Comparing Models for Epistemic Game Theory

Paolo Galeazzi

Institute for Logic, Language and Computation
University of Amsterdam



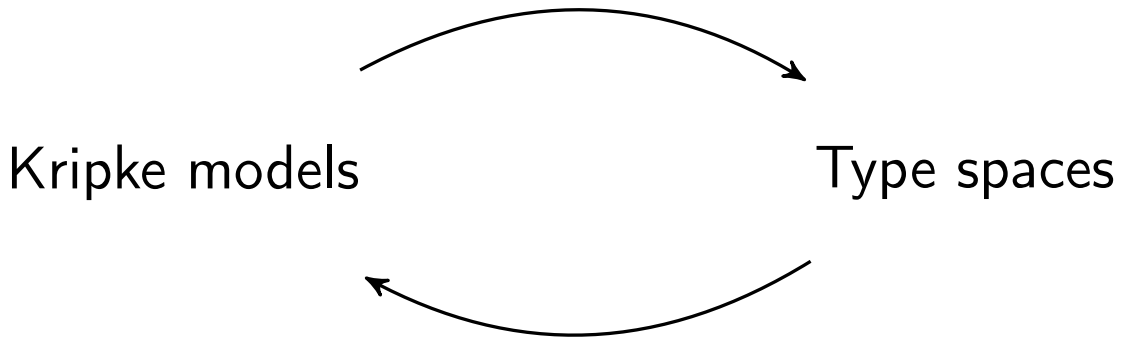
Introduction

In recent years many game theorists have understood that the analysis of the epistemic situation of the players is a fundamental aspect for the description of a game, in particular if we want to focus on the concept of rationality.

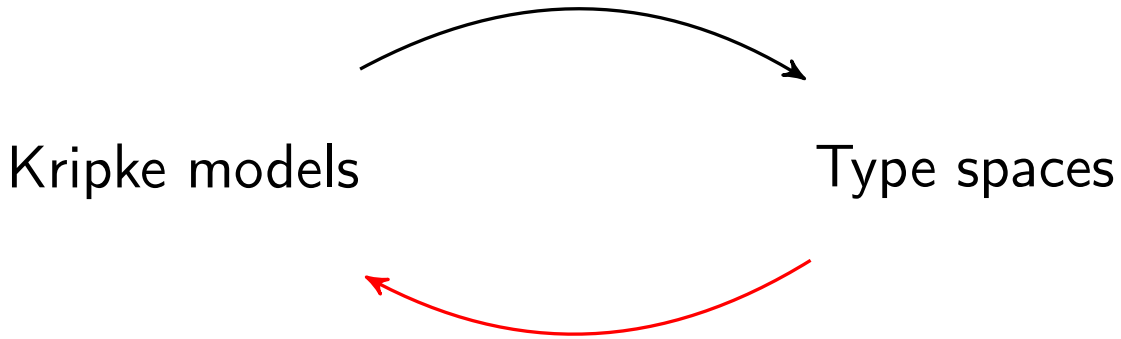
In the literature there are at least two main frameworks to model epistemic situations:

- Kripke models
- Type spaces

But what is the relation between the two?



The aim of this work is to show how we can translate type spaces into Kripke models, and to compare the expressivity of the two structures.



We will proceed as follows:

- Type space
- Plausibility model
- Translation
- Probabilistic model
- Comparison with other work and future work

Type space

A type space $\mathbf{T}$ for a game $\mathbf{G} = \langle N, S_1, ..., S_n, \pi_1, ..., \pi_n \rangle$ is a structure $\mathbf{T} = \langle S_1, ..., S_n, T_1, ..., T_n, \beta_1, ..., \beta_n \rangle$, where:

- S_i is the set of strategies of player i
- T_i is the set of types of player i
- $\beta_i : T_i \longrightarrow \Delta(S_{-i} \times T_{-i})$ is a belief function that associates each type of player i with a probability distribution over the Cartesian product of the types and the strategies of the other players.

Each element $(t, s) \in T \times S$ is a state. An event e is $e \subseteq T \times S$.
Belief operator:
$\mathscr{B}_i(e) = \{(t_i, t_{-i}, s) \in S \times T : \sum_{(t_{-i}, s_{-i}) \in e} \beta_i(t_i)(t_{-i}, s_{-i}) = 1\}$

An example

To keep things simple we consider a two player game

$$\mathbf{G} = \langle Ann, Bob; S_A = \{U, D\}, S_B = \{L, R\}; \pi_A, \pi_B \rangle$$

An example

An example of type space for **G** is the following **T**:

- $S_A = \{U, D\}$; $T_B = \{L, R\}$
- $T_A = \{1, 2\}$; $T_B = \{1, 2\}$
- $\beta_A(1)(1L) \mapsto \frac{1}{2}$, $\beta_A(1)(2L) \mapsto \frac{1}{2}$, $\beta_A(1)(1R) \mapsto 0$, $\beta_A(1)(2R) \mapsto 0$
  $\beta_A(2)(1L) \mapsto \frac{1}{3}$, $\beta_A(2)(2L) \mapsto 0$, $\beta_A(2)(1R) \mapsto \frac{2}{3}$, $\beta_A(2)(2R) \mapsto 0$
  $\beta_B(1)(1U) \mapsto \frac{1}{3}$, $\beta_B(1)(2U) \mapsto 0$, $\beta_B(1)(1D) \mapsto 0$, $\beta_B(1)(2D) \mapsto \frac{2}{3}$
  $\beta_B(2)(1U) \mapsto 1$, $\beta_B(2)(2U) \mapsto 0$, $\beta_B(2)(1D) \mapsto 0$, $\beta_B(2)(2D) \mapsto 0$

An example

Ann \ Bob	$1L$	$2L$	$1R$	$2R$
$1U$	$\frac{1}{2}, \frac{1}{3}$	$\frac{1}{2}, 1$	$0, \frac{1}{3}$	$0, 1$
$2U$	$\frac{1}{3}, 0$	$0, 0$	$\frac{2}{3}, 0$	$0, 0$
$1D$	$\frac{1}{2}, 0$	$\frac{1}{2}, 0$	$0, 0$	$0, 0$
$2D$	$\frac{1}{3}, \frac{2}{3}$	$0, 0$	$\frac{2}{3}, \frac{2}{3}$	$0, 0$

