Beyond Worlds and Accessibility

Part 1

Wesley Holliday
Department of Philosophy &
Group in Logic and the Methodology of Science
University of California, Berkeley

EASSLC 2014, July 2-8

Course Description

The starting point of this course is standard possible-world semantics for epistemic, doxastic, and other modal logics, based on models with "possible worlds and "accessibility relations."

Course Description

The starting point of this course is standard possible-world semantics for epistemic, doxastic, and other modal logics, based on models with "possible worlds and "accessibility relations."

We will explore a number of ways of enriching or departing from these models, motivated by philosophical considerations.

Course Description

The starting point of this course is standard possible-world semantics for epistemic, doxastic, and other modal logics, based on models with "possible worlds and "accessibility relations."

We will explore a number of ways of enriching or departing from these models, motivated by philosophical considerations.

Topics will include:

Course Description

The starting point of this course is standard possible-world semantics for epistemic, doxastic, and other modal logics, based on models with "possible worlds and "accessibility relations."

We will explore a number of ways of enriching or departing from these models, motivated by philosophical considerations.

Topics will include:

- models that add more structure to capture substantive theories of knowledge (in the style of Dretske, Nozick, and others);

Course Description

The starting point of this course is standard possible-world semantics for epistemic, doxastic, and other modal logics, based on models with "possible worlds and "accessibility relations."

We will explore a number of ways of enriching or departing from these models, motivated by philosophical considerations.

Topics will include:

- models that add more structure to capture substantive theories of knowledge (in the style of Dretske, Nozick, and others);
- models that replace total possible worlds with partial "possibilities" (in the style of Humberstone);

Course Description

The starting point of this course is standard possible-world semantics for epistemic, doxastic, and other modal logics, based on models with "possible worlds and "accessibility relations."

We will explore a number of ways of enriching or departing from these models, motivated by philosophical considerations.

Topics will include:

- models that add more structure to capture substantive theories of knowledge (in the style of Dretske, Nozick, and others);
- models that replace total possible worlds with partial "possibilities" (in the style of Humberstone);
- models that replace total possible worlds with partial "situations" (in the style of Barwise and Perry).

Course Description

The starting point of this course is standard possible-world semantics for epistemic, doxastic, and other modal logics, based on models with "possible worlds and "accessibility relations."

We will explore a number of ways of enriching or departing from these models, motivated by philosophical considerations.

Topics will include:

- models that add more structure to capture substantive theories of knowledge (in the style of Dretske, Nozick, and others);
- models that replace total possible worlds with partial "possibilities" (in the style of Humberstone);
- models that replace total possible worlds with partial "situations" (in the style of Barwise and Perry).

In each case, we will study the logical properties of these classes of models, illuminating the philosophical issues at stake.

Outline of the Course

- Day 1 - Introduction and Background
- Day 1-3 - RA and Subjunctivist Semantics
- Day 3 - Set-Selection Function Semantics
- Day 4 - Possibility Semantics
- Day 5 - Situation Semantics

For the slides and references for this course, go to

http://philosophy.berkeley.edu/holliday

then go to Short Courses – EASLLC 2014.

The slides are hyperlinked, so click on references to find them on the web.

For the slides and references for this course, go to

http://philosophy.berkeley.edu/holliday

then go to Short Courses – EASLLC 2014.

The slides are hyperlinked, so click on references to find them on the web.

The logical material today is primarily based on the following works:

- "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism," *Journal of Philosophical Logic*, 2014.
- "Epistemic Logic and Epistemology," *Handbook of Formal Philosophy*, forthcoming.

For the slides and references for this course, go to

http://philosophy.berkeley.edu/holliday

then go to Short Courses – EASLLC 2014.

The slides are hyperlinked, so click on references to find them on the web.

The logical material today is primarily based on the following works:

- "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism," *Journal of Philosophical Logic*, 2014.
- "Epistemic Logic and Epistemology," *Handbook of Formal Philosophy*, forthcoming.

Recent work on related topics comes from researchers in **China**! See, e.g., papers by Zhaoqing Xu and Chenwei Shi on the webpage given above.

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

(2) it can handle concrete examples discussed in the philosophical literature;

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

(2) it can handle concrete examples discussed in the philosophical literature;

(3) it goes beyond particular examples to provide a systematic and general view of the topic;

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

(2) it can handle concrete examples discussed in the philosophical literature;

(3) it goes beyond particular examples to provide a systematic and general view of the topic;

(4) it leads to philosophically-relevant discoveries that would be difficult to make by non-formal methods alone;

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

(2) it can handle concrete examples discussed in the philosophical literature;

(3) it goes beyond particular examples to provide a systematic and general view of the topic;

(4) it leads to philosophically-relevant discoveries that would be difficult to make by non-formal methods alone;

(5) it leads to the development of new views that solve previous philosophical problems.

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

(2) it can handle concrete examples discussed in the philosophical literature;

(3) it goes beyond particular examples to provide a systematic and general view of the topic;

(4) it leads to philosophically-relevant discoveries that would be difficult to make by non-formal methods alone;

(5) it leads to the development of new views that solve previous philosophical problems.

(1) and (2) address the worry that by formalizing we may "change the subject."

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

(2) it can handle concrete examples discussed in the philosophical literature;

(3) it goes beyond particular examples to provide a systematic and general view of the topic;

(4) it leads to philosophically-relevant discoveries that would be difficult to make by non-formal methods alone;

(5) it leads to the development of new views that solve previous philosophical problems.

(1) and (2) address the worry that by formalizing we may "change the subject." (3) - (5) address the question, "What do we get out of this?"

From my "Summary of "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism"," *Logic Across the University*:

An application of **formal methods in philosophy** gains value insofar as:

(1) it is faithful to the philosophical views being formalized;

(2) it can handle concrete examples discussed in the philosophical literature;

(3) it goes beyond particular examples to provide a systematic and general view of the topic;

(4) it leads to philosophically-relevant discoveries that would be difficult to make by non-formal methods alone;

(5) it leads to the development of new views that solve previous philosophical problems.

(1) and (2) address the worry that by formalizing we may "change the subject." (3) - (5) address the question, "What do we get out of this?" Of course, there are other features that contribute to the value of an application of formal methods, as well as features that detract from it.

Basic Epistemic Logic & What One Knows

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

A: $v \in R_a(w)$ means that everything a knows in w is true in v.

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

A: $v \in R_a(w)$ means that everything a knows in w is true in v.

Q: So you're explaining when '$K_a\varphi$' is true in terms of *knowledge* itself?

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

A: $v \in R_a(w)$ means that everything a knows in w is true in v.

Q: So you're explaining when '$K_a\varphi$' is true in terms of *knowledge* itself?

A: Yes.

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

A: $v \in R_a(w)$ means that everything a knows in w is true in v.

Q: So you're explaining when '$K_a\varphi$' is true in terms of *knowledge* itself?

A: Yes.

Q: Isn't that a problem? Somehow cheating or conceptually circular?

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

A: $v \in R_a(w)$ means that everything a knows in w is true in v.

Q: So you're explaining when '$K_a\varphi$' is true in terms of *knowledge* itself?

A: Yes.

Q: Isn't that a problem? Somehow cheating or conceptually circular?

A: No!

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

A: $v \in R_a(w)$ means that everything a knows in w is true in v.

Q: So you're explaining when '$K_a\varphi$' is true in terms of *knowledge* itself?

A: Yes.

Q: Isn't that a problem? Somehow cheating or conceptually circular?

A: No! The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

Basic Epistemic Logic & What One Knows

Model what an agent knows by what is true throughout a set $R(w)$ of possibilities that are "compatible" with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

A: The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

Basic Epistemic Logic & What One Knows

Model what an agent knows by what is true throughout a set $R(w)$ of possibilities that are "compatible" with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

A: The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

Qualification: two aspects of these models are not neutral with respect to different substantive accounts of what it takes to know something.

Basic Epistemic Logic & What One Knows

Model what an agent knows by what is true throughout a set $R(w)$ of possibilities that are "compatible" with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

A: The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

Qualification: two aspects of these models are not neutral with respect to different substantive accounts of what it takes to know something.

Most obviously, if we impose conditions on R_a beyond reflexivity $[w \in R_a(w)]$ such as symmetry $[v \in R_a(w) \Rightarrow w \in R_a(v)]$, we build in substantive views (in this case, that $\neg\varphi \to K_a\neg K_a\varphi$ should be valid) that conflict with some accounts of knowledge (ditto for 'information').

Basic Epistemic Logic & What One Knows

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Less obviously, while the $\Rightarrow$ direction of (1) follows immediately from the stipulation that $\underline{v \in R_a(w) \text{ means everything } a \text{ knows in } w \text{ is true in } v}$, the $\Leftarrow$ direction of (1) does *not* follow just from that stipulation.

Basic Epistemic Logic & What One Knows

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Less obviously, while the $\Rightarrow$ direction of (1) follows immediately from the stipulation that $\underline{v \in R_a(w) \text{ means everything } a \text{ knows in } w \text{ is true in } v}$, the $\Leftarrow$ direction of (1) does *not* follow just from that stipulation.

In fact, to add the $\Leftarrow$ direction is just to assume the principle that **knowledge is closed under known implication**:

$$(K_a\varphi \wedge K_a(\varphi \to \psi)) \to K_a\psi$$

Basic Epistemic Logic & What One Knows

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Less obviously, while the $\Rightarrow$ direction of (1) follows immediately from the stipulation that $\underline{v \in R_a(w) \text{ means everything } a \text{ knows in } w \text{ is true in } v}$, the $\Leftarrow$ direction of (1) does *not* follow just from that stipulation.

In fact, to add the $\Leftarrow$ direction is just to assume the principle that **knowledge is closed under known implication**:

$$(K_a\varphi \wedge K_a(\varphi \to \psi)) \to K_a\psi$$

If $\mathcal{M}, w \vDash K_a\varphi$ and $\mathcal{M}, w \vDash K_a(\varphi \to \psi)$, then by the $\Rightarrow$ direction of (1), $\forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi$ and $\mathcal{M}, v \vDash \varphi \to \psi$, so $\mathcal{M}, v \vDash \psi$.

Basic Epistemic Logic & What One Knows

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Less obviously, while the $\Rightarrow$ direction of (1) follows immediately from the stipulation that $\underline{v \in R_a(w) \text{ means everything } a \text{ knows in } w \text{ is true in } v}$, the $\Leftarrow$ direction of (1) does *not* follow just from that stipulation.

In fact, to add the $\Leftarrow$ direction is just to assume the principle that **knowledge is closed under known implication**:

$$(K_a\varphi \wedge K_a(\varphi \to \psi)) \to K_a\psi$$

If $\mathcal{M}, w \vDash K_a\varphi$ and $\mathcal{M}, w \vDash K_a(\varphi \to \psi)$, then by the $\Rightarrow$ direction of (1), $\forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi$ and $\mathcal{M}, v \vDash \varphi \to \psi$, so $\mathcal{M}, v \vDash \psi$. The $\Leftarrow$ direction then yields the conclusion that $\mathcal{M}, w \vDash K_a\psi$.

Basic Epistemic Logic & What One Knows

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Less obviously, while the $\Rightarrow$ direction of (1) follows immediately from the stipulation that $\underline{v \in R_a(w) \text{ means everything } a \text{ knows in } w \text{ is true in } v}$, the $\Leftarrow$ direction of (1) does *not* follow just from that stipulation.

In fact, to add the $\Leftarrow$ direction is just to assume the principle that **knowledge is closed under known implication**:

$$(K_a\varphi \wedge K_a(\varphi \to \psi)) \to K_a\psi$$

If $\mathcal{M}, w \vDash K_a\varphi$ and $\mathcal{M}, w \vDash K_a(\varphi \to \psi)$, then by the $\Rightarrow$ direction of (1), $\forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi$ and $\mathcal{M}, v \vDash \varphi \to \psi$, so $\mathcal{M}, v \vDash \psi$. The $\Leftarrow$ direction then yields the conclusion that $\mathcal{M}, w \vDash K_a\psi$.

The $\Leftarrow$ direction conflicts with some accounts of knowledge I will discuss.

Epistemic Logic & What One Knows

A: The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

Those who criticize basic epistemic models for not giving more substantive insight into what it takes to know something about the world (or what it takes to get information from the world) are missing the point.

Epistemic Logic & What One Knows

A: The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

The beauty of it is that basic epistemic logic can nevertheless give illuminating analyses of a wide range of epistemic phenomena, especially in *multi-agent* settings with information change (think Muddy Children).

Epistemic Logic & What One Knows

A: The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

The beauty of it is that basic epistemic logic can nevertheless give illuminating analyses of a wide range of epistemic phenomena, especially in *multi-agent* settings with information change (think Muddy Children).

For many examples, see:

R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. 1995. *Reasoning about Knowledge.*

J.J.C. Meyer and W. van der Hoek. 1995. *Epistemic Logic for AI and Computer Science.*

H. van Ditmarsch, W. van der Hoek, and B. Kooi. 2008. *Dynamic Epistemic Logic.*

J. van Benthem. 2011. *Logical Dynamics of Information and Interaction.*

Epistemic Accessibility vs. Indistinguishability

Distinction: accessibility vs. other notions of **indistinguishability**.

Epistemic Accessibility vs. Indistinguishability

Distinction: accessibility vs. other notions of **indistinguishability**.

Suppose that we replace R by a binary relation E on W, where our intuitive interpretation is that wEv holds "iff the subject's perceptual experience and memory" in scenario v "exactly match his perceptual experience and memory" in scenario w (Lewis, "Elusive Knowledge").

Epistemic Accessibility vs. Indistinguishability

Distinction: accessibility vs. other notions of **indistinguishability**.

Suppose that we replace R by a binary relation E on W, where our intuitive interpretation is that wEv holds "iff the subject's perceptual experience and memory" in scenario v "exactly match his perceptual experience and memory" in scenario w (Lewis, "Elusive Knowledge").

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in E_a(w) : \mathcal{M}, v \vDash \varphi. \tag{2}$$

Hence the agent knows φ in w just in case φ is true in all scenarios that are experientially indistinguishable from w for the agent.

Epistemic Accessibility & Indistinguishability

Important differences between the picture with E and the one with R:

Epistemic Accessibility & Indistinguishability

Important differences between the picture with E and the one with R:

- The epistemic model with E does not simply represent the content of one's knowledge; rather, it commits us to a particular view of the conditions under which an agent has knowledge, specified in terms of perceptual experience and memory.

Epistemic Accessibility & Indistinguishability

Important differences between the picture with E and the one with R:

- The epistemic model with E does not simply represent the content of one's knowledge; rather, it commits us to a particular view of the conditions under which an agent has knowledge, specified in terms of perceptual experience and memory.
- It is plausible that E has certain properties, such as *symmetry* (wEv iff vEw), which are questionable as properties of R.

