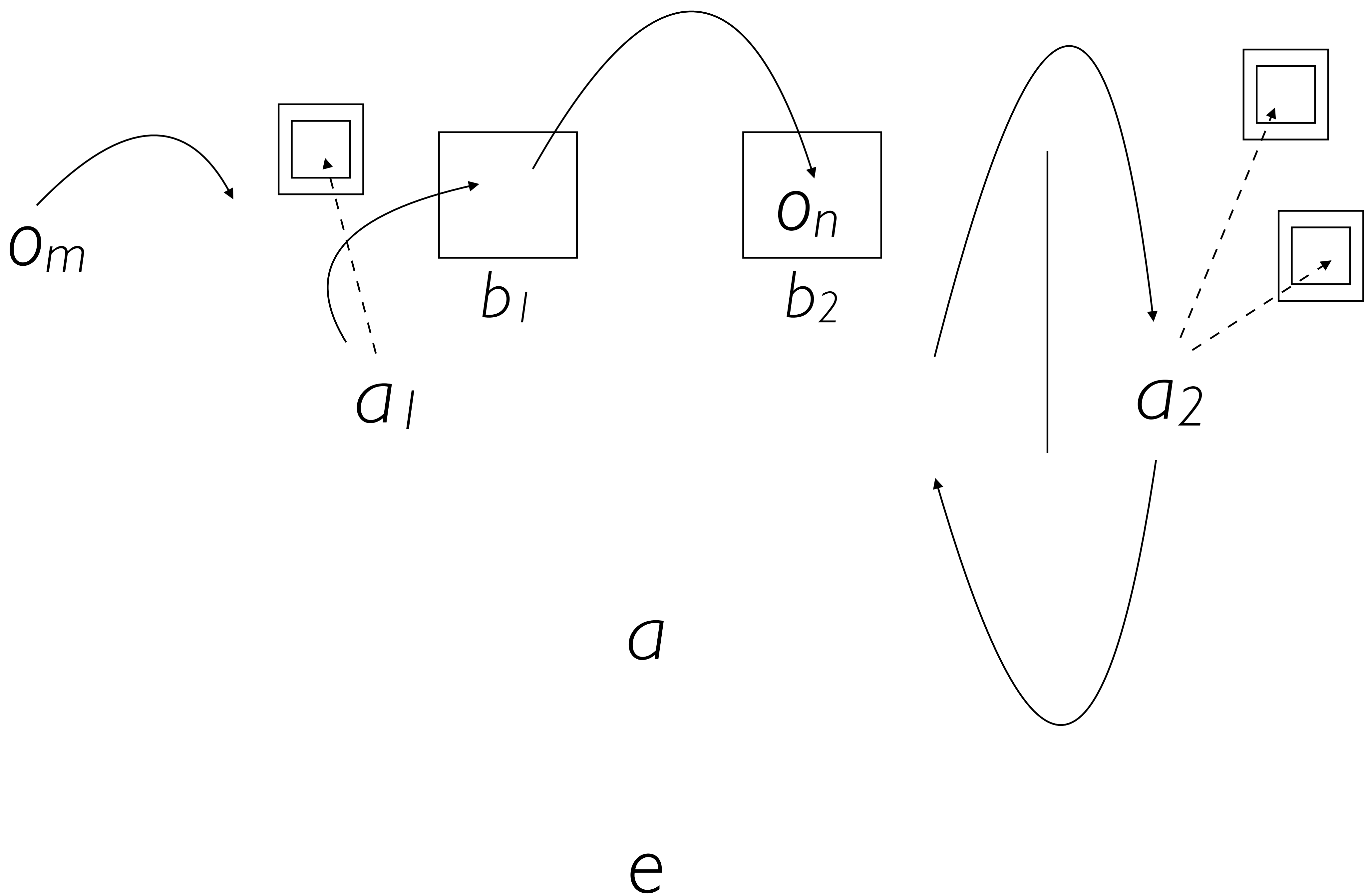
(nine timepoints)



a

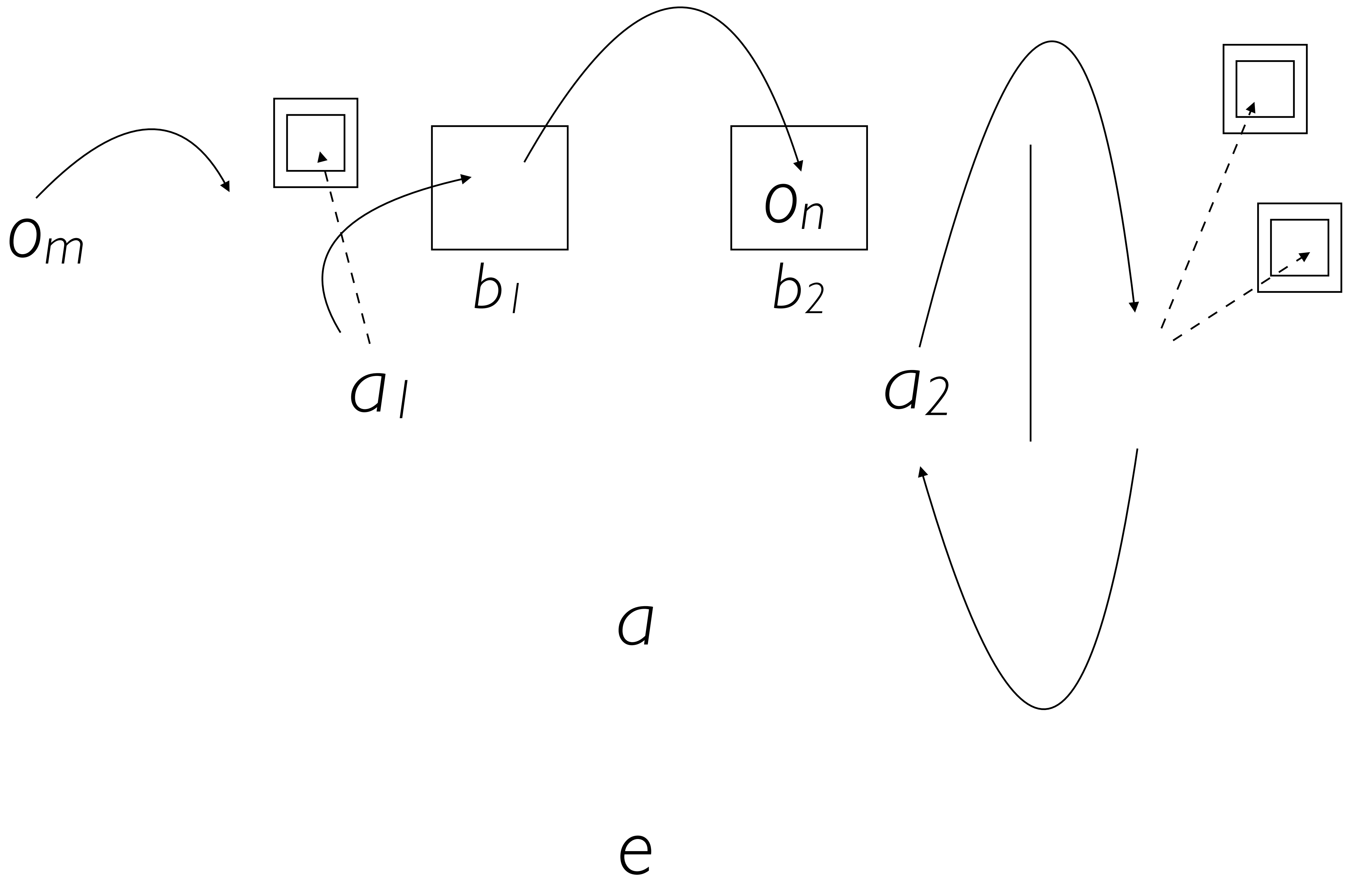
e

Framework for FBT$^1{}_4$
(nine timepoints)

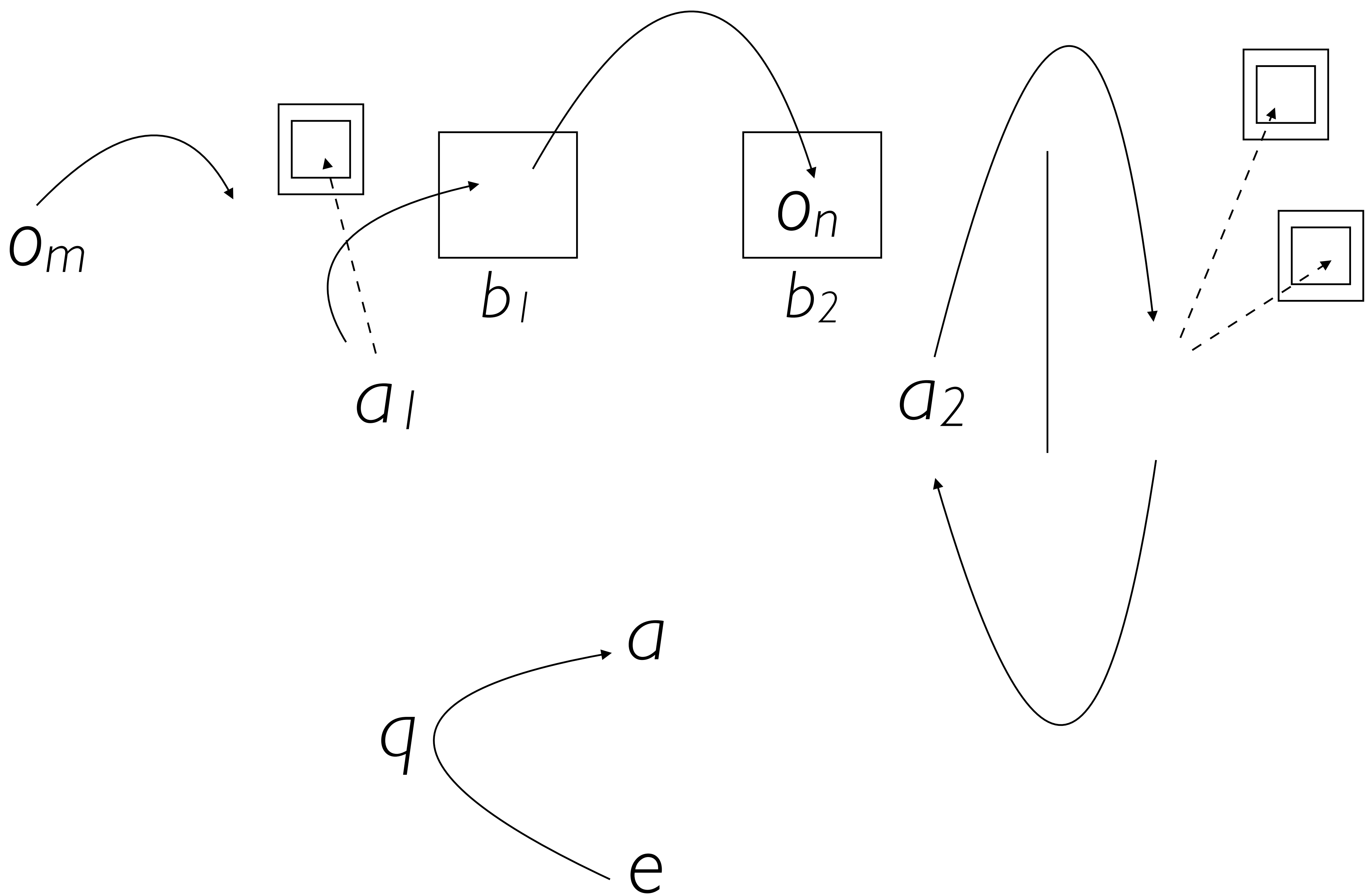


Framework for FBT^{1}_{4}

(nine timepoints)