An example

- $\mathscr{B}_{Ann}(1L, 2L) = \{(1U, 1L), (1U, 2L), (1U, 1R), (1U, 2R), (1D, 1L), (1D, 2L), (1D, 1R), (1D, 2R)\}$
- $\mathscr{B}_{Ann}(1L, 1R) = \{(2U, 1L), (2U, 2L), (2U, 1R), (2U, 2R), (2D, 1L), (2D, 2L), (2D, 1R), (2D, 2R)\}$
- $\mathscr{B}_{Bob}(1U, 1D) = \{(1U, 2L), (2U, 2L), (1D, 2L), (2D, 2L), (1U, 2R), (2U, 2R), (1D, 2R), (2D, 2R)\}$
- $\mathscr{B}_{Bob}(1U) = \{(1U, 2L), (2U, 2L), (1D, 2L), (2D, 2L), (1U, 2R), (2U, 2R), (1D, 2R), (2D, 2R)\}$
- $\mathscr{B}_{Bob}\mathscr{B}_{Ann}(1L, 2L) = \{(1U, 2L), (2U, 2L), (1D, 2L), (2D, 2L), (1U, 2R), (2U, 2R), (1D, 2R), (2D, 2R)\}$

Plausibility model

A plausibility model for N agents is a structure $\mathbf{M} = \langle N, W, \succeq_1, ..., \succeq_n, \upsilon \rangle$, where

- N is the set of agents
- W is a set of possible worlds
- $\succeq_i$ is a preorder, i.e. a reflexive and transitive relation on W
- $\upsilon : \Phi \longrightarrow \wp(W)$ is a valuation function, i.e. a function that assigns to each primitive proposition $p \in \Phi$ a set of possible worlds in which p holds

From preorders $\succeq_i$ we can define an equivalence relation $\sim_i$ over W in the following way:

$$\forall w, w' \; w \sim_i w' \text{ iff } w \succeq_i w' \text{ or } w' \succeq_i w$$

Operators

- $w \vDash K_i\varphi$ iff $v \vDash \varphi$ for all v s.t. $v \sim_i w$
- $w \vDash B_i\varphi$ iff $v \vDash \varphi$ for all $v \in Sup_{\geq_i}([w]_{\sim_i})$

Translation

We define the plausibility state model $\mathbf{TM}$ corresponding to a given type space $\mathbf{T}$ as follows:

$$\mathbf{TM} = \langle N, W, \sim_1, ..., \sim_n, \succeq_1, ..., \succeq_n, \upsilon \rangle$$

where

- $W = S \times T$ is the set of worlds;
- $\upsilon : \Phi \longrightarrow \wp(W)$ is the valuation function, where $\Phi = \{S_1, ..., S_n\}$ and υ s.t. $w \in \upsilon(s_i)$ iff $w \equiv (t, s_i, s_{-i})$;
- $\sim_i$ is the accessibility relation of player i, given by: $w \sim_i w'$ iff $t_i(w) = t_i(w')$. Then $\sim_i$ determines a partition over W;
- $\succeq_i$ is the plausibility ordering of player i, that satisfies: $\forall w, w'$ $w \succeq_i w'$ or $w' \succeq_i w$ iff $w \sim_i w'$.

Two different plausibility orderings

Since we are translating a probabilistic type space into a plausibility model, we will not get a quantitative probability over worlds. Indeed, we obtain a plausibility ordering $\succeq_i$, that we can define at least in two different ways:

- orderings preserving: $w \succeq_i w'$ iff $\beta_i(t_i)(t_{-i}(w), s_{-i}(w)) \geq \beta_i(t_i)(t_{-i}(w'), s_{-i}(w'))$
- operators preserving: $w \succeq_i w'$ and $w' \not\succeq_i w$ iff $\beta_i(t_i)(t_{-i}(w), s_{-i}(w)) > 0$ and $\beta_i(t_i)(t_{-i}(w'), s_{-i}(w')) = 0$

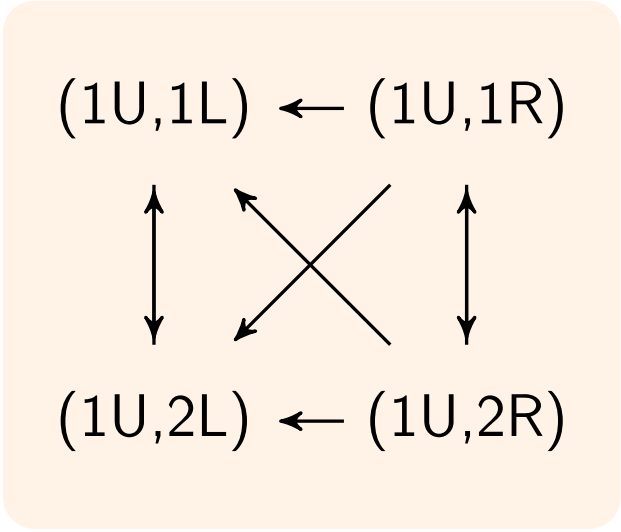
The ex-interim condition

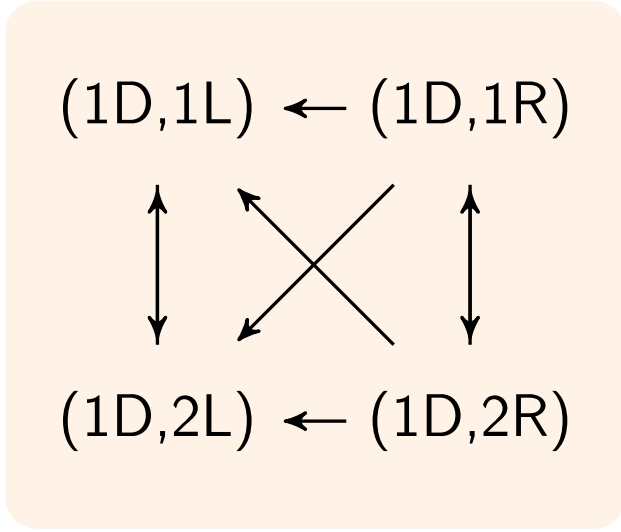
For the sake of simplicity suppose now that we want to model the *ex interim* situation of a game, i.e. a situation where the players not only know their own types, but have also decided and know their own actions/strategies. This can be easily modelled in our framework by adding the following *ex interim* condition:

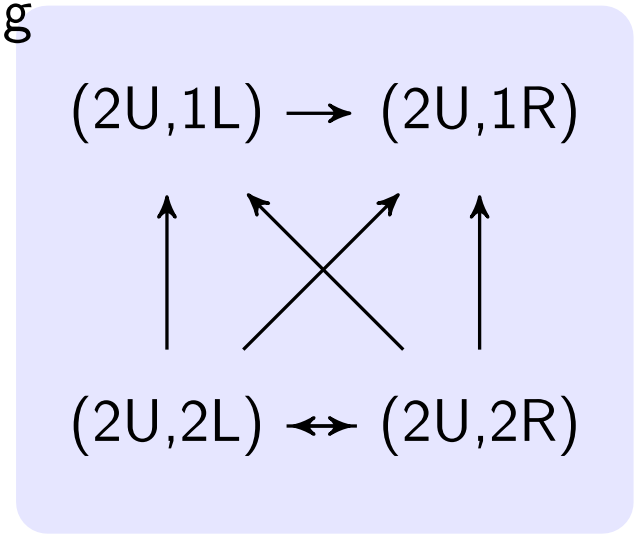
$$w \sim_i w' \text{ iff } t_i(w) = t_i(w') \text{ and } s_i(w) = s_i(w')$$