Epistemic Accessibility & Indistinguishability

Important differences between the picture with E and the one with R:

- The epistemic model with E does not simply represent the content of one's knowledge; rather, it commits us to a particular view of the conditions under which an agent has knowledge, specified in terms of perceptual experience and memory.
- It is plausible that E has certain properties, such as *symmetry* (wEv iff vEw), which are questionable as properties of R.

  Since the properties of the relation will determine valid principles for the knowledge operator K, we must be clear about which interpretation of the relation we adopt.

"Considering Possible"

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

While one may read wRv as

- "**for all the agent knows** in w, scenario v is the scenario she is in,"

one should *not* read wRv as

- "in w, the agent **considers** scenario v **possible**,"

where the latter suggest a subjective psychological notion.

"Considering Possible"

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

While one may read wRv as

- "**for all the agent knows** in w, scenario v is the scenario she is in,"

one should *not* read wRv as

- "in w, the agent **considers** scenario v **possible**,"

where the latter suggest a subjective psychological notion.

An agent may not subjectively consider it possible that his friend, whom he has regarded for years as his most trusted ally, has betrayed him.

"Considering Possible"

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

While one may read wRv as

- "**for all the agent knows** in w, scenario v is the scenario she is in,"

one should *not* read wRv as

- "in w, the agent **considers** scenario v **possible**,"

where the latter suggest a subjective psychological notion.

An agent may not subjectively consider it possible that his friend, whom he has regarded for years as his most trusted ally, has betrayed him. It obviously does not follow that he *knows* that his friend has not betrayed him, as it would according to the subjective reading of R with (1).

Basic Epistemic Logic & What One Knows

Model what an agent a knows by what is true throughout a set $R_a(w)$ of possibilities that are compatible with her knowledge in world w:

$$\mathcal{M}, w \vDash K_a\varphi \Leftrightarrow \forall v \in R_a(w) : \mathcal{M}, v \vDash \varphi. \tag{1}$$

Q: what exactly does it mean for v to be "compatible"?

A: $v \in R_a(w)$ means that everything a knows in w is true in v.

Q: So you're explaining when '$K_a\varphi$' is true in terms of *knowledge* itself?

A: Yes.

Q: Isn't that a problem? Somehow cheating or conceptually circular?

A: No! The point of basic epistemic models is to represent the *content* of an agent's knowledge—*what the agent knows*—not to give a substantive account of *what it takes to know something* in other terms.

Epistemology & What It Takes to Know

Notice that the way of modeling the *content* of an agent's knowledge in basic epistemic-logical models gives no instructions re the question:

> In general, if one wants to *know* that something is true—not just believe or wish it to be true—what does it take to achieve that?

Epistemology & What It Takes to Know

Notice that the way of modeling the *content* of an agent's knowledge in basic epistemic-logical models gives no instructions re the question:

> In general, if one wants to *know* that something is true—not just believe or wish it to be true—what does it take to achieve that?

Philosophers have long tried to provide some guidance here.

Epistemology & What It Takes to Know

Notice that the way of modeling the *content* of an agent's knowledge in basic epistemic-logical models gives no instructions re the question:

> In general, if one wants to *know* that something is true—not just believe or wish it to be true—what does it take to achieve that?

Philosophers have long tried to provide some guidance here. (Let us consider empirical knowledge.)

Epistemology & What It Takes to Know

Notice that the way of modeling the *content* of an agent's knowledge in basic epistemic-logical models gives no instructions re the question:

> In general, if one wants to *know* that something is true—not just believe or wish it to be true—what does it take to achieve that?

Philosophers have long tried to provide some guidance here. (Let us consider empirical knowledge.) Here are two *unpromising* answers:

Epistemology & What It Takes to Know

Notice that the way of modeling the *content* of an agent's knowledge in basic epistemic-logical models gives no instructions re the question:

> In general, if one wants to *know* that something is true—not just believe or wish it to be true—what does it take to achieve that?

Philosophers have long tried to provide some guidance here. (Let us consider empirical knowledge.) Here are two *unpromising* answers:

- The agent should keep making observations of the world until she can eliminate *every* possible way in which p could be false, which is to say: until her experience is so extensive that it is *impossible* for things to appear to her as they do in a world where p is false.

Epistemology & What It Takes to Know

Notice that the way of modeling the *content* of an agent's knowledge in basic epistemic-logical models gives no instructions re the question:

> In general, if one wants to *know* that something is true—not just believe or wish it to be true—what does it take to achieve that?

Philosophers have long tried to provide some guidance here. (Let us consider empirical knowledge.) Here are two *unpromising* answers:

- The agent should keep making observations of the world until she can eliminate *every* possible way in which p could be false, which is to say: until her experience is so extensive that it is *impossible* for things to appear to her as they do in a world where p is false.
- She should keep investigating the world until her belief about p becomes so strongly correlated with the truth of p that it would be *impossible* for her to believe p as she does in a world where p is false.

Epistemology & What It Takes to Know

Notice that the way of modeling the *content* of an agent's knowledge in basic epistemic-logical models gives no instructions re the question:

> In general, if one wants to *know* that something is true—not just believe or wish it to be true—what does it take to achieve that?

Philosophers have long tried to provide some guidance here. (Let us consider empirical knowledge.) Here are two *unpromising* answers:

- The agent should keep making observations of the world until she can eliminate *every* possible way in which p could be false, which is to say: until her experience is so extensive that it is *impossible* for things to appear to her as they do in a world where p is false.
- She should keep investigating the world until her belief about p becomes so strongly correlated with the truth of p that it would be *impossible* for her to believe p as she does in a world where p is false.

Problem: according to these answers, nobody ever knows anything.

Problem: according to these answers, nobody ever knows anything.

Many philosophers have realized this, e.g.:

J.L. Austin. 1946. "Other Minds."

Alvin Goldman. 1976. "Discrimination and Perceptual Knowledge."

Fred Dretske. 1981. "The Pragmatic Dimension of Knowledge."

David Lewis. 1996. "Elusive Knowledge."

"If knowledge required the elimination of all logically possible alternatives, there would be no knowledge (at least of contingent truths)."
(Goldman 1976, 775)

Problem: according to these answers, nobody ever knows anything.

Many philosophers have realized this, e.g.:

J.L. Austin. 1946. "Other Minds."

Alvin Goldman. 1976. "Discrimination and Perceptual Knowledge."

Fred Dretske. 1981. "The Pragmatic Dimension of Knowledge."

David Lewis. 1996. "Elusive Knowledge."

"There are always, it seems, possibilities that our evidence is powerless to eliminate... If knowledge...requires the elimination of all competing possibilities (possibilities that contrast with what is known), then, clearly we seldom, if ever, satisfy the conditions for applying the concept."

(Dretske 1981, 365)

Relevant Alternatives and Subjunctivism

Austin, Goldman, Dretske, Lewis, and others propose a different idea:

According to **relevant alternatives** theories of knowledge, the agent should keep making observations of the world until she eliminates every *relevant* $\neg p$-possibility; an RA theorist then gives an account of what factors determine which possibilities are relevant and which are not.

Relevant Alternatives and Subjunctivism

Austin, Goldman, Dretske, Lewis, and others propose a different idea:

According to **relevant alternatives** theories of knowledge, the agent should keep making observations of the world until she eliminates every *relevant* $\neg p$-possibility; an RA theorist then gives an account of what factors determine which possibilities are relevant and which are not.

Robert Nozick, Dretske, Ernest Sosa, and others propose related ideas:

According to **subjunctivist** theories of knowledge, the agent should keep investigating the world until her belief about p becomes sufficiently correlated with the truth of p that, e.g., the following counterfactual is true: if p were false (in the "closest worlds"), she wouldn't believe it.

Relevant Alternatives and Subjunctivism

Austin, Goldman, Dretske, Lewis, and others propose a different idea:

According to **relevant alternatives** theories of knowledge, the agent should keep making observations of the world until she eliminates every *relevant* $\neg p$-possibility; an RA theorist then gives an account of what factors determine which possibilities are relevant and which are not.

Robert Nozick, Dretske, Ernest Sosa, and others propose related ideas:

According to **subjunctivist** theories of knowledge, the agent should keep investigating the world until her belief about p becomes sufficiently correlated with the truth of p that, e.g., the following counterfactual is true: if p were false (in the "closest worlds"), she wouldn't believe it.

Robert Nozick. 1981. *Philosophical Explanations*.

Fred Dretske. 1971. "Conclusive Reasons."

Ernest Sosa. 1999. "How to Defeat Opposition to Moore."

Relevant Alternatives and Subjunctivism

According to **relevant alternatives** theories of knowledge, the agent should keep making observations of the world until she eliminates every *relevant* $\neg p$-possibility; an RA theorist then gives a story about what factors determine which possibilities are relevant and which are not.

According to **subjunctivist** theories of knowledge, the agent should keep investigating the world until her belief about p becomes sufficiently correlated with the truth of p that, e.g, the following counterfactual is true: if p were false (in the "closest worlds"), she wouldn't believe it.

Claim: these ideas are not just the work of philosophers' imaginations, detached from real life. To the contrary, arguably considerations of relevance and counterfactuals can help to explain why people are actually willing or unwilling to attribute knowledge to other people *in practice*.

Relevant Alternatives and Subjunctivism

According to **relevant alternatives** theories of knowledge, the agent should keep making observations of the world until she eliminates every *relevant* $\neg p$-possibility; an RA theorist then gives a story about what factors determine which possibilities are relevant and which are not.

According to **subjunctivist** theories of knowledge, the agent should keep investigating the world until her belief about p becomes sufficiently correlated with the truth of p that, e.g, the following counterfactual is true: if p were false (in the "closest worlds"), she wouldn't believe it.

What do these theories have to do with epistemic logic?

Relevant Alternatives and Subjunctivism

According to **relevant alternatives** theories of knowledge, the agent should keep making observations of the world until she eliminates every *relevant* $\neg p$-possibility; an RA theorist then gives a story about what factors determine which possibilities are relevant and which are not.

According to **subjunctivist** theories of knowledge, the agent should keep investigating the world until her belief about p becomes sufficiently correlated with the truth of p that, e.g, the following counterfactual is true: if p were false (in the "closest worlds"), she wouldn't believe it.

What do these theories have to do with epistemic logic? Answer: they bear on the correctness of core principles of epistemic logic.

Logical Omniscience vs. Epistemic Closure

Let's distinguish two very different objections to

$$\mathrm{K} \quad (K\varphi \wedge K(\varphi \to \psi)) \to K\psi.$$

Logical Omniscience vs. Epistemic Closure

Let's distinguish two very different objections to

$$\mathrm{K} \quad (K\varphi \wedge K(\varphi \to \psi)) \to K\psi.$$

First, the **problem of logical omniscience**: real (finite, boundedly rational) agents don't always "put two and two together."

Logical Omniscience vs. Epistemic Closure

Let's distinguish two very different objections to

$$\mathrm{K} \quad (K\varphi \wedge K(\varphi \to \psi)) \to K\psi.$$

First, the **problem of logical omniscience**: real (finite, boundedly rational) agents don't always "put two and two together."

(Aside: one way to bypass the logical omniscience problem is to read $K\varphi$ as "φ is entailed by what the agent knows." Then K is fine.)

Logical Omniscience vs. Epistemic Closure

Let's distinguish two very different objections to

$$\mathrm{K} \quad (K\varphi \wedge K(\varphi \to \psi)) \to K\psi.$$

First, the **problem of logical omniscience**: real (finite, boundedly rational) agents don't always "put two and two together."

(Aside: one way to bypass the logical omniscience problem is to read $K\varphi$ as "φ is entailed by what the agent knows." Then K is fine.)

Second, the **denial of epistemic closure**: even if agents did always put two and two together, K should still be rejected in its fullest generality for knowledge (though for circumscribed applications it might be fine).

Fallibilism, Closure, and Contradiction

According to Goldman, Dretske, and many others, the **infallibilist** view that for every proposition p, one must eliminate *all* $\neg p$-possibilities in order to know p entails **skepticism**: *we don't know anything*.

Fallibilism, Closure, and Contradiction

According to Goldman, Dretske, and many others, the **infallibilist** view that for every proposition p, one must eliminate *all* $\neg p$-possibilities in order to know p entails **skepticism**: *we don't know anything*.

Rejecting infallibilism, they hold the **fallibilist** view that for at least some p, there are some $\neg p$-possibilities that one does not have to eliminate in order to know p.

Fallibilism, Closure, and Contradiction

According to Goldman, Dretske, and many others, the **infallibilist** view that for every proposition p, one must eliminate *all* $\neg p$-possibilities in order to know p entails **skepticism**: *we don't know anything*.

Rejecting infallibilism, they hold the **fallibilist** view that for at least some p, there are some $\neg p$-possibilities that one does not have to eliminate in order to know p. Let s pick out such a set of possibilities, so knowing p does not require the agent to eliminate s-possibilities.

Fallibilism, Closure, and Contradiction

According to Goldman, Dretske, and many others, the **infallibilist** view that for every proposition p, one must eliminate *all* $\neg p$-possibilities in order to know p entails **skepticism**: *we don't know anything*.

Rejecting infallibilism, they hold the **fallibilist** view that for at least some p, there are some $\neg p$-possibilities that one does not have to eliminate in order to know p. Let s pick out such a set of possibilities, so knowing p does not require the agent to eliminate s-possibilities. But by K:

$$(Kp \wedge K(p \to \neg s)) \to K\neg s,$$

which says that knowing p *does* require knowing $\neg s$, which requires eliminating s-possibilities (assuming that knowing something requires eliminating at least *some* counter-possibilities).

Fallibilism, Closure, and Contradiction

According to Goldman, Dretske, and many others, the **infallibilist** view that for every proposition p, one must eliminate *all* $\neg p$-possibilities in order to know p entails **skepticism**: *we don't know anything*.

Rejecting infallibilism, they hold the **fallibilist** view that for at least some p, there are some $\neg p$-possibilities that one does not have to eliminate in order to know p. Let s pick out such a set of possibilities, so knowing p does not require the agent to eliminate s-possibilities. But by K:

$$(Kp \wedge K(p \to \neg s)) \to K\neg s,$$

which says that knowing p *does* require knowing $\neg s$, which requires eliminating s-possibilities (assuming that knowing something requires eliminating at least *some* counter-possibilities). *Contradiction*?

Relevant Alternatives and Subjunctivism

Two differences between these philosophical views of *what it takes to know* and the basic epistemic-logical models of *what an agent knows*:

Relevant Alternatives and Subjunctivism

Two differences between these philosophical views of *what it takes to know* and the basic epistemic-logical models of *what an agent knows*:

- The philosophical theories appeal to additional structure for relevance or for counterfactuals, not in basic epistemic models;

Relevant Alternatives and Subjunctivism

Two differences between these philosophical views of *what it takes to know* and the basic epistemic-logical models of *what an agent knows*:

- The philosophical theories appeal to additional structure for relevance or for counterfactuals, not in basic epistemic models;
- The philosophical theories are presented informally. They are not investigated mathematically as models for epistemic logic are.