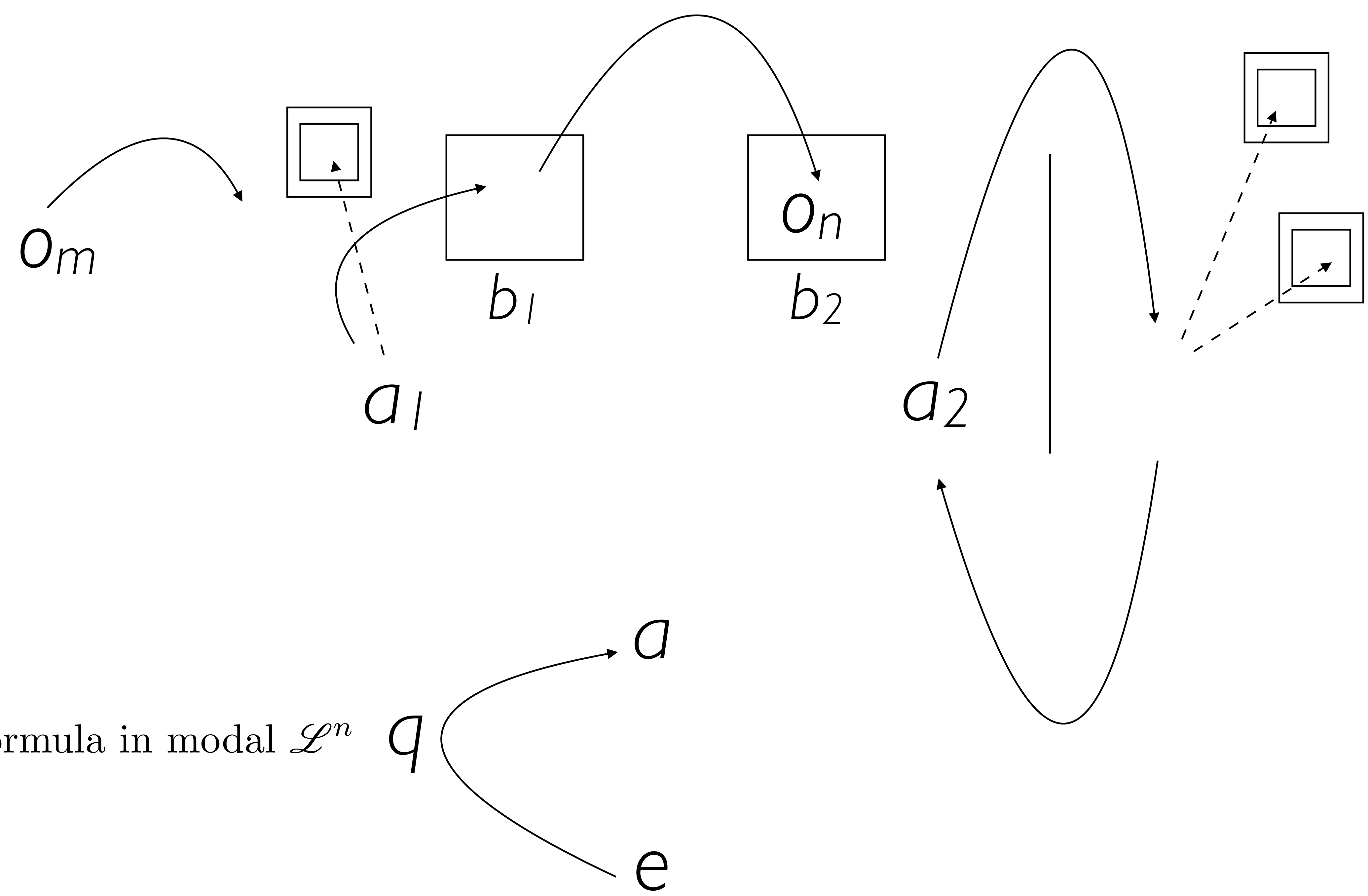


Framework for FBT$^1{}_4$
(nine timepoints)



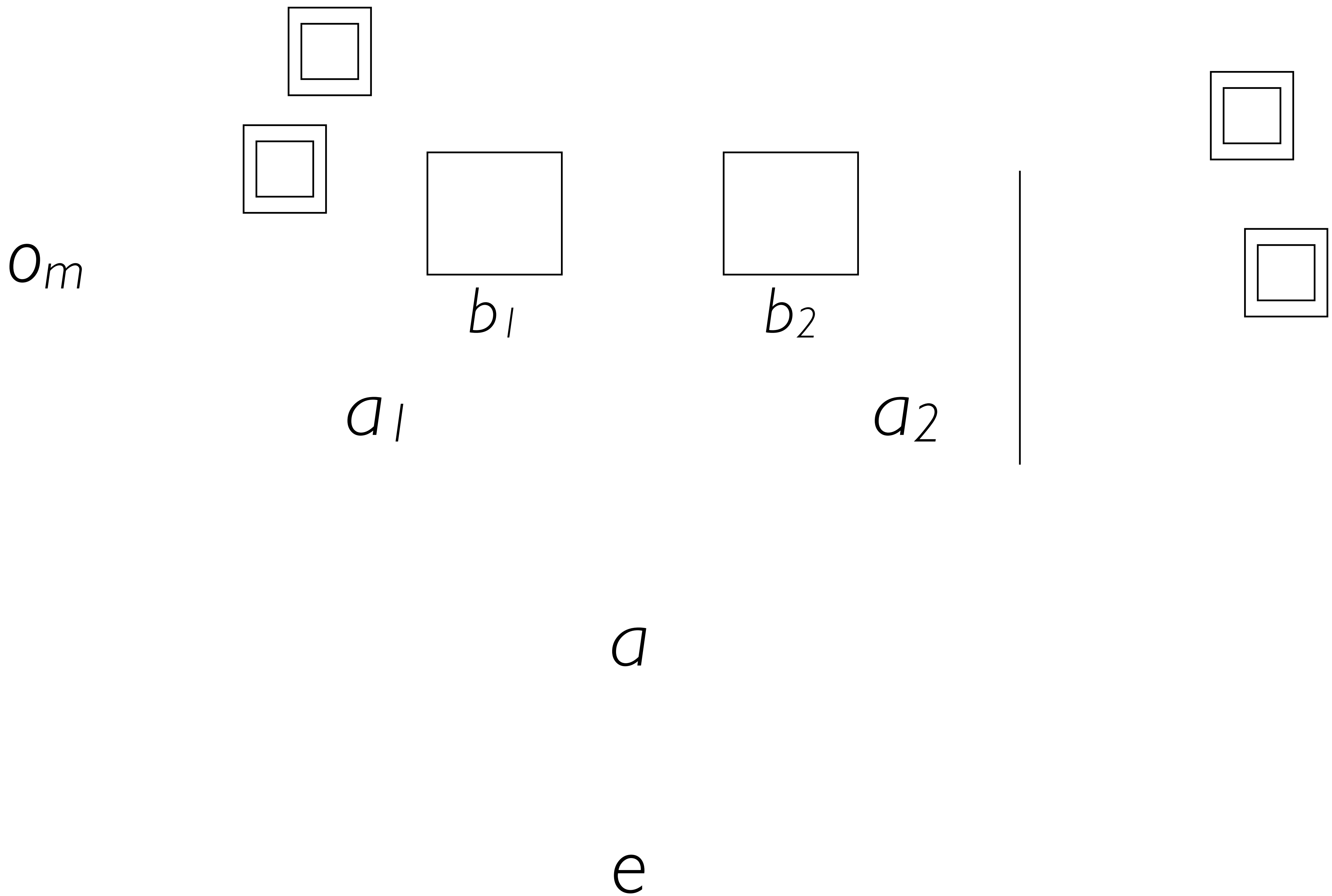
Framework for FBT^1_4

(nine timepoints)