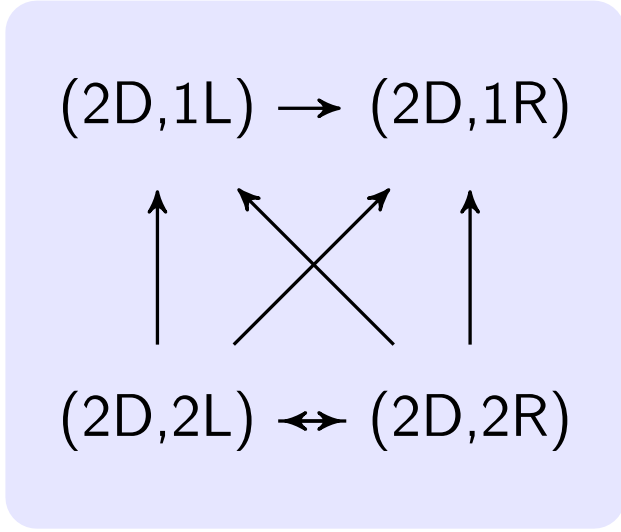
Orderings preserving translation: Ann

orderings preserving Ann



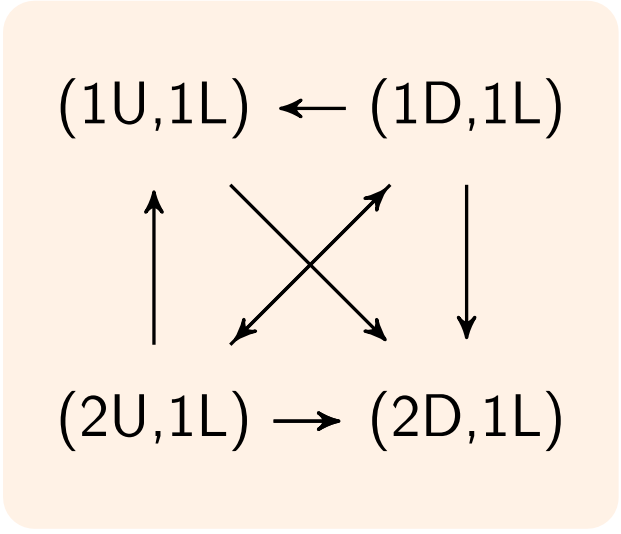


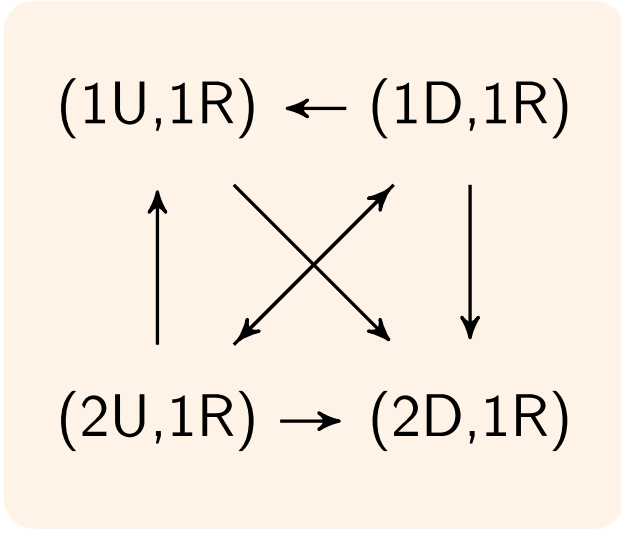


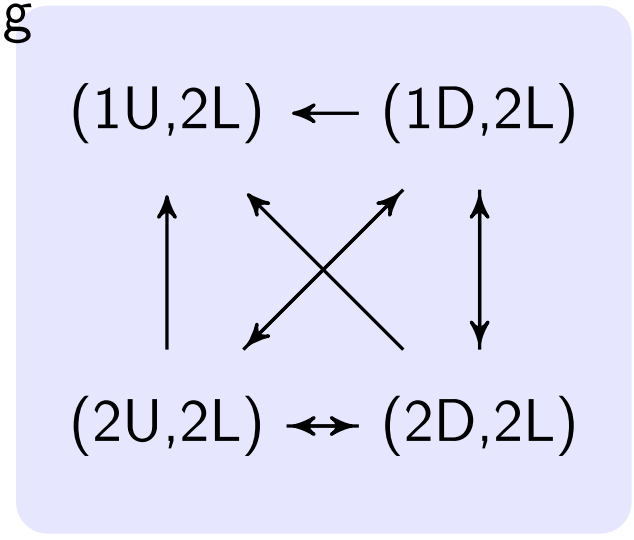


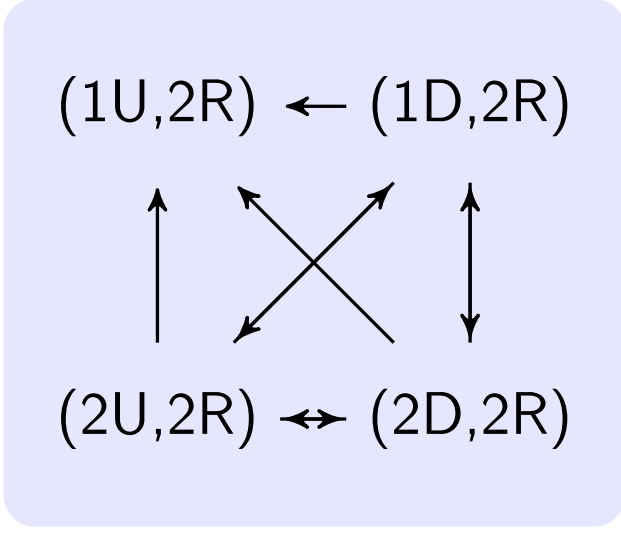
Orderings preserving translation: Bob

orderings preserving Bob



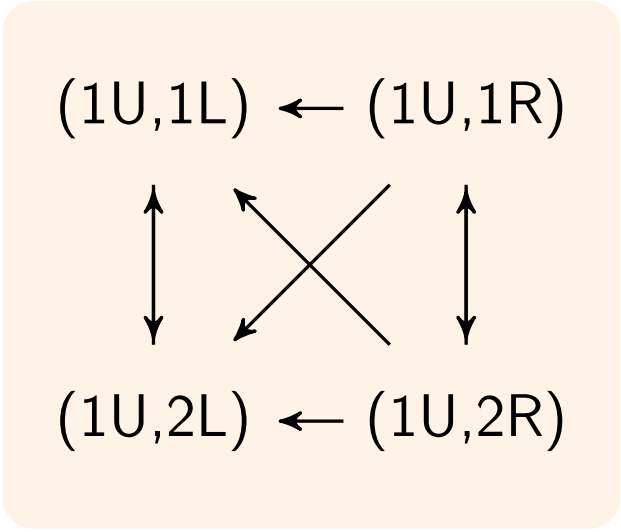


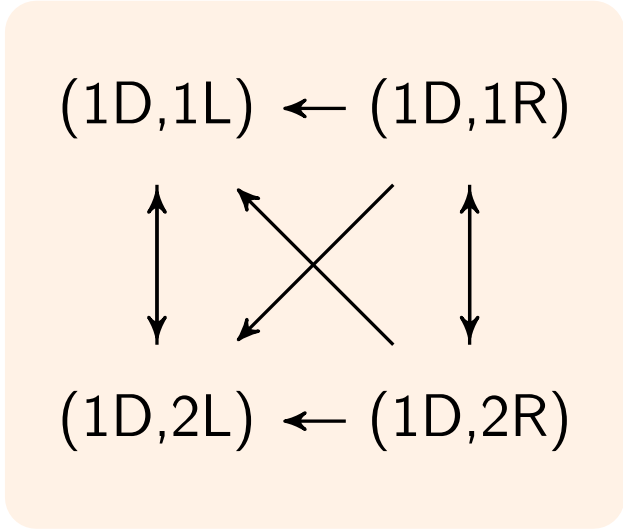


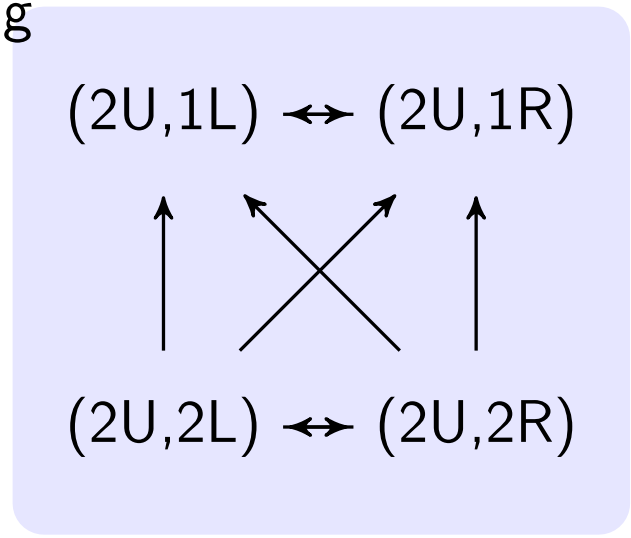


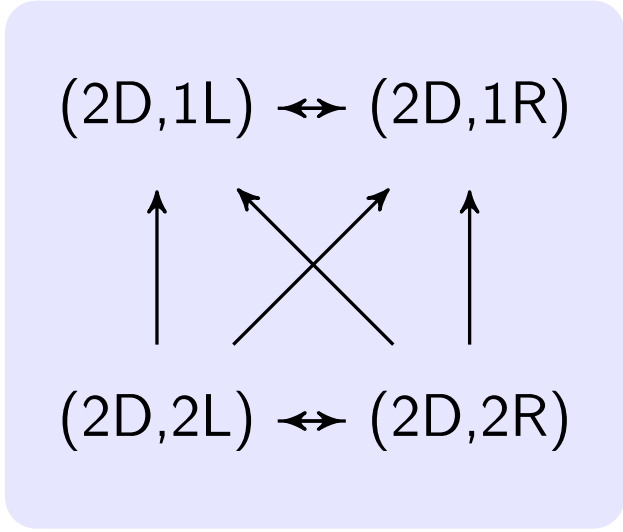
Operators preserving translation: Ann

operators preserving Ann



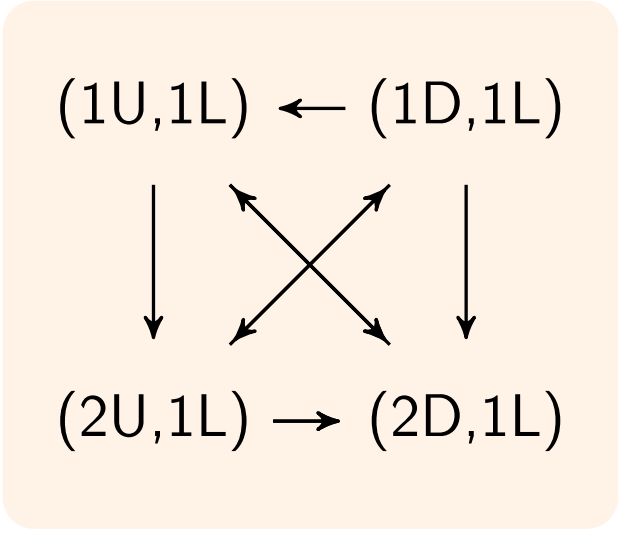


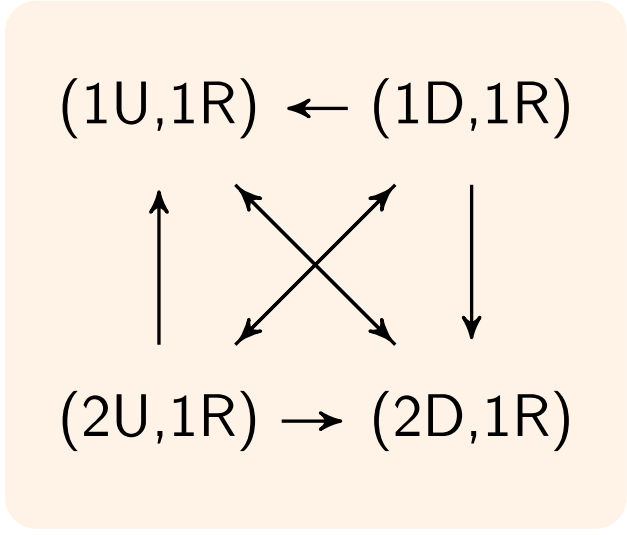


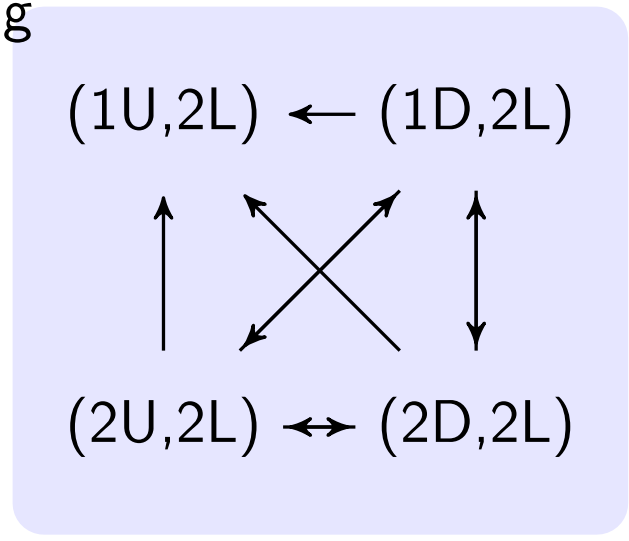


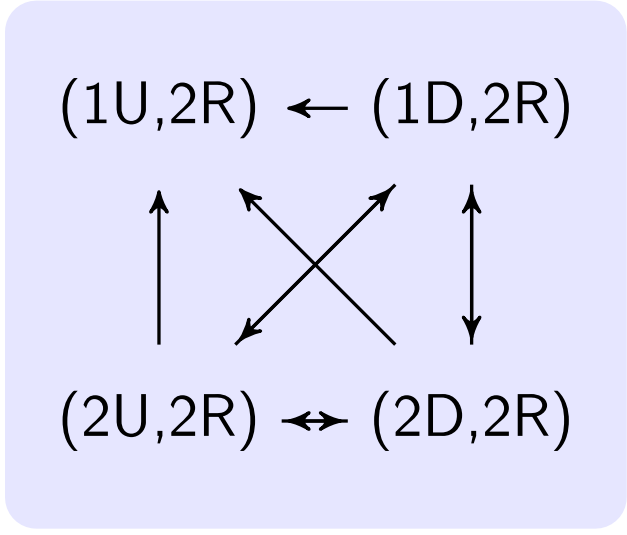
Operators preserving translation: Bob

operators preserving Bob









Operators preserving translation

- $(B_{Ann}L)^{TM} = \{(1U,1L),(1U,2L),(1U,1R),(1U,2R),$
  $(1D,1L),(1D,2L),(1D,1R),(1D,2R)\}$
- $(B_{Ann}(1L \vee 1R))^{TM} = \{(2U,1L),(2U,2L),(2U,1R),(2U,2R),$
  $(2D,1L),(2D,2L),(2D,1R),(2D,2R)\}$
- $(B_{Bob}(1U \vee 1D))^{TM} = \{(1U,2L),(2U,2L),(1D,2L),(2D,2L),$
  $(1U,2R),(2U,2R),(1D,2R),(2D,2R)\}$
- $(B_{Bob}U)^{TM} = \{(1U,2L),(2U,2L),(1D,2L),(2D,2L),$
  $(1U,2R),(2U,2R),(1D,2R),(2D,2R)\}$
- $(B_{Bob}B_{Ann}L)^{TM} = \{(1U,2L),(2U,2L),(1D,2L),(2D,2L),$