Relevant Alternatives and Subjunctivism

Two differences between these philosophical views of *what it takes to know* and the basic epistemic-logical models of *what an agent knows*:

- The philosophical theories appeal to additional structure for relevance or for counterfactuals, not in basic epistemic models;
- The philosophical theories are presented informally. They are not investigated mathematically as models for epistemic logic are.

But we can do something about this: add the extra structure to basic epistemic models and then investigate the theories mathematically.

A Family of Theories

I will begin by providing model-theoretic, logical formalizations of a family of *relevant alternatives* (RA) and *subjunctivist* theories of knowledge:

- L the RA theory of **David Lewis** (1996);
- D one way of interpreting the RA theory of **Fred Dretske** (1981);
- H the RA theory of **Mark Heller** (1989, 1999);
- N the basic tracking theory of **Robert Nozick** (1981);
- S and the safety theory of **Ernest Sosa** (1999).

Example (Medical Diagnosis)

Two medical students, A and B, are subjected to a test. Their professor introduces them to the same patient, who presents various symptoms, and the students are to make a diagnosis of the patient's condition. After some independent investigation, both students conclude that the patient has a common condition c. In fact, they are both correct.

Example (Medical Diagnosis)

Two medical students, A and B, are subjected to a test. Their professor introduces them to the same patient, who presents various symptoms, and the students are to make a diagnosis of the patient's condition. After some independent investigation, both students conclude that the patient has a common condition c. In fact, they are both correct.

Yet only student A passes the test.

Example (Medical Diagnosis)

Two medical students, A and B, are subjected to a test. Their professor introduces them to the same patient, who presents various symptoms, and the students are to make a diagnosis of the patient's condition. After some independent investigation, both students conclude that the patient has a common condition c. In fact, they are both correct.

Yet only student A passes the test.

The professor wished to see if the students would check for another common condition c', which causes the same visible symptoms as c. While student A ran laboratory tests to rule out c' before making the diagnosis of c, the student B made the diagnosis of c after only a physical exam, having never considered the possibility of c'.

RA-style analysis:

- Although both students gave the correct diagnosis of c, student B did not *know* that the patient's condition was c, since he did not rule out the alternative of c'.

RA-style analysis:

- Although both students gave the correct diagnosis of c, student B did not *know* that the patient's condition was c, since he did not rule out the alternative of c'.
- Student B *got lucky*. Had the patient's condition been c', student B might still have made the diagnosis of c, since the physical exam would not have revealed a difference.

RA-style analysis:

- Although both students gave the correct diagnosis of c, student B did not *know* that the patient's condition was c, since he did not rule out the alternative of c'.
- Student B *got lucky*. Had the patient's condition been c', student B might still have made the diagnosis of c, since the physical exam would not have revealed a difference.
- Student A secured against this possibility of error by running the laboratory tests.

RA-style analysis:

- Although both students gave the correct diagnosis of c, student B did not *know* that the patient's condition was c, since he did not rule out the alternative of c'.
- Student B *got lucky*. Had the patient's condition been c', student B might still have made the diagnosis of c, since the physical exam would not have revealed a difference.
- Student A secured against this possibility of error by running the laboratory tests.

Of course, student A did not secure against *every* possibility of error...

Example (Medical Diagnosis continued)

Suppose there is an extremely rare disease x[1] such that people with disease x appear to have c on many laboratory tests, even though people with x are completely *immune* to c, and only extensive further testing can detect the presence or absence of x in its early stages.

[1] Maybe never-before-documented, but a possibility dreamed up by a hypochondriac.

Example (Medical Diagnosis continued)

Suppose there is an extremely rare disease x[1] such that people with disease x appear to have c on many laboratory tests, even though people with x are completely *immune* to c, and only extensive further testing can detect the presence or absence of x in its early stages.

Should we say that student `A` did not know that the patient's condition was c after all, since she did not rule out the possibility of x?

[1] Maybe never-before-documented, but a possibility dreamed up by a hypochondriac.

RA-style analysis cont.:

- The requirement that one rule out *all* possibilities of error makes knowledge impossible, since there are always some possibilities of error—however remote and far-fetched—that are not eliminated by one's evidence and experience.

RA-style analysis cont.:

- The requirement that one rule out *all* possibilities of error makes knowledge impossible, since there are always some possibilities of error—however remote and far-fetched—that are not eliminated by one's evidence and experience.
- If the student has no reason to think that the patient may have the rare disease x, then it should not be necessary to rule out such a remote possibility in order to know that the patient has some common condition.

- Suppose that all medical students know that people with x have complete immunity to c, as assumed above, so

$$K_{\mathtt{A}}\left(c \to \neg x\right).$$

- Suppose that all medical students know that people with x have complete immunity to c, as assumed above, so

$$K_{\mathtt{A}}\left(c \to \neg x\right).$$

- Since the student did not run any tests that could possibly detect the presence or absence of x, it would be unreasonable to claim that she knows that the patient does not have x, so

$$\neg K_{\mathtt{A}} \neg x.$$

- Suppose that all medical students know that people with x have complete immunity to c, as assumed above, so

$$K_{\mathtt{A}}\left(c \to \neg x\right).$$

- Since the student did not run any tests that could possibly detect the presence or absence of x, it would be unreasonable to claim that she knows that the patient does not have x, so

$$\neg K_{\mathtt{A}} \neg x.$$

- Then together with our judgment that the student knows that the patient has condition c,

$$K_{\mathtt{A}} c,$$

we have a clear violation of *closure under known implication*,

$$\left(K_{\mathtt{A}} c \wedge K_{\mathtt{A}}(c \to \neg x)\right) \to K_{\mathtt{A}} \neg x.$$

According to Dretske, the RA theory explains why closure fails:

According to Dretske, the RA theory explains why closure fails:

- In order to know the premises c and $c \to \neg x$, the agent must rule out certain relevant alternatives. In order to know the conclusion $\neg x$, the agent must also rule out certain relevant alternatives.

According to Dretske, the RA theory explains why closure fails:

- In order to know the premises c and $c \to \neg x$, the agent must rule out certain relevant alternatives. In order to know the conclusion $\neg x$, the agent must also rule out certain relevant alternatives.
- But the sets of relevant alternatives for the premises and for the conclusion *are not the same*.

According to Dretske, the RA theory explains why closure fails:

- In order to know the premises c and $c \to \neg x$, the agent must rule out certain relevant alternatives. In order to know the conclusion $\neg x$, the agent must also rule out certain relevant alternatives.
- But the sets of relevant alternatives for the premises and for the conclusion *are not the same*.

  We have already argued that x is not a relevant alternative that must be ruled out in order for $K_{\mathtt{A}} c$ to hold. But x certainly is a relevant alternative that must be ruled out in order for $K_{\mathtt{A}} \neg x$ to hold.

According to Dretske, the RA theory explains why closure fails:

- In order to know the premises c and $c \to \neg x$, the agent must rule out certain relevant alternatives. In order to know the conclusion $\neg x$, the agent must also rule out certain relevant alternatives.
- But the sets of relevant alternatives for the premises and for the conclusion *are not the same*.

  We have already argued that x is not a relevant alternative that must be ruled out in order for $K_{\mathtt{A}} c$ to hold. But x certainly is a relevant alternative that must be ruled out in order for $K_{\mathtt{A}} \neg x$ to hold.
- It is because the relevant alternatives may be different for the premises and the conclusion that closure under known implication does not hold in general.

According to Dretske, the RA theory explains why closure fails:

- In order to know the premises c and $c \to \neg x$, the agent must rule out certain relevant alternatives. In order to know the conclusion $\neg x$, the agent must also rule out certain relevant alternatives.
- But the sets of relevant alternatives for the premises and for the conclusion *are not the same*.

  We have already argued that x is not a relevant alternative that must be ruled out in order for $K_{\mathtt{A}} c$ to hold. But x certainly is a relevant alternative that must be ruled out in order for $K_{\mathtt{A}} \neg x$ to hold.
- It is because the relevant alternatives may be different for the premises and the conclusion that closure under known implication does not hold in general.

Question: which other closure principles fail (or hold) on the RA theory?

Validity Omniscience and Doxastic Closure

Dretske and Nozick argue that closure under known implication would fail even for agents with ideal rationality.

Validity Omniscience and Doxastic Closure

Dretske and Nozick argue that closure under known implication would fail even for agents with ideal rationality.

According to Dretske, "Were we all ideally astute logicians, were we all fully appraised of all the necessary consequences...of every proposition, perhaps then the epistemic operators would . . . [satisfy **K**]. It is this . . . claim that I mean to reject" ("Epistemic Operators," 1010).

Validity Omniscience and Doxastic Closure

Dretske and Nozick argue that closure under known implication would fail even for agents with ideal rationality.

According to Dretske, "Were we all ideally astute logicians, were we all fully appraised of all the necessary consequences...of every proposition, perhaps then the epistemic operators would . . . [satisfy **K**]. It is this . . . claim that I mean to reject" ("Epistemic Operators," 1010).

- **Validity omniscience**: one knows all valid logical principles;

Validity Omniscience and Doxastic Closure

Dretske and Nozick argue that closure under known implication would fail even for agents with ideal rationality.

According to Dretske, "Were we all ideally astute logicians, were we all fully appraised of all the necessary consequences...of every proposition, perhaps then the epistemic operators would . . . [satisfy **K**]. It is this . . . claim that I mean to reject" ("Epistemic Operators," 1010).

- **Validity omniscience**: one knows all valid logical principles;
- **Full doxastic closure**: one believes all the logical consequences of the (set of) propositions one believes.

Validity Omniscience and Doxastic Closure

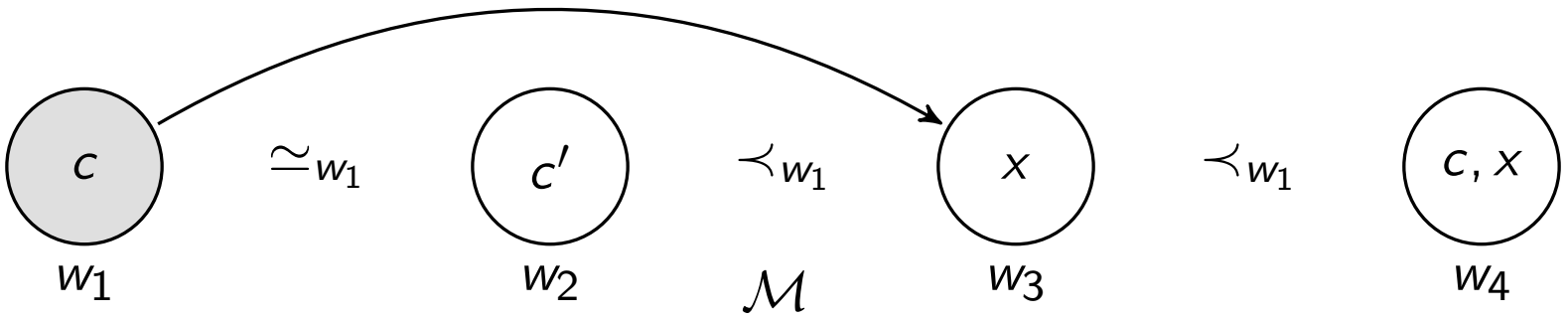
Dretske and Nozick argue that closure under known implication would fail even for agents with ideal rationality.

According to Dretske, "Were we all ideally astute logicians, were we all fully appraised of all the necessary consequences...of every proposition, perhaps then the epistemic operators would . . . [satisfy **K**]. It is this . . . claim that I mean to reject" ("Epistemic Operators," 1010).

- **Validity omniscience**: one knows all valid logical principles;
- **Full doxastic closure**: one believes all the logical consequences of the (set of) propositions one believes.

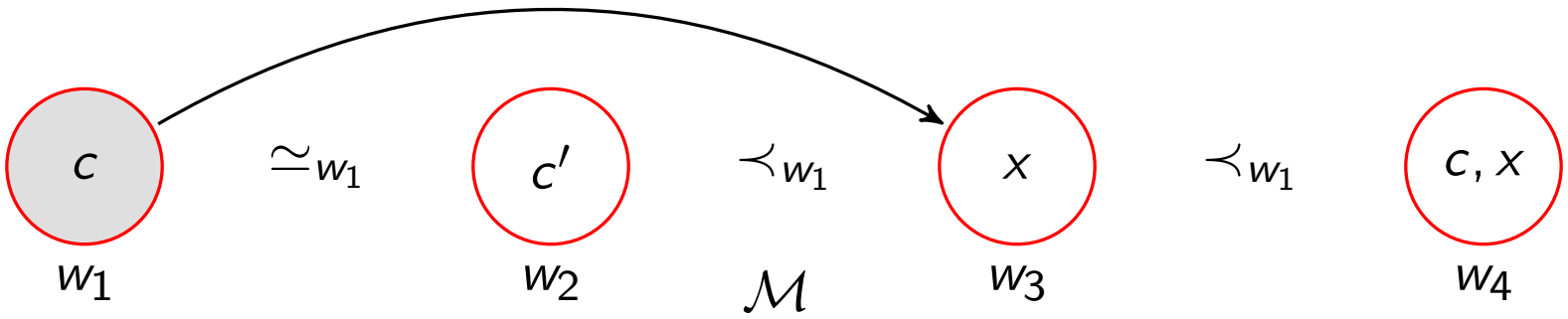
One may satisfy the conditions for knowledge (ruling out the relevant alternatives, tracking the truth, etc.) with respect to some propositions and yet not with respect to all logical consequences of the set of those propositions, *even if* one has explicitly deduced all of these consequences.

Relevant Alternatives Models



A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

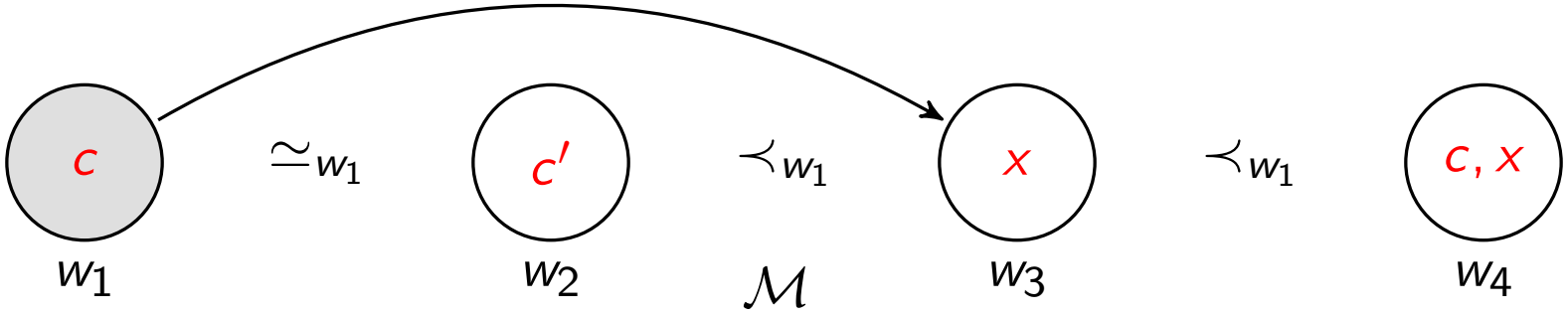
Relevant Alternatives Models



A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $W = \{w, v, u, \dots\}$ is a non-empty set of "worlds" or "possibilities."

