Multi-Agent Simulative Belief Ascription

Min Cheol Seo*

*Ph.D. Candidate
Department of Philosophy
Sungkyunkwan University
Seoul, South Korea

Joint Conference of APPSA and LMPST Taiwan 2025

Table of Contents

- In our everyday life, we often predict, explain, and coordinate another's behaviour by attributing beliefs, desires, intentions, etc.

- In our everyday life, we often predict, explain, and coordinate another's behaviour by attributing beliefs, desires, intentions, etc.
- The question is how our mind manages this often called *mind-reading*. And broadly, two answers are predominately considered: (i) **Theory-Theory**, and (ii) **Simulation Theory**.

- In our everyday life, we often predict, explain, and coordinate another's behaviour by attributing beliefs, desires, intentions, etc.
- The question is how our mind manages this often called *mind-reading*. And broadly, two answers are predominately considered: (i) **Theory-Theory**, and (ii) **Simulation Theory**.
- Consider the following scenario:

- In our everyday life, we often predict, explain, and coordinate another's behaviour by attributing beliefs, desires, intentions, etc.
- The question is how our mind manages this often called *mind-reading*. And broadly, two answers are predominately considered: (i) **Theory-Theory**, and (ii) **Simulation Theory**.
- Consider the following scenario:

  A : "I do not like those who make the room messy".

  B : 'A does not like people who make the room messy, and I am one of them'.

  'So A does not like me'.

  B : Says to C, "A does not like me".

- In our everyday life, we often predict, explain, and coordinate another's behaviour by attributing beliefs, desires, intentions, etc.
- The question is how our mind manages this often called *mind-reading*. And broadly, two answers are predominately considered: (i) **Theory-Theory**, and (ii) **Simulation Theory**.
- Consider the following scenario:

  A : "I do not like those who make the room messy".
  B : 'A does not like people who make the room messy, and I am one of them'.
  'So A does not like me'.
  B : Says to C, "A does not like me".

Claim

Mental simulation is central to mind-reading.

With the presented scenario, we can run with an *informal* definition:

With the presented scenario, we can run with an *informal* definition:

> "*What A would believe if A were me*".

With the presented scenario, we can run with an *informal* definition:

> "*What A would believe if A were me*".

Definition (Simulative Belief)

B simulatively believe that *A* believes *P* iff

1. *B* sets aside his own beliefs and adopt *A*'s perceived beliefs,
2. *B* let his reasoning machinery run on those stand-in states under given circumstances,
3. In that pretend perspective, *P* turns out true; therefore, *B* reports *A believes that P*.

Why Studying Simulative Beliefs?

- **Philosophy**.
  - **Empathy & moral appraisal**. Feeling from the inside, not calculating from rules/theories.
  - **Confabulation**. Self-projection errors when we cannot quarantine our own beliefs.

- **Philosophy**.
  - **Empathy & moral appraisal**. Feeling from the inside, not calculating from rules/theories.
  - **Confabulation**. Self-projection errors when we cannot quarantine our own beliefs.
- **AIs**.

- **Philosophy**.
  - **Empathy & moral appraisal**. Feeling from the inside, not calculating from rules/theories.
  - **Confabulation**. Self-projection errors when we cannot quarantine our own beliefs.
- **AIs**.
  - **ToMB**. GPT-4* class models now clear classic false-belief tests in Theory-of-Mind Benchmark. Recent studies suggest including Simulation Theory to expand and improve its accuracy.[1] [13, 2023] [12, 2024]

[1] The benchmark includes: false-belief, unexpected-contents, but most importantly, *multi-agent reasoning*.

- **Philosophy**.
  - **Empathy & moral appraisal**. Feeling from the inside, not calculating from rules/theories.
  - **Confabulation**. Self-projection errors when we cannot quarantine our own beliefs.
- **AIs**.
  - **ToMB**. GPT-4* class models now clear classic false-belief tests in Theory-of-Mind Benchmark. Recent studies suggest including Simulation Theory to expand and improve its accuracy.[1] [13, 2023] [12, 2024]
  - **SARs**. Embedding a lightweight simulation module enables SARs to predict whether a vocal cue is a request vs. comment, boosting turn-taking fluency. [6, 2024]

[1]The benchmark includes: false-belief, unexpected-contents, but most importantly, *multi-agent reasoning*.

