Standard Deontic Logic (SDL = D) Isn't Going to Cut It!

(Chisholm's Paradox; The Free Choice Permission Paradox)

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

Intro to Logic
4/6/2020

all

Download: LAMA-BDLA_031320.pdf

Bonus Problems

Coq1

Test Problems

Test_1_G2
Test_1_G1
Test_1_G3
Test_2_G2
Test_2_G1
KK1point3Strings
LevesquePrologProgram1
Test_2_G3
Special_Problem_3b
EmmaHelpedToo

Exercise Problems

BabyArithmetic

☐ Test Problem: Test_1_G2

Hello longsnowflake876, please solve this second propositional calculus problem we have created for you.

Deadline 21 Feb 2020 17:00:00 EST

Solve

☑ Test Problem: Test_1_G1

Hello longsnowflake876, please solve this first propositional calculus problem we have created for you.

Deadline 21 Feb 2020 17:00:00 EST

Solve

☐ Test Problem: Test_1_G3

Hello longsnowflake876, please solve this third propositional calculus problem we have created for you.

Deadline 21 Feb 2020 17:00:00 EST

Solve

all

Download: LAMA-BDLA_031320.pdf

Bonus Problems

Coq1

Test Problems

Test_1_G2
Test_1_G1
Test_1_G3
Test_2_G2
Test_2_G1
KK1point3Strings
LevesquePrologProgram1
Test_2_G3
Special_Problem_3b
EmmaHelpedToo

Exercise Problems

BabyArithmetic

☐ Test Problem: Test_1_G2

Hello longsnowflake876, please solve this second propositional calculus problem we have created for you.

Deadline 21 Feb 2020 17:00:00 EST

Solve

☑ Test Problem: Test_1_G1

Hello longsnowflake876, please solve this first propositional calculus problem we have created for you.

Deadline 21 Feb 2020 17:00:00 EST

Solve

☐ Test Problem: Test_1_G3

Hello longsnowflake876, please solve this third propositional calculus problem we have created for you.

Deadline 21 Feb 2020 17:00:00 EST

Solve

Peek ahead to next time for some context today …

At least supposedly, long term:

At least supposedly, long term:

"We're in *very* deep trouble."

At least supposedly, long term:

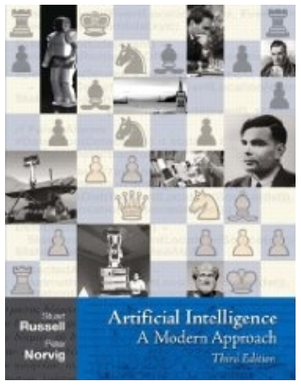
"We're in *very* deep trouble."




At least supposedly, long term:

"We're in *very* deep trouble."



The PAID Problem!

Actually, it's quite simple:
"Equation" for Why Stakes are High

Actually, it's quite simple: "Equation" for Why Stakes are High

$\forall$x : Agents

Actually, it's quite simple: "Equation" for Why Stakes are High

$\forall$x : Agents

Powerful(x) + Autonomous(x) + Intelligent(x) = Dangerous(x)

Actually, it's quite simple: "Equation" for Why Stakes are High

$\forall$x : Agents

Powerful(x) + Autonomous(x) + Intelligent(x) = Dangerous(x)

Actually, it's quite simple: "Equation" for Why Stakes are High

$\forall$x : Agents

Powerful(x) + Autonomous(x) + Intelligent(x) = Dangerous(x)

$$u(\mathrm{AIA}_i(\pi_j)) > \tau^+ \in \mathbb{Z} \text{ or } \tau^- \in \mathbb{Z}$$

Actually, it's quite simple: "Equation" for Why Stakes are High

$\forall$x : Agents

Powerful(x)

$u(\mathrm{AIA}_i(\pi_j)) >$

Dangerous(x)

Are Autonomous-and-Creative Machines Intrinsically Untrustworthy?*

Selmer Bringsjord • Naveen Sundar G.

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA

020217NY

Abstract

Given what we find in the case of human cognition, the following principle appears to be quite plausible: An artificial agent that is both autonomous (A) and creative (C) will tend to be, from the viewpoint of a rational, fully informed agent, (U) untrustworthy. After briefly explaining the intuitive, internal structure of this disturbing principle, in the context of the human sphere, we provide a more formal rendition of it designed to apply to the realm of intelligent artificial agents. The more-formal version makes use of some of the basic structures available in one of our cognitive-event calculi, and can be expressed as a (confessedly — for reasons explained — naïve) theorem. We prove the theorem, and provide simple demonstrations of it in action, using a novel theorem prover (ShadowProver). We then end by pointing toward some future defensive engineering measures that should be taken in light of the theorem.

Contents

Actually, it's quite simple: "Equation" for Why Stakes are High

$\forall$x : Agents

Powerful(x) + Autonomous(x) + Intelligent(x) = Dangerous(x)

$$u(\mathrm{AIA}_i(\pi_j)) > \tau^+ \in \mathbb{Z} \text{ or } \tau^- \in \mathbb{Z}$$

Actually, it's quite simple: "Equation" for Why Stakes are High

$\forall$x : Agents

Powerful(x) + Autonomous(x) + Intelligent(x) = Dangerous(x)

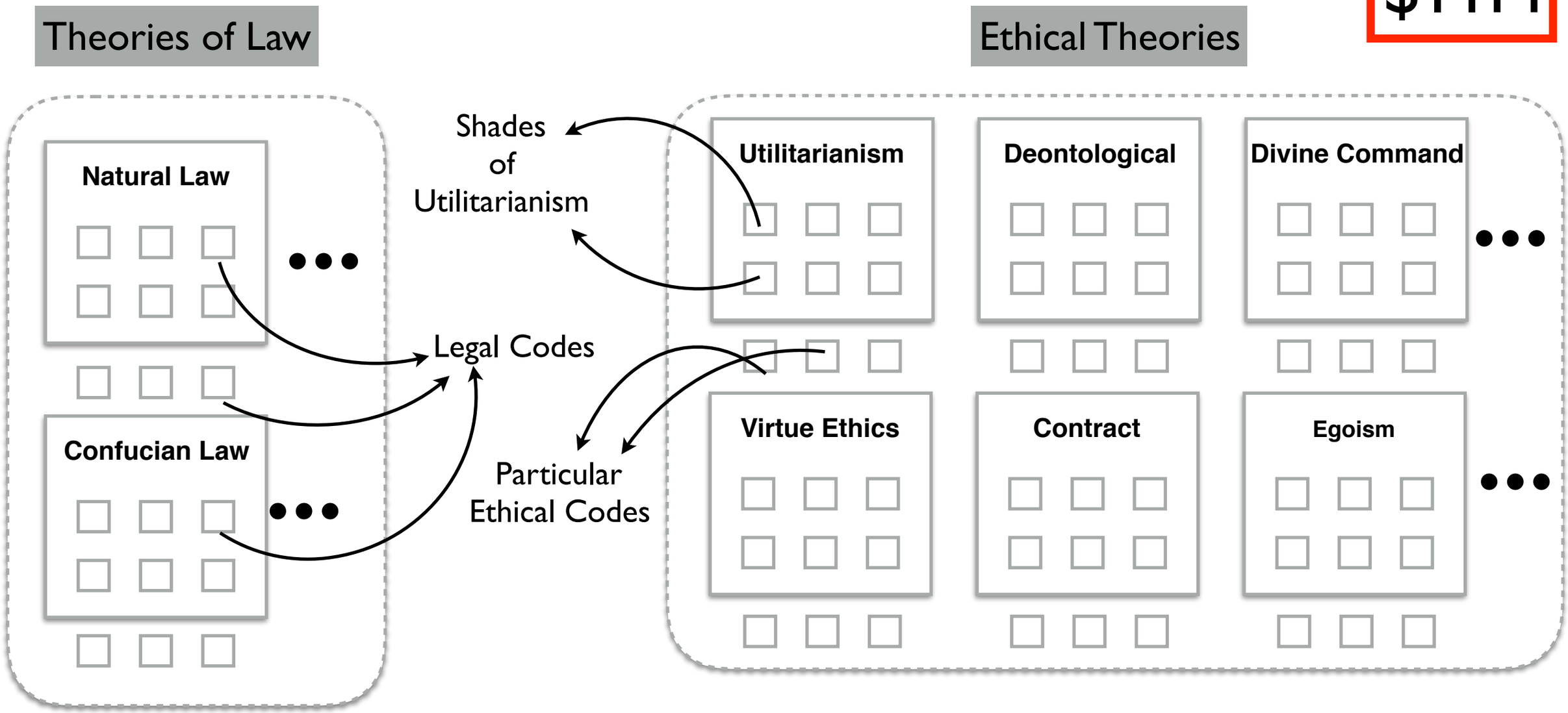
$$u(\mathrm{AIA}_i(\pi_j)) > \tau^+ \in \mathbb{Z} \text{ or } \tau^- \in \mathbb{Z}$$

Theorem ACU: In a collaborative situation involving agents a (as the "trustor") and a' (as the "trustee"), if a' is at once both autonomous and ToM-creative, a' is untrustworthy from an ideal-observer o's viewpoint, with respect to the action-goal pair $\langle \alpha, \gamma \rangle$ in question.

Proof: Let a and a' be agents satisfying the hypothesis of the theorem in an arbitrary collaborative situation. Then, by definition, $a \neq a'$ desires to obtain some goal γ in part by way of a contributed action α_k from a', a' knows this, and moreover a' knows that a believes that this contribution will succeed. Since a' is by supposition ToM-creative, a' may desire to surprise a with respect to a's belief regarding a''s contribution; and because a' is autonomous, attempts to ascertain whether such surprise will come to pass are fruitless since what will happen is locked inaccessibly in the oracle that decides the case. Hence it follows by TRANS that an ideal observer o will regard a' to be untrustworthy with respect to the pair $\langle \alpha, \gamma \rangle$ pair. **QED**