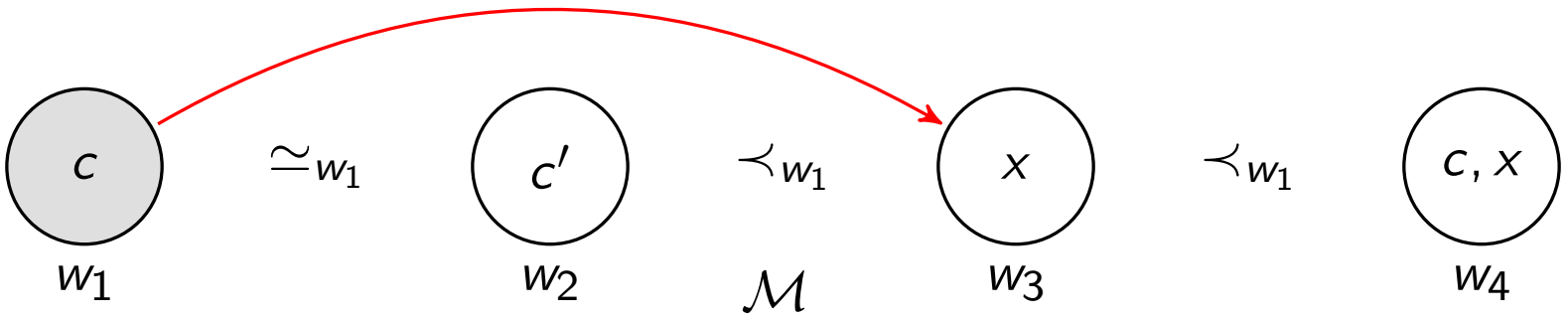
Relevant Alternatives Models



A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- V: $\mathrm{At} \to \mathcal{P}(W)$, where $\mathrm{At} = \{p, q, r, \ldots\}$ is a fixed set of atomic sentence symbols, the same for all models.

Relevant Alternatives Models

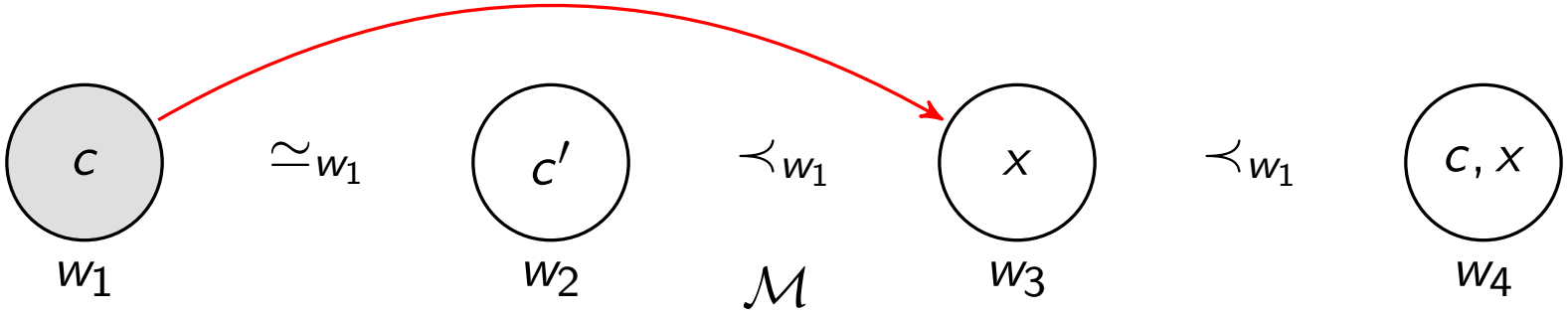


A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\rightarrowtriangle$ is a reflexive relation on W (note omitted loops in diagrams).

  $w \rightarrowtriangle v$ means possibility v is **uneliminated** for the agent in w.

Relevant Alternatives Models



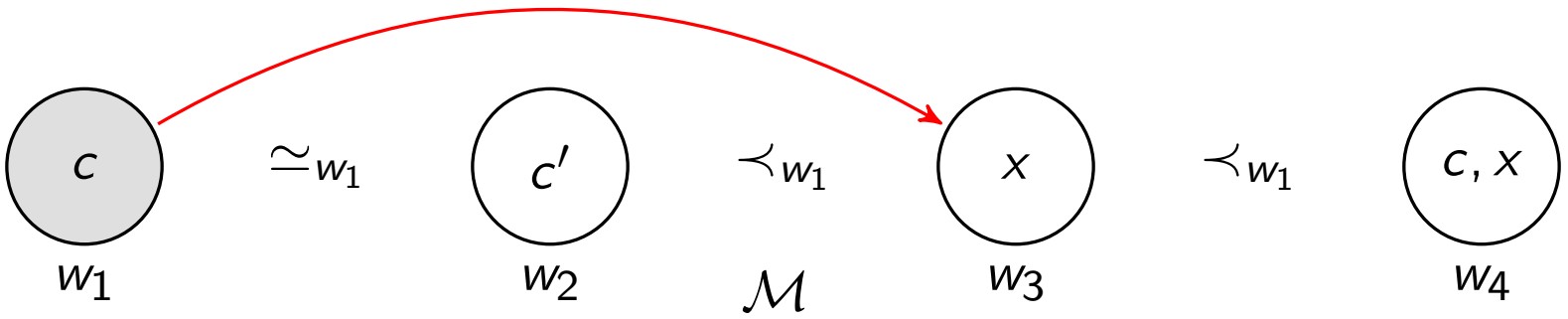
A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\rightarrowtriangle$ is a reflexive relation on W (note omitted loops in diagrams).

  Lewis (1996) proposes that "a possibility ... [v] ... is *uneliminated* iff the subject's perceptual experience and memory in ... [v] ... exactly match his perceptual experience and memory in actuality" (553). Thus, for Lewis $\rightarrowtriangle$ is an equivalence relation.

  David Lewis. 1996. "Elusive Knowledge."

Relevant Alternatives Models



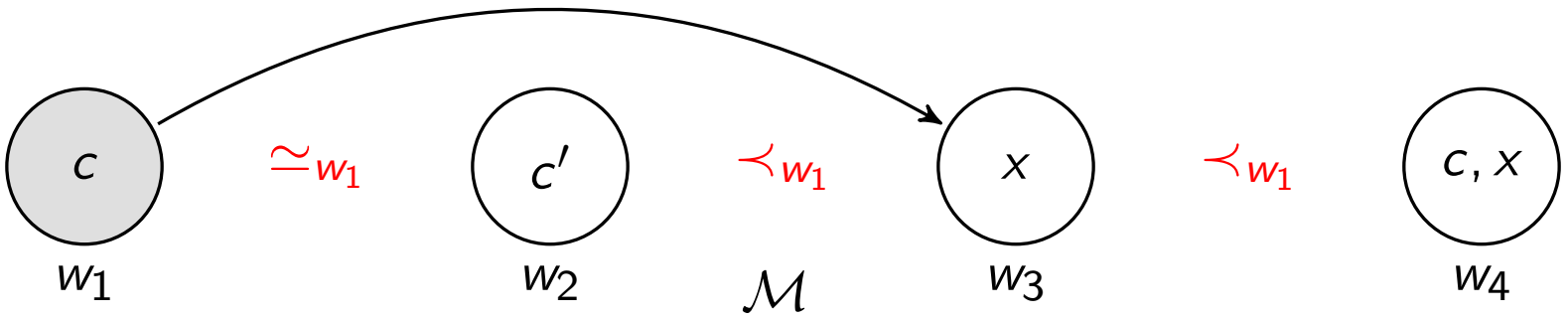
A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\rightarrowtriangle$ is a reflexive relation on W (note omitted loops in diagrams).

  $w \rightarrowtriangle v$ means possibility v is **uneliminated** for the agent in w.

  Whether we assume transitivity, symmetry, Euclideaness, etc., in addition to reflexivity will not affect our main results.

Relevant Alternatives Models

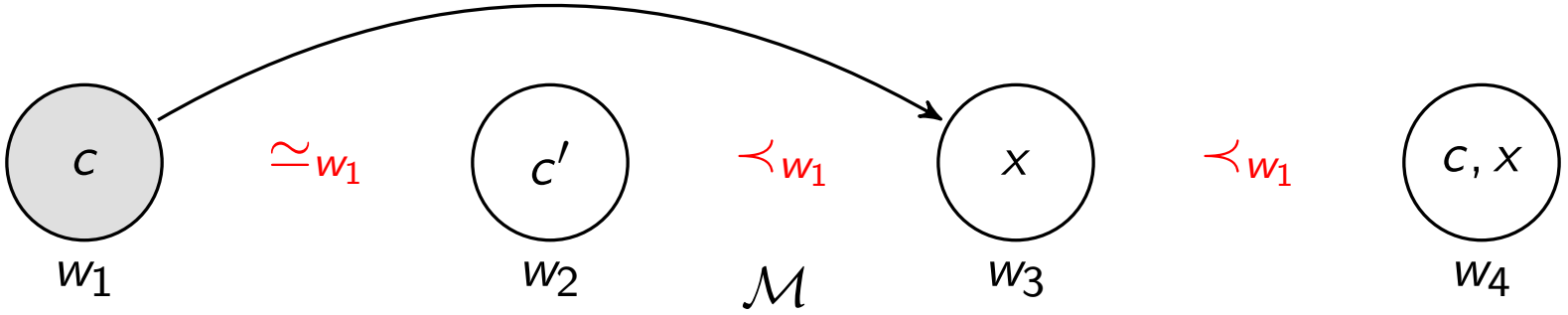


A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
  1. $\preceq_w$ is reflexive and transitive (preorder);
  2. $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering).

  $u \preceq_w v$ means u is **at least as relevant** (at w) as v is;

Relevant Alternatives Models

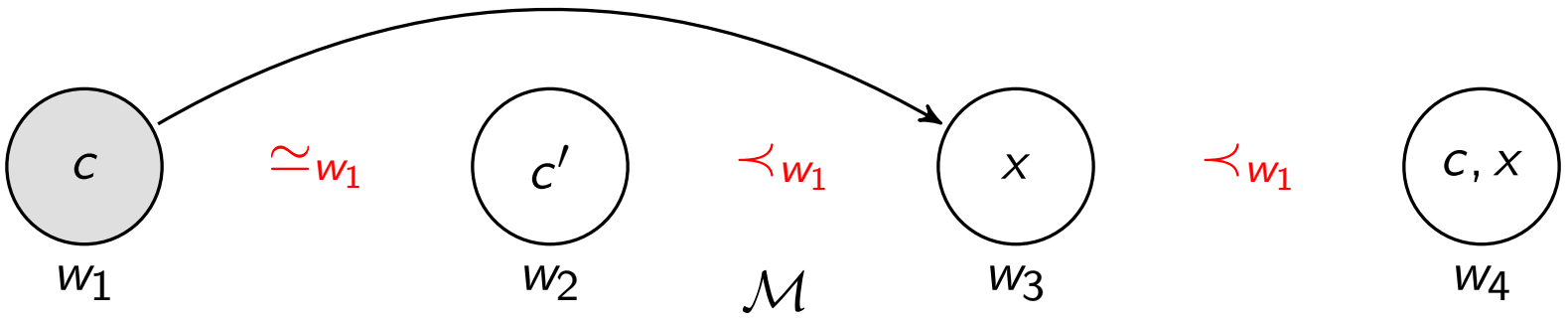


A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
  1. $\preceq_w$ is reflexive and transitive (preorder);
  2. $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering).

$u \preceq_w v$ means u is **at least as relevant** (at w) as v is;
$u \prec_w v$ iff $u \preceq_w v$ and not $v \preceq_w u$;

Relevant Alternatives Models

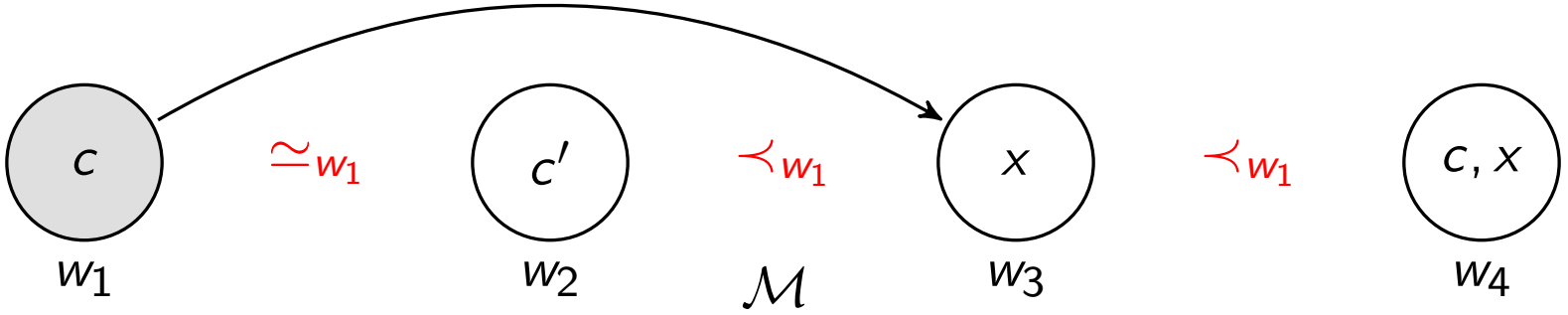


A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
  1. $\preceq_w$ is reflexive and transitive (preorder);
  2. $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering).

$u \preceq_w v$ means u is **at least as relevant** (at w) as v is;
$u \prec_w v$ iff $u \preceq_w v$ and not $v \preceq_w u$;
$u \simeq_w v$ iff $u \preceq_w v$ and $v \preceq_w u$.

Relevant Alternatives Models

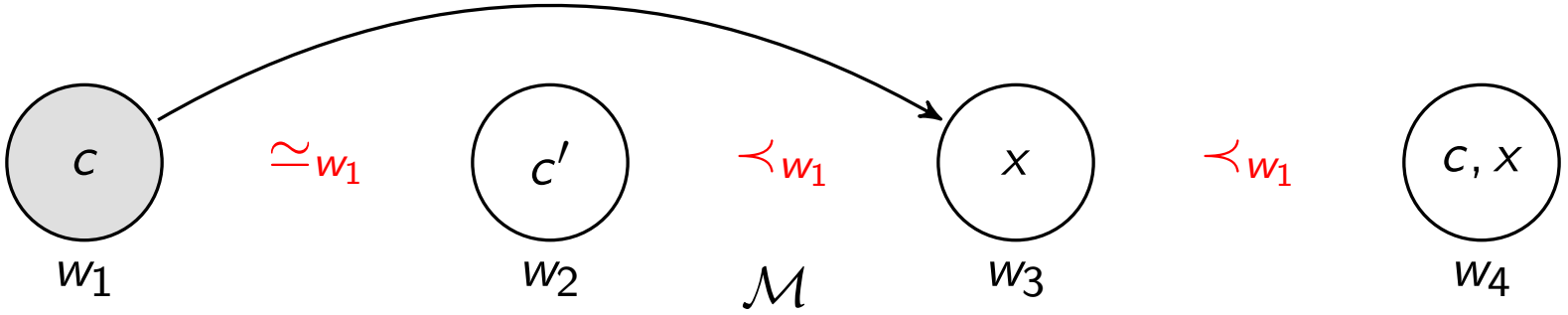


A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$ ...