  $(1U,2R),(2U,2R),(1D,2R),(2D,2R)\}$

We can formally prove that $\succeq_i^{op}$ preserves the operators.

We have defined $\Phi = \{S_1, ..., S_n\}$. Consequently we have a proposition s_i in TM corresponding to the event e_{s_i} in TS, where $e_{s_i} = \{(t, s) \in T \times S : s_i(t, s) = s_i\}$.

From this we can naturally define other propositions corresponding to events represented in a TS:

- $s_i \wedge s_j$ the proposition that agent i plays her action s_i and agent j plays his action s_j, corresponding to the event $e_{s_i} \cap e_{s_j}$;
- $\neg s_i$ the proposition that agent i does not play s_i, corresponding to the event $(T - e_{s_i})$;
- $B_i s_j$ the proposition that agent i believes that agent j plays s_j, corresponding to the event $\mathscr{B}_i(e_{s_j})$.

Theorem

Generally given an event e_φ in TS and φ the proposition expressing that event in TM and $\varphi^{TM} = \{w \in W : (TM, w) \vDash \varphi\}$.
We can now state the following theorem.

Theorem

Let $(t, s) \in T \times S$ be a state in TS, $w \in W$ the corresponding world in TM and $e_\varphi \subseteq T \times S$ an event in TS, then $(t, s) \in e_\varphi$ in TS iff $(TM, w) \vDash \varphi$, or equivalently $(t, s) \in e_\varphi$ iff $w \in \varphi^{TM}$.

Proof

Proof. By induction on the structure of φ. We prove only some cases.