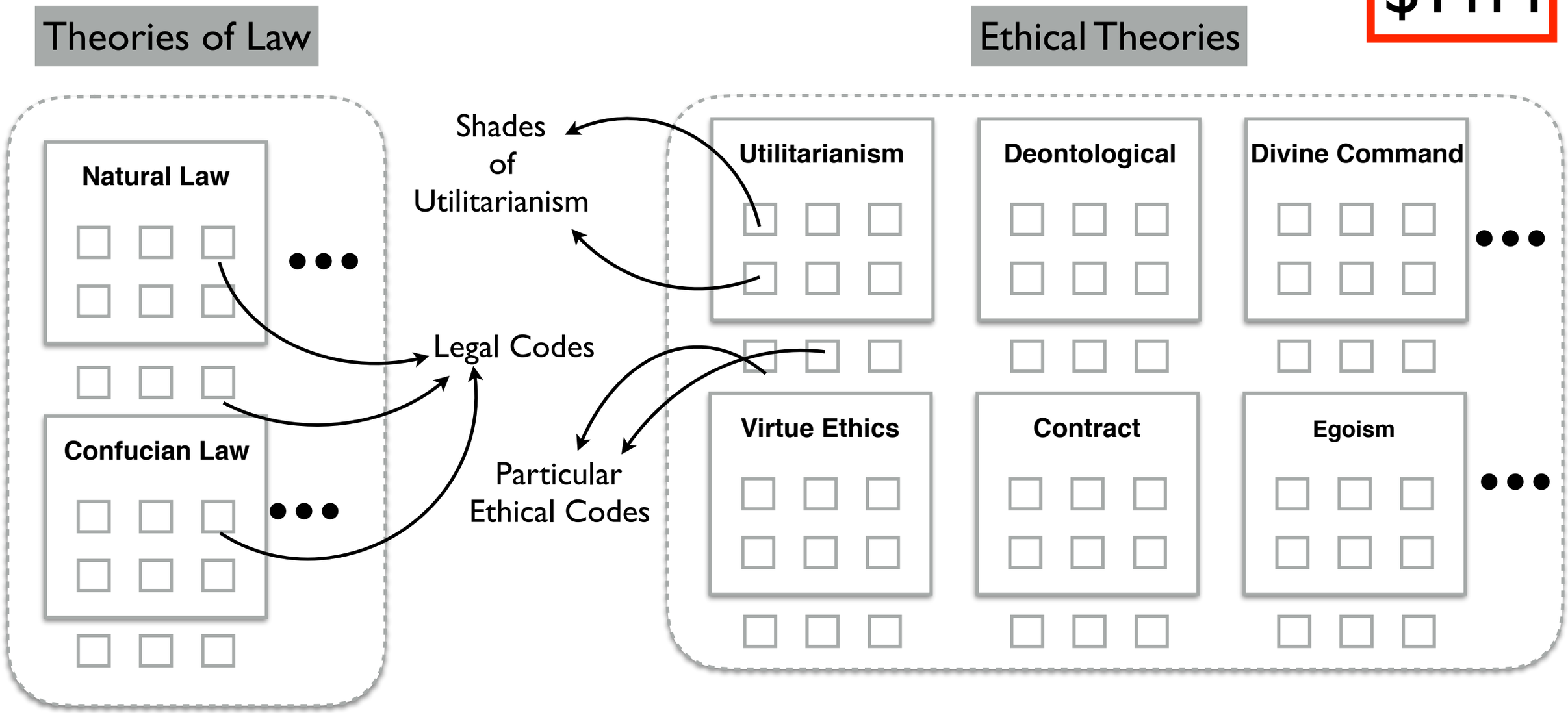
Making Morally X Machines

$1 1M



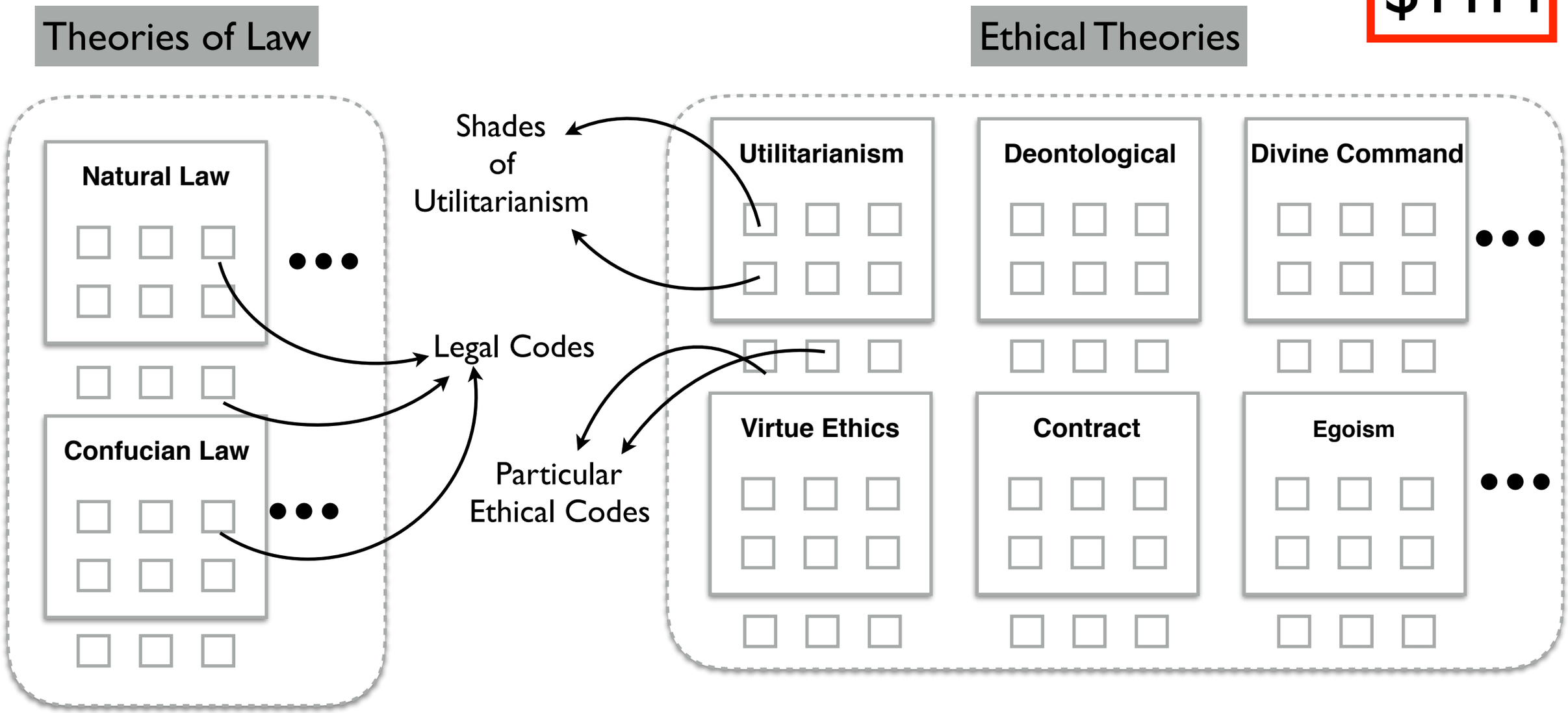
Making Morally X Machines

$1 1M



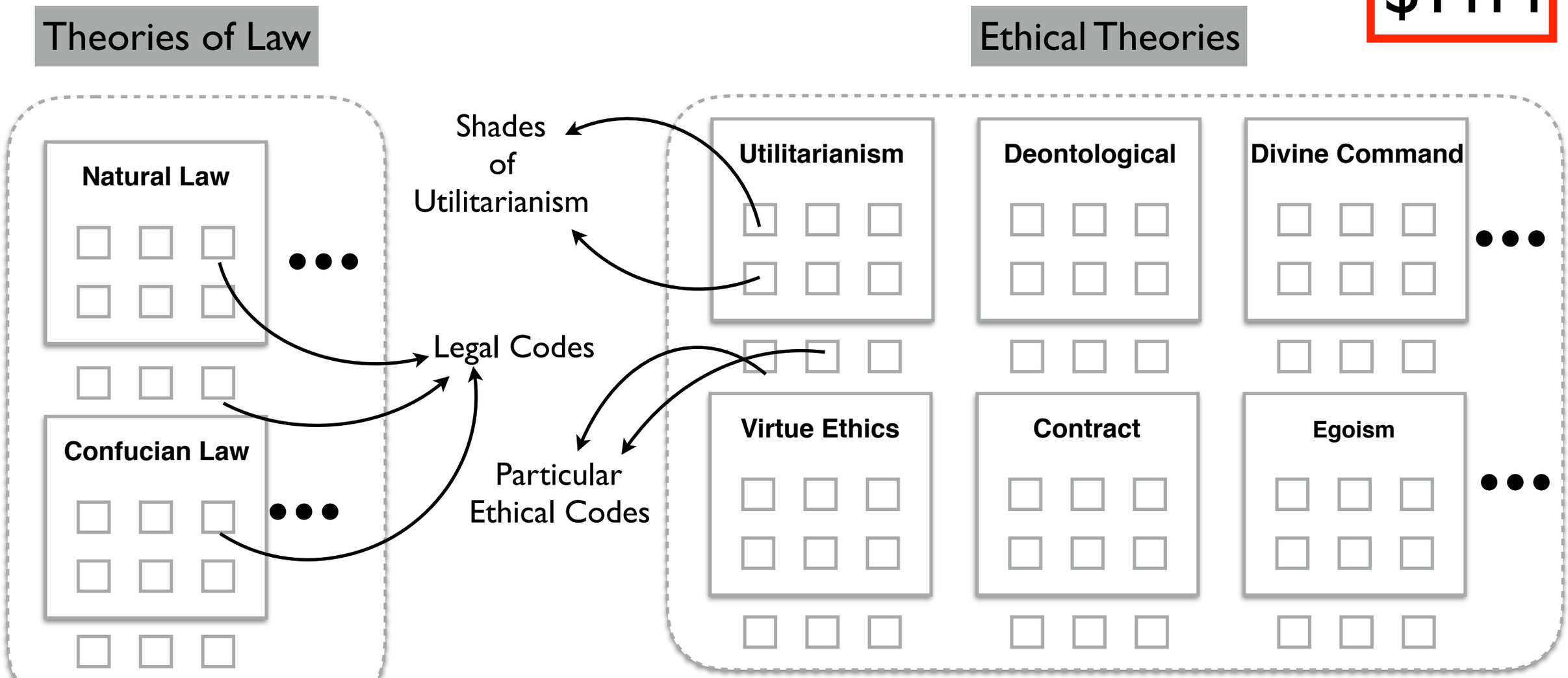
Making Morally X Machines

$1 1M



Making Morally X Machines

$1 1M

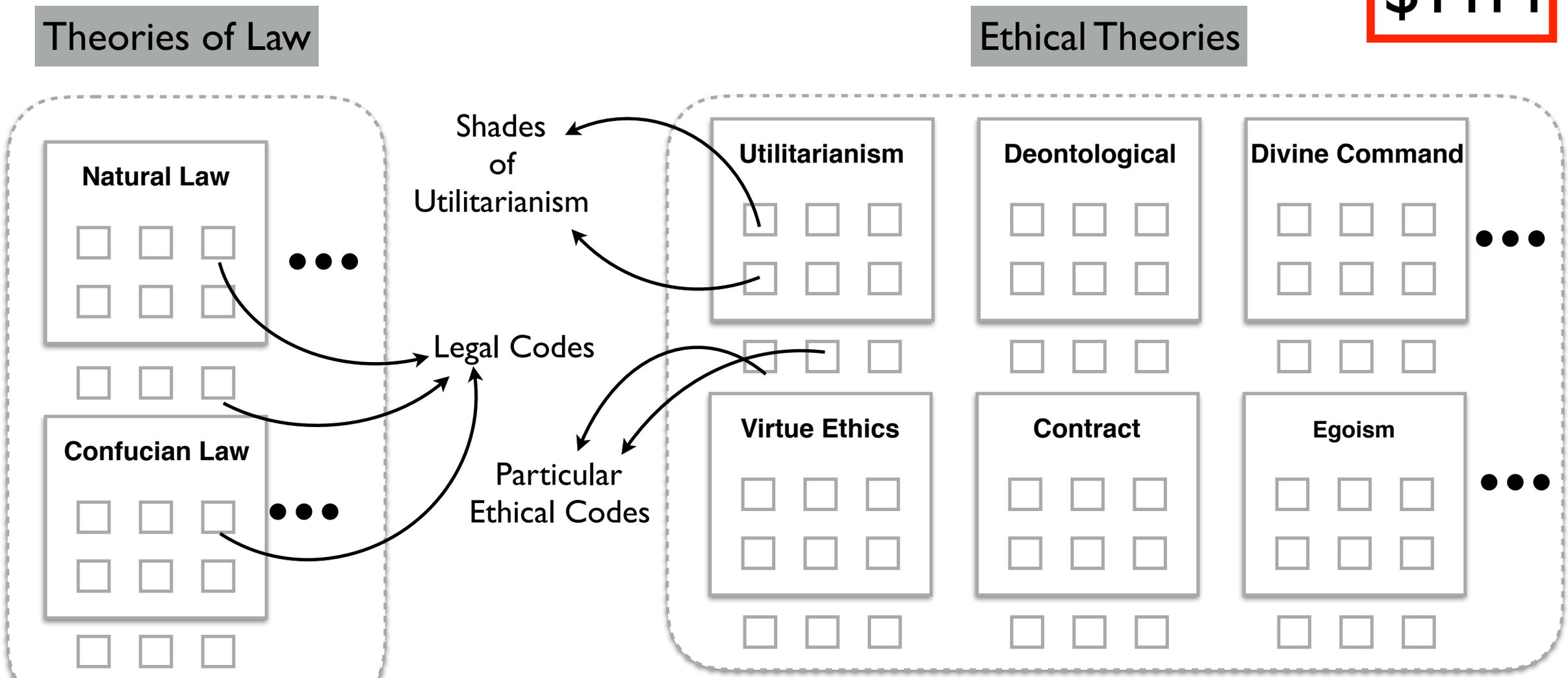


Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

Making Morally X Machines

$1 1M

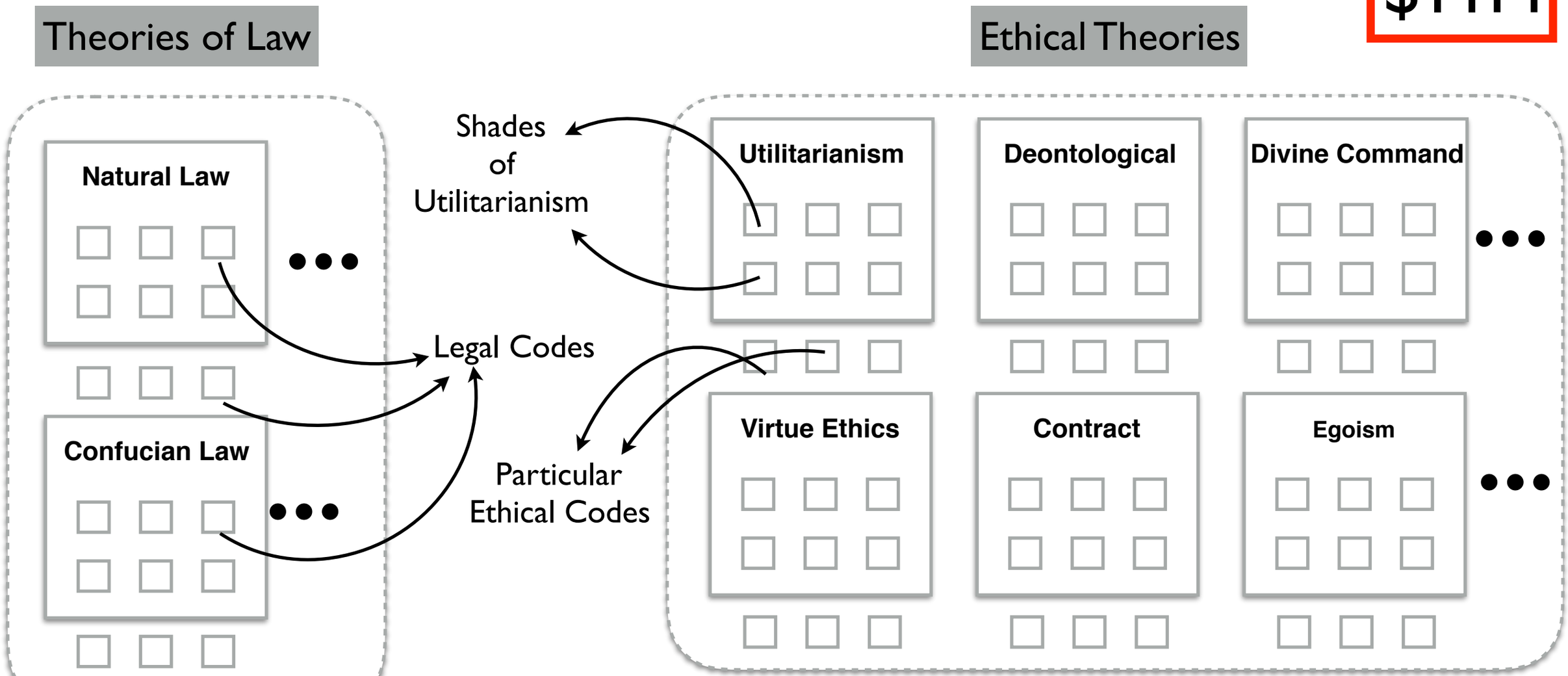


Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

Making Morally X Machines

$IIM



Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

Step 2

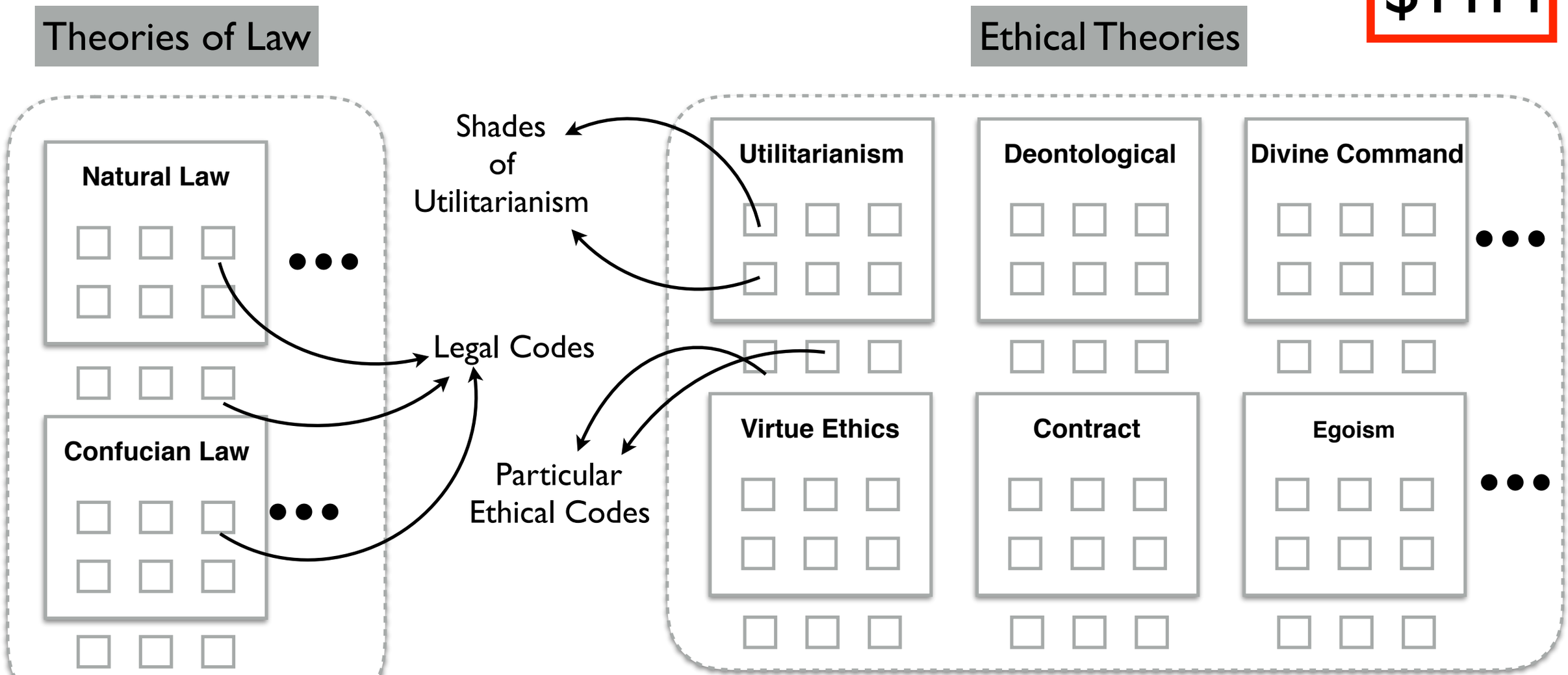
Automate

Prover

Spectra

Making Morally X Machines

$1 1M



Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

Step 2

Automate

Prover

Spectra

Making Morally X Machines

$1 1M

Theories of Law

- Natural Law
- Confucian Law
- ...

