Day 1: Epistemic Logic

NASSLLI 2022

Adam Bjorndahl

Carnegie Mellon University

Aces & Eights

- Three players are dealt two cards each from a deck consisting of 4 aces and 4 eights.
  - Two cards remain undealt.
- No one is allowed to look at their own cards; however, each player shows their two cards to everyone else.
- The object of the game is to figure out what cards you are holding.
  - Suit doesn't matter!
  - It's either two aces, two eights, or one of each.
- Players take turns: if they don't know their own cards, they announce their ignorance and play passes to the next player.
  - Continue cycling through the players until somebody wins or everybody gives up.

Scenario 1

Assume the order of play is Alice, then Bob, then Carl, then Alice again, etc.

- Suppose Alice holds two aces and Bob holds two eights.
- Both declare they do not know which cards they are holding.
- Now it's Carl's turn; can he determine what cards he is holding?

Scenario 2

- Now suppose that Bob holds two eights and Carl holds an ace and an eight.
- No one can figure out their own cards on the first round.
- Play returns to Alice: can she figure out her own cards?

Scenario 3

- Now suppose that Alice and Carl both hold an ace and an eight.
- No one can figure out their own cards on the first round, and Alice still can't figure out her cards on her second turn.
- Can Bob deduce his hand?

Questions

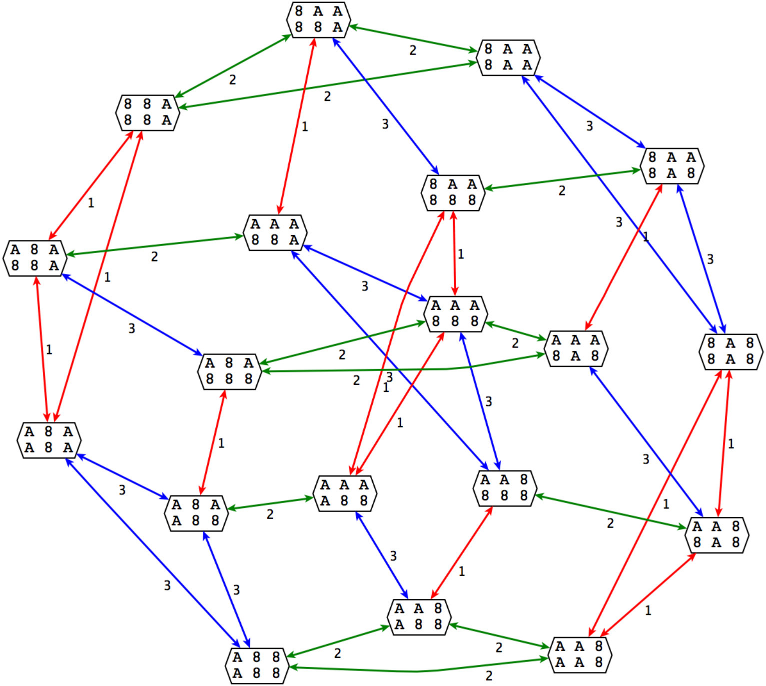
- Does someone always win?