Induction basis: $(\varphi \equiv s_i)$. *Only if:* suppose $(t, s) \in e_{s_i}$, then $(t, s) \equiv (t, s_i, s_{-i})$ and $(TM, w) \vDash s_i$ by definition of $\upsilon(s_i)$.
If: suppose $(TM, w) \vDash s_i$. By definition of $\upsilon(s_i)$ we have $w \equiv (t, s) \equiv (t, s_i, s_{-i})$. So $(t, s) \in e_{s_i}$.

Inductive steps:
$(\varphi \equiv \neg\psi)$. *Only if:* suppose $(t, s) \in (T \times S - e_\psi)$. By inductive hypothesis $w \in (W - \psi^{TM})$. Consequently $(TM, w) \nvDash \psi_e$ and $(TM, w) \vDash \neg\psi$.
If: suppose $(TM, w) \vDash \neg\psi$. So $w \in (W - \psi^{TM})$. By inductive hypothesis $(t, s) \in (T \times S - e_\psi)$.

Proof

($\varphi \equiv K_i\psi$). *Only if:* suppose $(t, s) \equiv (t_i, t_{-i}, s) \in \mathscr{K}_i(e_\psi)$. Consequently, $(t_i, t'_{-i}, s') \in e_\psi$ for all $(t'_{-i}, s') \in T_{-i} \times S$ and by inductive hypothesis $(TM, w') \vDash \psi$ for all $w' \sim_i w$. Then $(TM, w) \vDash K_i\psi$.
If: suppose $(TM, w) \vDash K_i\psi$. It follows that $(TM, w') \vDash \psi_e$ for all $w' \sim_i w \equiv (t, s) \equiv (t_i, t'_{-i}, s')$. By inductive hypothesis $(t_i, t'_{-i}, s') \in e_\psi$ for all $(t'_{-i}, s') \in T_{-i} \times S$. Thus $(t, s) \in \mathscr{K}_i(e_\psi)$.

Proof

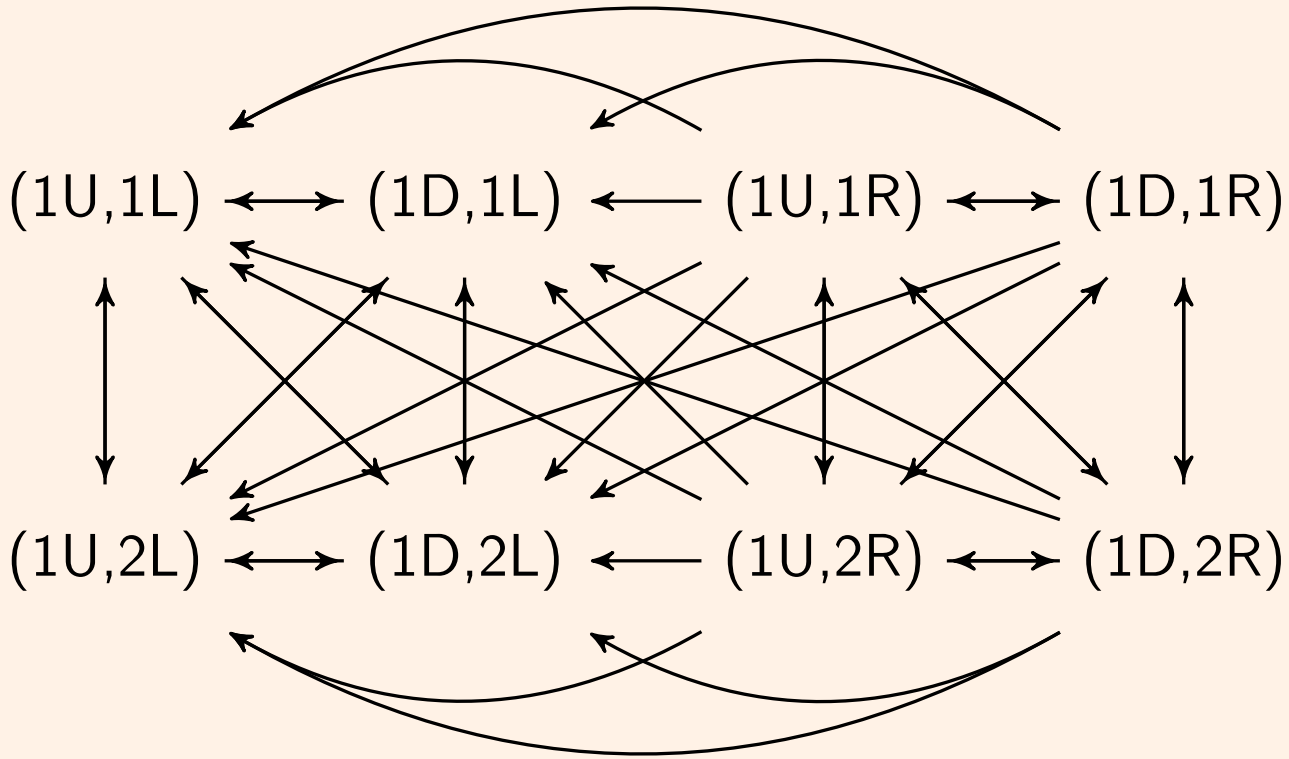
($\varphi \equiv B_i\psi$). *Only if:* suppose $(t, s) \equiv (t_i, t_{-i}, s) \in \mathscr{B}_i(e_\psi)$. By definition $(t'_{-i}, s'_{-i}) \in e_\psi$ for all $(t'_{-i}, s'_{-i}) \in (T_{-i} \times S_i$ s.t. $\beta_i(t_i)(t'_{-i}, s'_{-i}) > 0$ and $w' \in Sup_{\geq_i}^{op}([w]_{\sim_i})$ for all $w' \equiv (t'_{-i}, s'_{-i}) \in (T_{-i} \times S_i$ s.t. $\beta_i(t_i)(t'_{-i}, s'_{-i}) > 0$. By inductive hypothesis $(TM, w') \vDash \psi$ for all $w \in Sup_{\geq_i}^{op}([w]_{\sim_i})$. It follows that $(TM, w) \vDash B_i\psi$.
If: suppose $(TM, w) \vDash B_i\psi$. Then $(TM, w') \vDash \psi$ for all $w' \in Sup_{\geq_i}^{op}([w]_{\sim_i})$ and by inductive hypothesis $(t', s') \equiv (t_i, t'_{-i}, s') \in e_\psi$ for all $(t', s') \equiv w'$, for all $w' \in Sup_{\geq_i}^{op}([w]_{\sim_i})$. By definition $w' \in Sup_{\geq_i}^{op}([w]_{\sim_i})$ iff $(t', s') \equiv (t_i, t'_{-i}, s') \equiv w'$ s.t. $\beta_i(t_i)(t'_{-i}, s'_{-i}) > 0$. Thus $(t, s) \in \mathscr{B}_i(e_\psi)$. ■

Expressivity: the ex-interim condition

Notice that we could also represent (by dropping the ex interim condition) a situation in which players have not decided yet their actions and they do not know what their own action will be. In this way we can clearly express the two different stages of the game.