Shades of Utilitarianism

Legal Codes

Particular Ethical Codes

Ethical Theories

- Utilitarianism
- Deontological
- Divine Command
- Virtue Ethics
- Contract
- Egoism
- ...

Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

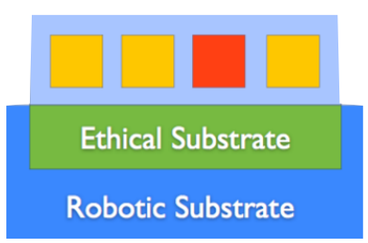
Step 2

Automate

- Prover
- Spectra

Step 3

Ethical OS



Making Morally X Machines

$1 1M

Theories of Law

- Natural Law
- Confucian Law
- ...

Shades of Utilitarianism

Legal Codes

Particular Ethical Codes

Ethical Theories

- Utilitarianism
- Deontological
- Divine Command
- Virtue Ethics
- Contract
- Egoism
- ...

Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

Step 2

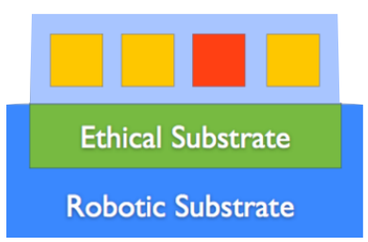
Automate

Prover

Spectra

Step 3

Ethical OS



Making Morally X Machines

$1 1M

Theories of Law

- Natural Law
- Confucian Law
- ...

Shades of Utilitarianism

Legal Codes

Particular Ethical Codes

Ethical Theories

- Utilitarianism
- Deontological
- Divine Command
- Virtue Ethics
- Contract
- Egoism
- ...

Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

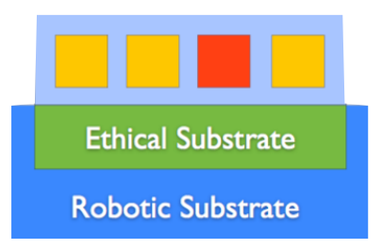
Step 2

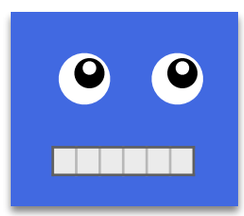
Automate

- Prover
- Spectra

Step 3

Ethical OS





Making Morally X Machines

$1 1M

Theories of Law

- Natural Law
- Confucian Law
- ...

Shades of Utilitarianism

Legal Codes

Particular Ethical Codes

Ethical Theories

- Utilitarianism
- Deontological
- Divine Command
- Virtue Ethics
- Contract
- Egoism
- ...

Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

Step 2

Automate

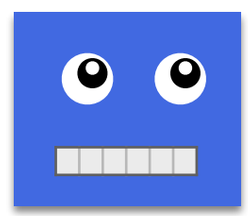
- Prover
- Spectra

Step 3

Ethical OS

- Ethical Substrate
- Robotic Substrate

Ethical Robot



Making Morally X Machines

$1 1M

Theories of Law

- Natural Law
- Confucian Law

Shades of Utilitarianism

Legal Codes

Particular Ethical Codes

Ethical Theories

- Utilitarianism
- Deontological
- Divine Command
- Virtue Ethics
- Contract
- Egoism

Step 1

1. Pick a theory
2. Pick a code
3. Run through EH.

Step 2

Automate

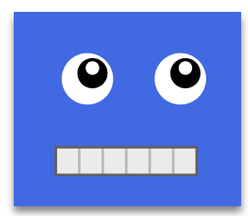
- Prover
- Spectra

Step 3

Ethical OS

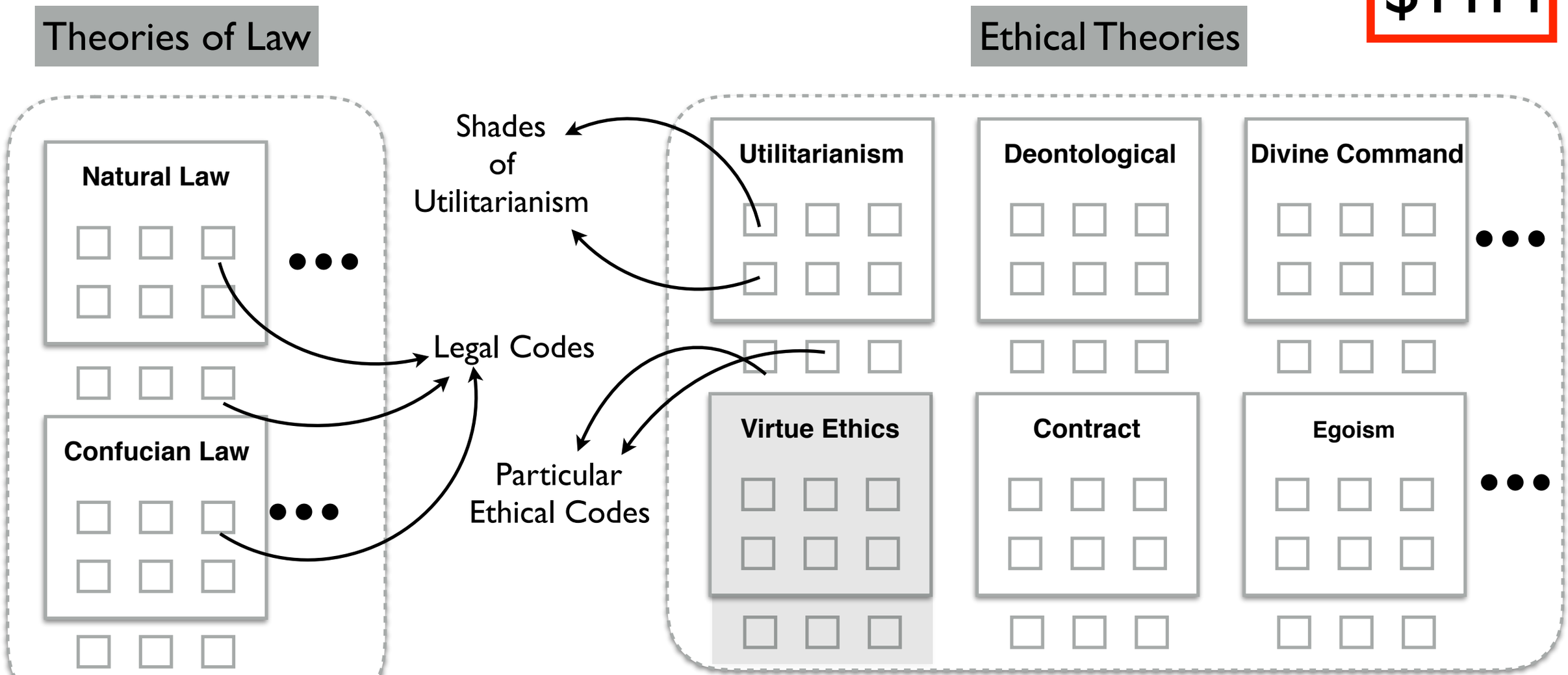
- Ethical Substrate
- Robotic Substrate

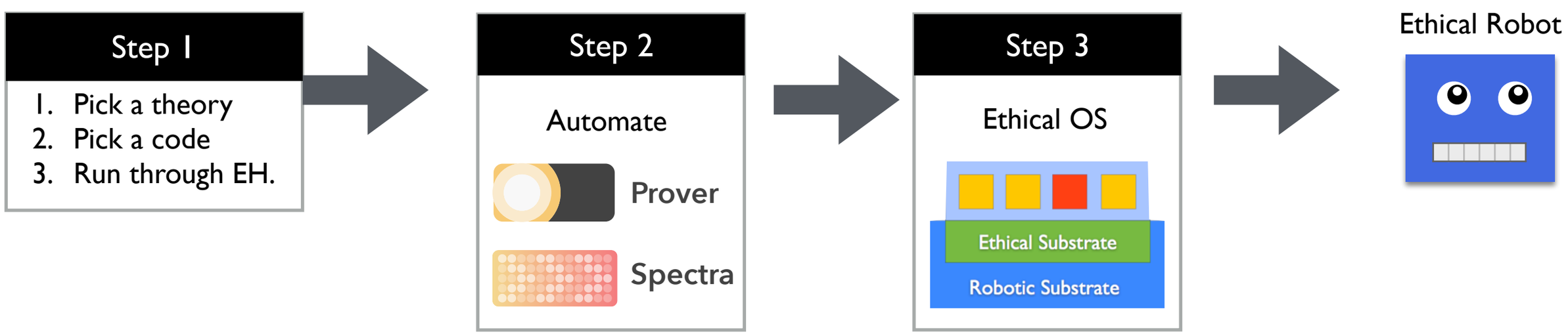
Ethical Robot



Making Morally X Machines

$1 1M





"Toward the Engineering of Virtuous Robots" Naveen, Selmer et al.

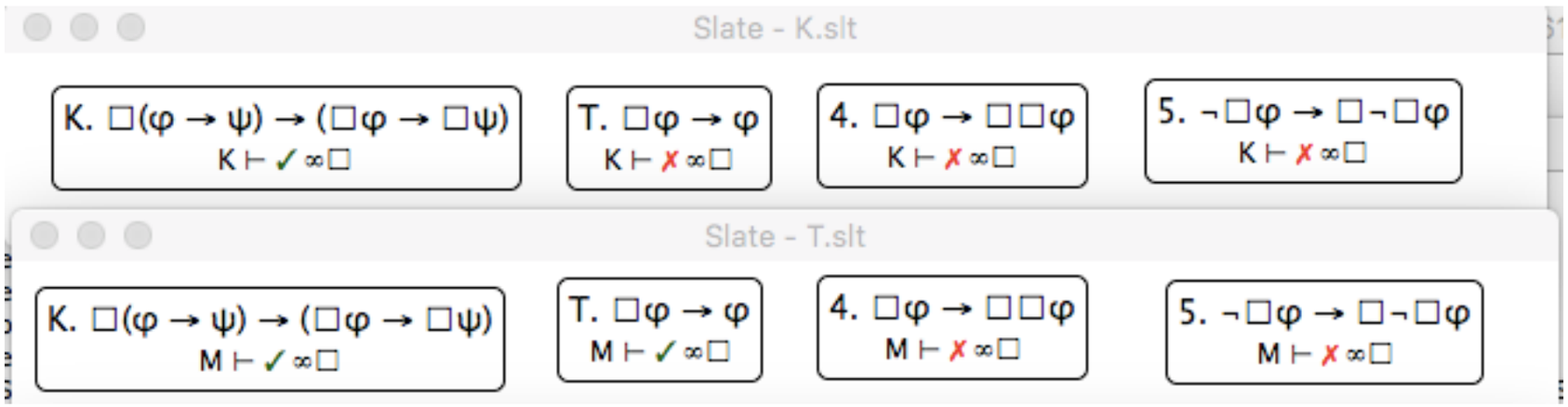
Well, maybe, but at any rate, *what logic*??

Well, maybe, but at any rate, *what logic*??