Many epistemologists state conditions for knowledge in terms of some kind of (possibly context-dependent) ordering on worlds: e.g., Heller's (1989) RA picture of "worlds surrounding the actual world ordered according to how realistic they are, so that those worlds that are more realistic are closer to the actual world than the less realistic ones" (25).

Mark Heller. 1989. "Relevant Alternatives."

Relevant Alternatives Models

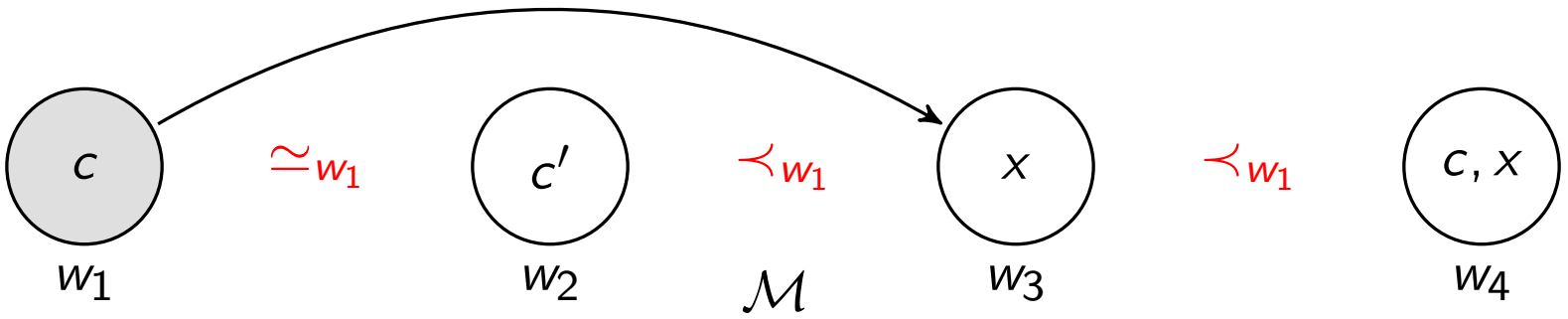


A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
  1. $\preceq_w$ is reflexive and transitive (preorder);
  2. $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering).

  The second condition corresponds to Lewis's (1996) Rule of Actuality, that "actuality is always a relevant alternative" (554).

Relevant Alternatives Models



A (single-agent) RA model is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, where:

- $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
  1. $\preceq_w$ is reflexive and transitive (preorder);
  2. $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering);
  3. if $\emptyset \neq S \subseteq W_w$, then $\mathrm{Min}_{\preceq_w}(S) \neq \emptyset$ (well-founded).

For $S \subseteq W$, $\mathrm{Min}_{\preceq_w}(S) = \{v \in S \cap W_w \mid \neg\exists u \in S : u \prec_w v\}$.

Definition (RA Model)

A *relevant alternatives model* is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$ where:

1. W is a non-empty set;
2. $\rightarrowtriangle$ is a reflexive binary relation on W;
3. $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
   - 3.1 $\preceq_w$ is reflexive and transitive (preorder);
   - 3.2 $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering);
   - 3.3 if $\emptyset \neq S \subseteq W_w$, then $\mathrm{Min}_{\preceq_w}(S) \neq \emptyset$ (well-founded);
4. $V : \mathrm{At} \to \mathcal{P}(W)$.

Definition (RA Model)

A *relevant alternatives model* is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$ where:

1. W is a non-empty set;
2. $\rightarrowtriangle$ is a reflexive binary relation on W;
3. $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
   - 3.1 $\preceq_w$ is reflexive and transitive (preorder);
   - 3.2 $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering);
   - 3.3 if $\emptyset \neq S \subseteq W_w$, then $\mathrm{Min}_{\preceq_w}(S) \neq \emptyset$ (well-founded);
4. $V : \mathrm{At} \to \mathcal{P}(W)$.

We will also consider *total* RA models in which for all $w \in W$, $\preceq_w$ is a *total* preorder, i.e., $\forall u, v \in W_w$, $u \preceq_w v$ or $v \preceq_w u$.

Definition (RA Model)

A *relevant alternatives model* is a tuple $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$ where:

1. W is a non-empty set;
2. $\rightarrowtriangle$ is a reflexive binary relation on W;
3. $\preceq$ assigns to each $w \in W$ a binary relation $\preceq_w$ on some $W_w \subseteq W$;
   - 3.1 $\preceq_w$ is reflexive and transitive (preorder);
   - 3.2 $w \in W_w$, and for all $v \in W_w$, $w \preceq_w v$ (weak centering);
   - 3.3 if $\emptyset \neq S \subseteq W_w$, then $\mathrm{Min}_{\preceq_w}(S) \neq \emptyset$ (well-founded);
4. $V : \mathrm{At} \to \mathcal{P}(W)$.

We will also consider *total* RA models in which for all $w \in W$, $\preceq_w$ is a *total* preorder, i.e., $\forall u, v \in W_w$, $u \preceq_w v$ or $v \preceq_w u$.

These models represent a fixed context. For context change, see:

W. Holliday. 2012. "Epistemic Logic, Relevant Alternatives, and the Dynamics of Context."

New Directions in Logic, Language, and Computation, LNCS.

Language

We could use our RA models to give semantics for a variety of different formal languages, but we'll start with the basic modal language:

Definition (Epistemic Language)

The (single-agent) *epistemic language* $\mathcal{L}$ is defined inductively by

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K\varphi,$$

where $p \in \mathrm{At}$.

For simplicity in this talk, sometimes I will state results for the *flat* fragment in which only propositional formulas occur in the scope of K.

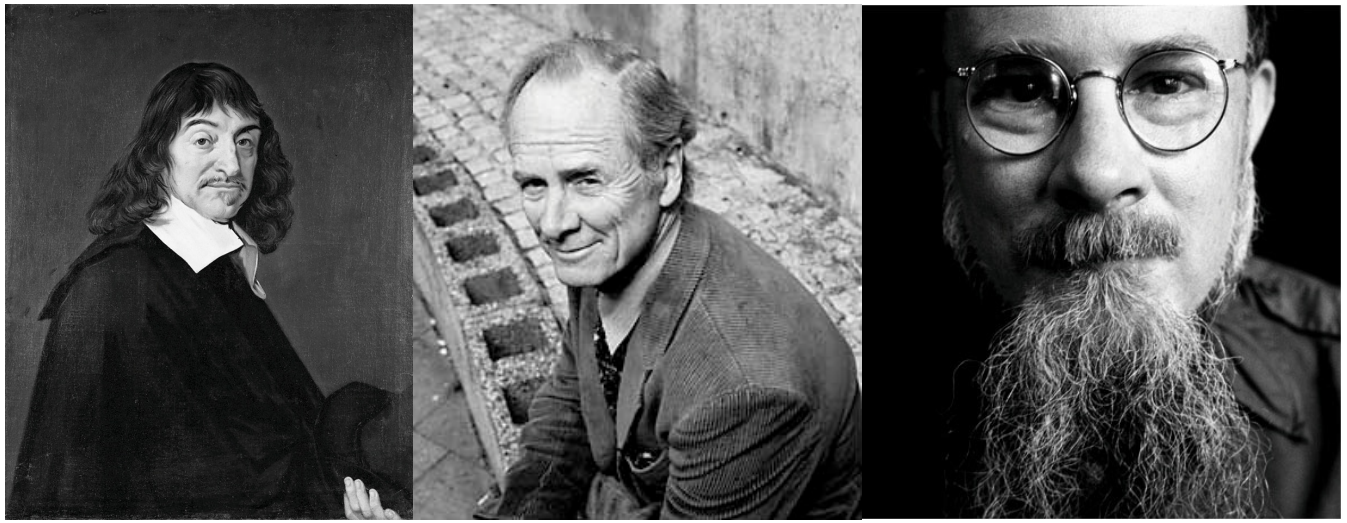
Semantics

We will consider three semantics for the K operator in RA models:

Semantics

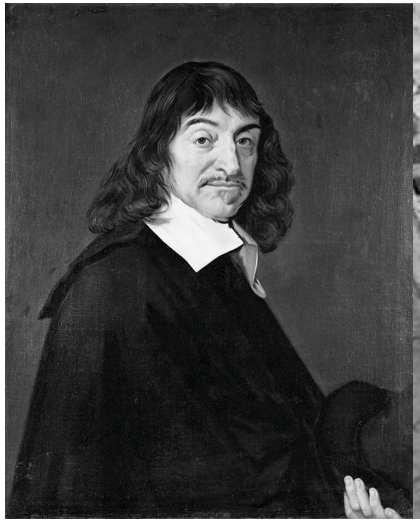
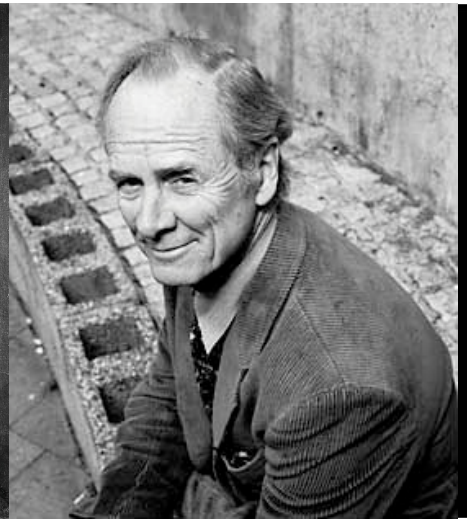
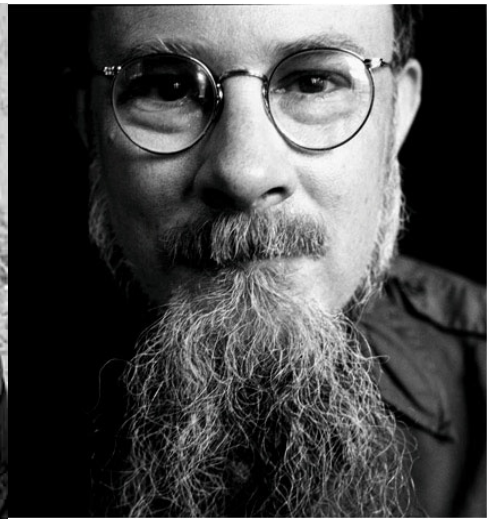
We will consider three semantics for the K operator in RA models:

- C-semantics for **C**artesian. C-semantics is not intended to capture Descartes' view of knowledge. Rather, it is supposed to reflect a high standard for the truth of knowledge claims—knowledge requires ruling out all possibilities of error, however remote—in the spirit of Descartes' worries about error in the First Meditation.



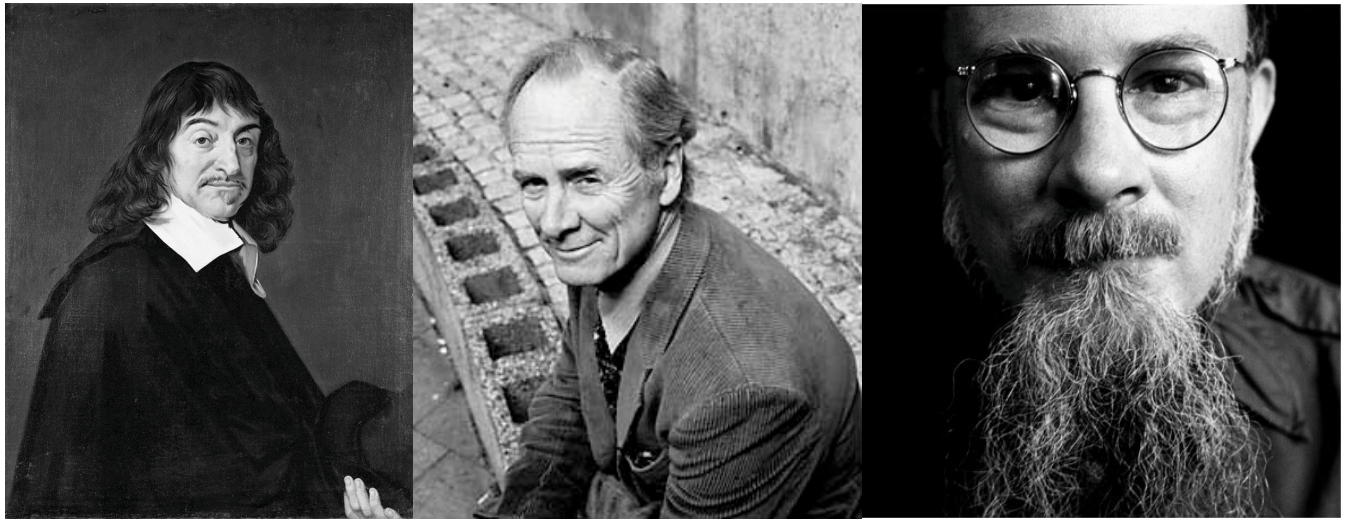
Semantics

- C-semantics for **C**artesian.
- D-semantics is one way (but not the only way or even the best way) of understanding **D**retske's (1981) RA theory, using Heller's (1989, 1999) picture of relevance orderings over possibilities.

Semantics

- C-semantics for **C**artesian.
- D-semantics is one way (but not the only way or even the best way) of understanding **D**retske's (1981) RA theory, using Heller's (1989, 1999) picture of relevance orderings over possibilities.
- L-semantics is for **L**ewis's (1999) RA theory (for a fixed context).