Ann

operators preserving type 1 before deciding an action



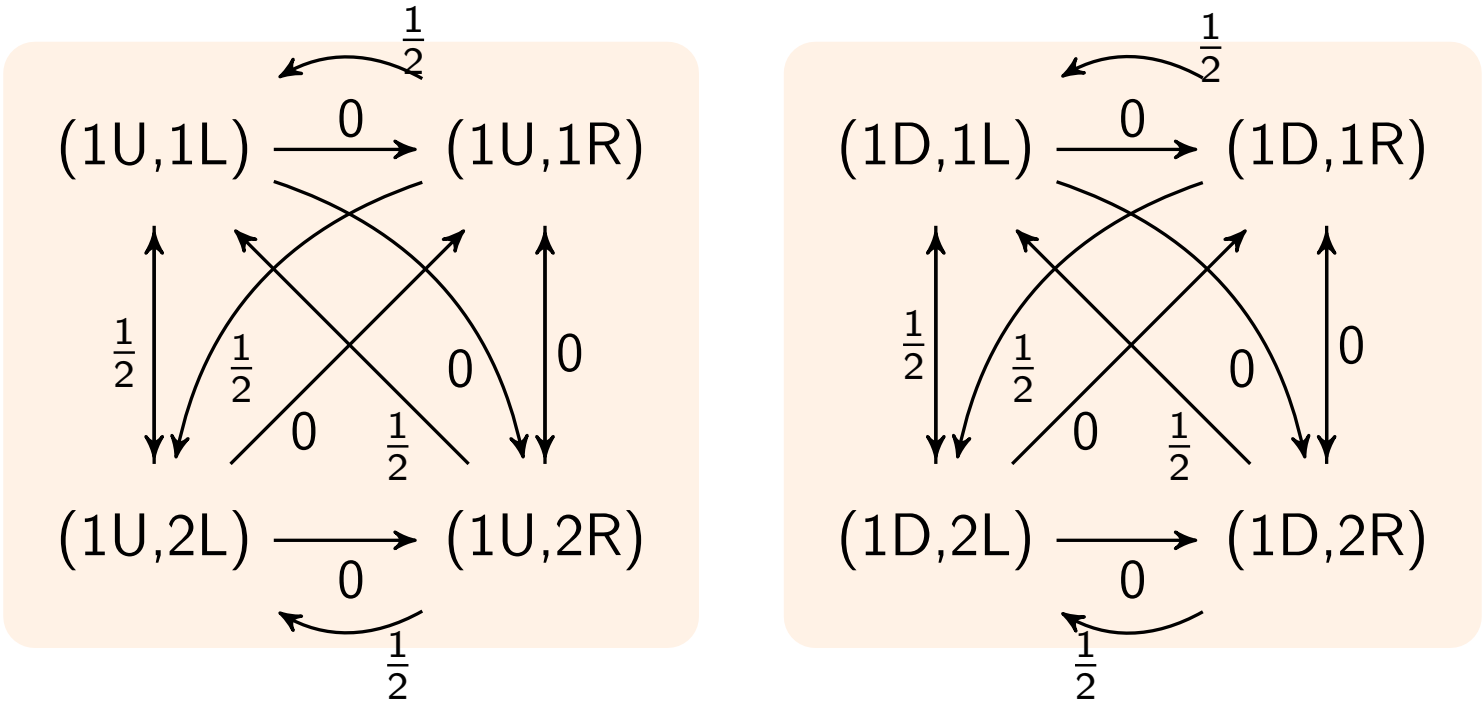
Let us write $\succeq_i^{or}$ for the orderings preserving relation and $\succeq_i^{op}$ for the operators preserving relation.

Since each type is uniquely determined by his beliefs about the others (i.e. if two types hold the same beliefs of any level about the others then they are simply the same type), $\succeq_i^{or}$ and $\succeq_i^{op}$ determine two different partitions over T_i s.t. $\frac{T_i}{\succeq_i^{or}}$ is finer than $\frac{T_i}{\succeq_i^{op}}$, i.e. $[t_i]_{\succeq_i^{or}} \subseteq [t_i]_{\succeq_i^{op}}$, for all $t_i \in T_i$.

Hi-fi translation

Obviously, if we aim to adhere closely to the original type space we have to use a probabilistic state model.