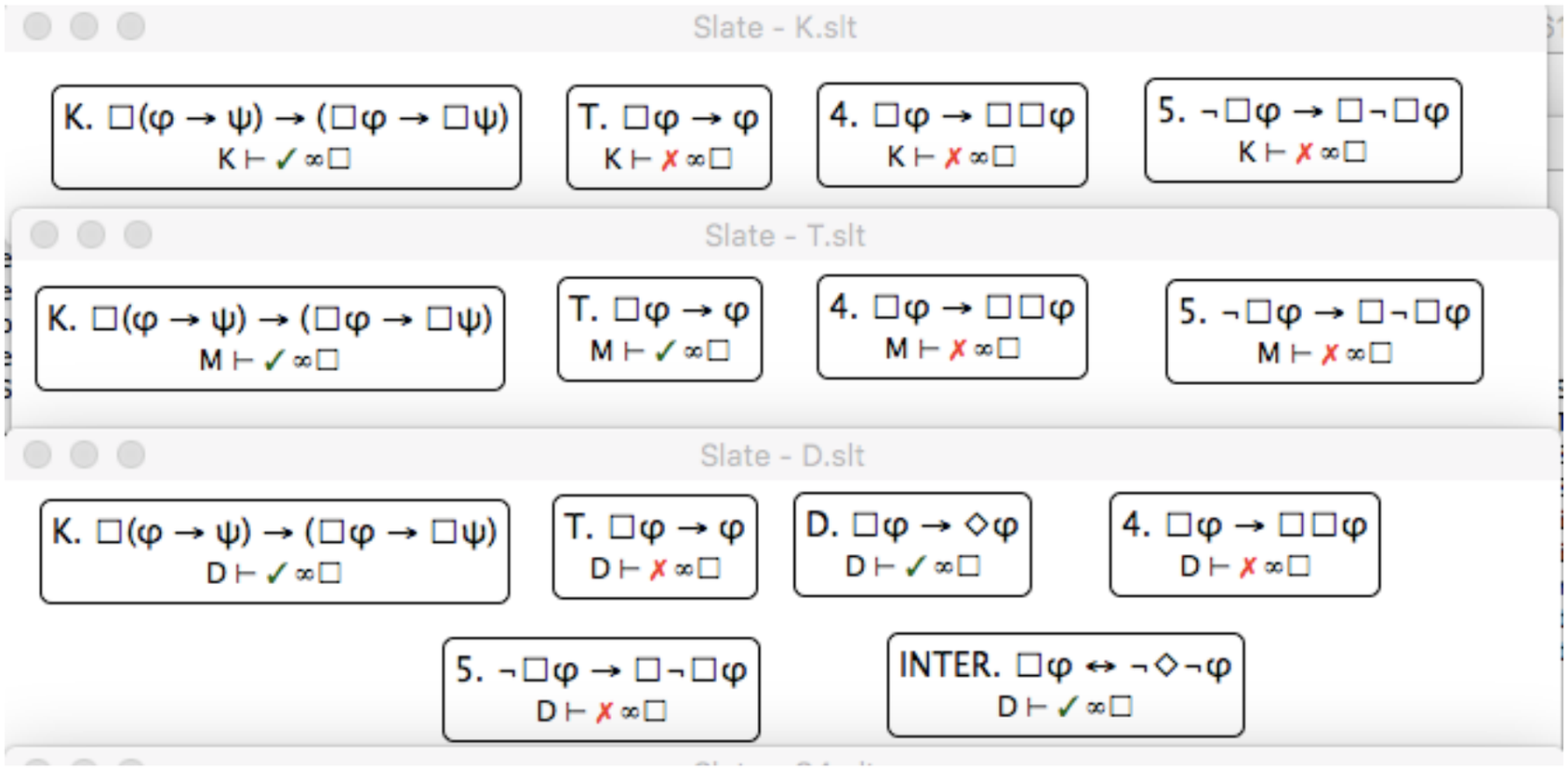
Perhaps **D** = **SDL**? …

Review: Encapsulation

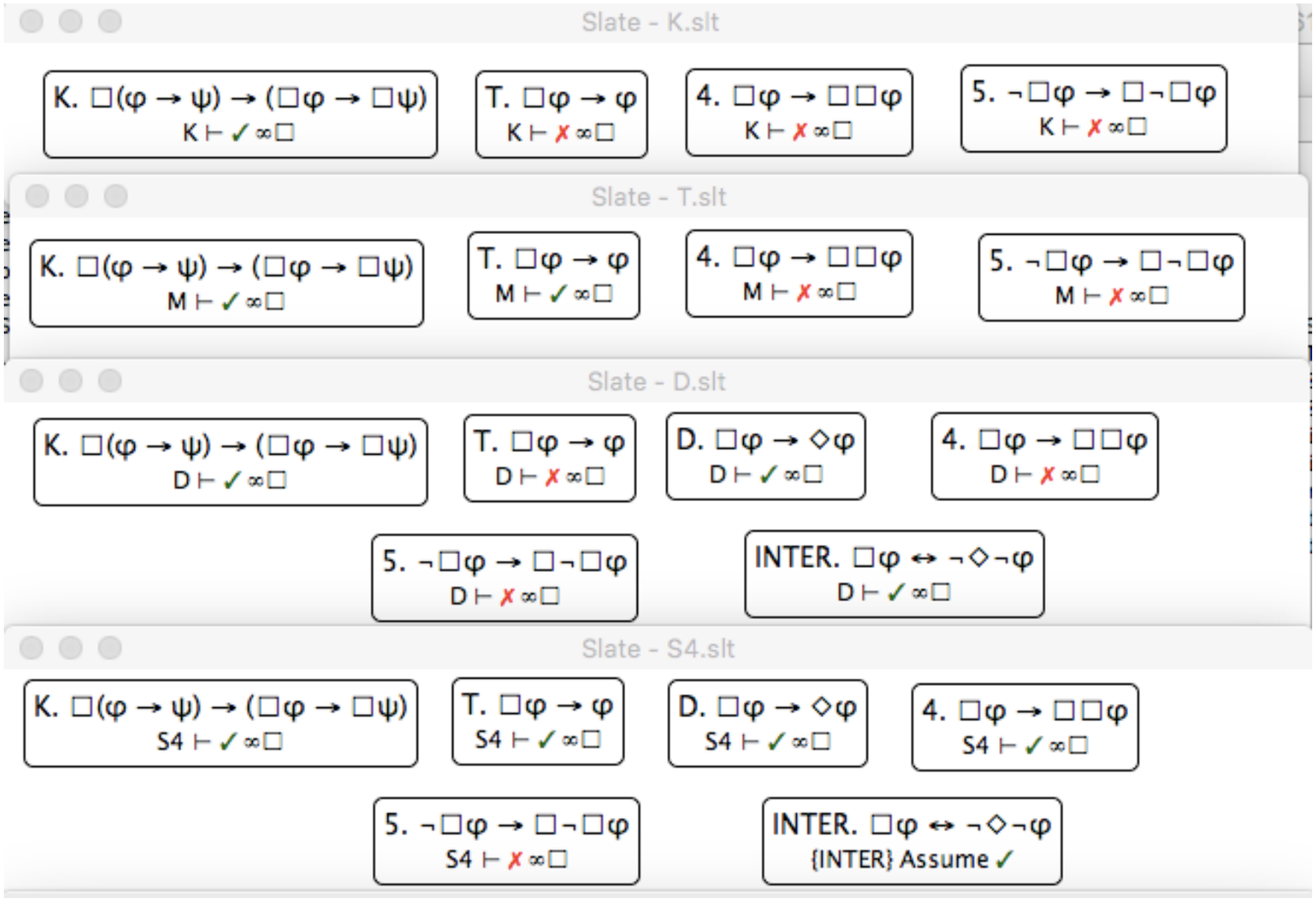
Review: Encapsulation



Review: Encapsulation



Review: Encapsulation



Review: Encapsulation

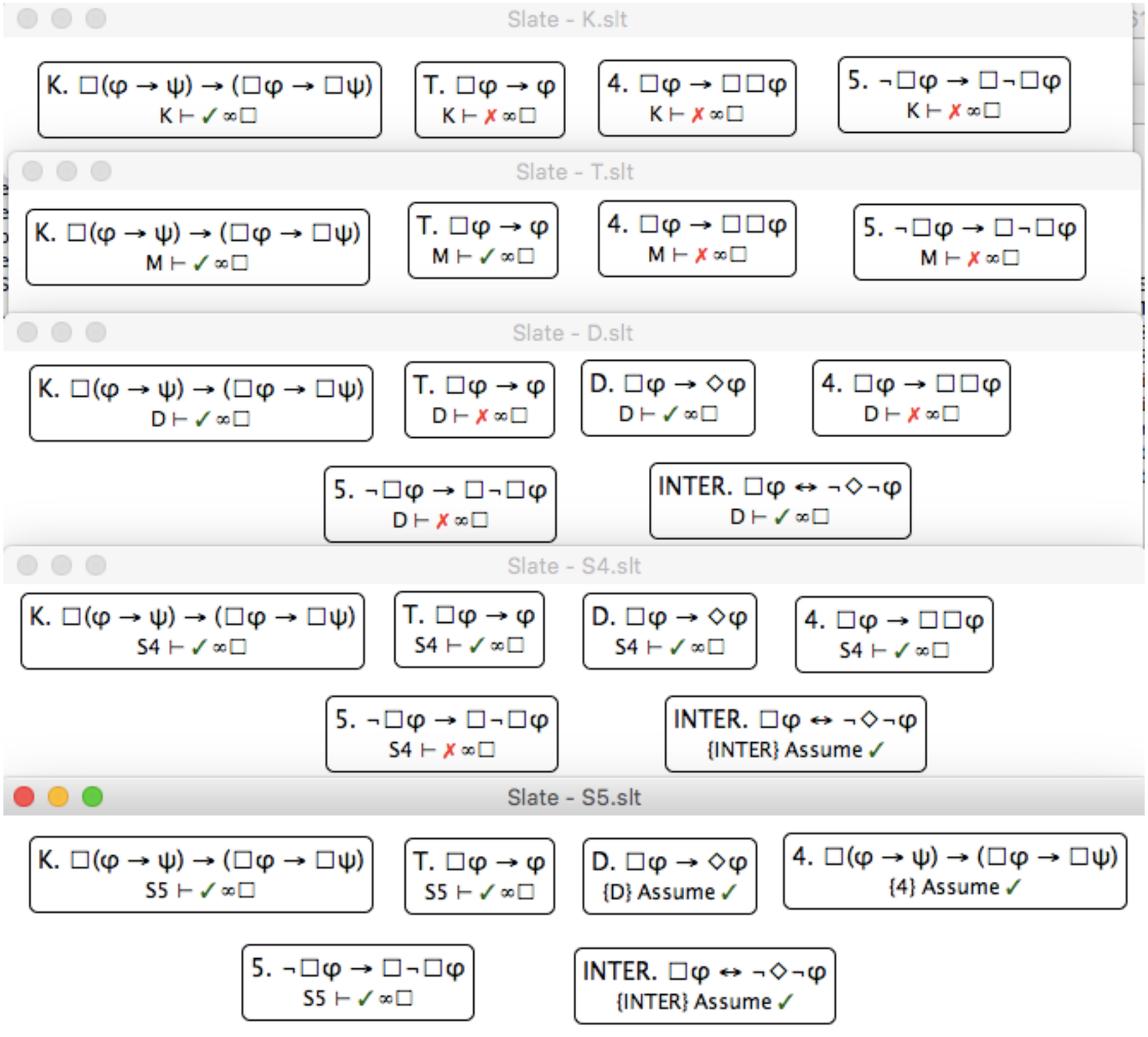
K

T

D

4 = S4

5 = S5



Review: Encapsulation

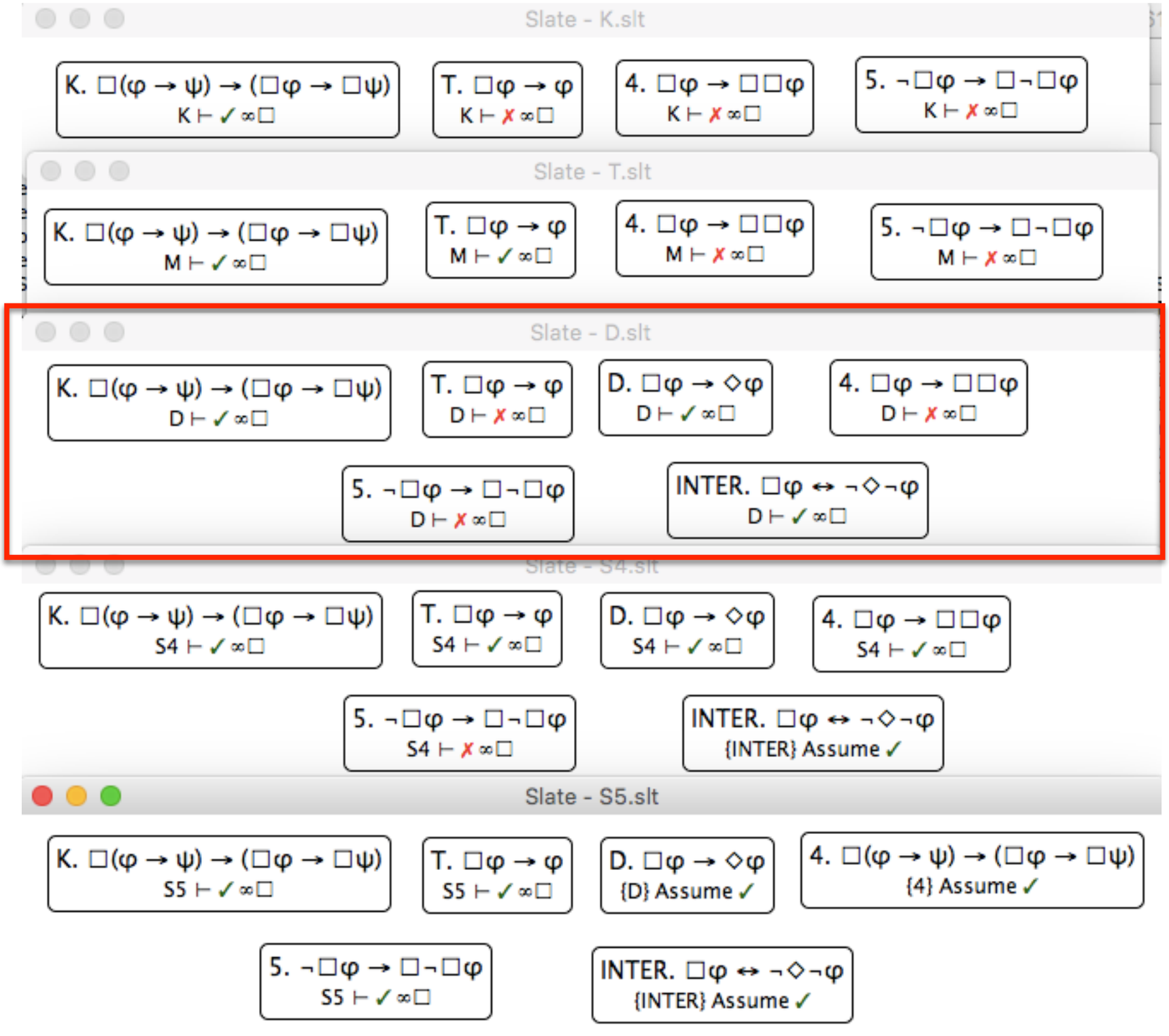
K

T

D

4 = S4

5 = S5



HyperSlate™

Review: Encapsulation

K

T

D

4 = S4

5 = S5

Slate - K.slt

K. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
K ⊢ ✓ ∞□

T. $\Box\varphi \to \varphi$
K ⊢ ✗ ∞□

4. $\Box\varphi \to \Box\Box\varphi$
K ⊢ ✗ ∞□

5. $\neg\Box\varphi \to \Box\neg\Box\varphi$
K ⊢ ✗ ∞□

Slate - T.slt

K. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
M ⊢ ✓ ∞□

T. $\Box\varphi \to \varphi$
M ⊢ ✓ ∞□

4. $\Box\varphi \to \Box\Box\varphi$
M ⊢ ✗ ∞□

5. $\neg\Box\varphi \to \Box\neg\Box\varphi$
M ⊢ ✗ ∞□

Slate - D.slt

K. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
D ⊢ ✓ ∞□

T. $\Box\varphi \to \varphi$
D ⊢ ✗ ∞□

D. $\Box\varphi \to \Diamond\varphi$
D ⊢ ✓ ∞□

4. $\Box\varphi \to \Box\Box\varphi$
D ⊢ ✗ ∞□

5. $\neg\Box\varphi \to \Box\neg\Box\varphi$
D ⊢ ✗ ∞□

INTER. $\Box\varphi \leftrightarrow \neg\Diamond\neg\varphi$
D ⊢ ✓ ∞□

Slate - S4.slt

K. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
S4 ⊢ ✓ ∞□

T. $\Box\varphi \to \varphi$
S4 ⊢ ✓ ∞□

D. $\Box\varphi \to \Diamond\varphi$
S4 ⊢ ✓ ∞□

4. $\Box\varphi \to \Box\Box\varphi$
S4 ⊢ ✓ ∞□

5. $\neg\Box\varphi \to \Box\neg\Box\varphi$
S4 ⊢ ✗ ∞□

INTER. $\Box\varphi \leftrightarrow \neg\Diamond\neg\varphi$
{INTER} Assume ✓

Slate - S5.slt

K. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
S5 ⊢ ✓ ∞□

T. $\Box\varphi \to \varphi$
S5 ⊢ ✓ ∞□

D. $\Box\varphi \to \Diamond\varphi$
{D} Assume ✓

4. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
{4} Assume ✓

5. $\neg\Box\varphi \to \Box\neg\Box\varphi$
S5 ⊢ ✓ ∞□

INTER. $\Box\varphi \leftrightarrow \neg\Diamond\neg\varphi$
{INTER} Assume ✓

4.4.4 D = SDL (= 'Standard Deontic Logic')

We here introduce what is known as 'Standard Deontic Logic' (**SDL**), which in Slate is the system **D**. Deontic logic is the sub-branch of logic devoted to formalizing the fundamental concepts of morality; for example, the concepts of *obligation*, *permissibility*, and *forbiddenness*. The first of these three concepts can apparently serve as a cornerstone, since to say that ϕ (a formulae representing some state-of-affairs) is permissible seems to amount to saying that it's not obligatory that it not be the case that ϕ (which shows permissibility can be defined in terms of obligation), and to say that ϕ is forbidden would seem to amount to it being obligatory that it not be the case that ϕ (which of course appears to show that forbiddenness buildable from obligation). This interconnected trio of ethical concepts is a triad explicitly invoked and analyzed since the end of the 18^{th} century, and the importance of the triad even to modern deontic logic would be quite hard to exaggerate.[9]

SDL is traditionally axiomatized by the following:[10]

SDL

TAUT All theorems of the propositional calculus.