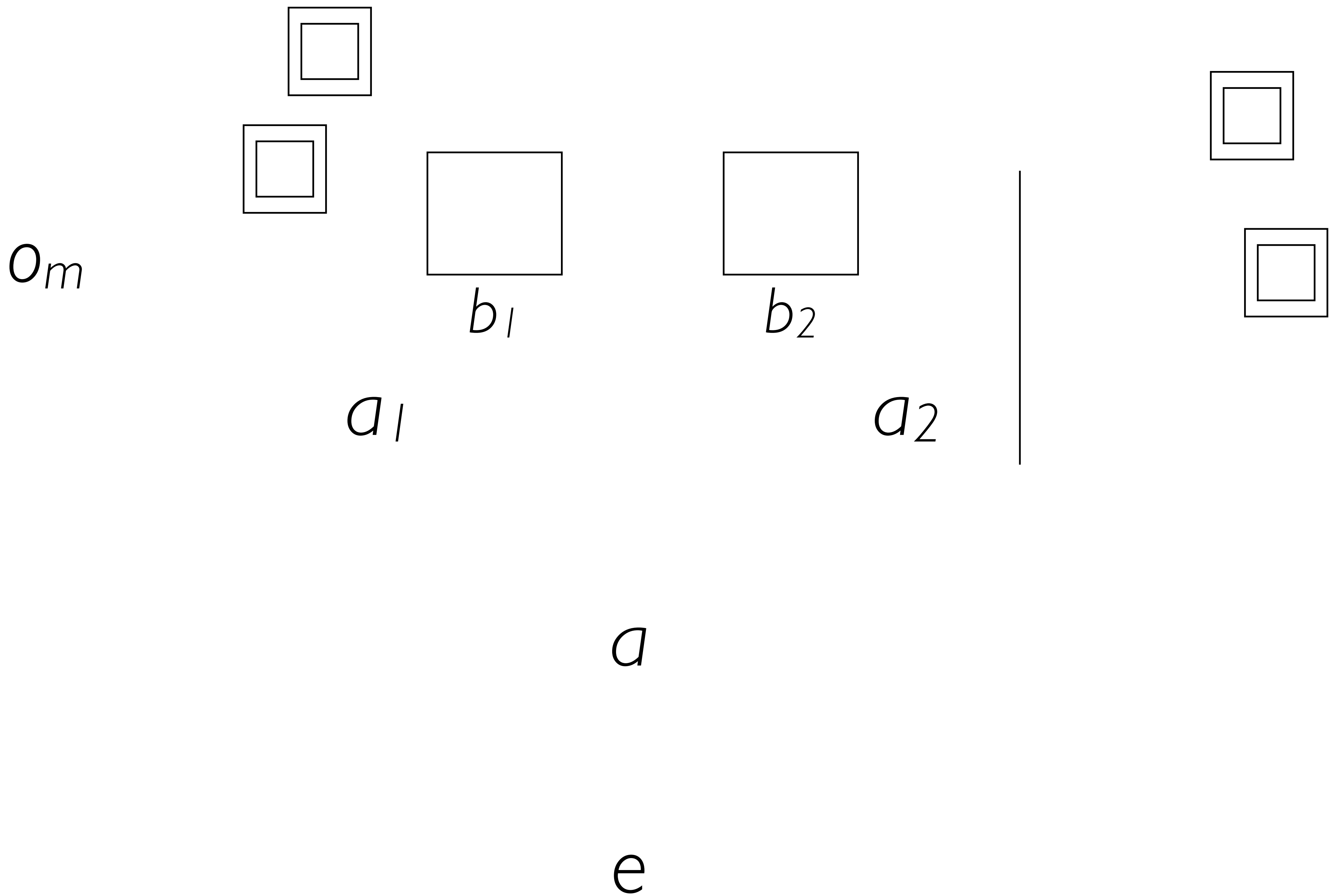
q a formula in modal $\mathscr{L}^n$

Framework for FBT$^1_5$



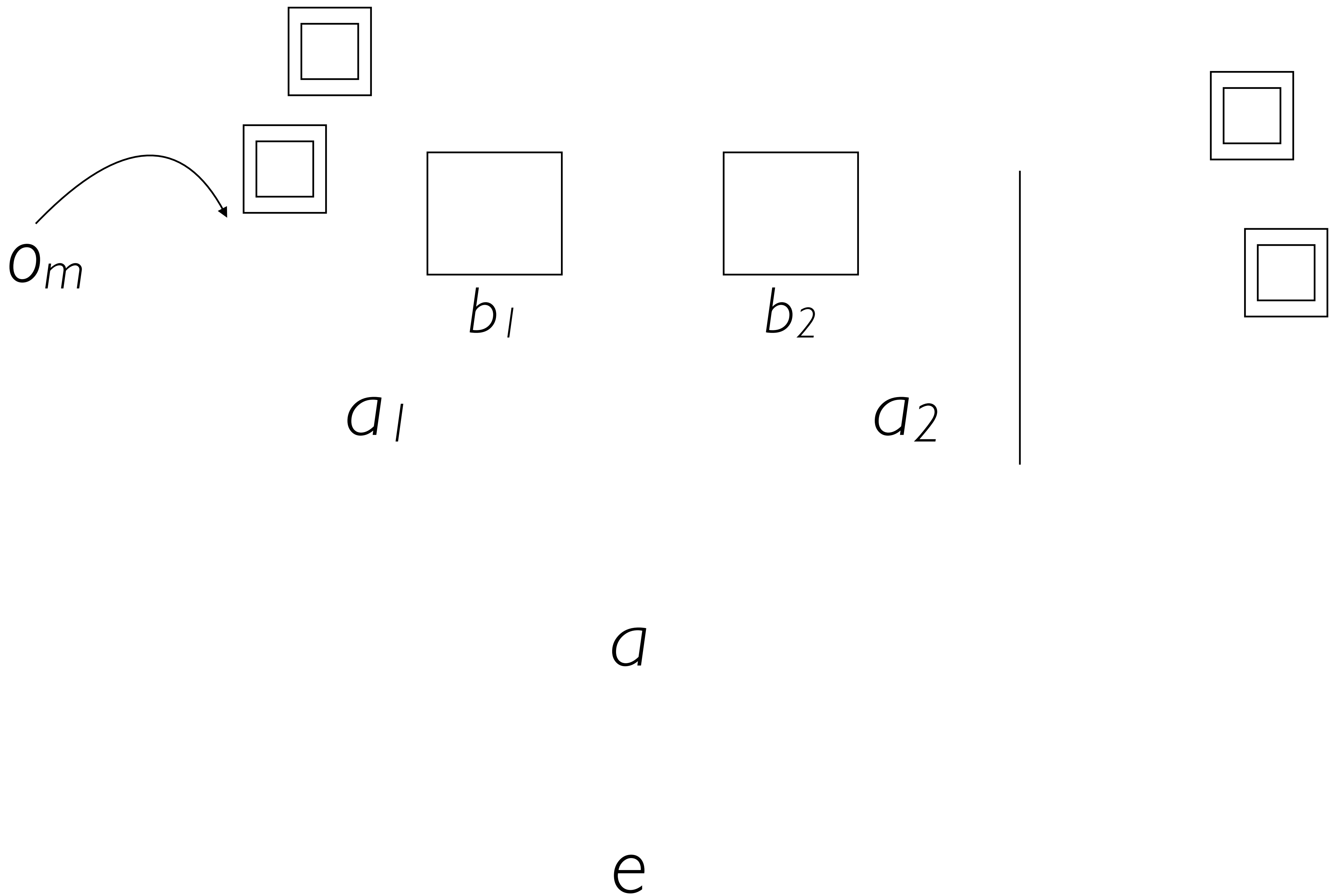
Framework for FBT^1_5

(ten timepoints)



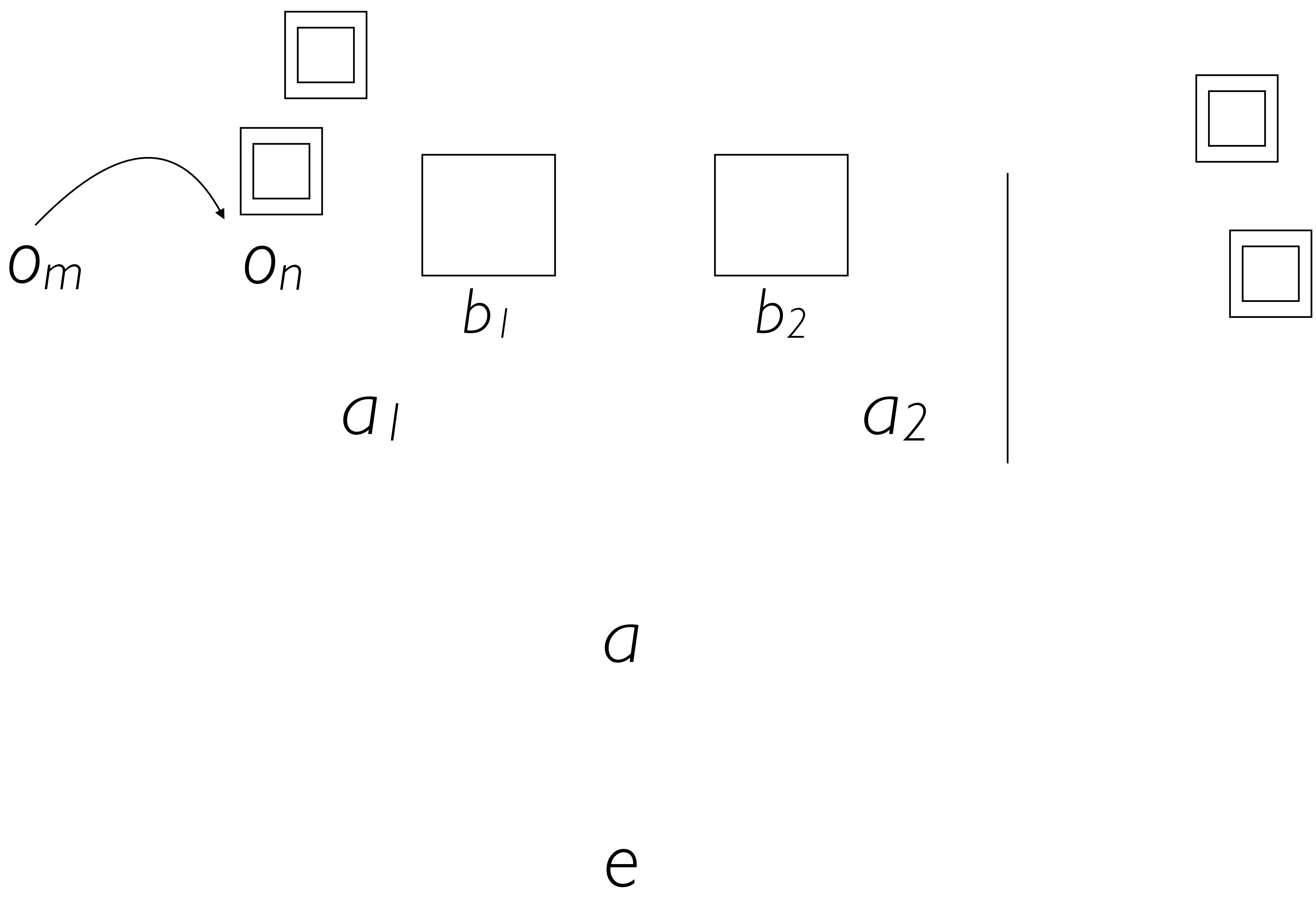
Framework for $FBT^1{}_5$

(ten timepoints)