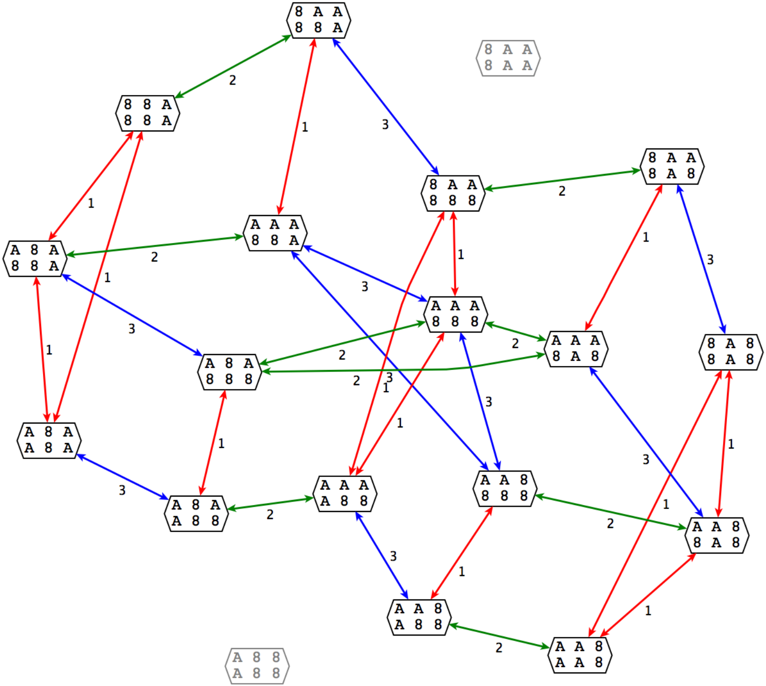
Questions

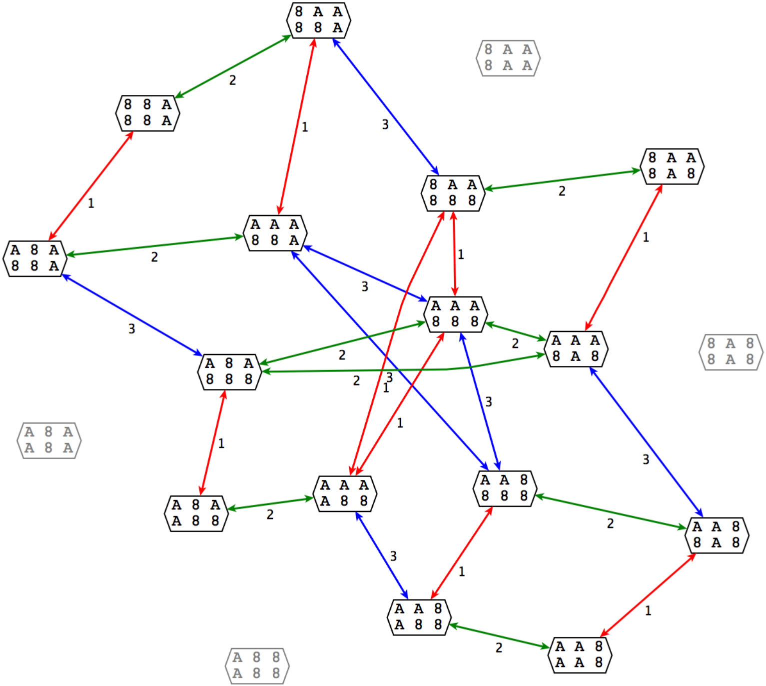
- Does someone always win?
- How long can a game last?