OB-K $\odot(\phi \rightarrow \psi) \rightarrow (\odot\phi \rightarrow \odot\psi)$

OB-D $\odot\phi \rightarrow \neg\odot\neg\phi$

MP If $\vdash \phi$ and $\vdash \phi \rightarrow \psi$, then $\vdash \quad \psi$

OB-NEC If $\vdash \phi$ then $\vdash \odot\phi$

4.4.4 D = SDL (= 'Standard Deontic Logic')

We here introduce what is known as 'Standard Deontic Logic' (**SDL**), which in Slate is the system **D**. Deontic logic is the sub-branch of logic devoted to formalizing the fundamental concepts of morality; for example, the concepts of *obligation*, *permissibility*, and *forbiddenness*. The first of these three concepts can apparently serve as a cornerstone, since to say that ϕ (a formulae representing some state-of-affairs) is permissible seems to amount to saying that i

CHAPTER 4. PROPOSITIONAL MODAL LOGIC 1

that ϕ (which shows permissibility can be
say that ϕ is forbidden would seem to amo
the case that ϕ (which of course appears to s
obligation). This interconnected trio of ethic
and analyzed since the end of the 18^{th} centu
to modern deontic logic would be quite hard

SDL is traditionally axiomatized by the fo

SDL

TAUT All theorems of the proposition

OB-K $\odot(\phi \rightarrow \psi) \rightarrow (\odot\phi \rightarrow \odot\psi)$

OB-D $\odot\phi \rightarrow \neg\odot\neg\phi$

MP If $\vdash \phi$ and $\vdash \phi \rightarrow \psi$, then $\vdash \quad \phi$

OB-NEC If $\vdash \phi$ then $\vdash \odot\phi$

OB-RE If $\vdash \phi \longleftrightarrow \psi$, then $\vdash \odot\phi \longleftrightarrow \odot\psi$.

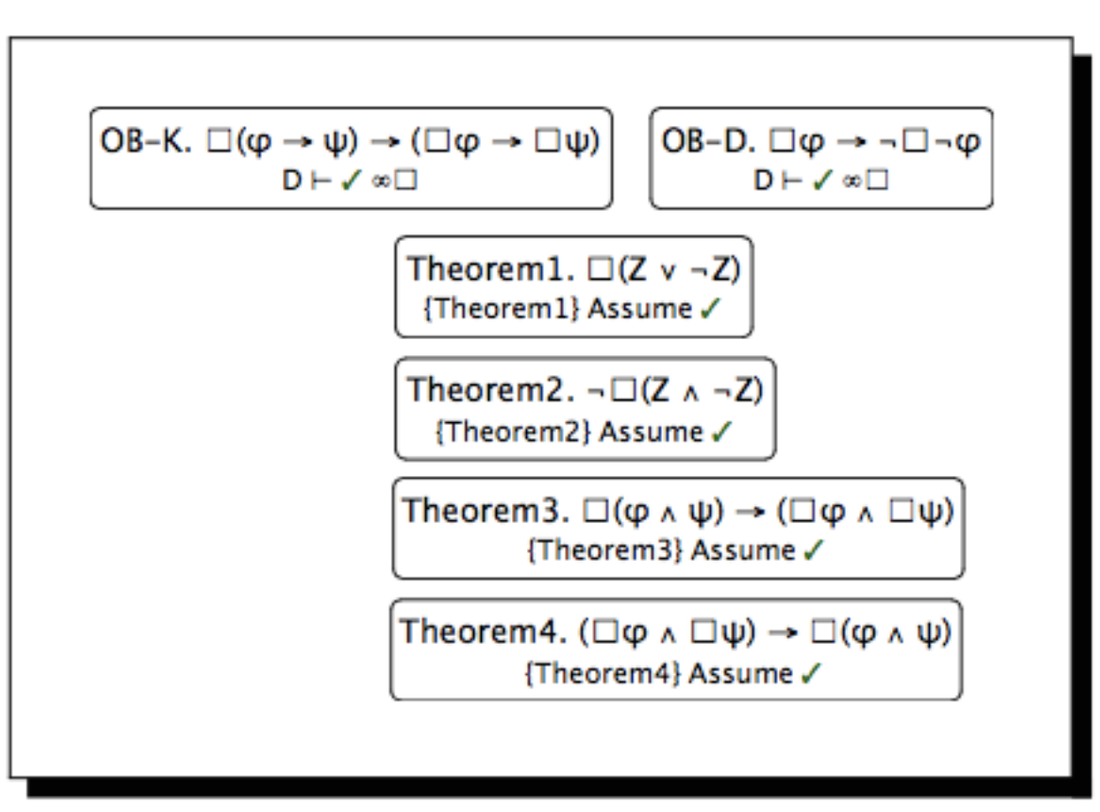


Figure 4.7: The Initial Configuration Upon Opening the File `SDL.slt`

4.4.4.1 Chisholm's Paradox and SDL

There are a host of problems that, together, constitute what is probably a fatal threat to **SDL** as a model of human-level ethical reasoning. We discuss in the present section the first of these problems to hit the "airwaves": Chisholm's Paradox (CP) (Chisholm 1963). CP can be generated in Slate, you we shall see. But before we get to the level of experimentation in Slate, let's understand the scenario that Chisholm's imagined.

Chisholm's clever scenario revolves around the character Jones.[11] It's given that Jones is obligated to go to assist his neighbors, in part because he has promised to do so. The second given fact is that it's obligatory that, if Jones goes to assist his neighbors, he tells them (in advance) that he is coming. In addiiton, and this is the third given, if Jones *doesn't* go to assist his neighbors, it's obligatory that he not tell

[11] We change some particulars to ease exposition; generally, again, follow, the *SEP* entry on deontic logic (recall footnote 10). The core logic mirrors (Chisholm 1963), the original publication.

them that he is coming. The fourth and final given fact is simply that Jones doesn't go to assist his neighbors. (On the way to do so, suppose he comes upon a serious vehicular accident, is proficient in emergency medicine, and (commendably!) seizes the opportunity to save the life (and subsequently monitor) of one of the victims in this accident.) These four givens have been represented in an obvious way within four formula nodes in a Slate file; see Figure 4.8. (Notice that $\Box$ is used in place of $\odot$.) The paradox arises from the fact that Chisholm's quartet of givens, which surely reflect situations that are common in everyday life, in conjunction with the axioms of **SDL**, entail outright contradictions (see Exercise 2 for $\mathbf{D} = \mathbf{SDL}$, in §4.4.4.2).

4.4.4.1 Chisholm's Paradox and SDL

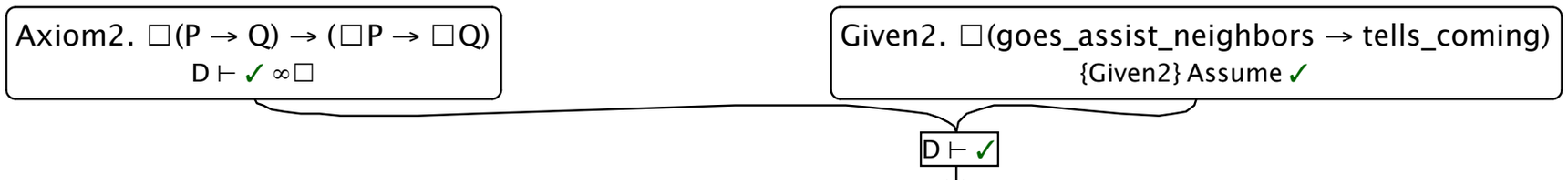
There are a host of problems that, together, constitute what is probably a fatal threat to **SDL** as a model of human-level ethical reasoning. We discuss in the present section the first of these problems to hit the "airwaves": Chisholm's Paradox (CP) (Chisholm 1963). CP can be generated in Slate, you we shall see. But before we get to the level of experimentation in Slate, let's understand the scenario that Chisholm's imagined.

Chisholm's clever scenario revolves around the character Jones.[11] It's given that Jones is obligated to go to assist his neighbors, in part because he has promised to do so. The second given fact is that it's obligatory that, if Jones goes to assist his neighbors, he tells them (in advance) that he is coming. In addiiton, and this is the third given, if Jones *doesn't* go to assist his neighbors, it's obligatory that he not tell

[11] We change some particulars to ease exposition; generally, again, follow, the *SEP* entry on deontic logic (recall footnote 10). The core logic mirrors (Chisholm 1963), the original publication.

them that he is coming. The fourth and final given fact is simply that Jones doesn't go to assist his neighbors. (On the way to do so, suppose he comes upon a serious vehicular accident, is proficient in emergency medicine, and (commendably!) seizes the opportunity to save the life (and subsequently monitor) of one of the victims in this accident.) These four givens have been represented in an obvious way within four formula nodes in a Slate file; see Figure 4.8. (Notice that $\Box$ is used in place of $\odot$.) The paradox arises from the fact that Chisholm's quartet of givens, which surely reflect situations that are common in everyday life, in conjunction with the axioms of **SDL**, entail outright contradictions (see Exercise 2 for $\mathbf{D} = \mathbf{SDL}$, in §4.4.4.2).

Chisholm's Paradox

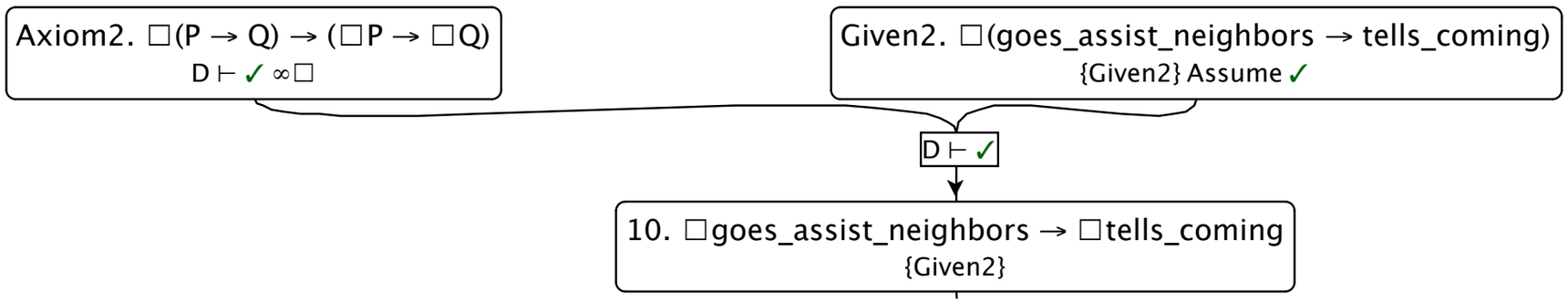


Axiom4. "Modus ponens for provability."
{Axiom4} Assume ✓

Axiom5. "Theorems are obligatory."
{Axiom5} Assume ✓

Axiom1. "All theorems of the propositional calculus."
{Axiom1} Assume ✓

Chisholm's Paradox

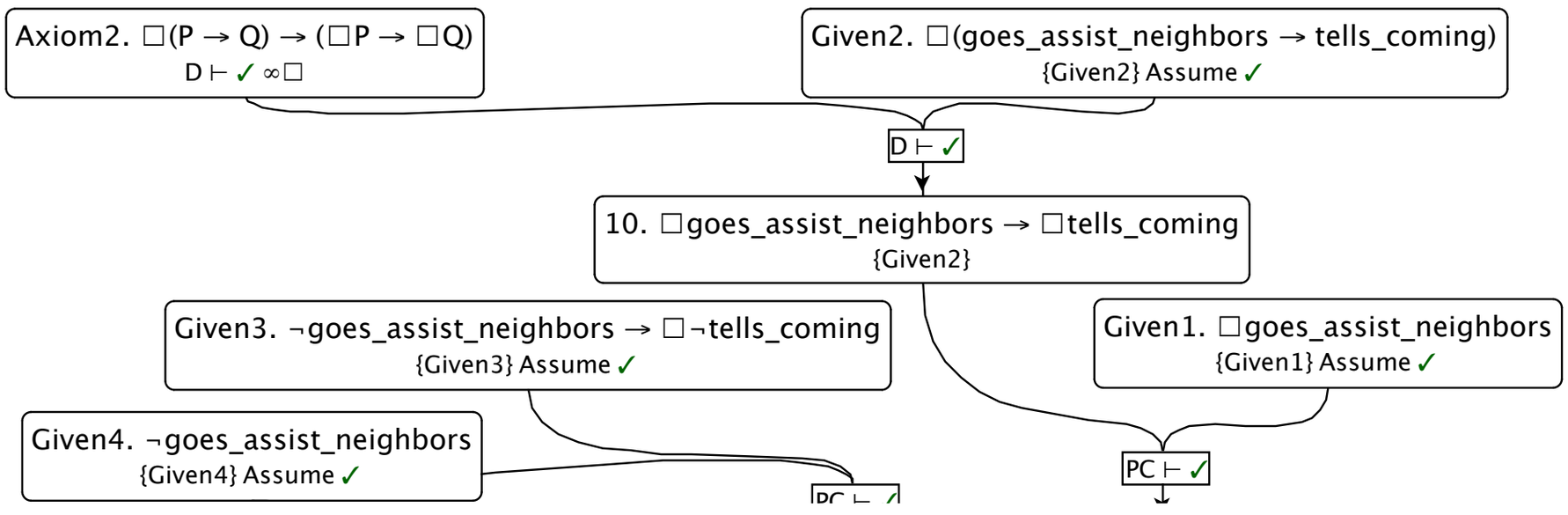


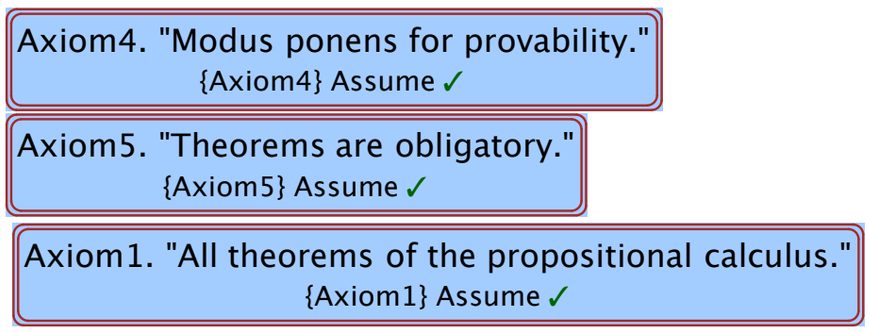
Axiom4. "Modus ponens for provability."
{Axiom4} Assume ✓