Shortcomings of current LLMs ToM Benchmarks:

Shortcomings of current LLMs ToM Benchmarks:

- **LLM ToM** $\neq$ **Simulation**. It is only reliable, when prompts explicitly create a surrogate belief context (In SIMToM, ToMB)

Shortcomings of current LLMs ToM Benchmarks:

- **LLM ToM $\neq$ Simulation**. It is only reliable, when prompts explicitly create a surrogate belief context (In SIMToM, ToMB)
- **Formal Gap**. We still lack a stable mapping from prompt tokens to a well-behaved relation, R^{sim}; without it completeness, decidability, and safety proofs fail.

Shortcomings of current LLMs ToM Benchmarks:

- **LLM ToM $\neq$ Simulation**. It is only reliable, when prompts explicitly create a surrogate belief context (In SIMToM, ToMB)
- **Formal Gap**. We still lack a stable mapping from prompt tokens to a well-behaved relation, R^{sim}; without it completeness, decidability, and safety proofs fail.
- **Depth, Tags, Fusion and Verification**. Each adds a modal/complexity layer, thereby generating theoretic friction that current AI tool-chains don't address.

As you have already suspected, there are difficulties surrounding formalisation of simulative beliefs:

1. **Dual Perspectives**. real v. surrogate.

As you have already suspected, there are difficulties surrounding formalisation of simulative beliefs:

1. **Dual Perspectives**. real v. surrogate.
2. **Copy & Revise**. AGM?

As you have already suspected, there are difficulties surrounding formalisation of simulative beliefs:

1. **Dual Perspectives**. real v. surrogate.
2. **Copy & Revise**. AGM?
3. **Introspection Gap**. Which axioms?

As you have already suspected, there are difficulties surrounding formalisation of simulative beliefs:

1. **Dual Perspectives**. real v. surrogate.
2. **Copy & Revise**. AGM?
3. **Introspection Gap**. Which axioms?
4. **Layer Explosion**. Nested beliefs

As you have already suspected, there are difficulties surrounding formalisation of simulative beliefs:

1. **Dual Perspectives**. real v. surrogate.
2. **Copy & Revise**. AGM?
3. **Introspection Gap**. Which axioms?
4. **Layer Explosion**. Nested beliefs
5. **Verification**. Safety proofs turn undecidable.

As you have already suspected, there are difficulties surrounding formalisation of simulative beliefs:

1. **Dual Perspectives**. real v. surrogate.
2. **Copy & Revise**. AGM?
3. **Introspection Gap**. Which axioms?
4. **Layer Explosion**. Nested beliefs
5. **Verification**. Safety proofs turn undecidable.

just to name a few.

A Brief Sketch of Simulation Theory

- In describing Mental Simulation in ST, we have two rival pictures describe how simulation contributes when it is used:

A Brief Sketch of Simulation Theory

- In describing Mental Simulation in ST, we have two rival pictures describe how simulation contributes when it is used:
  1. **Constitution View**. The simulation itself *is* the representation of the other's state; nothing further is required.

A Brief Sketch of Simulation Theory

- In describing Mental Simulation in ST, we have two rival pictures describe how simulation contributes when it is used:
  1. **Constitution View**. The simulation itself *is* the representation of the other's state; nothing further is required.
  2. **Causation View**. Simulation merely provides causal inputs to a *separate* judgment that attributes the state.

A Brief Sketch of Simulation Theory

- In describing Mental Simulation in ST, we have two rival pictures describe how simulation contributes when it is used:
  1. **Constitution View**. The simulation itself *is* the representation of the other's state; nothing further is required.
  2. **Causation View**. Simulation merely provides causal inputs to a *separate* judgment that attributes the state.
- Before we move on, let us briefly consider above two views.

Dimension	**Goldman [9] (Three-Stage, Causation)**	**Gordon [10, 11] (Radical, Constitution)**
Process flow	Pretence $\rightarrow$ run $\rightarrow$ *introspect* + *judge*	Single perspective-shift; ask in that viewpoint whether *P* holds
Role of introspection	Needs an "inner sense"'to read off simulated output	No introspection (Evans-style ascent routine)
Conceptual load	Ends with judgment "A believes *P*" (concept-involving)	Core attributions can be non-conceptual
Status of simulation	Simulation *causes* attribution	Simulation *constitutes* attribution
Scope / centrality	A strong tool inside hybrid theories	Default mechanism in everyday mind-reading

Causation v. *Constitution*, Conceptual v. Non-conceptual, Layered v. Lean–two paths to understanding other minds.