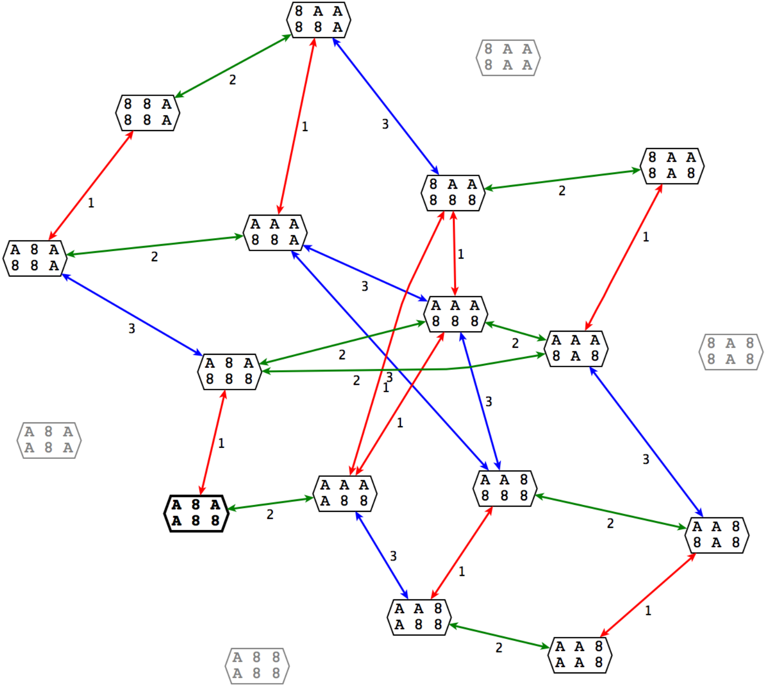
Questions

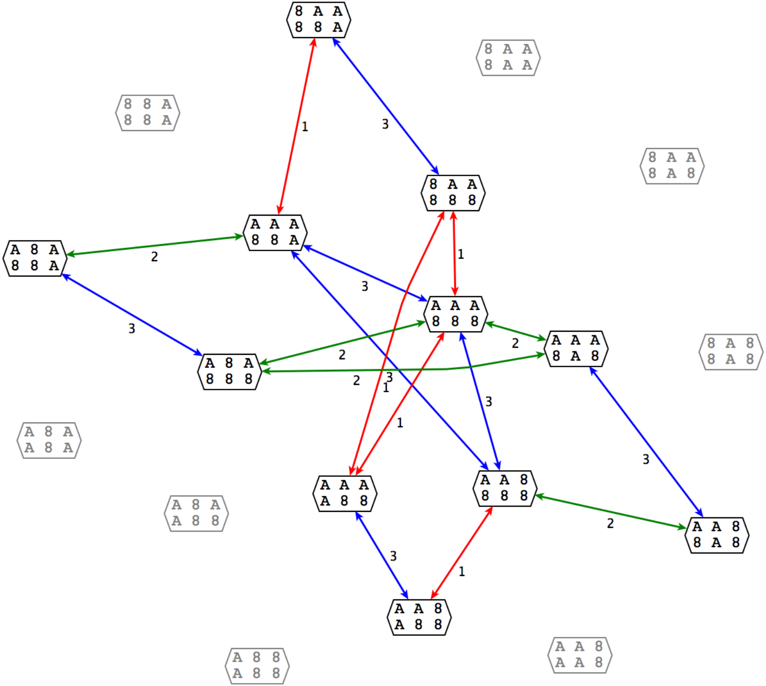
- Does someone always win?
- How long can a game last?
- Is it possible for one player to figure out their own cards, but the other two remain forever ignorant, no matter how long the game is played?

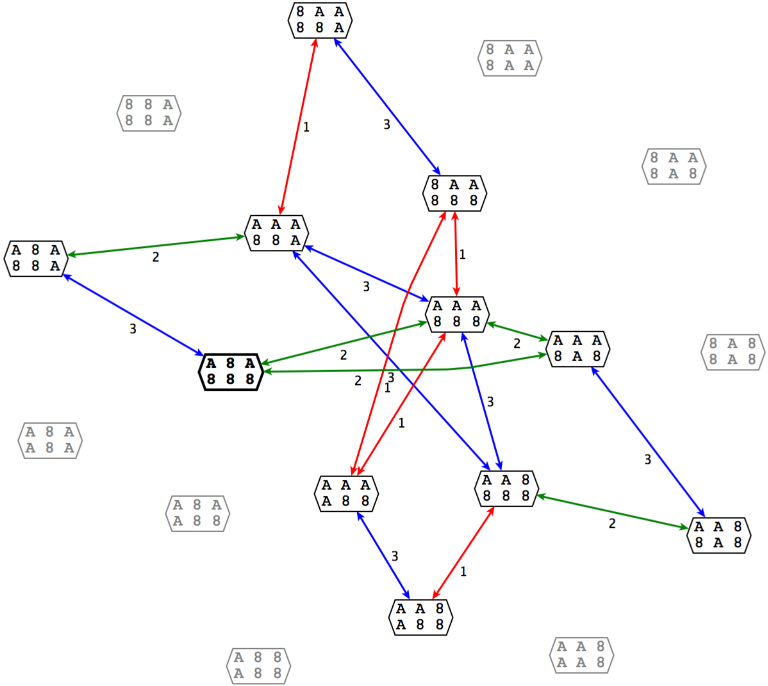

8 A A
8 8 A
8 A A
8 A A
8 8 A
8 8 A
8 A A
8 A 8
8 A A
8 8 8
A A A
8 8 A
A 8 A
8 8 A
A A A
8 8 8
A A A
8 A 8
8 A 8
8 A 8
A 8 A
8 8 8
A 8 A
A 8 A
A A A
A 8 8
A A 8
8 8 8
A 8 A
A 8 8
A A 8
8 A 8
A A 8
A 8 8
A 8 8
A 8 8
A A 8
A A 8