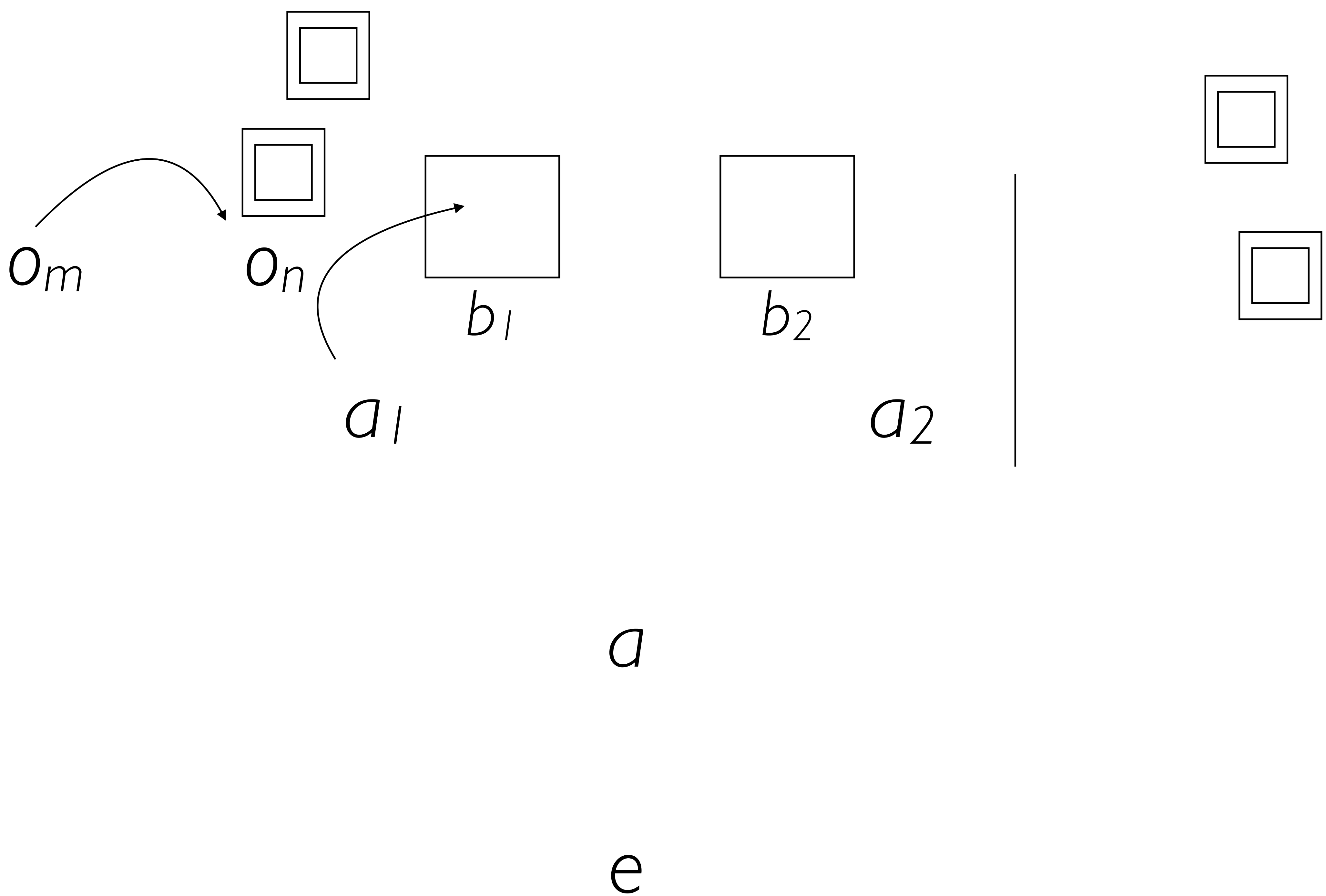


Framework for FBT^1_5

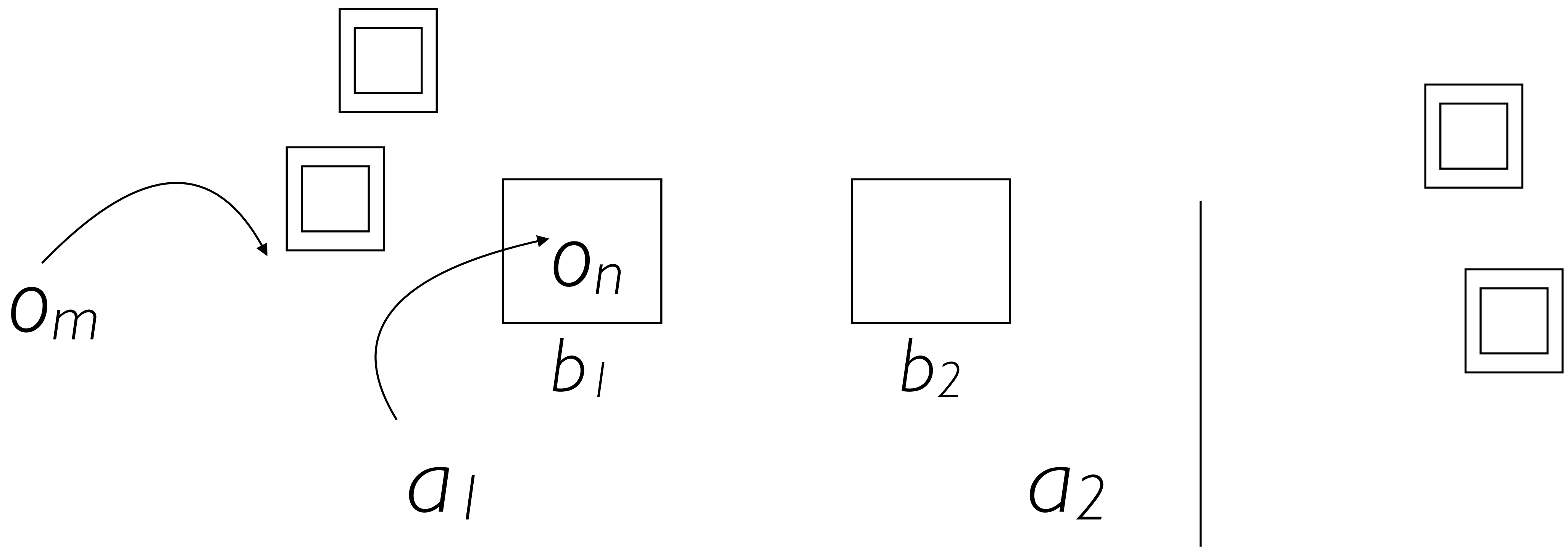
(ten timepoints)



Framework for FBT^1_5
(ten timepoints)



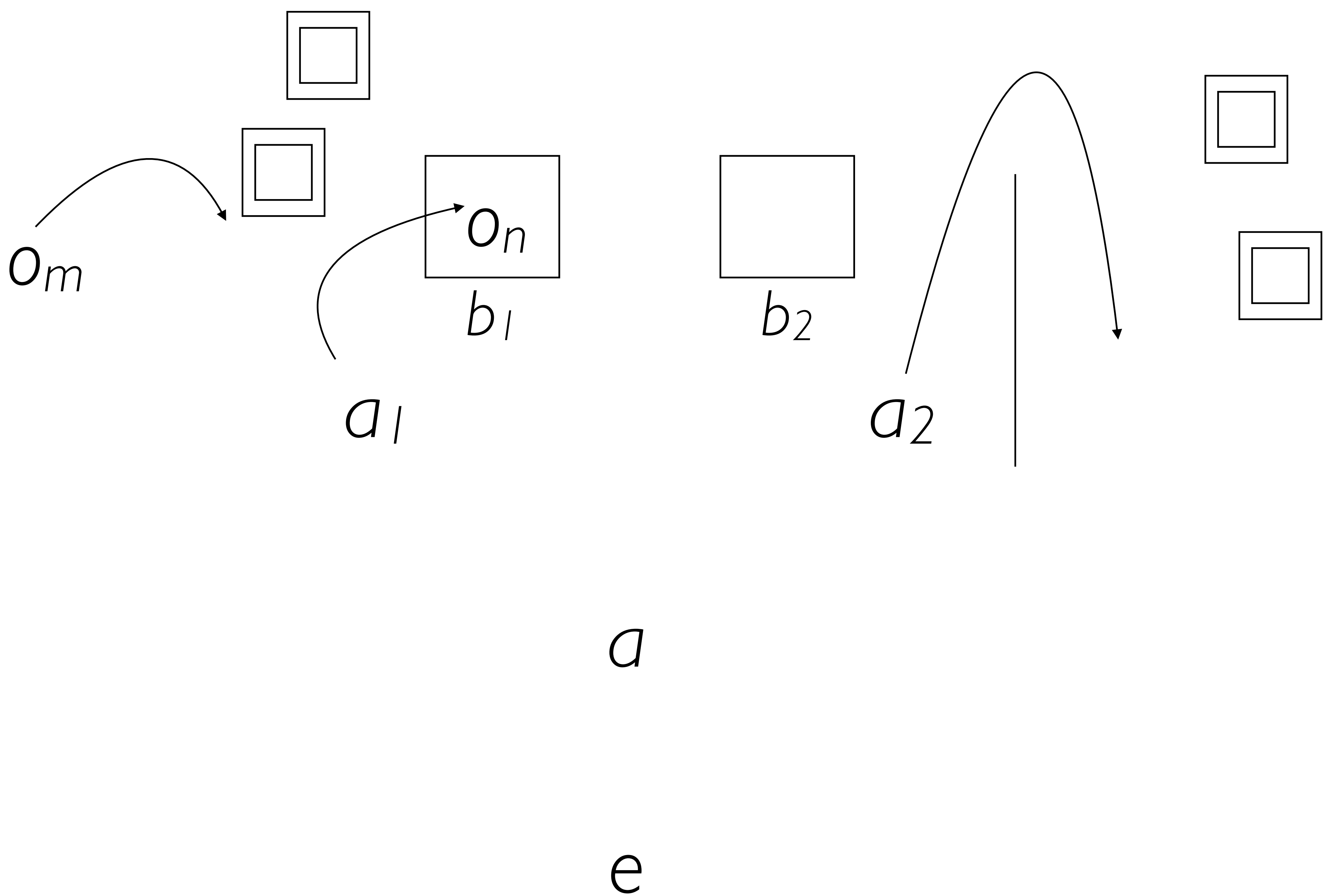
Framework for $FBT^1{}_5$
(ten timepoints)



a

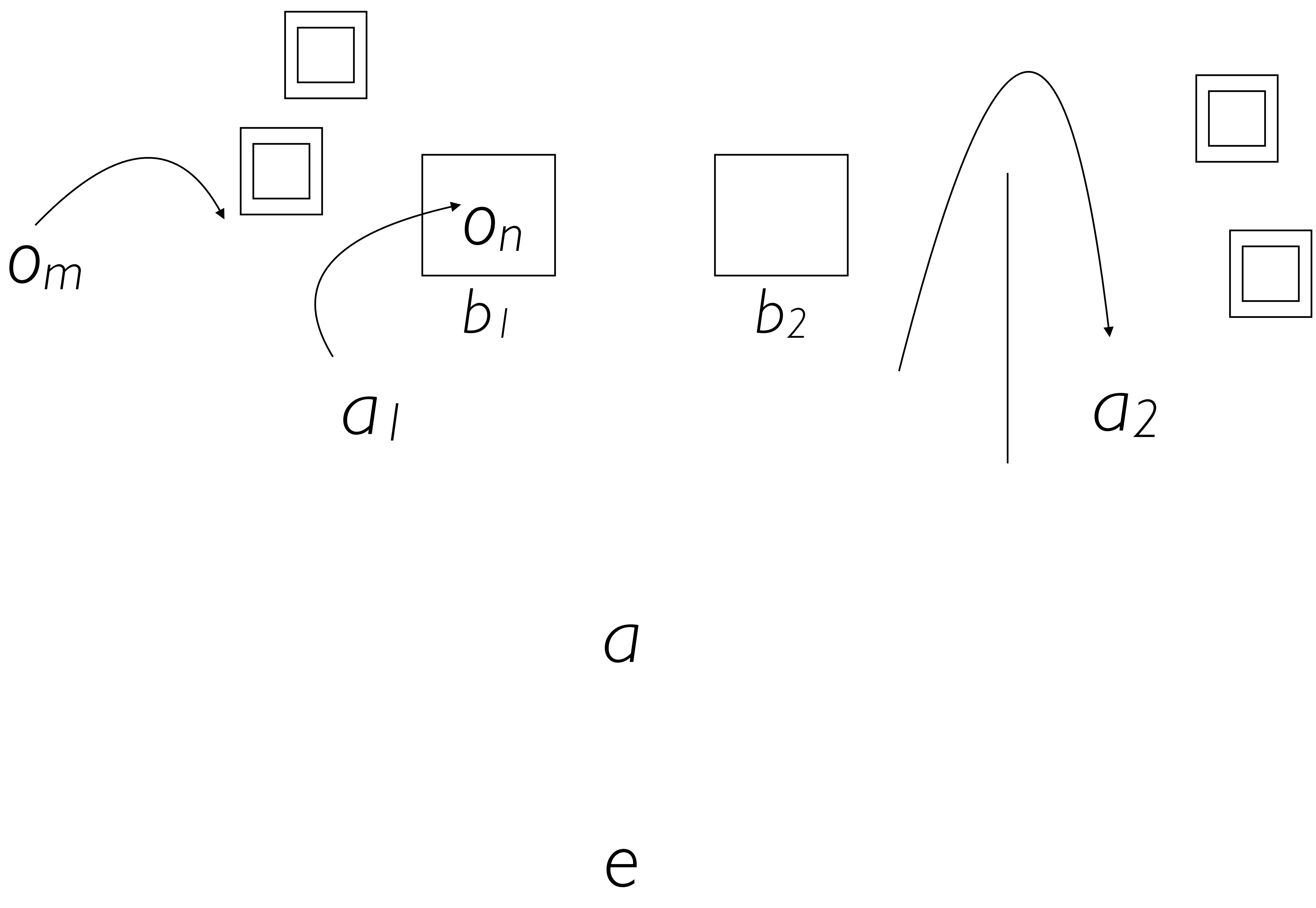
e

Framework for FBT$^1{}_5$
(ten timepoints)