From this, we give a skeleton formalisation of simulative belief and its framework:

$$\varphi ::= (p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i\varphi \mid B^{sh}_{i\to j}\varphi \mid B^{sim}_{i,j}\varphi\ ),$$

$$B_i\varphi = i \text{ believes (proper) } \varphi,$$
$$B^{sh}_{i\to j}\varphi = i\text{'s surrogate for } j,\ \varphi \text{ holds},$$
$$B^{sim}_{i,j}\varphi = \text{After introspection, } i \text{ judges that } j \text{ believes } \varphi.$$

With the previous resources, I now can give staged Kripke semantics:

Symbol	**Construction**	**Gloss**
Stage 1 - Pretence	$Base_j(w) := \{\psi \mid \mathcal{M}, w \vDash B_j\psi\}$	
Stage 2 - Enactment/Update	$B^{sh}_{i\to j}(w) := Cn(Base_j(w)) *_i (Info_{shared}(w))$	AGM-style revision operates on surrogate with *common info* known by i,
Stage 2-Relation	$wR^{sh}_{i\to j}v$ iff $v \vDash B^{sh}_{i\to j}(w)$	worlds compatible with surrogate,
Stage 3 - Introspection	$\mathcal{M}, w \vDash B^{sim}_{i,j}\varphi$ iff $\mathcal{M}, w \vDash B^{sh}_{i\to j}\varphi$	i reads off of the surrogate output.

Given the stage semantics I have provided, we assume the following issues that the standard Kripke-Hintikka style semantics can face:

Given the stage semantics I have provided, we assume the following issues that the standard Kripke-Hintikka style semantics can face:

1. **Layer Complexity**. Too much layers are involved, making the framework inherently complex.

Given the stage semantics I have provided, we assume the following issues that the standard Kripke-Hintikka style semantics can face:

1. **Layer Complexity**. Too much layers are involved, making the framework inherently complex.
2. **Pretence**. How exactly we do it, and what if we are wrong about the pretence?

Given the stage semantics I have provided, we assume the following issues that the standard Kripke-Hintikka style semantics can face:

1. **Layer Complexity**. Too much layers are involved, making the framework inherently complex.
2. **Pretence**. How exactly we do it, and what if we are wrong about the pretence?
3. **Shared Information**. How do we decide what are shared information, and where we ground such information?

Given the stage semantics I have provided, we assume the following issues that the standard Kripke-Hintikka style semantics can face:

1. **Layer Complexity**. Too much layers are involved, making the framework inherently complex.
2. **Pretence**. How exactly we do it, and what if we are wrong about the pretence?
3. **Shared Information**. How do we decide what are shared information, and where we ground such information?
4. **Introspection**. Should full introspection be granted for simulative beliefs?

Given the stage semantics I have provided, we assume the following issues that the standard Kripke-Hintikka style semantics can face:

1. **Layer Complexity**. Too much layers are involved, making the framework inherently complex.
2. **Pretence**. How exactly we do it, and what if we are wrong about the pretence?
3. **Shared Information**. How do we decide what are shared information, and where we ground such information?
4. **Introspection**. Should full introspection be granted for simulative beliefs?
5. **Update**. How should we formalise *updating*, in light of new information?

In the standard **Kripke-Hintikka** style (multi-agent) epistemic/doxastic logics,

In the standard **Kripke-Hintikka** style (multi-agent) epistemic/doxastic logics, an agent's beliefs are represented by an accessibility relation R on a set of possible worlds, $W = \{w_1, w_2, \ldots, w_n\}$.

In the standard **Kripke-Hintikka** style (multi-agent) epistemic/doxastic logics, an agent's beliefs are represented by an accessibility relation R on a set of possible worlds, $W = \{w_1, w_2, \ldots, w_n\}$.

"Agent i believes p" is true at world w if p holds in all R_i-accessible worlds from w.

In the standard **Kripke-Hintikka** style (multi-agent) epistemic/doxastic logics, an agent's beliefs are represented by an accessibility relation R on a set of possible worlds, $W = \{w_1, w_2, \ldots, w_n\}$.

"Agent i believes p" is true at world w if p holds in all R_i-accessible worlds from w.

Problems:

In the standard **Kripke-Hintikka** style (multi-agent) epistemic/doxastic logics, an agent's beliefs are represented by an accessibility relation R on a set of possible worlds, $W = \{w_1, w_2, \ldots, w_n\}$.