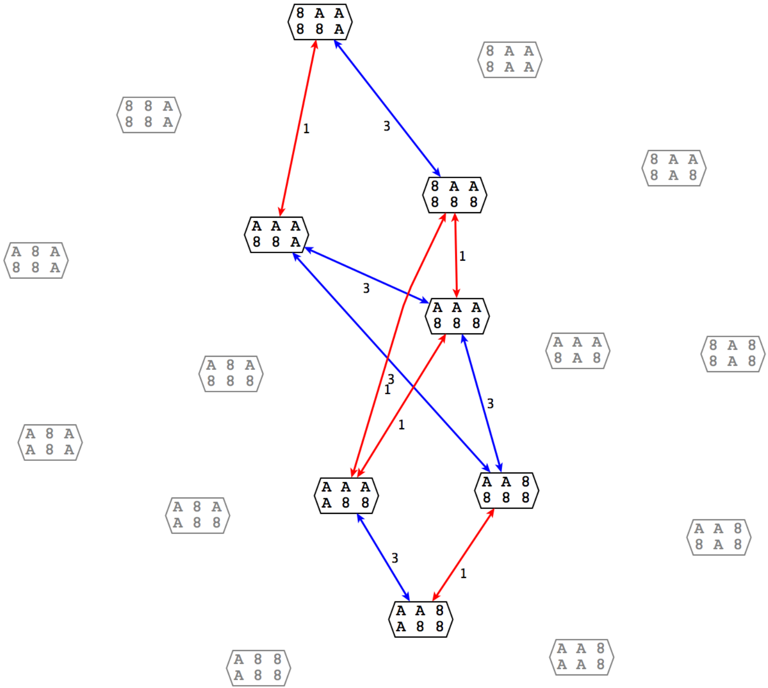

8 A A
8 8 A
8 A A
8 A A
8 8 A
8 8 A
8 A A
8 A 8
8 A A
8 8 8
A A A
8 8 A
A 8 A
8 8 A
A A A
8 8 8
A A A
8 A 8
8 A 8
8 A 8
A 8 A
8 8 8
A 8 A
A 8 A
A 8 A
A 8 8
A A A
A 8 8
A A 8
8 8 8
A A 8
8 A 8
A A 8
A 8 8
A A 8
A A 8
A 8 8
A 8 8

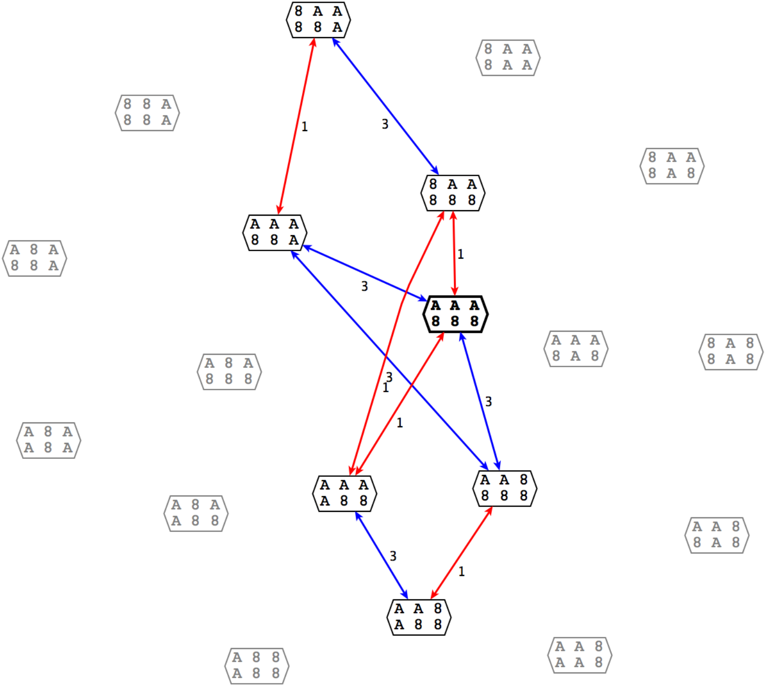

8 A A 8 8 A
8 A A 8 A A
8 8 A 8 8 A
8 A A 8 A 8
8 A A 8 8 8
A A A 8 8 A
A 8 A 8 8 A
A A A 8 8 8
A A A 8 A 8
8 A 8 8 A 8
A 8 A 8 8 8
A 8 A A 8 A
A A 8 8 8 8
A A A A 8 8
A 8 A A 8 8
A A 8 8 A 8
A A 8 A 8 8
A A 8 A A 8
A 8 8 A 8 8


8 A A
8 8 A
8 A A
8 A A
8 8 A
8 8 A
8 A A
8 A 8
8 A A
8 8 8
A A A
8 8 A
A 8 A
8 8 A
A A A
8 8 8
A A A
8 A 8
8 A 8
8 A 8
A 8 A
8 8 8
A 8 A
A 8 A
A 8 A
A 8 8
A A A
A 8 8
A A 8
8 8 8
A A 8
8 A 8
A A 8
A 8 8
A A 8
A A 8
A 8 8
A 8 8