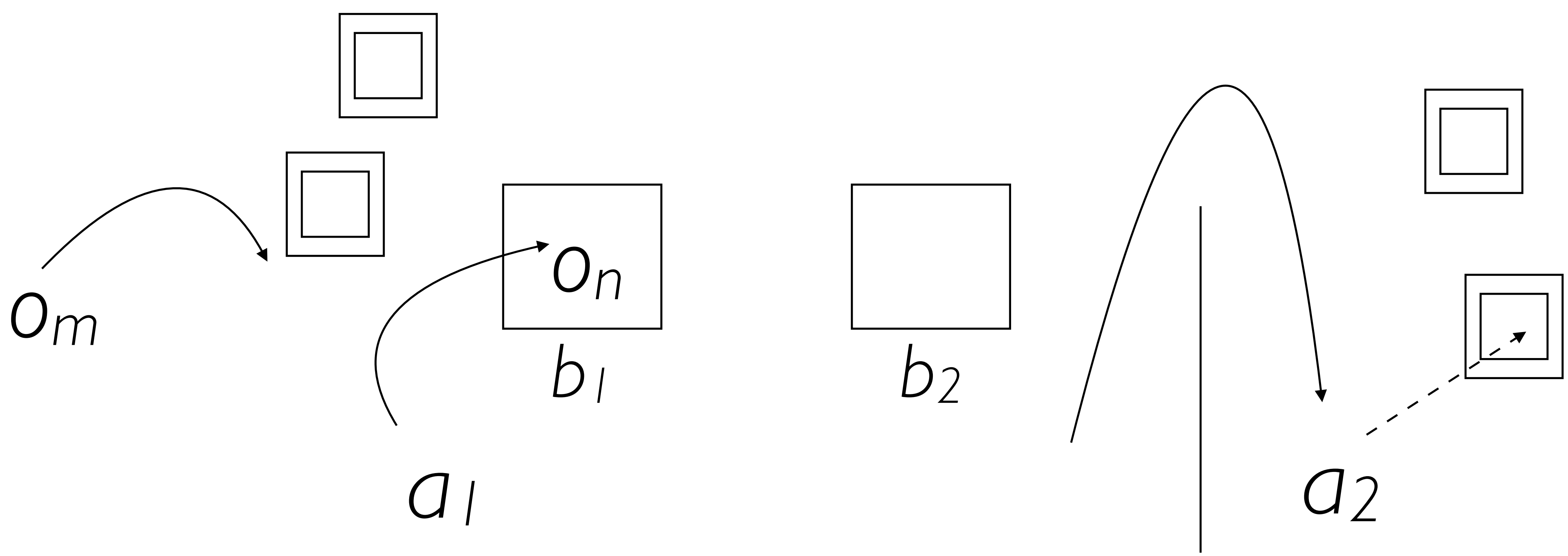
Framework for $FBT^1{}_5$

(ten timepoints)



Framework for FBT^{1}_{5}

(ten timepoints)