"Agent i believes p" is true at world w if p holds in all R_i-accessible worlds from w.

Problems:

1. **Simulative Operation**: No formal distinction between an agent's *actual* beliefs and *simulative* beliefs the ascriber imposes.

In the standard **Kripke-Hintikka** style (multi-agent) epistemic/doxastic logics, an agent's beliefs are represented by an accessibility relation R on a set of possible worlds, $W = \{w_1, w_2, \ldots, w_n\}$.

"Agent i believes p" is true at world w if p holds in all R_i-accessible worlds from w.

Problems:

1. **Simulative Operation**: No formal distinction between an agent's *actual* beliefs and *simulative* beliefs the ascriber imposes.
2. **Fixed Access Relation**: The agent's doxastic possibilities are typically held fixed in a single model.

In the standard **Kripke-Hintikka** style (multi-agent) epistemic/doxastic logics, an agent's beliefs are represented by an accessibility relation R on a set of possible worlds, $W = \{w_1, w_2, \ldots, w_n\}$.

"Agent i believes p" is true at world w if p holds in all R_i-accessible worlds from w.

Problems:

1. **Simulative Operation**: No formal distinction between an agent's *actual* beliefs and *simulative* beliefs the ascriber imposes.
2. **Fixed Access Relation**: The agent's doxastic possibilities are typically held fixed in a single model.
3. **Introspection and Revision**: Revising an agent's beliefs requires building a new (or globally modified) accessibility relation, or a new model altogether.

Multi-Agent AGM offers a robust framework that captures the *dynamic* aspects of belief interaction. [3, 2010][2]

[2] For a general introduction to AGM, see [2].

Multi-Agent AGM offers a robust framework that captures the *dynamic* aspects of belief interaction. [3, 2010][2]

Problems:

[2]For a general introduction to AGM, see [2].

Multi-Agent AGM offers a robust framework that captures the *dynamic* aspects of belief interaction. [3, 2010][2]

Problems:

1. **Simulative Operation**: Again, AGM is geared towards *genuine* beliefs, not *simulative* ones.

[2] For a general introduction to AGM, see [2].

Multi-Agent AGM offers a robust framework that captures the *dynamic* aspects of belief interaction. [3, 2010][2]

Problems:

1. **Simulative Operation**: Again, AGM is geared towards *genuine* beliefs, not *simulative* ones.
2. **Iterated Belief**: AGM primarily handles one-shot revision. It does not prescribe how beliefs evolve across multiple or nested updates.

[2] For a general introduction to AGM, see [2].

Gerbrandy and Groeneveld [8, 1997], (also, Gerbrandy [7, 1999]) offered an n-agent framework which addresses *iteration* via a modular approach. In their setting, a world w is a triple $\langle u, b_i, b_j \rangle$.

Gerbrandy and Groeneveld [8, 1997], (also, Gerbrandy [7, 1999]) offered an *n*-agent framework which addresses *iteration* via a modular approach. In their setting, a world w is a triple $\langle u, b_i, b_j \rangle$.

Here, $u \in U$ determines the belief-independent features of the world, and b_i is a set of *worlds* validating agent i's belief state.

Gerbrandy and Groeneveld [8, 1997], (also, Gerbrandy [7, 1999]) offered an *n*-agent framework which addresses *iteration* via a modular approach. In their setting, a world w is a triple $\langle u, b_i, b_j \rangle$.

Here, $u \in U$ determines the belief-independent features of the world, and b_i is a set of *worlds* validating agent i's belief state.

Problem(s):

Gerbrandy and Groeneveld [8, 1997], (also, Gerbrandy [7, 1999]) offered an n-agent framework which addresses *iteration* via a modular approach. In their setting, a world w is a triple $\langle u, b_i, b_j \rangle$.

Here, $u \in U$ determines the belief-independent features of the world, and b_i is a set of *worlds* validating agent i's belief state.

Problem(s):

1. b_i is a set of *worlds*, which may even contain w itself.

Gerbrandy and Groeneveld [8, 1997], (also, Gerbrandy [7, 1999]) offered an n-agent framework which addresses *iteration* via a modular approach. In their setting, a world w is a triple $\langle u, b_i, b_j \rangle$.

Here, $u \in U$ determines the belief-independent features of the world, and b_i is a set of *worlds* validating agent i's belief state.