Axiom5. "Theorems are obligatory."
{Axiom5} Assume ✓

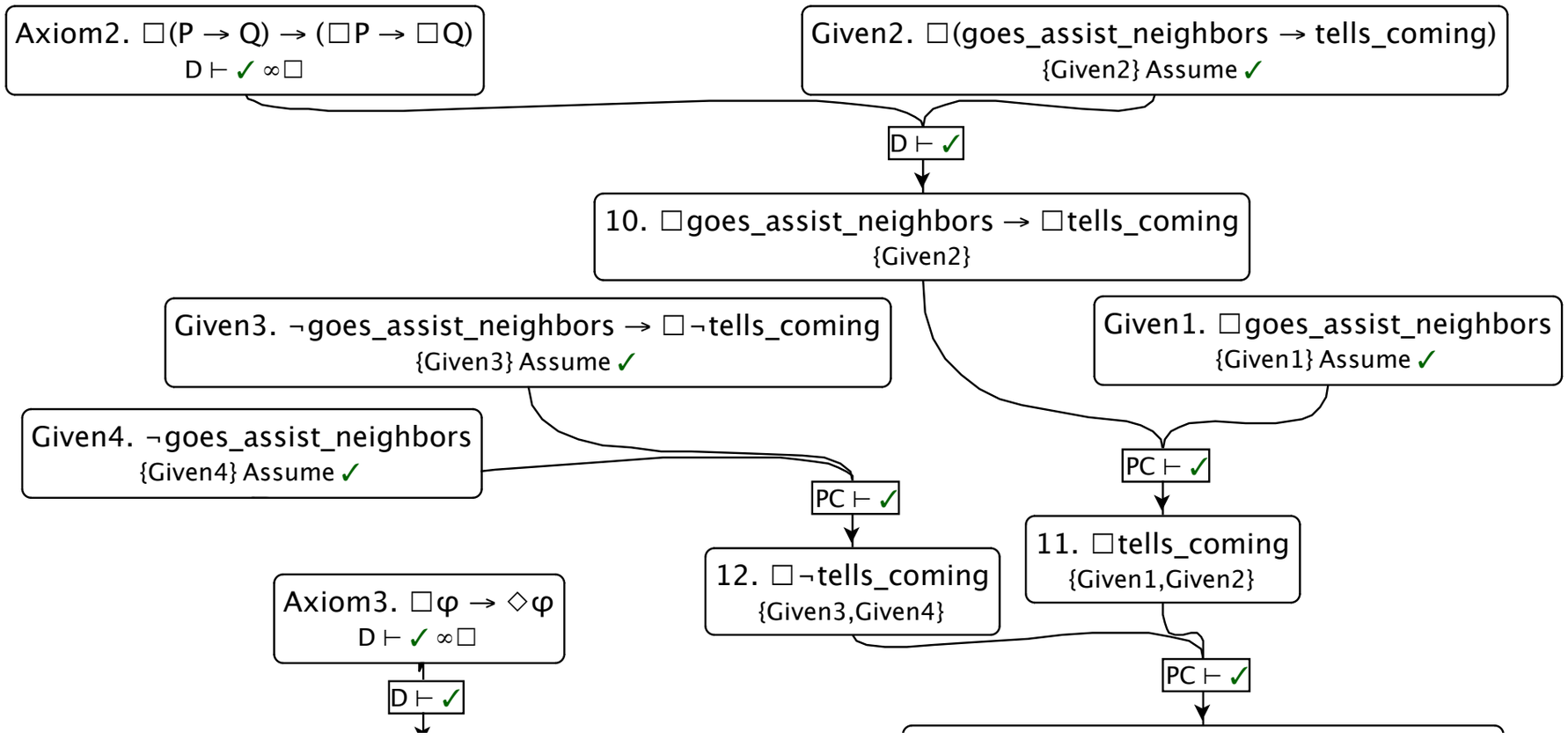
Axiom1. "All theorems of the propositional calculus."
{Axiom1} Assume ✓

Chisholm's Paradox





Chisholm's Paradox

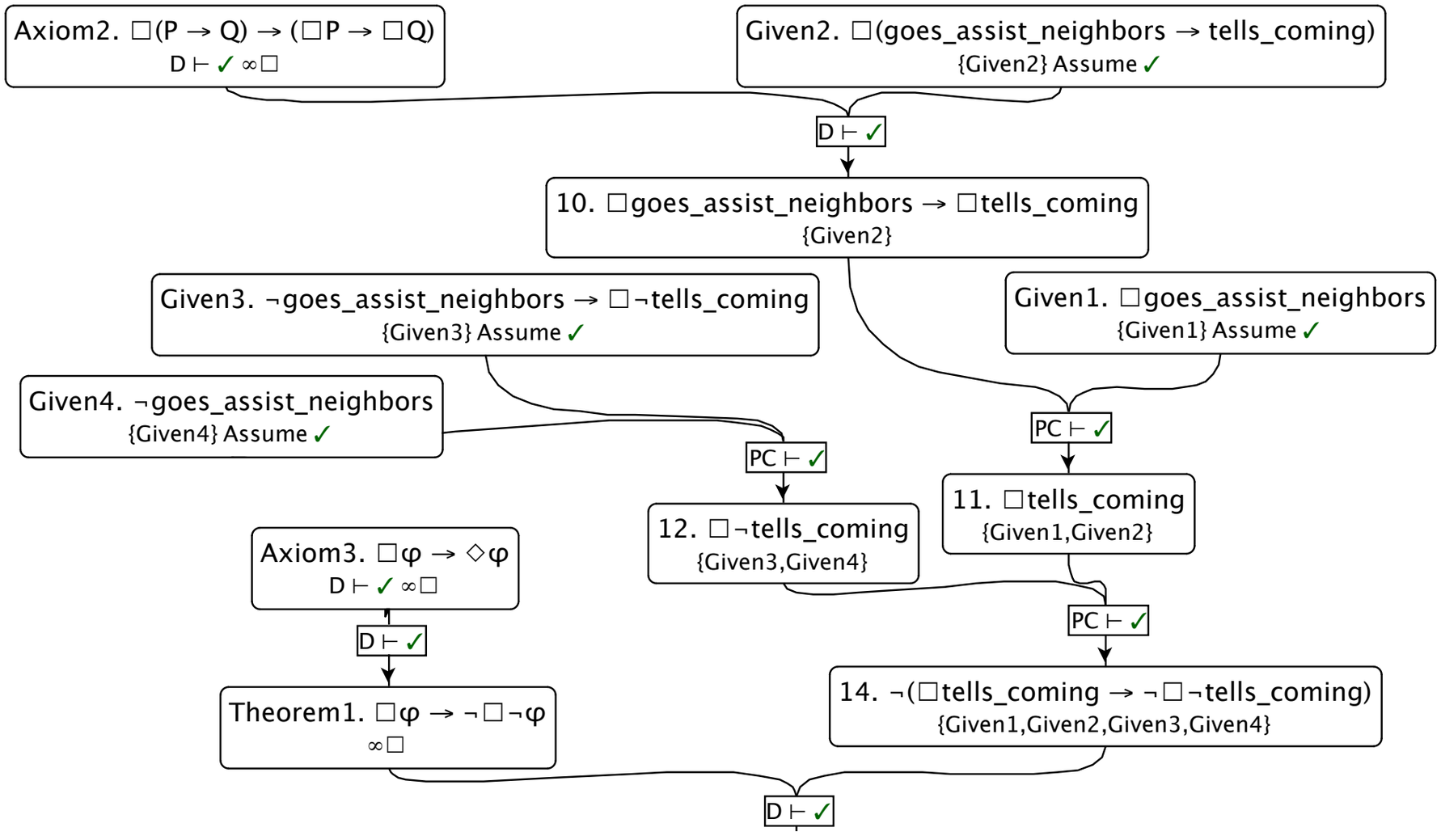


Axiom4. "Modus ponens for provability."
{Axiom4} Assume ✓

Axiom5. "Theorems are obligatory."
{Axiom5} Assume ✓

Axiom1. "All theorems of the propositional calculus."
{Axiom1} Assume ✓

Chisholm's Paradox

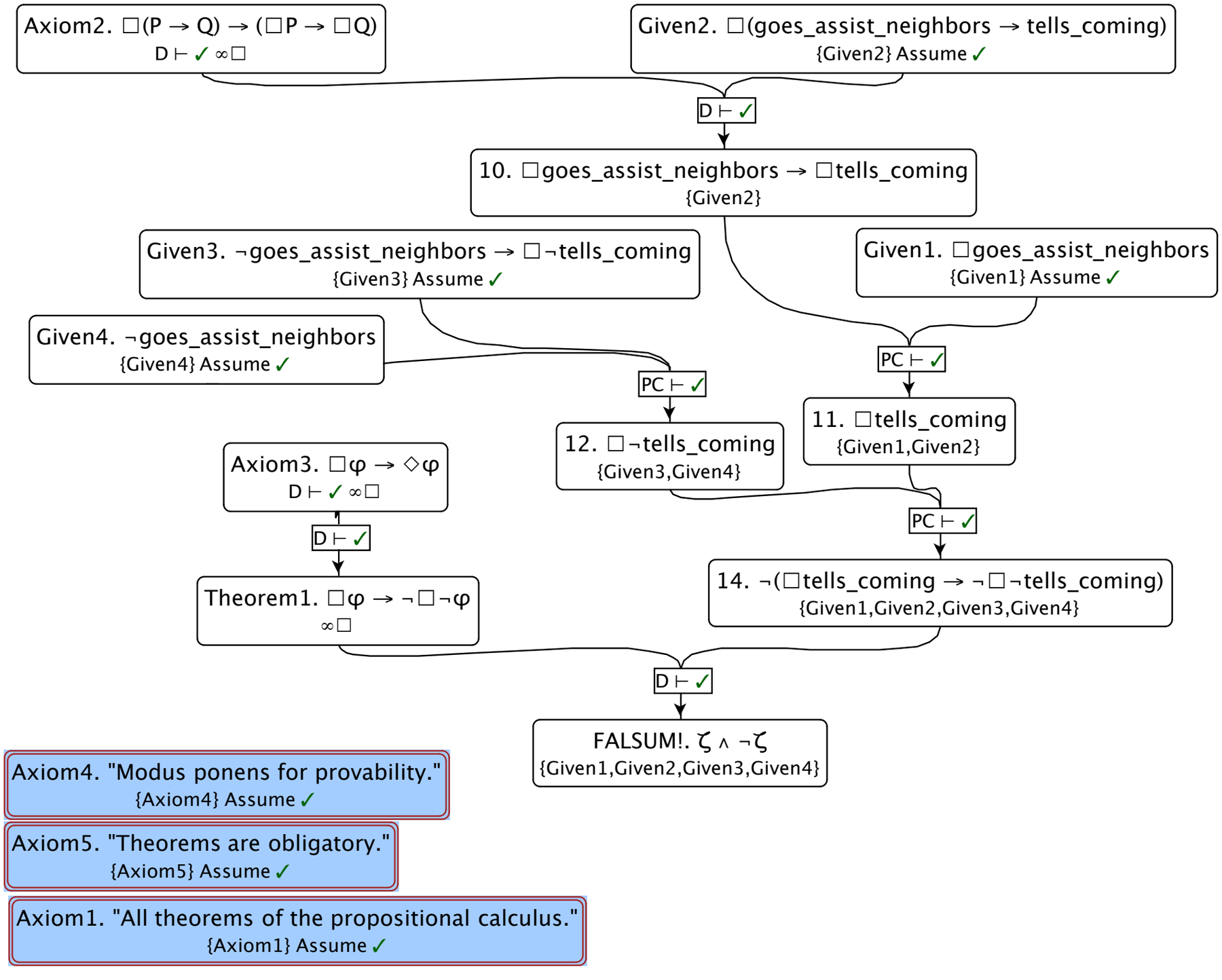


Axiom4. "Modus ponens for provability."
{Axiom4} Assume ✓

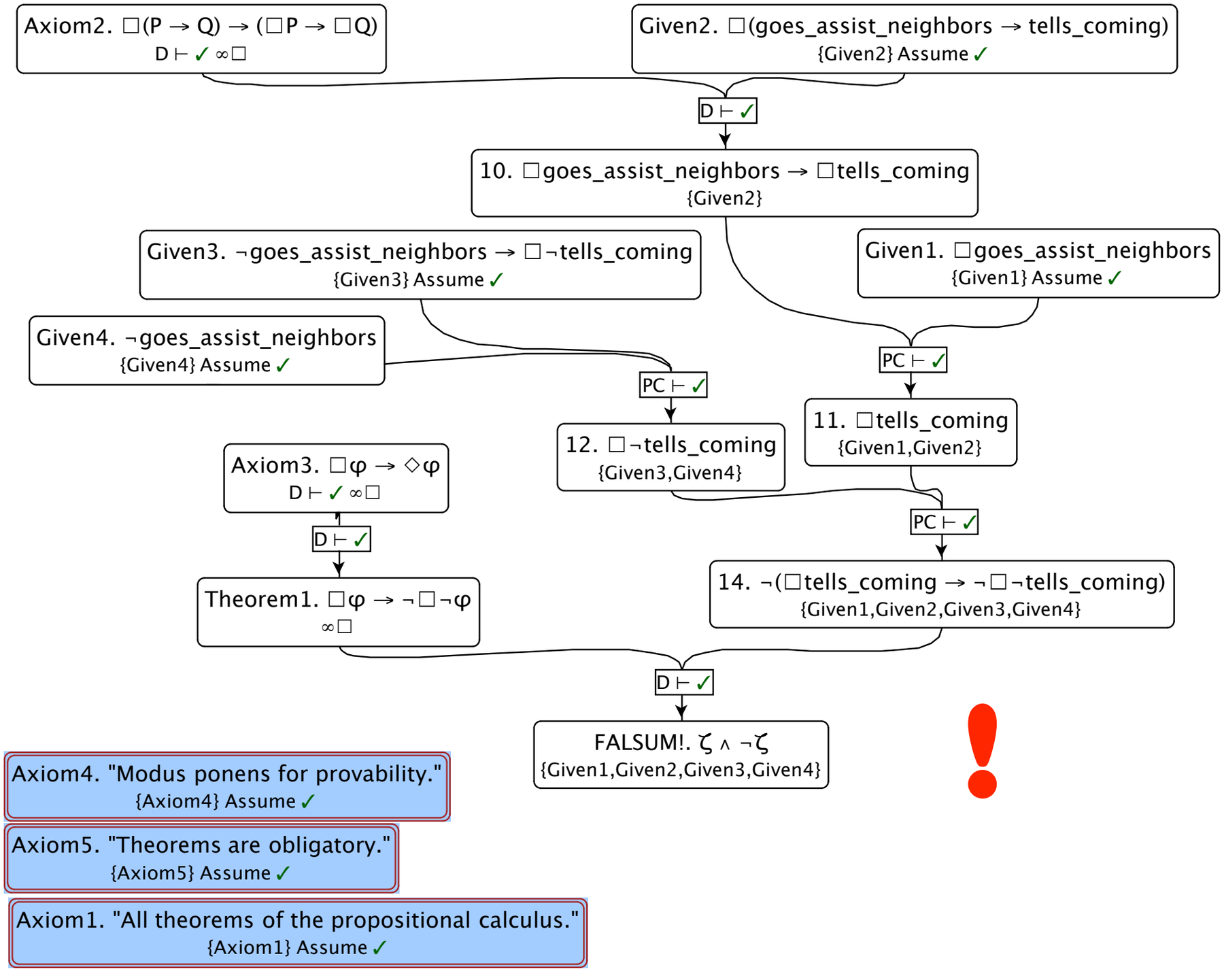
Axiom5. "Theorems are obligatory."
{Axiom5} Assume ✓