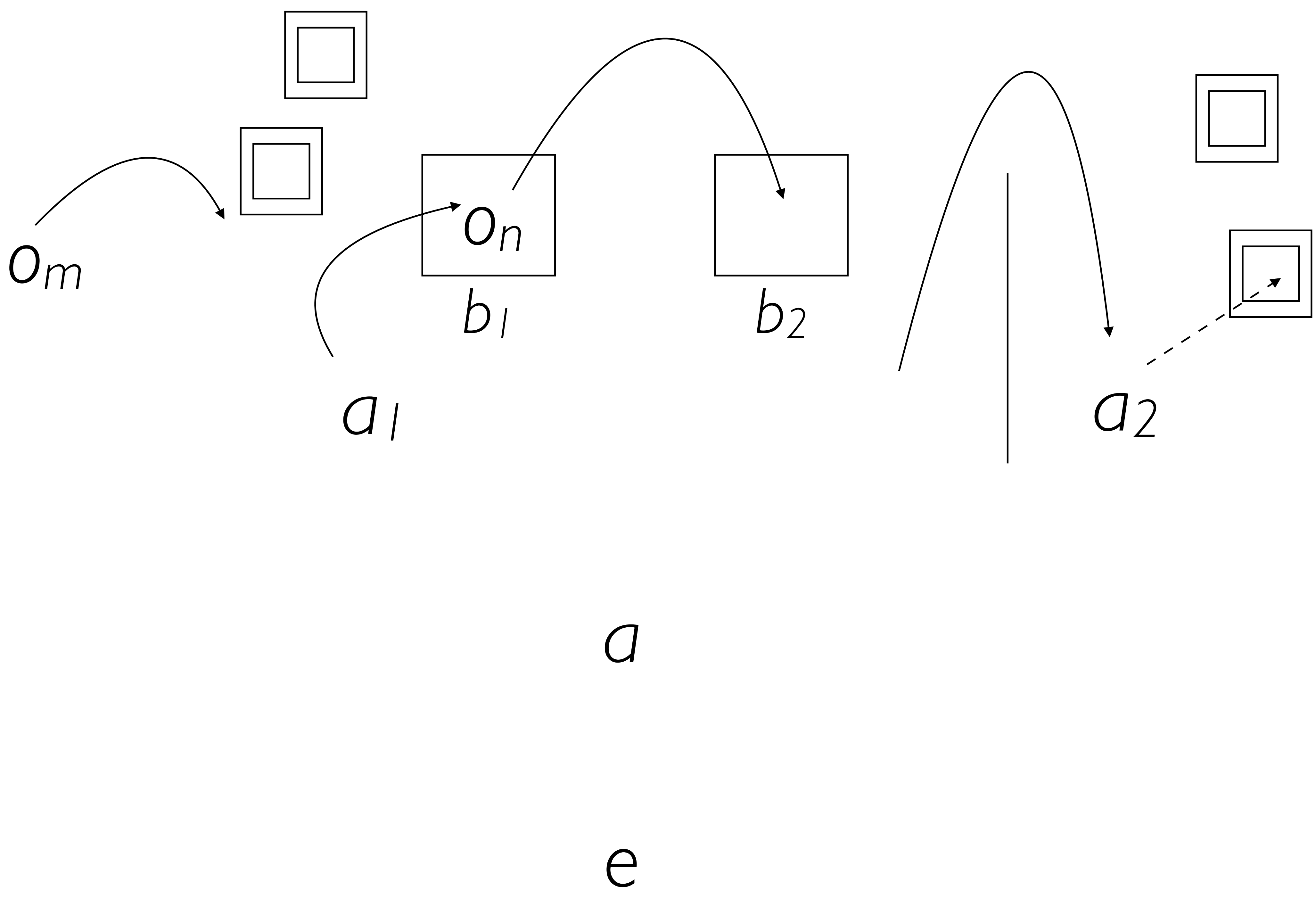


a

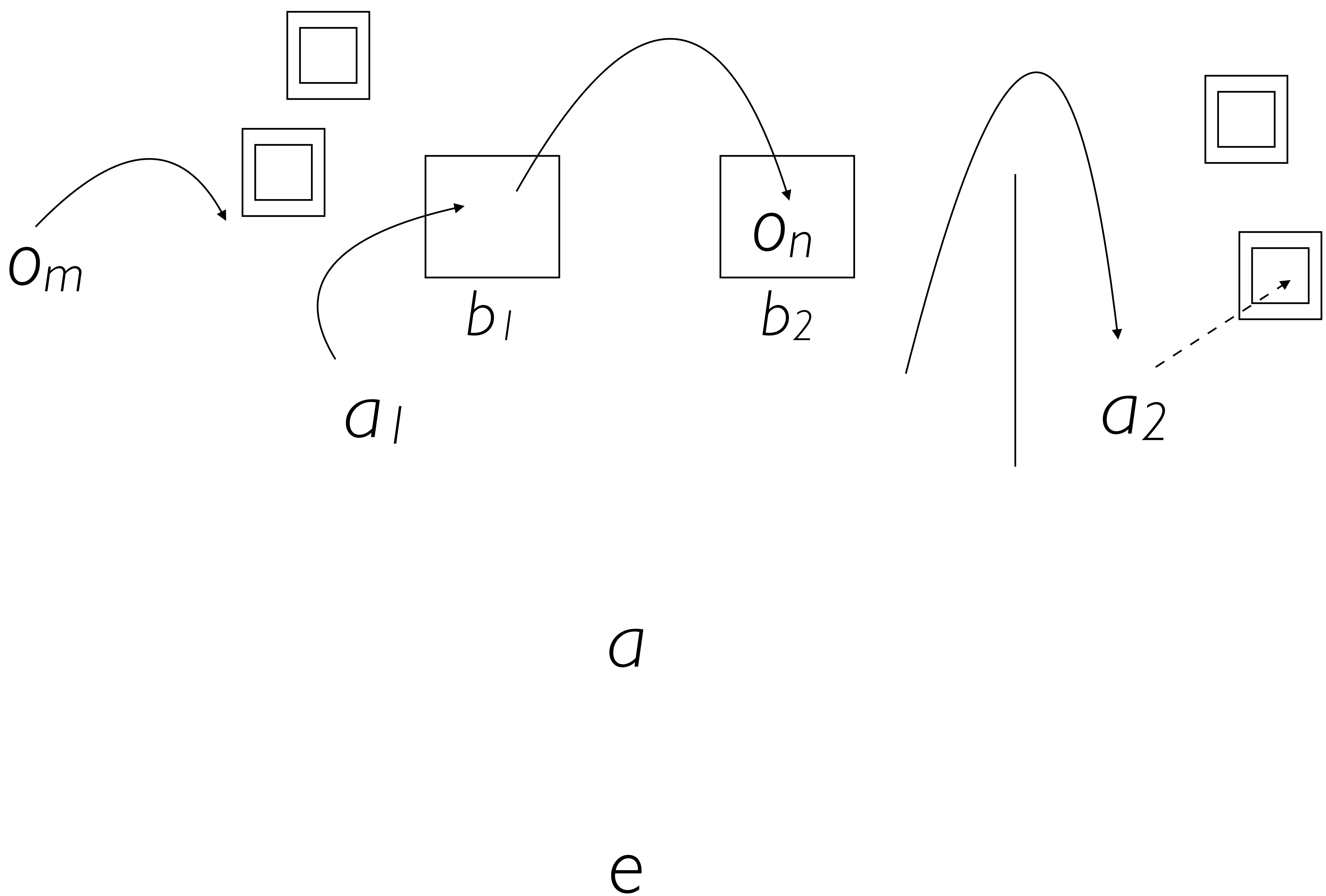
e

Framework for $FBT^1{}_5$

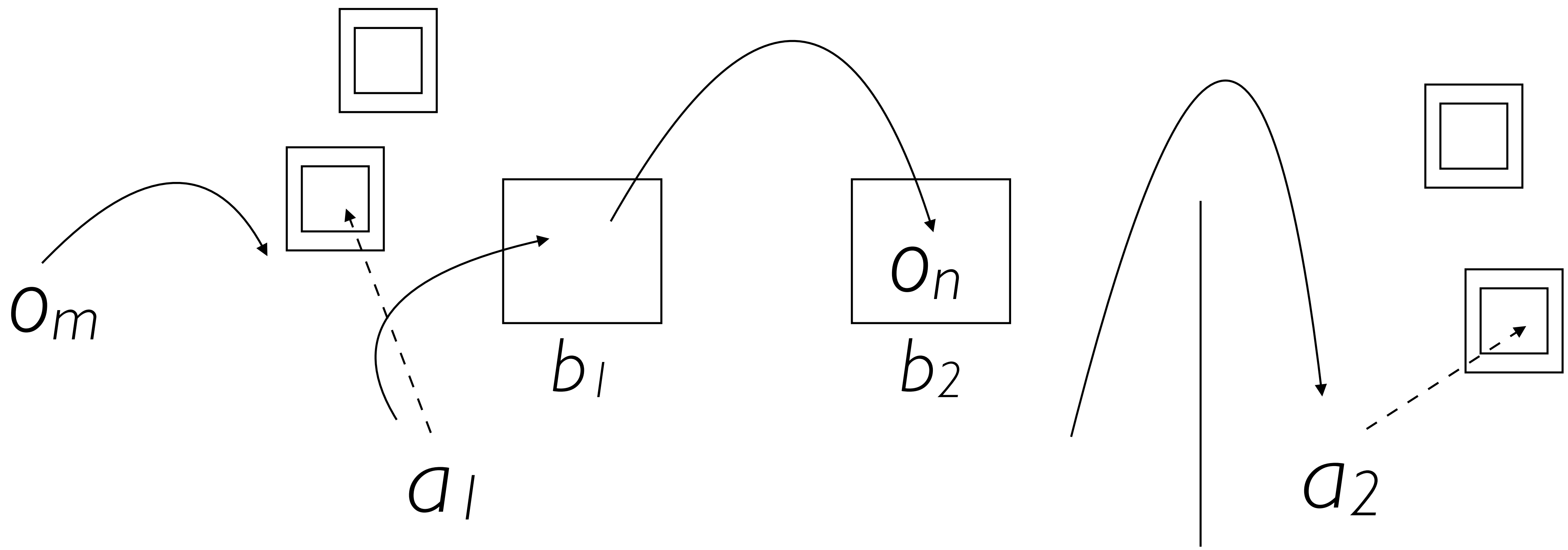
(ten timepoints)



Framework for $FBT^1{}_5$
(ten timepoints)



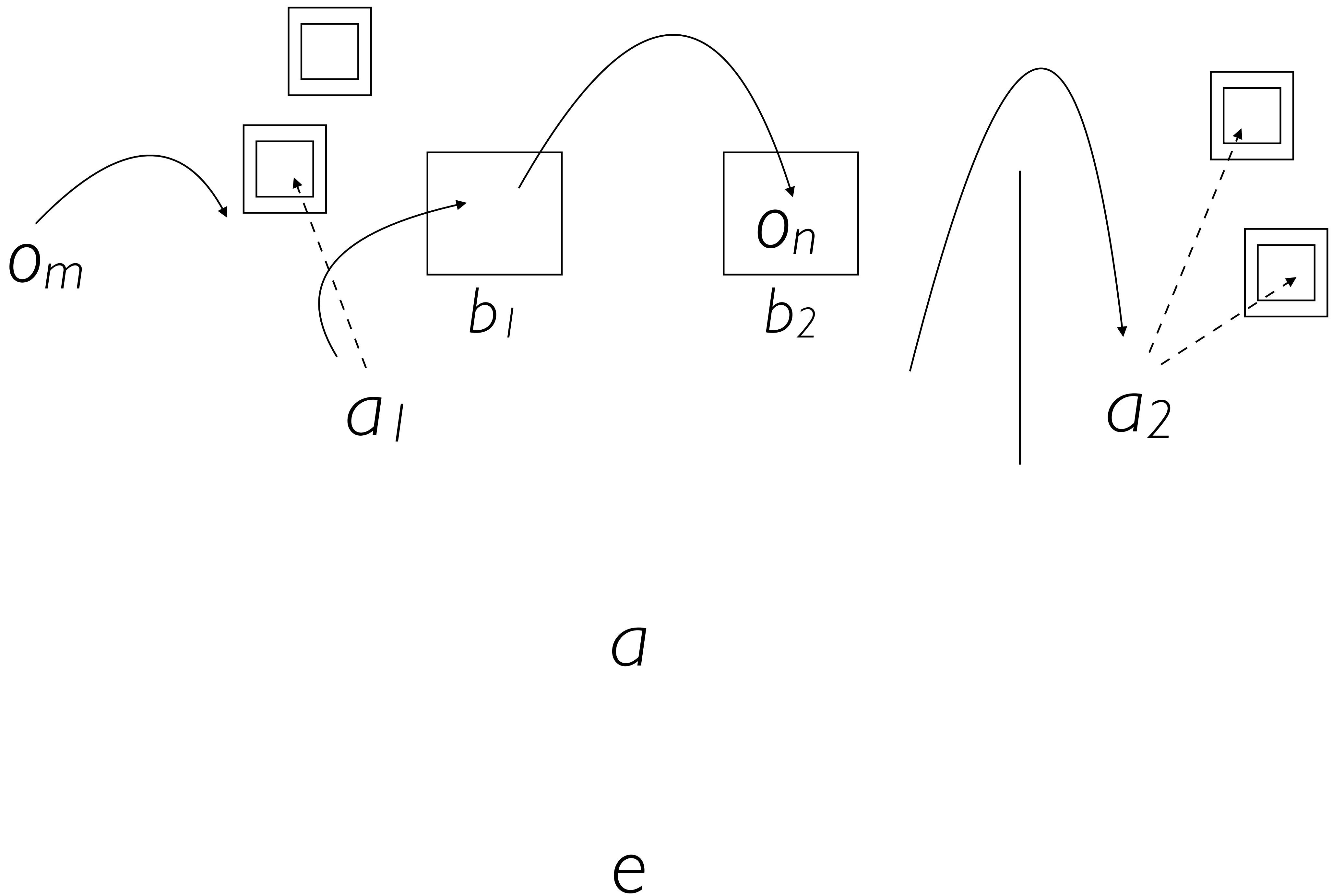
Framework for FBT^1_5
(ten timepoints)