Problem(s):

1. b_i is a set of *worlds*, which may even contain w itself.

Solutions:

Gerbrandy and Groeneveld [8, 1997], (also, Gerbrandy [7, 1999]) offered an n-agent framework which addresses *iteration* via a modular approach. In their setting, a world w is a triple $\langle u, b_i, b_j \rangle$.

Here, $u \in U$ determines the belief-independent features of the world, and b_i is a set of *worlds* validating agent i's belief state.

Problem(s):

1. b_i is a set of *worlds*, which may even contain w itself.

Solutions:

1. Aczel's *Anti-Foundation Axiom* [1, 1988](non-wellfounded set theory).

Gerbrandy and Groeneveld [8, 1997], (also, Gerbrandy [7, 1999]) offered an *n*-agent framework which addresses *iteration* via a modular approach. In their setting, a world w is a triple $\langle u, b_i, b_j \rangle$.

Here, $u \in U$ determines the belief-independent features of the world, and b_i is a set of *worlds* validating agent i's belief state.

Problem(s):

1. b_i is a set of *worlds*, which may even contain w itself.

Solutions:

1. Aczel's *Anti-Foundation Axiom* [1, 1988](non-wellfounded set theory).
2. *Bisimilarity* to the Kripke-Hintikka model.

Cantwell [4, 2005] (and [5, 2007]) adopted Gerbrandy and Groeneveld's idea but developed a framework that does not rely on *non-wellfounded sets*. Crucially, the framework preserves a *modular representation* of possible worlds as $(n+1)$-tuples, $\langle u, b_1, b_2, \ldots, b_n \rangle$, where u determines belief-independent facts, and $b_1, \ldots, b_n$ represent each agent's belief state.

Cantwell [4, 2005] (and [5, 2007]) adopted Gerbrandy and Groeneveld's idea but developed a framework that does not rely on *non-wellfounded sets*. Crucially, the framework preserves a *modular representation* of possible worlds as $(n+1)$-tuples, $\langle u, b_1, b_2, \ldots, b_n \rangle$, where u determines belief-independent facts, and $b_1, \ldots, b_n$ represent each agent's belief state.

This neatly represents *local changes* in the belief state of a single agent, e.g. from $\langle u, b_1, b_2, b_3 \rangle$ to $\langle u, b_1', b_2, b_3 \rangle$, without altering u (the belief-external facts) or other agents' states.

A quick rundown of the n-agent framework $\mathcal{F}$:

A quick rundown of the n-agent framework $\mathcal{F}$:

$\mathcal{A}$ is the set of agents, labelled $1, \ldots, n \in \mathcal{A}$,

A quick rundown of the n-agent framework $\mathcal{F}$:

$\mathcal{A}$ is the set of agents, labelled $1, \ldots, n \in \mathcal{A}$,

U is the set of belief-independent states of the world,

A quick rundown of the n-agent framework $\mathcal{F}$:

$\mathcal{A}$ is the set of agents, labelled $1, \ldots, n \in \mathcal{A}$,

U is the set of belief-independent states of the world,

$\mathcal{B}_i$ is the set of possible belief states for agent i,[3]

[3] Belief states are *not* possible worlds, but are taken to be independent entities.

A quick rundown of the n-agent framework $\mathcal{F}$:

- $\mathcal{A}$ is the set of agents, labelled $1, \ldots, n \in \mathcal{A}$,
- U is the set of belief-independent states of the world,
- $\mathcal{B}_i$ is the set of possible belief states for agent i,[3]
- A *possible world* $w \in W$ is an ordered $(n+1)$-tuple $w = \langle u, b_1, \ldots, b_n \rangle$, with $u \in U$, and $b_i \in \mathcal{B}_i$ for each i,

[3] Belief states are *not* possible worlds, but are taken to be independent entities.

A quick rundown of the n-agent framework $\mathcal{F}$:

- $\mathcal{A}$ is the set of agents, labelled $1, \ldots, n \in \mathcal{A}$,
- U is the set of belief-independent states of the world,
- $\mathcal{B}_i$ is the set of possible belief states for agent i,[3]
- A *possible world* $w \in W$ is an ordered $(n+1)$-tuple $w = \langle u, b_1, \ldots, b_n \rangle$, with $u \in U$, and $b_i \in \mathcal{B}_i$ for each i,
- $\mathcal{C}$ is a function returning, for any agent i and $b \in \mathcal{B}_i$, a set of possible worlds.