8 A A
8 8 A
8 A A
8 A A
8 8 A
8 8 A
8 A A
8 A 8
8 A A
8 8 8
A A A
8 8 A
A 8 A
8 8 A
A A A
8 8 8
A A A
8 A 8
8 A 8
8 A 8
A 8 A
8 8 8
A 8 A
A 8 A
A A A
A 8 8
A A 8
8 8 8
A 8 A
A 8 8
A A 8
8 A 8
A A 8
A 8 8
A 8 8
A 8 8
A A 8
A A 8
1
2
3


8 A A
8 8 A
8 A A
8 A A
8 8 A
8 8 A
8 A A
8 A 8
8 A A
8 8 8
A A A
8 8 A
A 8 A
8 8 A
A A A
8 8 8
A A A
8 A 8
8 A 8
8 A 8
A 8 A
8 8 8
A 8 A
A 8 A
A A A
A 8 8
A A 8
8 8 8
A 8 A
A 8 8
A A 8
8 A 8
A A 8
A 8 8
A 8 8
A 8 8
A A 8
A A 8


8 A A
8 8 A
8 A A
8 A A
8 8 A
8 8 A
8 A A
8 A 8
8 A A
8 8 8
A A A
8 8 A
A 8 A
8 8 A
A A A
8 8 8
A A A
8 A 8
8 A 8
8 A 8
A 8 A
8 8 8
A 8 A
A 8 A
A A A
A 8 8
A A 8
8 8 8
A 8 A
A 8 8
A A 8
8 A 8
A A 8
A 8 8
A 8 8
A 8 8
A A 8
A A 8
1
3
1
3
3
1
1
3
3
1


8 A A
8 8 A
8 A A
8 A A
8 8 A
8 8 A
8 A A
8 A 8
1
3
8 A A
8 8 8
A A A
8 8 A
A 8 A
8 8 A
1
3
A A A
8 8 8
A A A
8 A 8
8 A 8
8 A 8
A 8 A
8 8 8
3
1
3
1
A 8 A
A 8 A
A A A
A 8 8
A A 8
8 8 8
A 8 A
A 8 8
A A 8
8 A 8
3
1
A A 8
A 8 8
A 8 8
A 8 8
A A 8
A A 8

Epistemic logic

The **basic epistemic language** $\mathcal{EL}$ is generated by:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi,$$

where $p \in \textsc{prop}$ (a countable set of *primitive propositions*), and $K\varphi$ is read "the agent knows φ".

Epistemic logic

The **basic epistemic language** $\mathcal{EL}$ is generated by:

$$\varphi ::= p \,|\, \neg\varphi \,|\, \varphi \wedge \psi \,|\, K\varphi,$$

where $p \in \textsc{prop}$ (a countable set of *primitive propositions*), and $K\varphi$ is read "the agent knows φ".

Formulas of $\mathcal{EL}$ are interpreted in **epistemic models** $M = (X, R, v)$, where:

- X is a (nonempty) set of *worlds* or *states*
- $R \subseteq X^2$ is a binary relation on X called the *accessibility relation*
  - when xRy we say that y is *accessible* from x
  - $R(x) = \{y \in X \,:\, xRy\}$ denotes the set of all worlds accessible from x
- $v : \textsc{prop} \to 2^X$ is a function called a *valuation*

Epistemic logic

The set $v(p) \subseteq X$ is called the *extension* or *truth set* of p; intuitively, it is the set of worlds where p is true.

This notion of truth is extended to all formulas of $\mathcal{EL}$ by defining $[\![\varphi]\!] \subseteq X$ recursively as follows:

$$
\begin{aligned}
[\![p]\!] &= v(p) \\
[\![\neg\varphi]\!] &= X \setminus [\![\varphi]\!] \\
[\![\varphi \wedge \psi]\!] &= [\![\varphi]\!] \cap [\![\psi]\!] \\
[\![K\varphi]\!] &= \{x \in X \; : \; (\forall y \in X)(xRy \Rightarrow y \in [\![\varphi]\!])\} \\
&= \{x \in X \; : \; R(x) \subseteq [\![\varphi]\!]\}.
\end{aligned}
$$

Epistemic logic

The set $v(p) \subseteq X$ is called the *extension* or *truth set* of p; intuitively, it is the set of worlds where p is true.