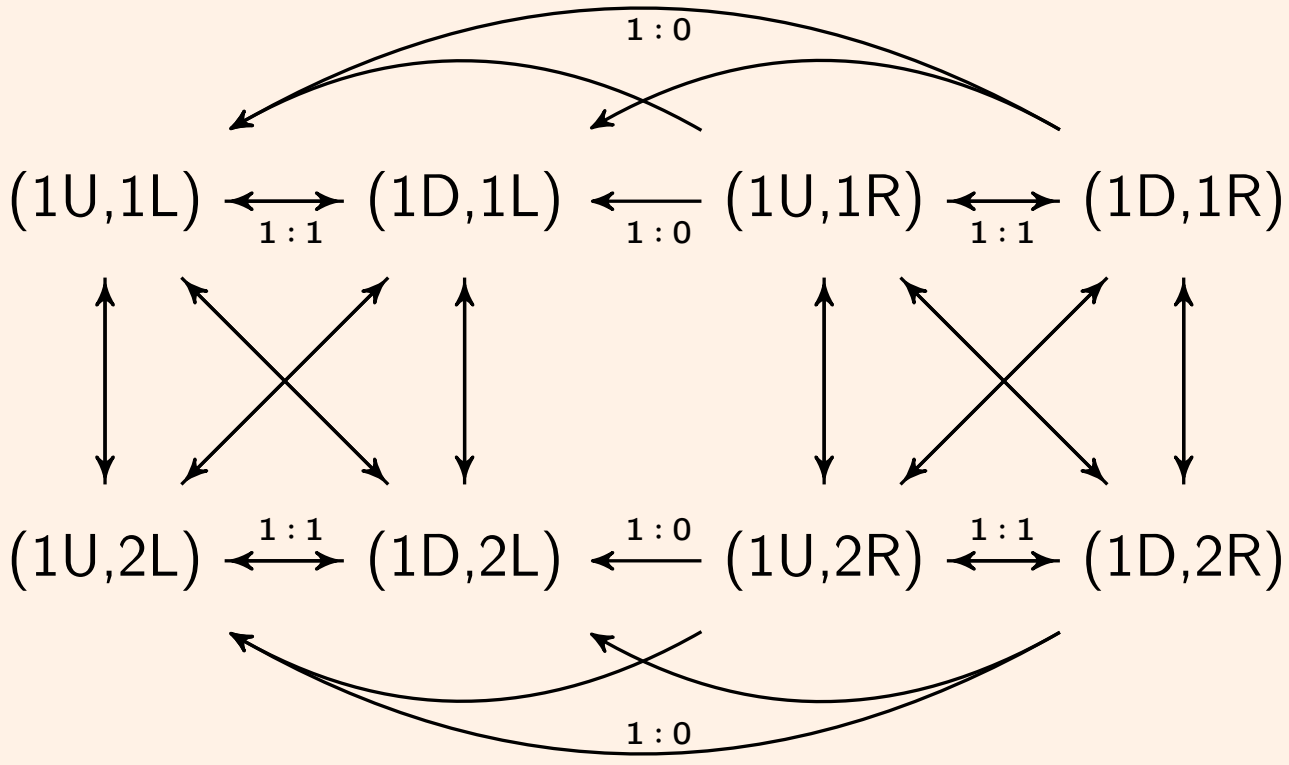
Ann
type 1
ex interim



Ann

type 1

before deciding an action



Soundness and completeness

In logic, soundness and completeness are two notions that connect syntax and semantics for a given logical system.

$$\vdash \varphi \longleftrightarrow \models \varphi$$

Given a logical system, i.e. a set of axioms together with inference rules, and given a semantics for it, we say that the logical system is sound if and only if its axioms and inference rules prove only formulas of the language that are valid with respect to the semantics: $\vdash \varphi \longrightarrow \models \varphi$

Vice versa, a logical system is complete if and only if all the formulas that are valid with respect to the semantics are provable from its axioms and inference rules: $\vdash \varphi \longleftarrow \models \varphi$

It is interesting to note that for Kripke models we have presented thus far, i.e. both for plausibility models and probability models, there is a logical system that is sound and complete.

Consequently together with the translation we get a sound and complete logic for epistemic game theory.

Logical system for plausibility models: axioms

Axioms

- Axioms for propositional logic
- Axioms for K_i:
  - K: $(K_i\varphi \wedge K_i(\varphi \longrightarrow \psi)) \longrightarrow K_i\psi$
  - T: $K_i\varphi \longrightarrow \varphi$
  - 4: $K_i\varphi \longrightarrow K_iK_i\varphi$
  - 5: $\neg K_i\varphi \longrightarrow K_i\neg K_i\varphi$
- Axioms for B_i:
  - K: $(B_i\varphi \wedge B_i(\varphi \longrightarrow \psi)) \longrightarrow B_i\psi$
  - D: $\neg B_i\bot$
  - 4: $B_i\varphi \longrightarrow B_iB_i\varphi$
  - 5: $\neg B_i\varphi \longrightarrow B_i\neg B_i\varphi$

Logical system for plausibility models: mixed axioms and inference rules

Axioms

- Mixed:

SPI: $B_i\varphi \longrightarrow K_i B_i \varphi$

SNI: $\neg B_i\varphi \longrightarrow K_i \neg B_i \varphi$

KB: $K_i\varphi \longrightarrow B_i \varphi$

Inference rules

MP: from $\vdash \varphi$ and $\vdash \varphi \longrightarrow \psi$ infer $\vdash \psi$

NR: from $\vdash \varphi$ infer $\vdash K_i\varphi$

Logical system for probability models

References

- R. Fagin and J. Halpern (1994)
- B. Kooi (2003)

Comparison with other work

- Zvesper (2010): using plausibility models instead of simple relational models provides us with a richer and more expressive framework for our translation. Indeed, in plausibility models we can express a type as a partition cell, given by the $\sim_i$ relation, without having any specific proposition for types in the language, where we only have propositions for actions/strategies. This seems to be conceptually closer to the spirit of type spaces, where each player is considered to *know* her own type.

Comparison with other work

- Brandenburger (2008): using different partitions we can easily express the fact that the players know or do not know their own strategies/actions. In state models we could have partitions where the player is taken to know/have decided her strategy, or where she believes she will play a certain strategy, or she has no idea about which strategy to play. Since in type spaces at every state a strategy is specified for each player, it is not clear how to distinguish situations in which players have decided how to play from situations in which they have not yet.

Some remarks and future work

Extension to mutual and common belief.

Qualitative vs quantitative: there are epistemic characterizations for solution concepts that we can express in the qualitative framework of plausibility models. (cf. Baltag, Smets, Zvesper 2009)
To what extent?

There are other epistemic notions expressible in plausibility models, some of them seem to have a straightforward counterpart in type spaces, for others it is difficult to identify the corresponding notion in type spaces. And vice versa.
Goal: formally studying the relation between these notions.

Studying this correspondence to better the understanding of epistemic game theory and formal epistemology in general: why splitting the epistemic community in two separeted parts?

Thank you

Thank you!

Glossary

- $t_i(w) = t_i$ s.t. $(t_i, t_{-i}, s) \equiv w$
- $t_{-i}(w) = t_{-i}$ s.t. $(t_{-i}, t_i, s) \equiv w$
- $s_i(w) = s_i$ s.t. $(t, s_i, s_{-i}) \equiv w$
- $s_{-i}(w) = s_{-i}$ s.t. $(t, s_{-i}, s_i) \equiv w$
- $[t_i]_{\geq_i^{or}} = \{t_i' \in T_i : \forall w, w'(t_i, t_{-i}(w), s(w)) \geq_i^{or}$
  $(t_i, t_{-i}(w'), s(w'))$ iff
  $(t_i', t_{-i}(w), s(w)) \geq_i^{or} (t_i', t_{-i}(w'), s(w'))\}$