[3] Belief states are *not* possible worlds, but are taken to be independent entities.

For a world $w = \langle u, b_1, \ldots, b_n \rangle$,

$$\mathrm{wst}(w) = u \quad \text{(gives the } \textit{world-state} \text{ of } w\text{)},$$
$$\mathrm{bst}_i(w) = b_i \quad \text{(gives the } \textit{belief state} \text{ of agent } i \text{ in } w\text{)}.$$

For a world $w = \langle u, b_1, \ldots, b_n \rangle$,

$$\mathrm{wst}(w) = u \quad \text{(gives the } \textit{world-state} \text{ of } w\text{)},$$
$$\mathrm{bst}_i(w) = b_i \quad \text{(gives the } \textit{belief state} \text{ of agent } i \text{ in } w\text{)}.$$

A full-introspection postulate:

$$\text{If } b \in \mathcal{B}_i \text{ and } w \in \mathcal{C}(b), \text{ then } \mathrm{bst}_i(w) = b.$$

For a world $w = \langle u, b_1, \ldots, b_n \rangle$,

$$\mathrm{wst}(w) = u \quad \text{(gives the \textit{world-state} of } w),$$
$$\mathrm{bst}_i(w) = b_i \quad \text{(gives the \textit{belief state} of agent } i \text{ in } w).$$

A full-introspection postulate:

$$\text{If } b \in \mathcal{B}_i \text{ and } w \in \mathcal{C}(b), \text{ then } \mathrm{bst}_i(w) = b.$$

An n-agent frame $\mathcal{F}$ can be defined as a tuple

$$\langle W, U, \{\mathcal{B}_i\}_{1 \leq i \leq n}, \mathcal{C} \rangle.$$

For a world $w = \langle u, b_1, \dots, b_n \rangle$,

$$\begin{aligned} \text{wst}(w) &= u \quad \text{(gives the } \textit{world-state} \text{ of } w), \\ \text{bst}_i(w) &= b_i \quad \text{(gives the } \textit{belief state} \text{ of agent } i \text{ in } w). \end{aligned}$$

A full-introspection postulate:

$$\text{If } b \in \mathcal{B}_i \text{ and } w \in \mathcal{C}(b), \text{ then } \text{bst}_i(w) = b.$$

An n-agent frame $\mathcal{F}$ can be defined as a tuple

$$\langle W, U, \{\mathcal{B}_i\}_{1 \leq i \leq n}, \mathcal{C} \rangle.$$

In his 2005 paper, Cantwell showed $\mathcal{F}$ can be represented by a standard Kripke system with n accessibility relations.

Following the AGM tradition, $\mathcal{F}$ incorporates agent-dependent belief dynamics and *common dynamics*.

Following the AGM tradition, $\mathcal{F}$ incorporates agent-dependent belief dynamics and *common dynamics*.

Expansion: $+_i(\phi, w) = w'$, adding ϕ to agent i's beliefs in w, moving to a new world w'.

Following the AGM tradition, $\mathcal{F}$ incorporates agent-dependent belief dynamics and *common dynamics*.

- **Expansion**: $+_i(\phi, w) = w'$, adding ϕ to agent i's beliefs in w, moving to a new world w'.
- **Selection**: $\gamma_b(\phi) \subseteq \phi$, choosing the most plausible ϕ-worlds consistent with b_i,

Following the AGM tradition, $\mathcal{F}$ incorporates agent-dependent belief dynamics and *common dynamics*.

- **Expansion**: $+_i(\phi, w) = w'$, adding ϕ to agent i's beliefs in w, moving to a new world w'.
- **Selection**: $\gamma_b(\phi) \subseteq \phi$, choosing the most plausible ϕ-worlds consistent with b_i,
- **Common Learning**: $\oplus_N(\phi, w)$, for a group $N \subseteq \{1, \ldots, n\}$, so they all learn ϕ, each updating their own beliefs.

Following the AGM tradition, $\mathcal{F}$ incorporates agent-dependent belief dynamics and *common dynamics*.

- **Expansion**: $+_i(\phi, w) = w'$, adding ϕ to agent i's beliefs in w, moving to a new world w'.
- **Selection**: $\gamma_b(\phi) \subseteq \phi$, choosing the most plausible ϕ-worlds consistent with b_i,
- **Common Learning**: $\oplus_N(\phi, w)$, for a group $N \subseteq \{1, \ldots, n\}$, so they all learn ϕ, each updating their own beliefs.