Semantics

Definition (Truth in an RA Model)

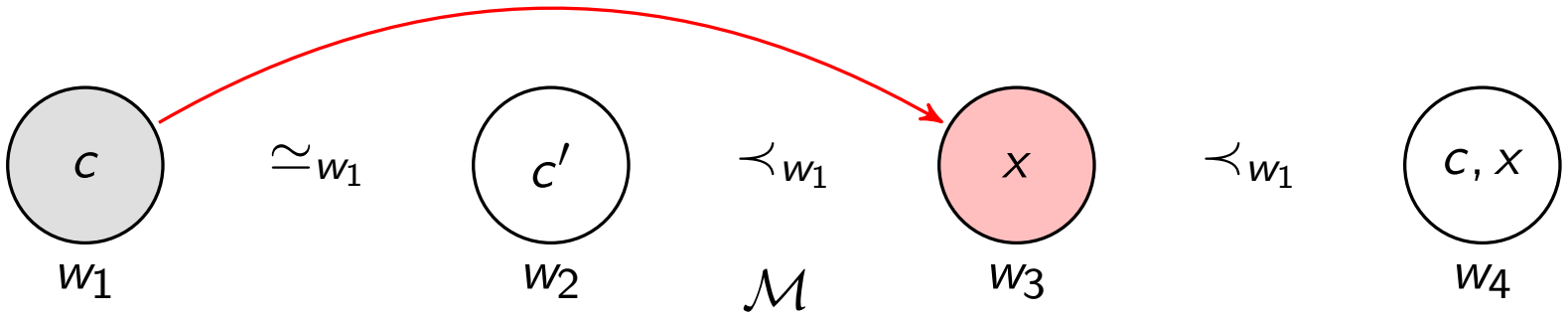
Given an RA model $\mathcal{M} = \langle W, \rightarrowtriangle, \preceq, V \rangle$, world $w \in W$, formula φ in the epistemic language, and $x \in \{c, d, l\}$, we define $\mathcal{M}, w \vDash_x \varphi$ and $[\![\varphi]\!]_x^{\mathcal{M}} = \{v \in W \mid \mathcal{M}, v \vDash_x \varphi\}$ as follows:

$$
\begin{array}{lll}
\mathcal{M}, w \vDash_x p & \text{iff} & w \in V(p); \\
\mathcal{M}, w \vDash_x \neg\varphi & \text{iff} & \mathcal{M}, w \nvDash_x \varphi; \\
\mathcal{M}, w \vDash_x (\varphi \wedge \psi) & \text{iff} & \mathcal{M}, w \vDash_x \varphi \text{ and } \mathcal{M}, w \vDash_X \psi. \\
\\
\mathcal{M}, w \vDash_c K\varphi & \text{iff} & \forall v \in \overline{[\![\varphi]\!]_c} \colon w \not\rightarrowtriangle v; \\
\mathcal{M}, w \vDash_d K\varphi & \text{iff} & \forall v \in \mathrm{Min}_{\preceq_w}\left(\overline{[\![\varphi]\!]_d}\right) \colon w \not\rightarrowtriangle v; \\
\mathcal{M}, w \vDash_l K\varphi & \text{iff} & \forall v \in \mathrm{Min}_{\preceq_w}(W) \cap \overline{[\![\varphi]\!]_l} \colon w \not\rightarrowtriangle v,
\end{array}
$$

where for $S \subseteq W$, $\overline{S} = \{v \in W \mid v \notin S\}$.

C-semantics and the Problem of Skepticism

$\mathcal{M}, w \vDash_c K\varphi$ iff $\forall v \in \overline{[\![\varphi]\!]_c}$: $w \not\rightarrow v$.

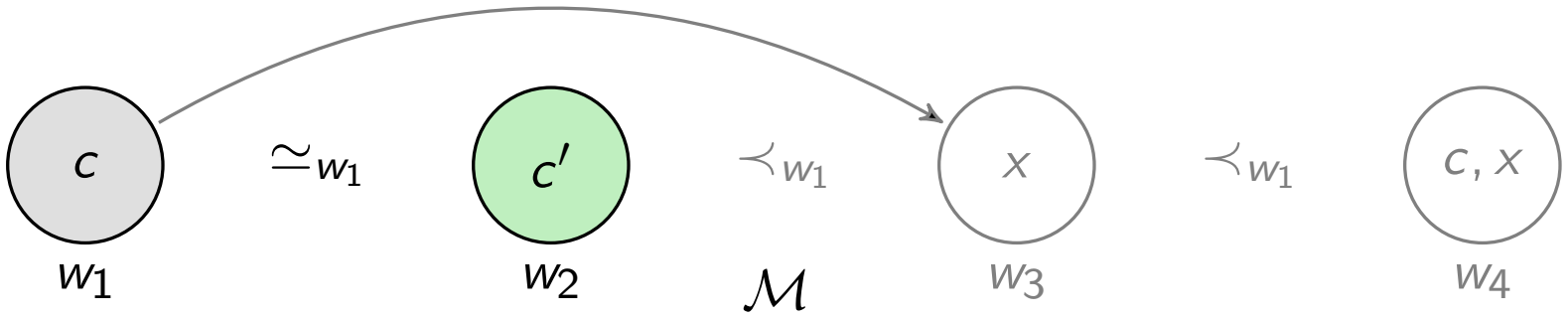


$\mathcal{M}, w_1 \nvDash_c Kc$

C-semantics looks like the standard epistemic semantics in the tradition of Hintikka (1962); but now that we are explicitly representing relevance, it formalizes the infallibilist idea that all possibilities of error, however remote and far-fetched, must be eliminated—a view that leads to **skepticism**.

L-semantics and Non-skepticism

$\mathcal{M}, w \vDash_l K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(W) \cap \overline{[\![\varphi]\!]_l}$: $w \not\rightarrow v$.

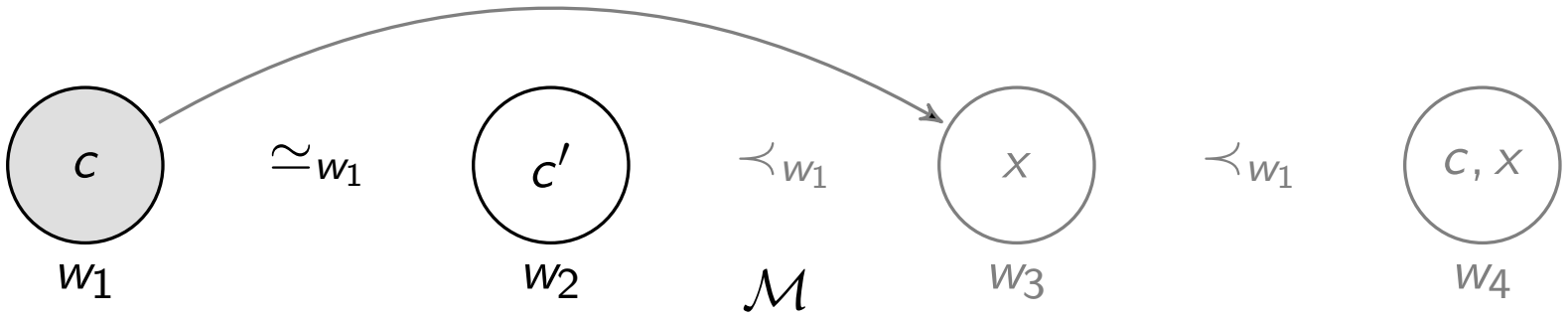


$\mathcal{M}, w_1 \vDash_l Kc$

L-semantics avoids skepticism by not requiring the elimination of every error possibility, but only those in the set $\mathrm{Min}_{\preceq_w}(W)$ of relevant worlds.

L-semantics and Non-skepticism

$\mathcal{M}, w \vDash_l K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(W) \cap \overline{[\![\varphi]\!]_l}$: $w \not\rightarrow v$.

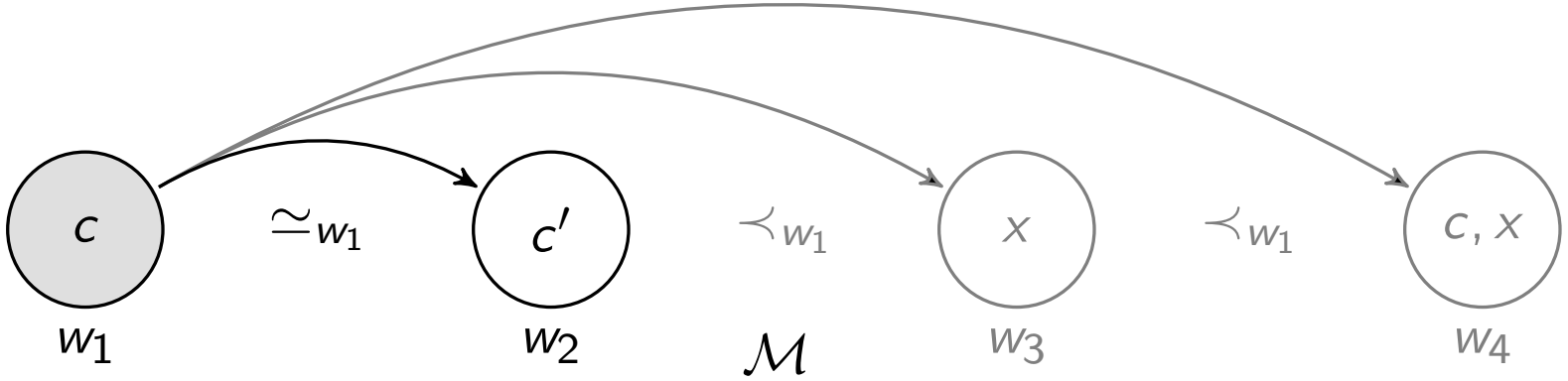


$\mathcal{M}, w_1 \vDash_l K\neg x$

L-semantics avoids skepticism by not requiring the elimination of every error possibility, but only those in the set $\mathrm{Min}_{\preceq_w}(W)$ of relevant worlds.

L-semantics and the Problem of Vacuous Knowledge

$\mathcal{M}, w \vDash_l K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(W) \cap \overline{[\![\varphi]\!]_l}$: $w \not\rightarrow v$.

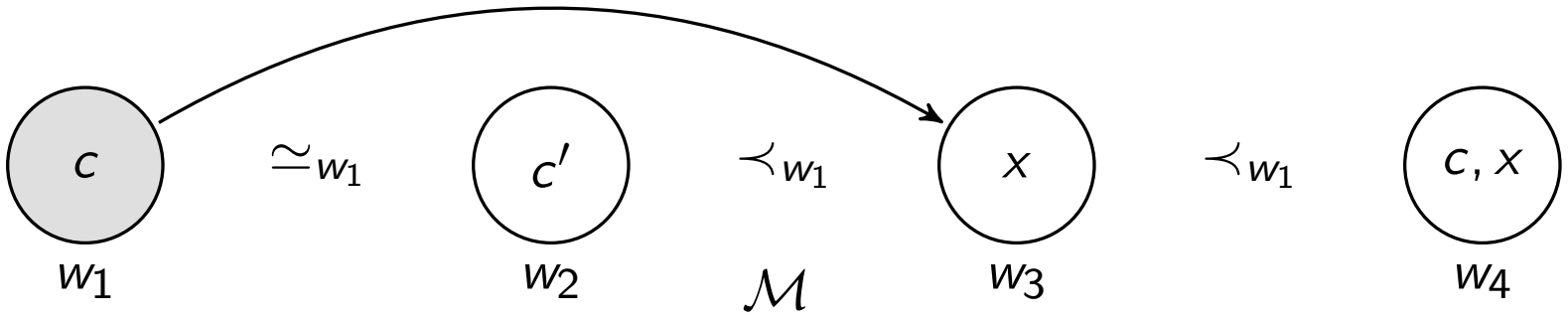


$\mathcal{M}, w_1 \vDash_l K\neg x$

With L-semantics, one can know $\neg x$ even if x is possible and one has not eliminated *any* x-possibilities—even if one has not eliminated any possibilities at all. This would be knowledge of *contingent empirical truths* despite not having empirically eliminated any possibilities—**vacuous knowledge**.

D-semantics and Closure

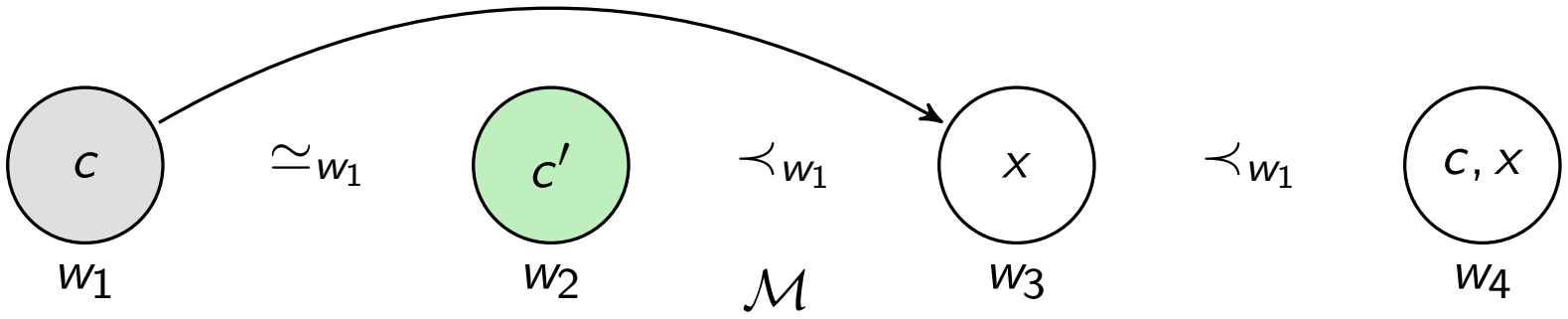
$\mathcal{M}, w \vDash_d K\varphi$ iff for all $v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$, $w \not\rightarrow v$.



D-semantics avoids the skepticism of C-semantics, since knowing φ requires only that the closest $\neg\varphi$-worlds be eliminated, and avoids the vacuous knowledge of L-semantics, since if $\neg\varphi$ is possible there will be some $\neg\varphi$-world(s) that must be eliminated. It also leads to **denial of closure**.

D-semantics and Closure

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

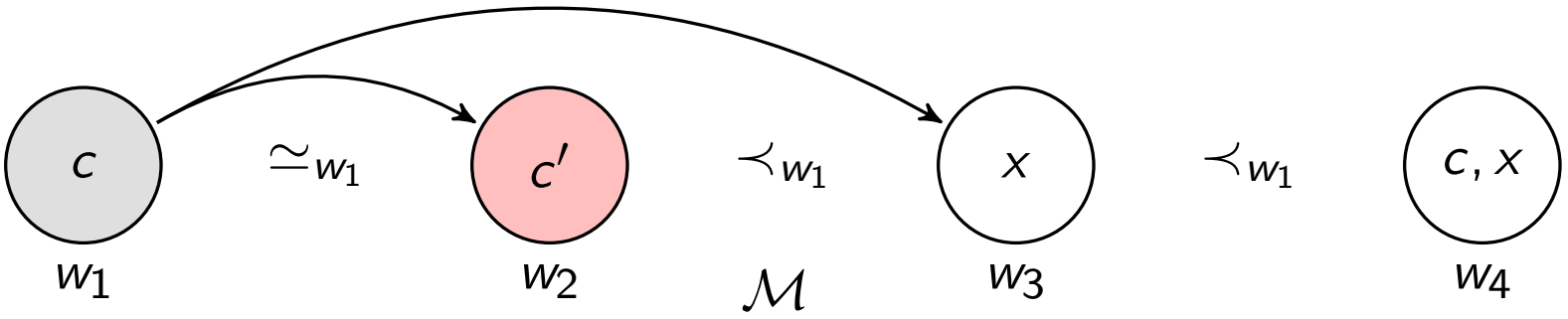


$\mathcal{M}, w_1 \vDash_d Kc$

D-semantics avoids the skepticism of C-semantics, since knowing φ requires only that the closest $\neg\varphi$-worlds be eliminated, and avoids the vacuous knowledge of L-semantics, since if $\neg\varphi$ is possible there will be some $\neg\varphi$-world(s) that must be eliminated. It also leads to **denial of closure**.