This notion of truth is extended to all formulas of $\mathcal{EL}$ by defining $[\![\varphi]\!] \subseteq X$ recursively as follows:

$$
\begin{aligned}
[\![p]\!] &= v(p) \\
[\![\neg\varphi]\!] &= X \setminus [\![\varphi]\!] \\
[\![\varphi \wedge \psi]\!] &= [\![\varphi]\!] \cap [\![\psi]\!] \\
[\![K\varphi]\!] &= \{x \in X : (\forall y \in X)(xRy \Rightarrow y \in [\![\varphi]\!])\} \\
&= \{x \in X : R(x) \subseteq [\![\varphi]\!]\}.
\end{aligned}
$$

When $x \in [\![\varphi]\!]$ we say φ *is true at* x *(in* M*)* and write $M, x \models \varphi$.

So $K\varphi$ is true at x just in case every accessible world is a φ-world.

- Intuitively: φ is true in every situation compatible with the agent's evidence at x, i.e., φ is guaranteed by the evidence.

Epistemic logic

The *dual* of the K modality is written $\hat{K}$ and is defined by

$$\hat{K}\varphi \equiv \neg K \neg \varphi.$$

So $\hat{K}\varphi$ reads "the agent does not know that φ is false".

Epistemic logic

The *dual* of the K modality is written $\hat{K}$ and is defined by

$$\hat{K}\varphi \equiv \neg K \neg \varphi.$$

So $\hat{K}\varphi$ reads "the agent does not know that φ is false".

$$\begin{aligned}
[\![\hat{K}\varphi]\!] &= [\![\neg K \neg \varphi]\!] \\
&= X \setminus [\![K \neg \varphi]\!] \\
&= X \setminus \{x \in X : R(x) \subseteq [\![\neg \varphi]\!]\} \\
&= X \setminus \{x \in X : R(x) \subseteq (X \setminus [\![\varphi]\!])\} \\
&= X \setminus \{x \in X : R(x) \cap [\![\varphi]\!] = \emptyset\} \\
&= \{x \in X : R(x) \cap [\![\varphi]\!] \neq \emptyset\}.
\end{aligned}$$

Thus, $\hat{K}\varphi$ is true at x iff some φ-world is accessible from x.

- Intuitively: φ is compatible with the agent's evidence.

Epistemic logic

A formula φ is called *valid* in a model $M = (X, R, v)$ if $[\![\varphi]\!] = X$.

Epistemic logic

A formula φ is called *valid* in a model $M = (X, R, v)$ if $[\![\varphi]\!] = X$.

- We typically restrict our attention to models where R is reflexive (i.e., $(\forall x \in X)(xRx)$). Any such model validates the "factivity" of knowledge:
  - $K\varphi \to \varphi$ is valid in reflexive models.

Epistemic logic

A formula φ is called *valid* in a model $M = (X, R, v)$ if $[\![\varphi]\!] = X$.

- We typically restrict our attention to models where R is reflexive (i.e., $(\forall x \in X)(xRx)$). Any such model validates the "factivity" of knowledge:
  - $K\varphi \to \varphi$ is valid in reflexive models.
- It is also often assumed that R is transitive (i.e., $(\forall x, y, z \in X)(xRy \,\&\, yRz \Rightarrow xRz)$). The corresponding property of knowledge is called "positive introspection":
  - $K\varphi \to KK\varphi$ is valid in transitive models.

Epistemic logic

A formula φ is called *valid* in a model $M = (X, R, v)$ if $[\![\varphi]\!] = X$.

- We typically restrict our attention to models where R is reflexive (i.e., $(\forall x \in X)(xRx)$). Any such model validates the "factivity" of knowledge:
  - $K\varphi \to \varphi$ is valid in reflexive models.
- It is also often assumed that R is transitive (i.e., $(\forall x, y, z \in X)(xRy \;\&\; yRz \Rightarrow xRz)$). The corresponding property of knowledge is called "positive introspection":
  - $K\varphi \to KK\varphi$ is valid in transitive models.
- When R is an equivalence relation—reflexive, transitive, and symmetric (i.e., $(\forall x, y \in X)(xRy \Rightarrow yRx)$)—the corresponding model validates factivity, positive introspection, and "negative introspection":
  - $\neg K\varphi \to K\neg K\varphi$ is valid when R is an equivalence relation.