The modular internal-world semantics for common learning is then combined with an AGM-style revision approach.

Masba is an extension of $\mathcal{F}$. The key addition is the *simulation layer*—"what j would believe if j were i":

Masba is an extension of $\mathcal{F}$. The key addition is the *simulation layer*—"what j would believe if j were i":

$$b^{sim}_{\langle i,j\rangle} \in \mathcal{B}^{sim}_{\langle i,j\rangle},$$

which denotes i's simulative belief states about j. In principle, when thinking of other agents we often simulate others based on the information that we already possess for ourselves:

$$w \xrightarrow{\text{Copy}(b_j)} w' \xrightarrow{\mathcal{B}^{sim}_{\langle i,j\rangle}} w''.$$

MASBA is an extension of $\mathcal{F}$. The key addition is the *simulation layer*—"what j would believe if j were i":

$$b^{sim}_{\langle i,j\rangle} \in \mathcal{B}^{sim}_{\langle i,j\rangle},$$

which denotes i's simulative belief states about j. In principle, when thinking of other agents we often simulate others based on the information that we already possess for ourselves:

$$w \xrightarrow{\text{Copy}(b_j)} w' \xrightarrow{\mathcal{B}^{sim}_{\langle i,j\rangle}} w''.$$

In addition to this, we would need what I shall call a *shared belief state*:

MASBA is an extension of $\mathcal{F}$. The key addition is the *simulation layer*—"what j would believe if j were i":

$$b^{sim}_{\langle i,j\rangle} \in \mathcal{B}^{sim}_{\langle i,j\rangle},$$

which denotes i's simulative belief states about j. In principle, when thinking of other agents we often simulate others based on the information that we already possess for ourselves:

$$w \xrightarrow{\text{Copy}(b_j)} w' \xrightarrow{\mathcal{B}^{sim}_{\langle i,j\rangle}} w''.$$

In addition to this, we would need what I shall call a *shared belief state*:

$$b^{sh}_{\langle j,i\rangle} \in \mathcal{B}^{sh}_{\langle j,i\rangle},$$

denoting *shared states* between j and i, *i.e.* i's belief about j's belief. Informally, "j believes that i believes such-and-such".

By introducing $\mathcal{B}^{sh}_{\langle j,i \rangle}$ and $\mathcal{B}^{sim}_{\langle i,j \rangle}$, the framework *localises* both shared and simulative beliefs by encapsulating them in separate compartments, preserving the integrity of each agent's actual belief state $\mathcal{B}_i$.

By introducing $\mathcal{B}^{sh}_{\langle j,i \rangle}$ and $\mathcal{B}^{sim}_{\langle i,j \rangle}$, the framework *localises* both shared and simulative beliefs by encapsulating them in separate compartments, preserving the integrity of each agent's actual belief state $\mathcal{B}_i$.

With this, we can define MASBA:

By introducing $\mathcal{B}^{sh}_{\langle j,i\rangle}$ and $\mathcal{B}^{sim}_{\langle i,j\rangle}$, the framework *localises* both shared and simulative beliefs by encapsulating them in separate compartments, preserving the integrity of each agent's actual belief state $\mathcal{B}_i$.

With this, we can define MASBA:

Definition (1)

MASBA is a tuple

$$\langle W, U, \{\mathcal{B}_i\}_{1\leq i\leq n}, \{\mathcal{B}^{sh}\}_{\langle j,i\rangle(1\leq i,j\leq n|i\neq j)}, \{\mathcal{B}^{sim}\}_{\langle i,j\rangle(1\leq i,j\leq n|i\neq j)}, \mathcal{C}\rangle.$$

As in $\mathcal{F}$, MASBA can also be represented in a standard Kripke framework via binary accessibility relations:

As in $\mathcal{F}$, Masba can also be represented in a standard Kripke framework via binary accessibility relations:

Definition (2)

Masba generates accessibility relations R_i $(1 \leq i \leq n)$, where R_i is a binary relation on W such that

$$vR_iw \iff w \in \mathcal{C}(\mathrm{bst}_i(v)).$$

As in $\mathcal{F}$, MASBA can also be represented in a standard Kripke framework via binary accessibility relations:

Definition (2)