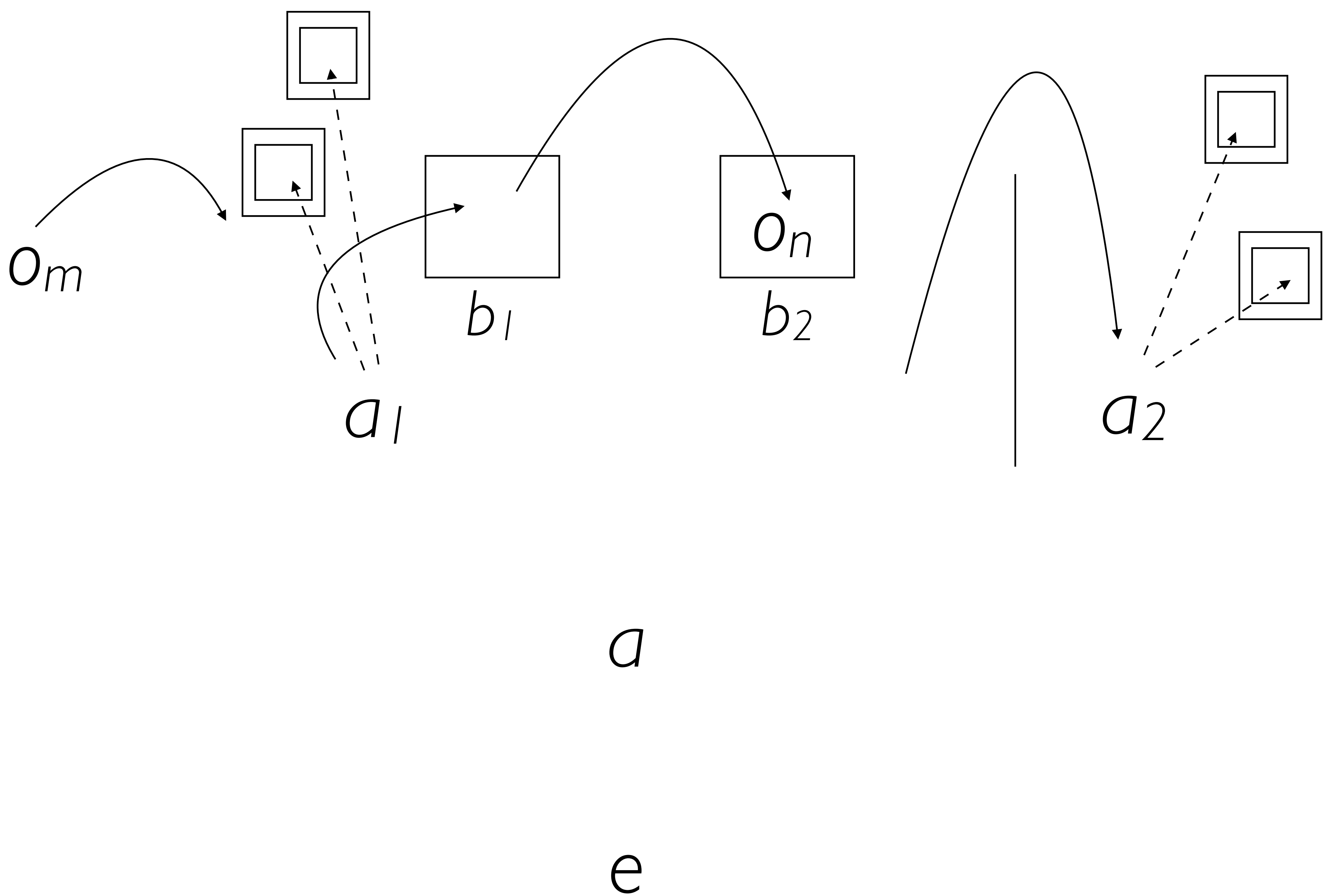
a

e

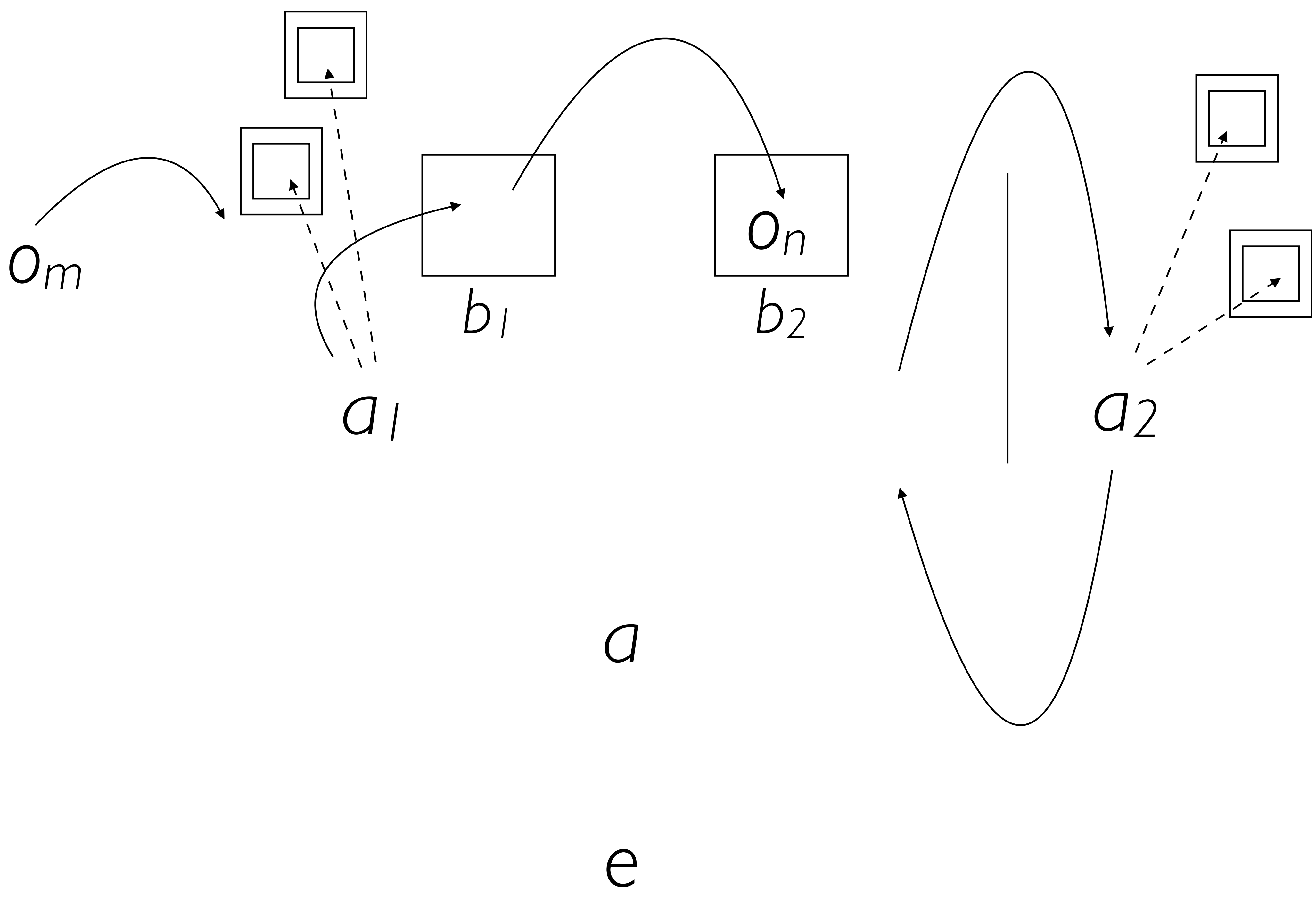
Framework for FBT^{1}_{5}
(ten timepoints)



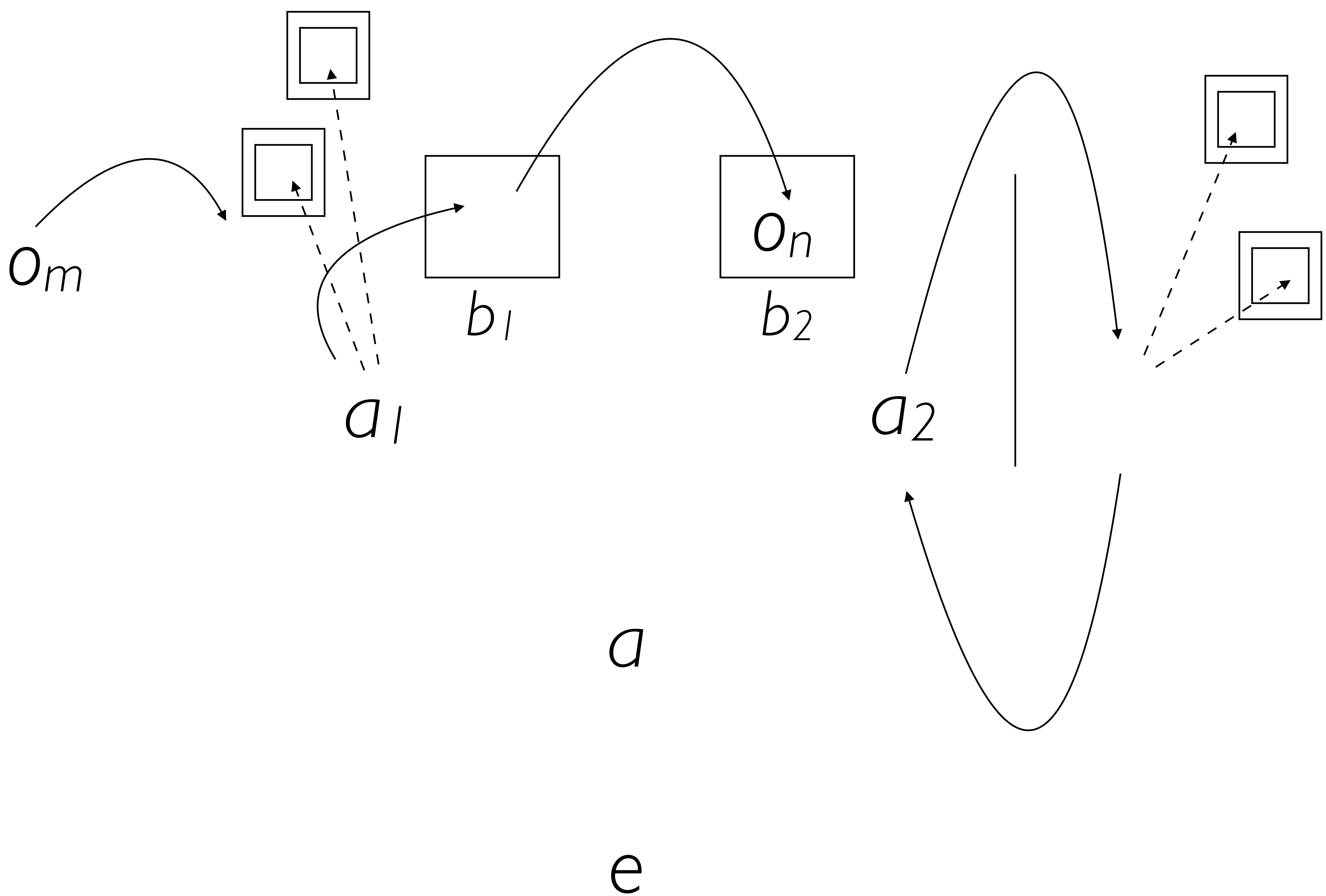
Framework for $FBT^1{}_5$
(ten timepoints)



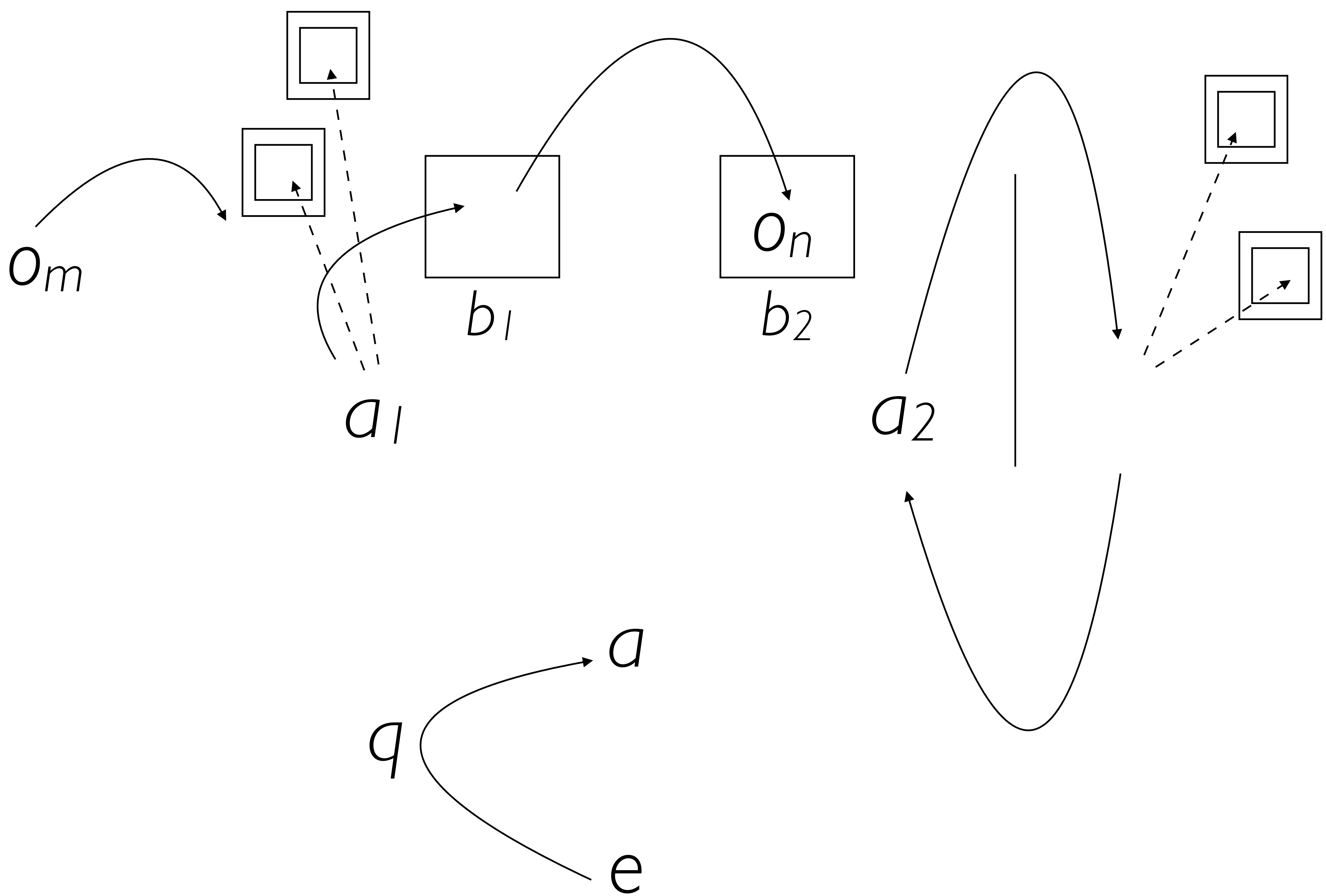
Framework for $FBT^1{}_5$
(ten timepoints)