D-semantics and Closure

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

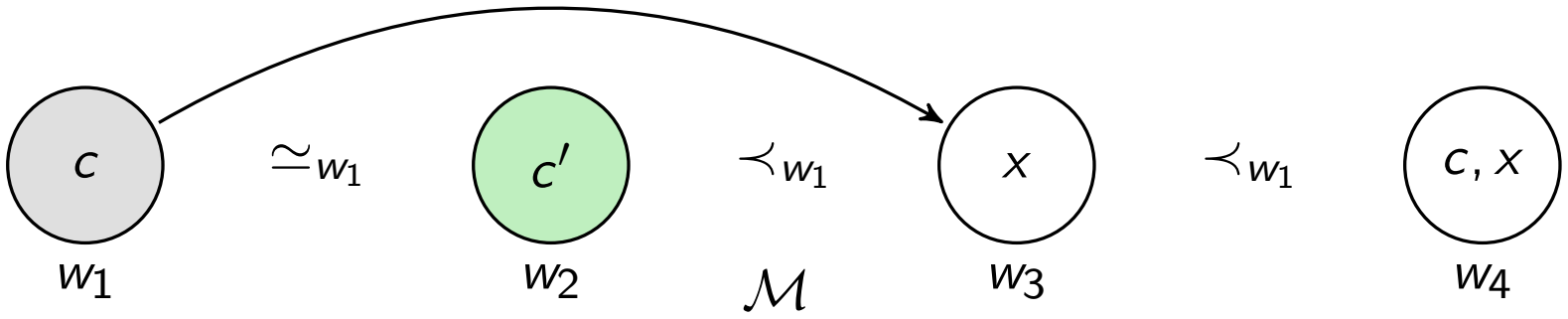


$\mathcal{M}, w_1 \not\vDash_d Kc$

D-semantics avoids the skepticism of C-semantics, since knowing φ requires only that the closest $\neg\varphi$-worlds be eliminated, and avoids the vacuous knowledge of L-semantics, since if $\neg\varphi$ is possible there will be some $\neg\varphi$-world(s) that must be eliminated. It also leads to **denial of closure**.

D-semantics and Closure

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

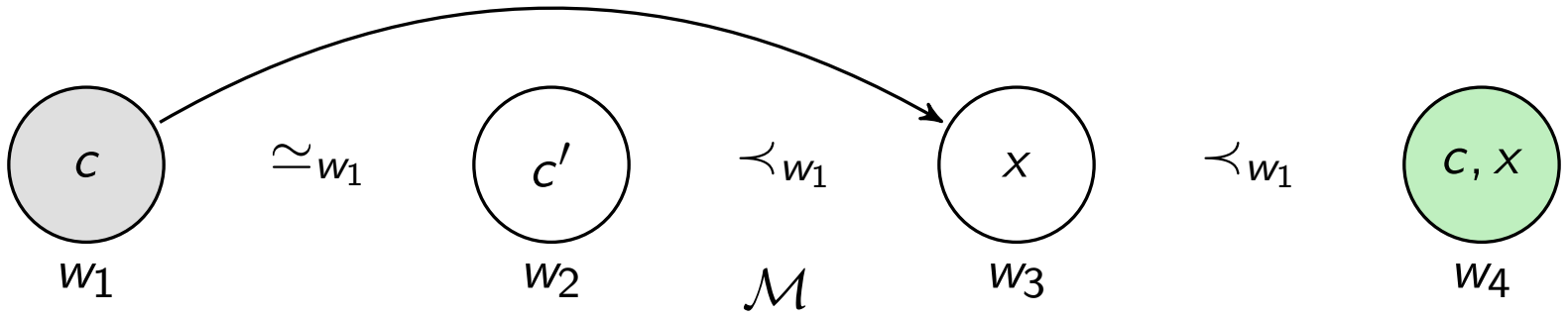


$\mathcal{M}, w_1 \vDash_d Kc$

D-semantics avoids the skepticism of C-semantics, since knowing φ requires only that the closest $\neg\varphi$-worlds be eliminated, and avoids the vacuous knowledge of L-semantics, since if $\neg\varphi$ is possible there will be some $\neg\varphi$-world(s) that must be eliminated. It also leads to **denial of closure**.

D-semantics and Closure

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrowtriangle v$.

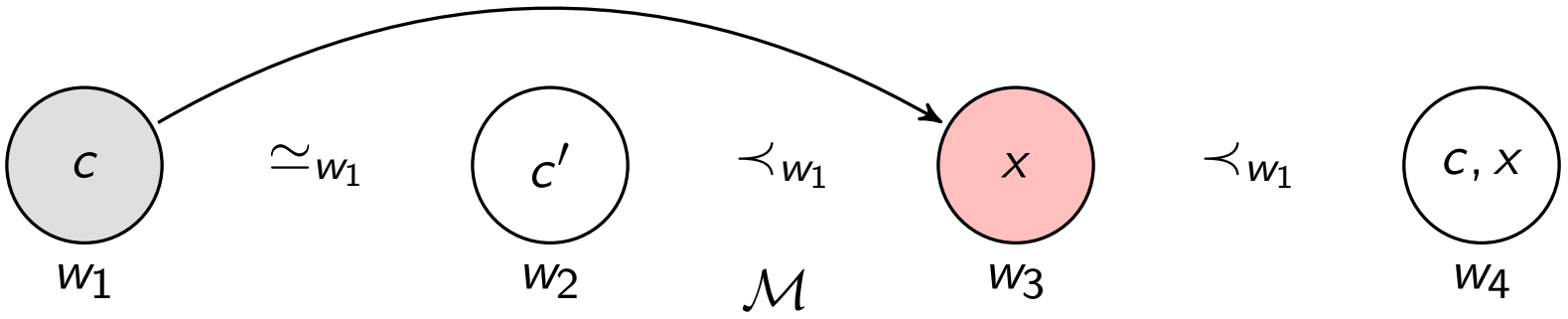


$\mathcal{M}, w_1 \vDash_d Kc$ $\qquad$ $\mathcal{M}, w_1 \vDash_d K(c \to \neg x)$

D-semantics avoids the skepticism of C-semantics, since knowing φ requires only that the closest $\neg\varphi$-worlds be eliminated, and avoids the vacuous knowledge of L-semantics, since if $\neg\varphi$ is possible there will be some $\neg\varphi$-world(s) that must be eliminated. It also leads to **denial of closure**.

D-semantics and Closure

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrowtriangle v$.

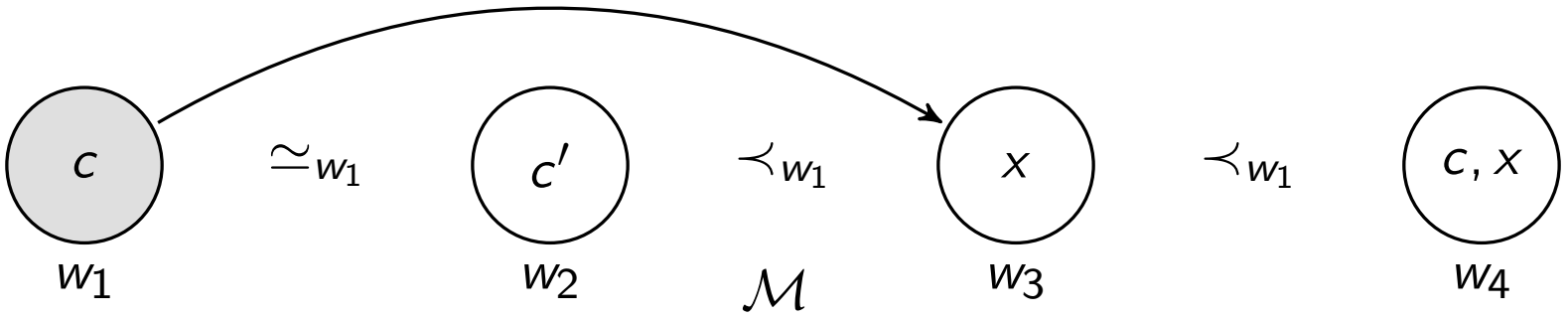


$\mathcal{M}, w_1 \vDash_d Kc$ $\quad$ $\mathcal{M}, w_1 \vDash_d K(c \to \neg x)$ $\quad$ $\mathcal{M}, w_1 \nvDash_d K\neg x$

D-semantics avoids the skepticism of C-semantics, since knowing φ requires only that the closest $\neg\varphi$-worlds be eliminated, and avoids the vacuous knowledge of L-semantics, since if $\neg\varphi$ is possible there will be some $\neg\varphi$-world(s) that must be eliminated. It also leads to **denial of closure**.

D-semantics and Closure

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrowtriangle v$.



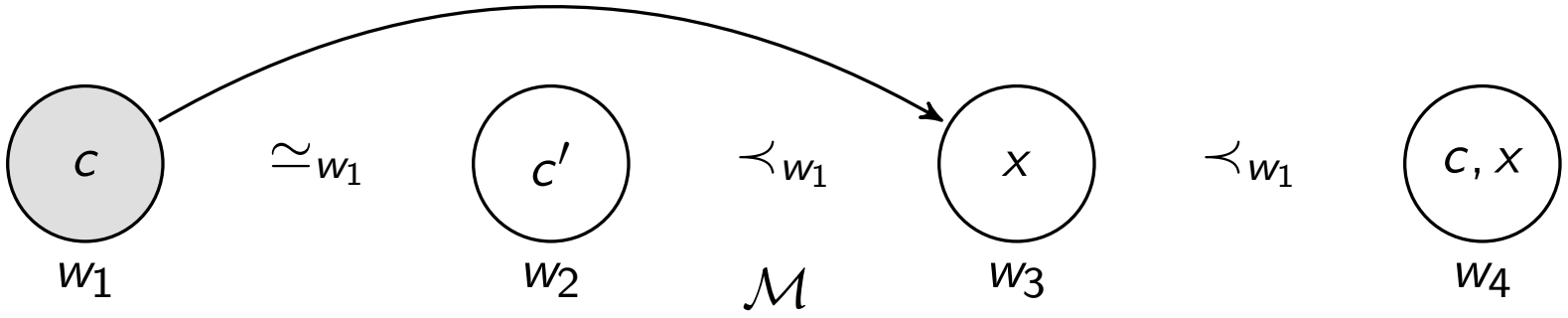
$\mathcal{M}, w_1 \not\vDash_d Kc \wedge K(c \rightarrow \neg x) \rightarrow K\neg x$

D-semantics avoids the skepticism of C-semantics, since knowing φ requires only that the closest $\neg\varphi$-worlds be eliminated, and avoids the vacuous knowledge of L-semantics, since if $\neg\varphi$ is possible there will be some $\neg\varphi$-world(s) that must be eliminated. It also leads to **denial of closure**.

D-Semantics vs. L-Semantics

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

$\mathcal{M}, w \vDash_l K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(W) \cap \overline{[\![\varphi]\!]_l}$: $w \not\rightarrow v$.



Fact (Known Implication)

$K\varphi \wedge K(\varphi \rightarrow \psi) \rightarrow K\psi$ is C/L-valid but not D-valid.

Fact (Known Implication)

$K\varphi \wedge K(\varphi \to \psi) \to K\psi$ *is C/L-valid but not D-valid.*

In the terminology of Dretske (1970), the knowledge operator in D-semantics is not a **fully penetrating** sentential operator.

Fred Dretske. 1970. "Epistemic Operators."

Fact (Known Implication)

$K\varphi \wedge K(\varphi \to \psi) \to K\psi$ *is C/L-valid but not D-valid.*

In the terminology of Dretske (1970), the knowledge operator in D-semantics is not a **fully penetrating** sentential operator.

Fred Dretske. 1970. "Epistemic Operators."

Yet Dretske wants knowledge to be **semi-penetrating**: "it seems to me fairly obvious that if someone knows that P and Q, he thereby knows that Q" and "If he knows that P is the case, he knows that P or Q is the case" (1009). This is the "trivial side" of Dretske's thesis.

Fact (Known Implication)

$K\varphi \wedge K(\varphi \to \psi) \to K\psi$ is C/L-valid but not D-valid.

In the terminology of Dretske (1970), the knowledge operator in D-semantics is not a **fully penetrating** sentential operator.

Fred Dretske. 1970. "Epistemic Operators."

Yet Dretske wants knowledge to be **semi-penetrating**: "it seems to me fairly obvious that if someone knows that P and Q, he thereby knows that Q" and "If he knows that P is the case, he knows that P or Q is the case" (1009). This is the "trivial side" of Dretske's thesis.

However, if we understand the RA theory according to D-semantics, based on Heller's picture, then even these closure principles fail.

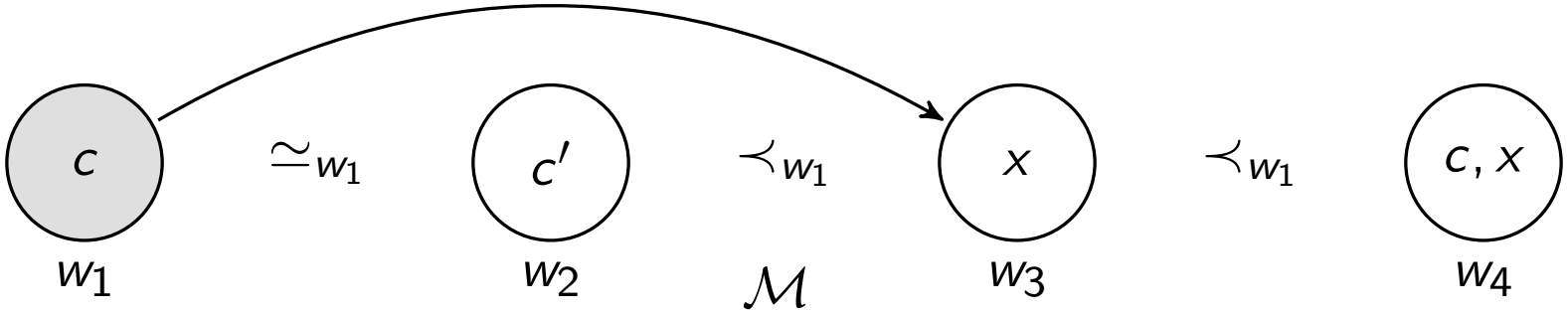
D-Semantics and the Problem of Containment

While some welcome the failure of closure under known implication, D-semantics faces a **problem of containment**: closure failures spread, and they spread to where no one wants them.

D-Semantics and the Problem of Containment

While some welcome the failure of closure under known implication, D-semantics faces a **problem of containment**: closure failures spread, and they spread to where no one wants them.