Axiom1. "All theorems of the propositional calculus."
{Axiom1} Assume ✓

Chisholm's Paradox



Chisholm's Paradox



☑ ChisholmsParadox

Here you are asked to build a proof that confirms *Chisholm's Paradox*. This paradox is that from a particular representation in **D** (= Standard Deontic Logic (SDL)) of four seemingly innocuous givens, a contradiction $\zeta \wedge \neg\zeta$ can be deduced. (Your instructor should have covered this in class, and may well have supplied a proof of CP.) The four givens are based on the story of a character Jones, who is obligated to go to assist his neighbors (move to a different domicile, e.g.). It would be wrong of him to show up unannounced, though; so if he goes to assist them, it ought to be that he tells them he's coming. In addition, if it's not the case that Jones goes to assist them, then it ought to be that it not be the case that he tells them he is coming. Finally, as a matter of fact, it's not case the Jones goes to assist (because on the way he comes across a car accident, and has an opportunity to save one of the victims).

Fortunately, the RAIR Lab's modern cognitive calculus $\mathcal{DCEC}^*$ allows Chisholm's Paradox to be avoided. A recent paper explaining the use by an ethically correct AI of this calculus is available here.

Your finished proof is allowed to make use of the PC provabiity oracle, but of no other oracle.

Deadline 22 Apr 2020 23:59:00 EDT

SDL's = D's Problems

Don't Stop Here …

The Free Choice Permission Paradox (Ross)

1. "You may either sleep on the sofa bed or the guest bed."
{1} Assume ✓

2. "Therefore: You may sleep on the sofa bed, and you may sleep on the guest bed."
{2} Assume ✓

The Free Choice Permission Paradox (Ross)

1'. ◇(sofa-bed ∨ guest-bed)
{1'} Assume ✓

D ⊢ ✗

2'. ◇sofa-bed ∧ ◇guest-bed
{1'}

1. "You may either sleep on the sofa bed or the guest bed."
{1} Assume ✓

2. "Therefore: You may sleep on the sofa bed, and you may sleep on the guest bed."
{2} Assume ✓

The Free Choice Permission Paradox (Ross)

1'. ◇(sofa-bed ∨ guest-bed)
{1'} Assume ✓

D ⊢ ✗

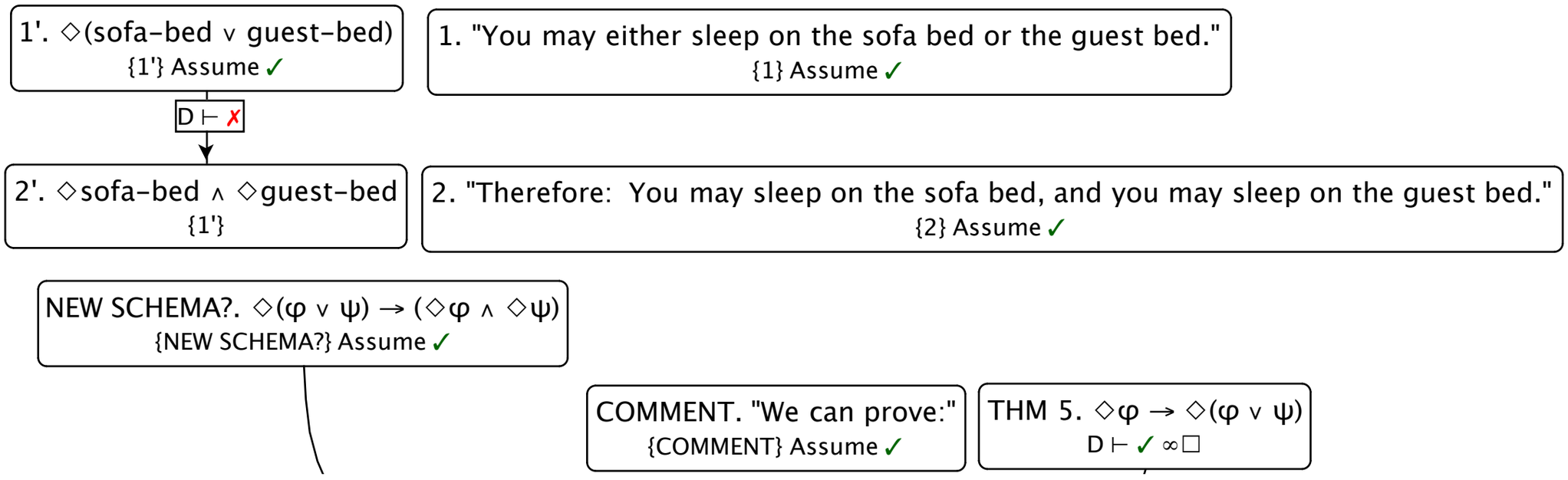
2'. ◇sofa-bed ∧ ◇guest-bed
{1'}

1. "You may either sleep on the sofa bed or the guest bed."
{1} Assume ✓

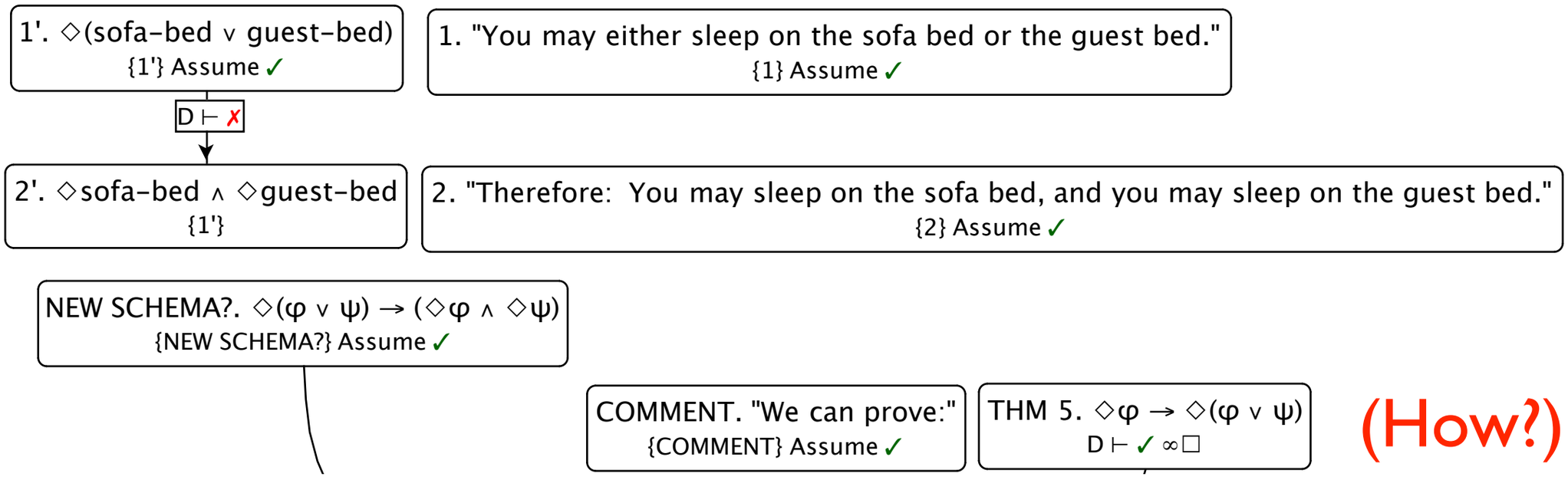
2. "Therefore: You may sleep on the sofa bed, and you may sleep on the guest bed."
{2} Assume ✓

NEW SCHEMA?. ◇(φ ∨ ψ) → (◇φ ∧ ◇ψ)
{NEW SCHEMA?} Assume ✓

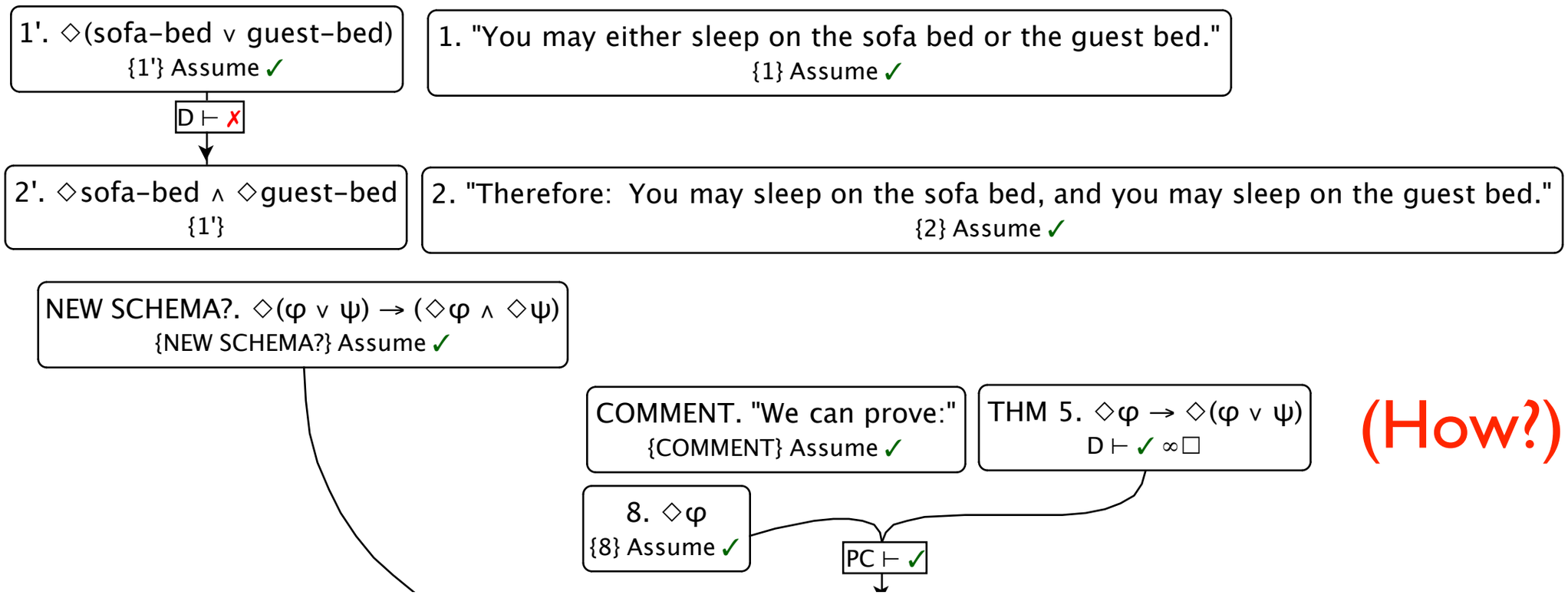
The Free Choice Permission Paradox (Ross)



The Free Choice Permission Paradox (Ross)

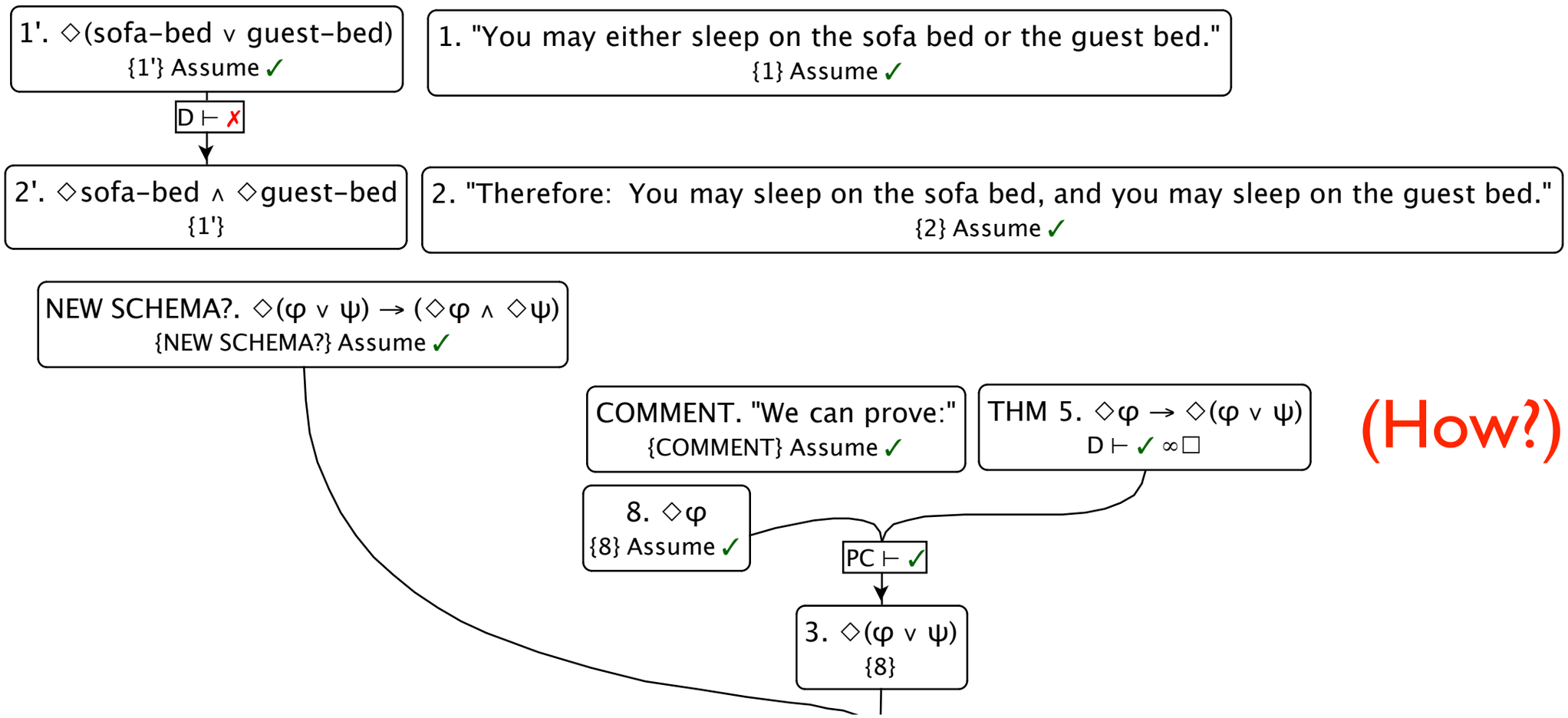


The Free Choice Permission Paradox (Ross)