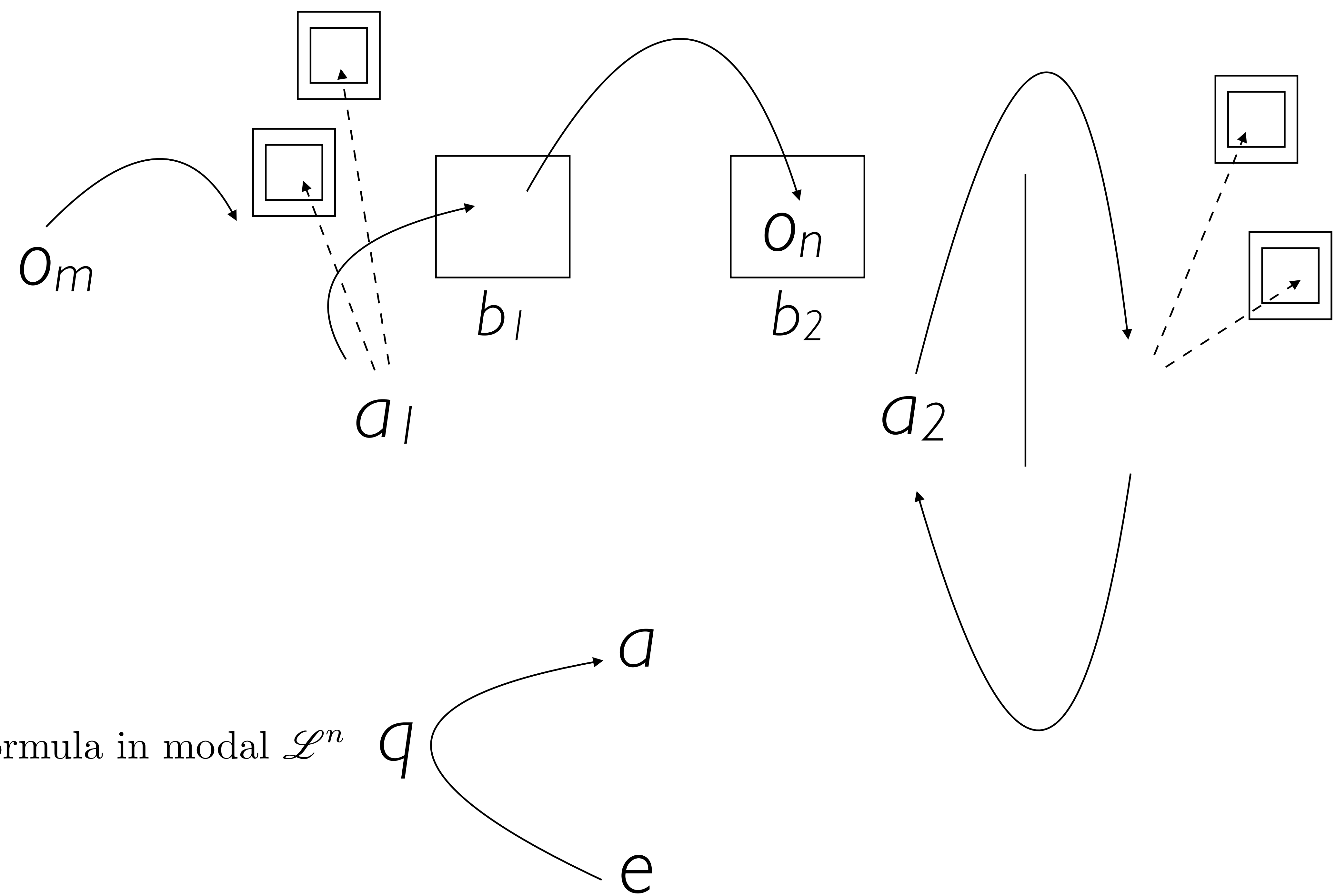
Framework for $FBT^1{}_5$
(ten timepoints)



Framework for $FBT^1{}_5$
(ten timepoints)



Framework for FBT$^1{}_5$
(ten timepoints)



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

Humans Can Succeed

Neurobiologically normal, nurtured, educated, and sufficiently motivated humans can correctly answer any relevant query *q* for the infinite progression, and prove that their answer is correct. For the obvious subclass of queries (the form of which appear in the box below), they can prove and exploit the following lemma.

Lemma: Suppose $\mathrm{FBT}_k, k \in \mathbb{Z}^+$, holds; (i.e. that level k of FBT holds). Then, if k is even, $\mathbf{B}_2\mathbf{B}_1 \ldots \mathbf{B}_2\ \iota$, where there are $k+1$ iterated $\mathbf{B}_i$ operators; otherwise $\mathbf{B}_1\mathbf{B}_2 \ldots \mathbf{B}_1\mathbf{B}_2\ \iota$, where there again there are $k+1$ iterated $\mathbf{B}_i$ operators.

Passing to Probing Mastery of the Specific Subclass

Experimenter to *a*: “At level *k*, from which box will a_2 attempt to retrieve the objects o_n? Prove it!”

Theoretical Machine Success on Infinite FBT!

Theorem: $\forall q \in \mathscr{CC}, \mathfrak{M}$ can correctly answer and justify q.
I.e., $\mathfrak{M}$ can pass FBT_ω.

Ok, so this logic machine exists in the *mathematical* universe; but does there exist an *implemented* machine with this power?

Theoretical Machine Success on Infinite FBT!

Theorem: $\forall q \in \mathscr{CC}, \mathfrak{M}$ can correctly answer and justify q. I.e., $\mathfrak{M}$ can pass FBT_ω.

Ok, so this logic machine exists in the *mathematical* universe; but does there exist an *implemented* machine with this power?

Simulation Courtesy of …

ShadowProver!

Level I

```
: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))))))}
```

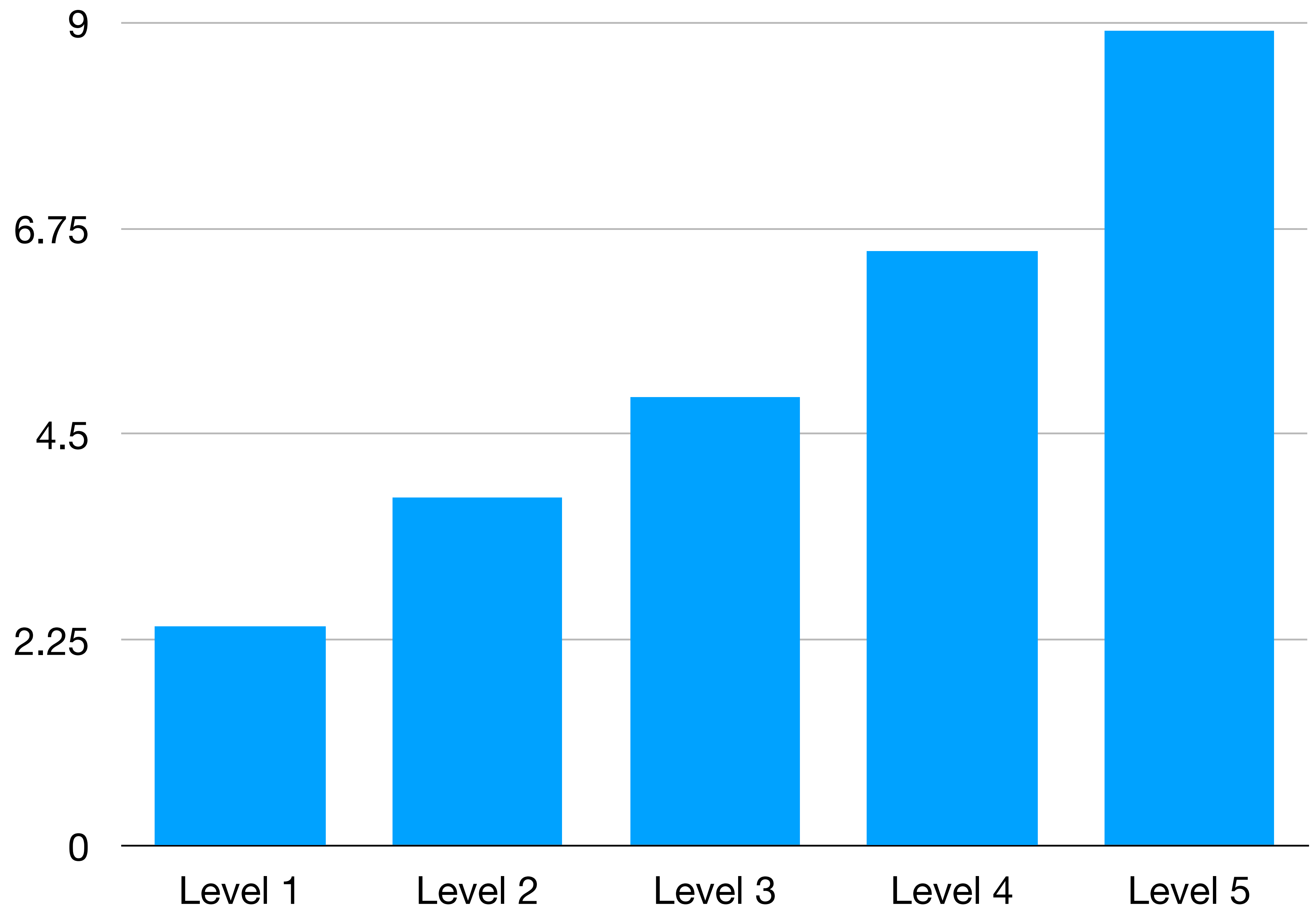
Level 5

```
{: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.

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

:goal
...l (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))))))))}
```

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

...1 (Common! t0
     (In o b1) t1)
...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 ...

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

:goal
     :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))))))))}
```

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 ----------------------
```