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

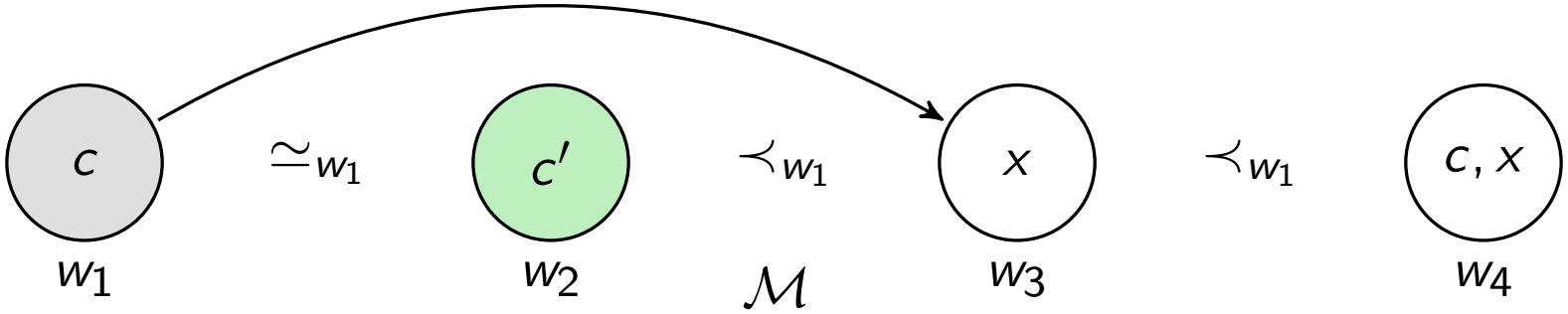


$\mathcal{M}, w_1 \not\vDash_d K(c \wedge \neg x) \rightarrow K\neg x$

D-Semantics and the Problem of Containment

While some welcome the failure of closure under known implication, D-semantics faces a **problem of containment**: closure failures spread, and they spread to where no one wants them.

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

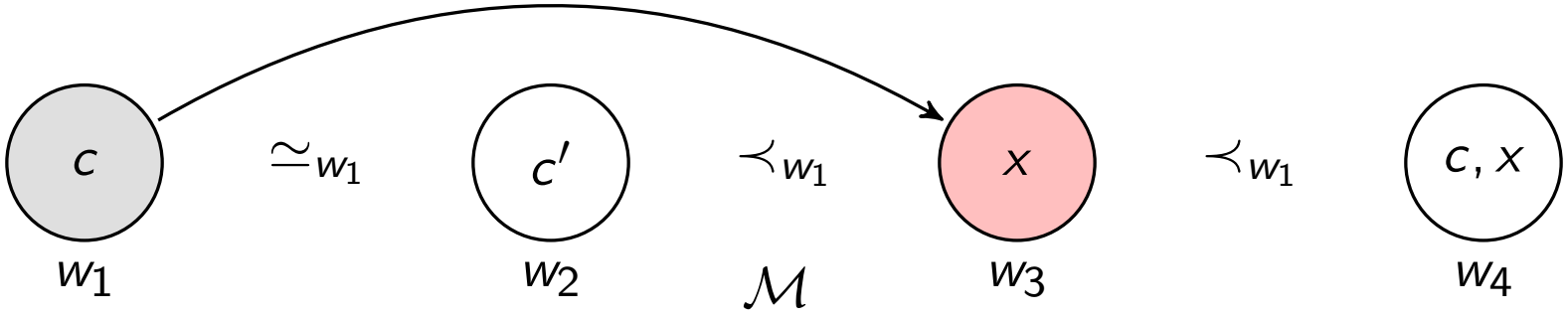


$\mathcal{M}, w_1 \nvDash_d K(c \wedge \neg x) \rightarrow K\neg x$
$\mathcal{M}, w_1 \vDash_d K(c \wedge \neg x)$

D-Semantics and the Problem of Containment

While some welcome the failure of closure under known implication, D-semantics faces a **problem of containment**: closure failures spread, and they spread to where no one wants them.

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

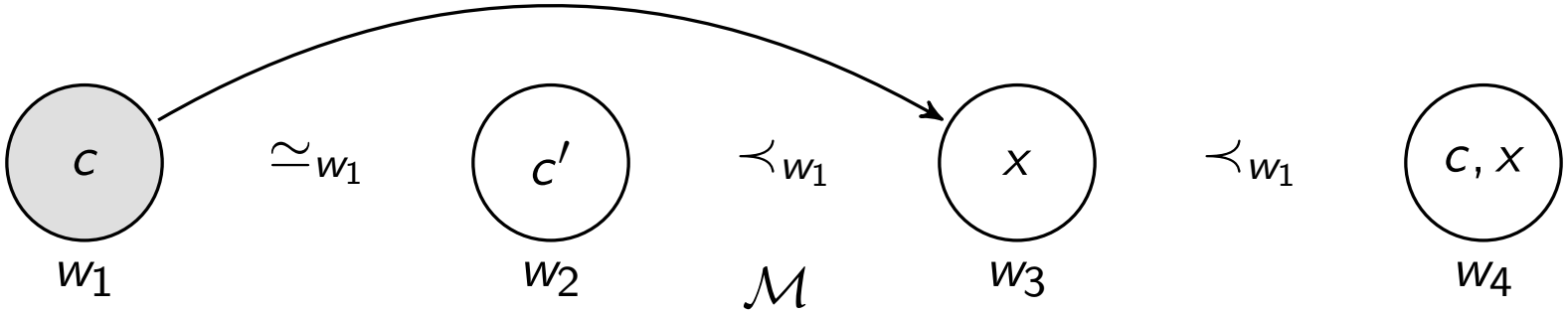


$\mathcal{M}, w_1 \not\vDash_d K(c \wedge \neg x) \rightarrow K\neg x$
$\mathcal{M}, w_1 \vDash_d K(c \wedge \neg x)$
$\mathcal{M}, w_1 \not\vDash_d K\neg x$

D-Semantics and the Problem of Containment

While some welcome the failure of closure under known implication, D-semantics faces a **problem of containment**: closure failures spread, and they spread to where no one wants them.

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

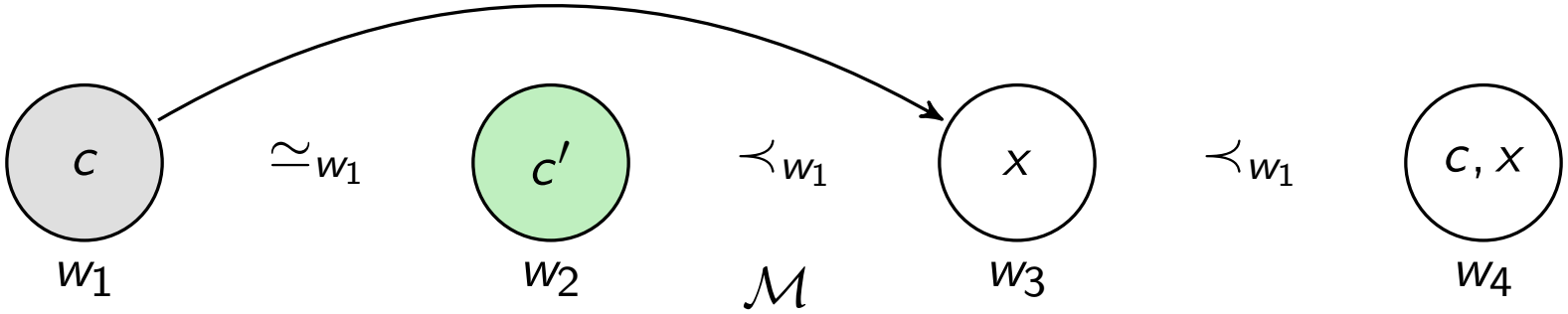


$\mathcal{M}, w_1 \not\vDash_d Kc \rightarrow K(c \vee \neg x)$

D-Semantics and the Problem of Containment

While some welcome the failure of closure under known implication, D-semantics faces a **problem of containment**: closure failures spread, and they spread to where no one wants them.

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.

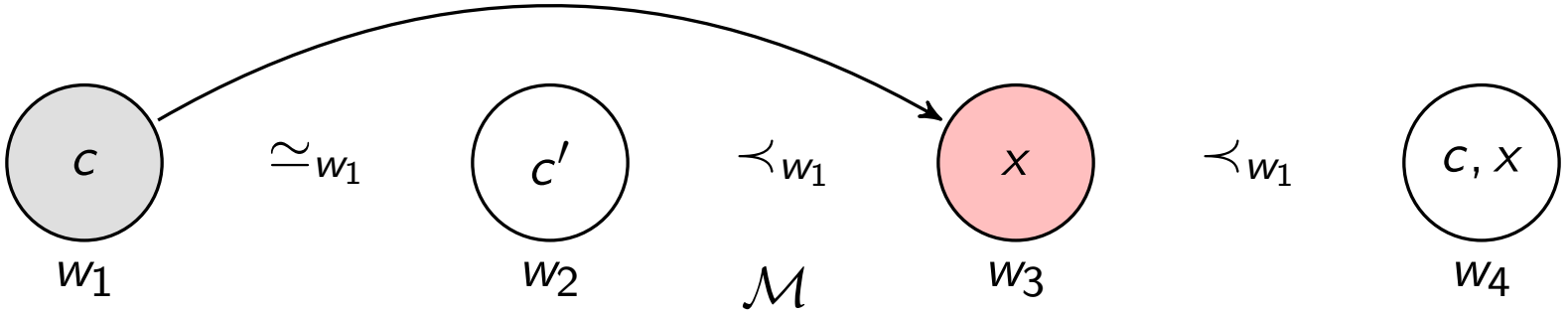


$\mathcal{M}, w_1 \nvDash_d Kc \rightarrow K(c \vee \neg x)$
$\mathcal{M}, w_1 \vDash_d Kc$

D-Semantics and the Problem of Containment

While some welcome the failure of closure under known implication, D-semantics faces a **problem of containment**: closure failures spread, and they spread to where no one wants them.

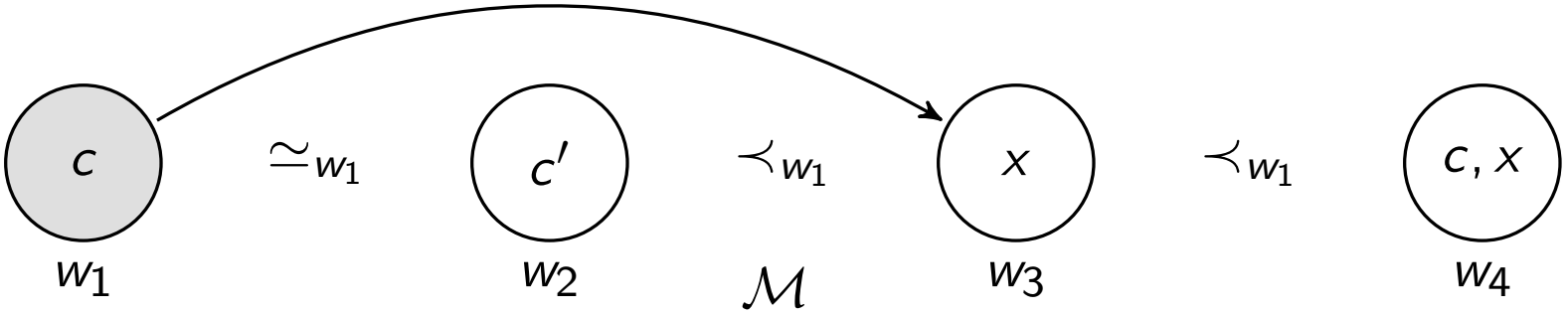
$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.



$\mathcal{M}, w_1 \nvDash_d Kc \rightarrow K(c \vee \neg x)$
$\mathcal{M}, w_1 \vDash_d Kc$
$\mathcal{M}, w_1 \nvDash_d K(c \vee \neg x)$

D-Semantics and the Problem of Containment

$\mathcal{M}, w \vDash_d K\varphi$ iff $\forall v \in \mathrm{Min}_{\preceq_w}(\overline{[\![\varphi]\!]_d})$: $w \not\rightarrow v$.



Fact (Distribution and Addition)

$K(\varphi \wedge \psi) \rightarrow K\varphi$ and $K\varphi \rightarrow K(\varphi \vee \psi)$ are C/L-valid, but not D-valid.

A Dretskean Dilemma

Fact (Distribution and Addition)

$K(\varphi \wedge \psi) \to K\varphi$ *and* $K\varphi \to K(\varphi \vee \psi)$ *are C/L-valid, but not D-valid.*

This is only the tip of the iceberg . . . Yet it already points to a **dilemma**:

A Dretskean Dilemma

Fact (Distribution and Addition)

$K(\varphi \wedge \psi) \to K\varphi$ and $K\varphi \to K(\varphi \vee \psi)$ are C/L-valid, but not D-valid.

This is only the tip of the iceberg . . . Yet it already points to a **dilemma**:

- If we understand RA theory according to D-semantics, then K is not semi-penetrating, contrary to the "trivial side" of Dretske's thesis.

A Dretskean Dilemma

Fact (Distribution and Addition)

$K(\varphi \wedge \psi) \to K\varphi$ and $K\varphi \to K(\varphi \vee \psi)$ are C/L-valid, but not D-valid.

This is only the tip of the iceberg . . . Yet it already points to a **dilemma**:

- If we understand RA theory according to D-semantics, then K is not semi-penetrating, contrary to the "trivial side" of Dretske's thesis.
- If we understand the theory according to L-semantics, then K is fully-penetrating, contrary to the non-trivial side of Dretske's thesis;

A Dretskean Dilemma

Fact (Distribution and Addition)

$K(\varphi \wedge \psi) \to K\varphi$ and $K\varphi \to K(\varphi \vee \psi)$ are C/L-valid, but not D-valid.

This is only the tip of the iceberg . . . Yet it already points to a **dilemma**:

- If we understand RA theory according to D-semantics, then K is not semi-penetrating, contrary to the "trivial side" of Dretske's thesis.
- If we understand the theory according to L-semantics, then K is fully-penetrating, contrary to the non-trivial side of Dretske's thesis; and we have the problem of vacuous knowledge.

A Dretskean Dilemma

Fact (Distribution and Addition)

$K(\varphi \wedge \psi) \to K\varphi$ and $K\varphi \to K(\varphi \vee \psi)$ are C/L-valid, but not D-valid.

This is only the tip of the iceberg . . . Yet it already points to a **dilemma**:

- If we understand RA theory according to D-semantics, then K is not semi-penetrating, contrary to the "trivial side" of Dretske's thesis.
- If we understand the theory according to L-semantics, then K is fully-penetrating, contrary to the non-trivial side of Dretske's thesis; and we have the problem of vacuous knowledge.

It is difficult to escape this dilemma while retaining the picture "of worlds arranged around the actual world in order of similarity, with those that are too far away from the actual world being irrelevant" (Heller 1999, 119).