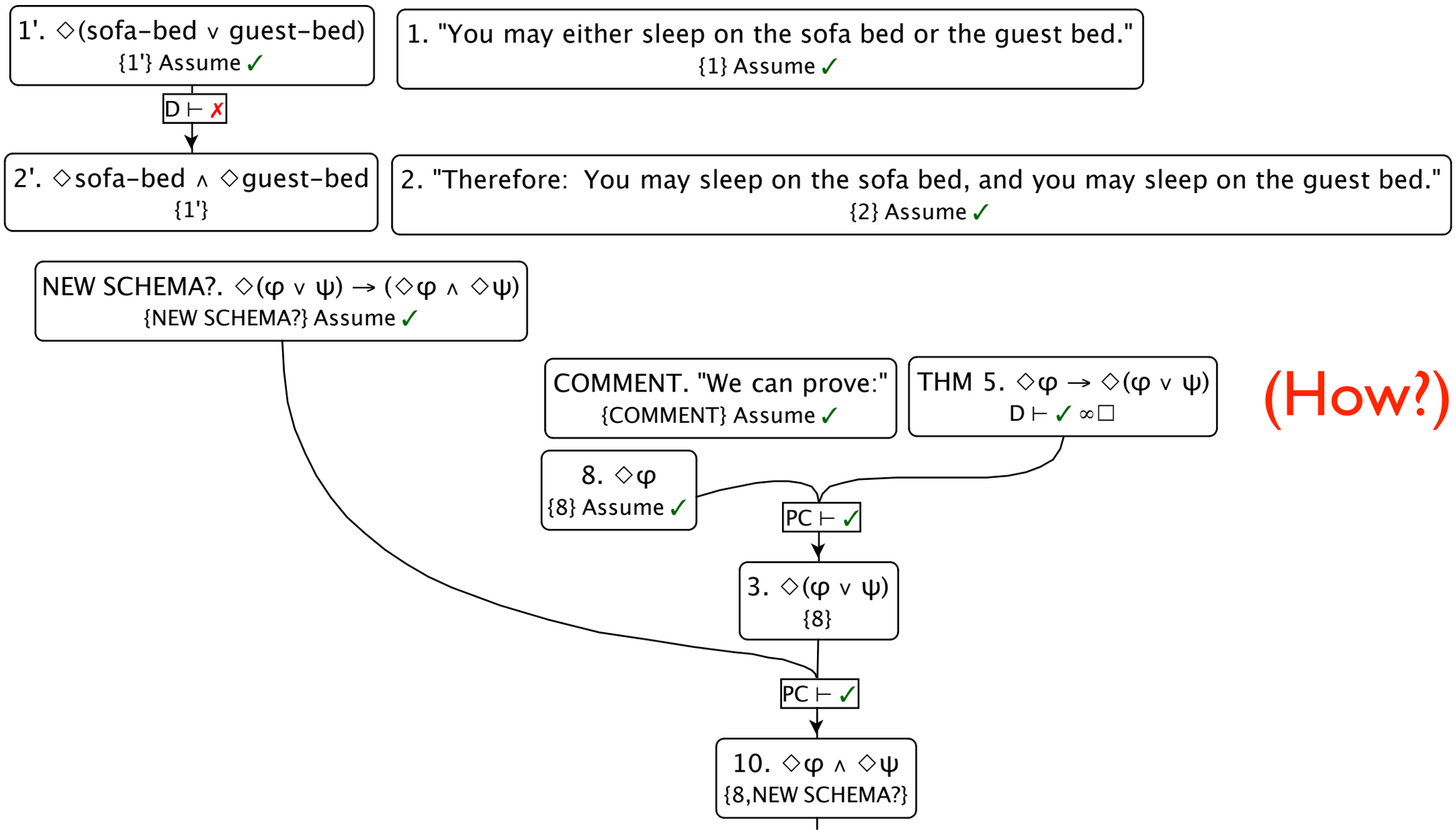
(How?)

The Free Choice Permission Paradox (Ross)



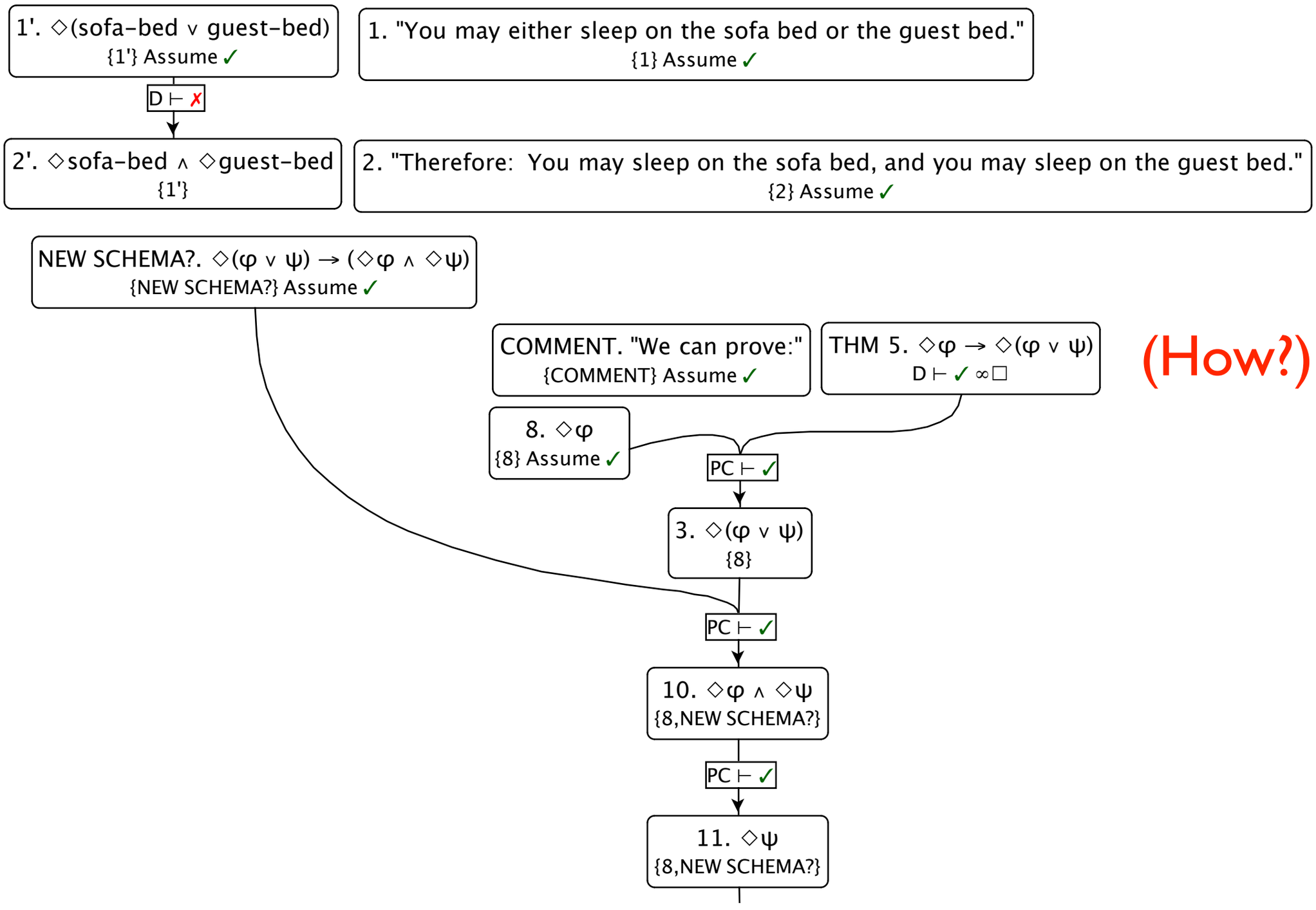
(How?)

The Free Choice Permission Paradox (Ross)



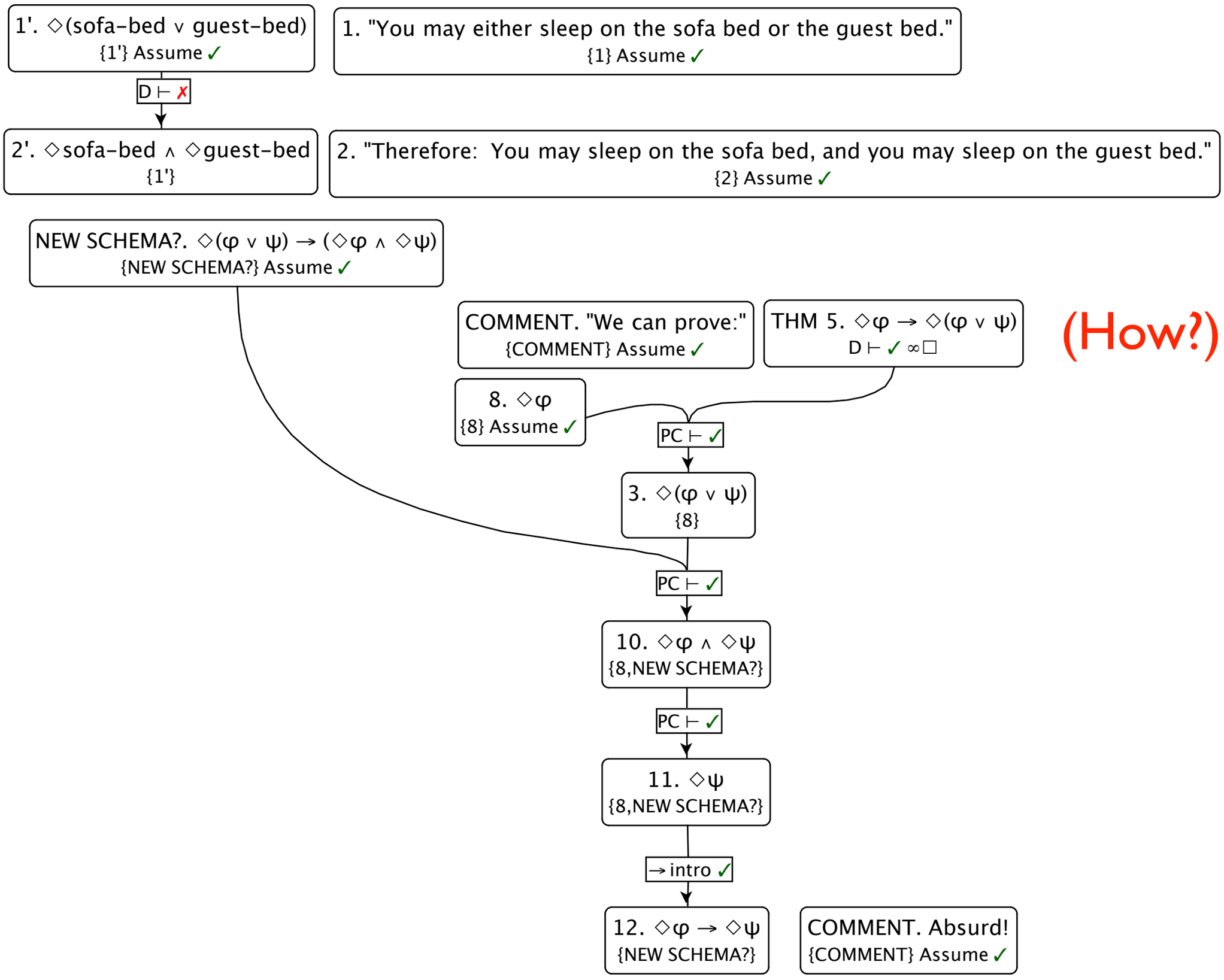
(How?)

The Free Choice Permission Paradox (Ross)



(How?)

The Free Choice Permission Paradox (Ross)



(How?)

“Computational logician, sorry, back to your drawing board to find a logic that works with The Four Steps!”

☐ TheFreeChoicePermissionParadox

Producing a valid proof in this problem will enable you to understand The Free Choice Permission Paradox (FCPP), discovered in 1941 by Ross ("Imperatives and Logic," *Theoria* **7**: 53–71). Given that the proof in question yields an absurdity, FCPP can be taken to show that **SDL** (Standard Deontic Logic) = **D** leads to inconsistency when applied; or, put in AI terms, you wouldn't want a robot to base its ethical decision-making on **D**! Fortunately, the RAIR Lab's modern cognitive calculus $\mathcal{DCEC}^*$ allows FCPP to be avoided. (A recent paper explaining the use by an ethically correct AI of this calculus is available here.)

Here's the paradox. Suppose that you travel to visit a friend, arrive late at night, and are weary. Your friend says hospitably: "You may either sleep on the sofa-bed or sleep on the guest-room bed." (1) From this statement it follows that you are permitted to sleep on the sofa-bed, and you are permitted to sleep on the guest-room bed. (2) In **D**, this pair gets symbolized like this:

(1')
$\Diamond(sofabed \vee guestbed)$

(2')
$\Diamond sofabed \wedge \Diamond guestbed$

But (2') doesn't follow deductively from (1') in **D**, as a call to the provability oracle for **D** in the HyperSlate™ file for this problem confirms. A suggested repair is to add to **D** the schema

$$\Diamond(\phi \vee \psi) \rightarrow (\Diamond\phi \wedge \Diamond\psi),$$

but as your proof will (hopefully) show, this addition allows a proof of the absurd theorem that if anything is morally perimssible, everything is!

Your finished proof is allowed to make use of the PC provabiity oracle, but of no other oracle. (No deadline for now.)