MASBA generates accessibility relations R_i $(1 \leq i \leq n)$, where R_i is a binary relation on W such that

$$vR_iw \iff w \in \mathcal{C}(\mathsf{bst}_i(v)).$$

Simulative (and shared) belief states can likewise be represented through analogous accessibility relations:

As in $\mathcal{F}$, MASBA can also be represented in a standard Kripke framework via binary accessibility relations:

Definition (2)

MASBA generates accessibility relations R_i $(1 \leq i \leq n)$, where R_i is a binary relation on W such that

$$vR_iw \iff w \in \mathcal{C}(\mathrm{bst}_i(v)).$$

Simulative (and shared) belief states can likewise be represented through analogous accessibility relations:

Definition (3)

In MASBA, the accessibility relation for simulative beliefs $R_{\langle i,j \rangle}$ is a binary relation on W:

$$vR^{sim}_{\langle i,j \rangle}w \iff w \in \mathcal{C}\left(\mathrm{bst}^{sim}_{\langle i,j \rangle}(v)\right).$$

The language of MASBA is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators $B_i,\ B^{sh}_{\langle j,i\rangle},\ B^{sim}_{\langle i,j\rangle}$.

The language of Masba is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i\rangle}$, $B^{sim}_{\langle i,j\rangle}$.

A model $\mathfrak{M}$ consists of a Masba structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

The language of MASBA is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i\rangle}$, $B^{sim}_{\langle i,j\rangle}$.

A model $\mathfrak{M}$ consists of a MASBA structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

1. $w \vDash p$ iff $\mathrm{wst}(w) \in V(p)$.

The language of Masba is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i \rangle}$, $B^{sim}_{\langle i,j \rangle}$.

A model $\mathfrak{M}$ consists of a Masba structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

1. $w \vDash p$ iff $\mathrm{wst}(w) \in V(p)$.
2. $w \vDash \phi \wedge \psi$ iff $w \vDash \phi$ and $w \vDash \psi$.

The language of Masba is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i \rangle}$, $B^{sim}_{\langle i,j \rangle}$.

A model $\mathfrak{M}$ consists of a Masba structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

1. $w \vDash p$ iff $\mathrm{wst}(w) \in V(p)$.
2. $w \vDash \phi \wedge \psi$ iff $w \vDash \phi$ and $w \vDash \psi$.
3. $w \vDash \neg\phi$ iff $w \nvDash \phi$.

The language of MASBA is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i\rangle}$, $B^{sim}_{\langle i,j\rangle}$.

A model $\mathfrak{M}$ consists of a MASBA structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

1. $w \vDash p$ iff $\mathrm{wst}(w) \in V(p)$.
2. $w \vDash \phi \wedge \psi$ iff $w \vDash \phi$ and $w \vDash \psi$.
3. $w \vDash \neg\phi$ iff $w \nvDash \phi$.
4. $w \vDash B_i\phi$ iff for each $w' \in \mathcal{C}(\mathrm{bst}_i(w))$, $w' \vDash \phi$.

The language of Masba is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i\rangle}$, $B^{sim}_{\langle i,j\rangle}$.

A model $\mathfrak{M}$ consists of a Masba structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

1. $w \vDash p$ iff $\mathrm{wst}(w) \in V(p)$.
2. $w \vDash \phi \wedge \psi$ iff $w \vDash \phi$ and $w \vDash \psi$.
3. $w \vDash \neg\phi$ iff $w \nvDash \phi$.
4. $w \vDash B_i\phi$ iff for each $w' \in \mathcal{C}(\mathrm{bst}_i(w))$, $w' \vDash \phi$.

The language of Masba is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i \rangle}$, $B^{sim}_{\langle i,j \rangle}$.

A model $\mathfrak{M}$ consists of a Masba structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

1. $w \vDash p$ iff $\mathrm{wst}(w) \in V(p)$.
2. $w \vDash \phi \wedge \psi$ iff $w \vDash \phi$ and $w \vDash \psi$.
3. $w \vDash \neg\phi$ iff $w \nvDash \phi$.
4. $w \vDash B_i\phi$ iff for each $w' \in \mathcal{C}(\mathrm{bst}_i(w))$, $w' \vDash \phi$.
5. $w \vDash B^{sim}_{\langle i,j \rangle}\phi$ iff for each $w' \in \mathcal{C}\left(\mathrm{bst}^{sim}_{\langle i,j \rangle}(w)\right)$, $w' \vDash \phi$.

The language of Masba is the usual classical propositional language $\mathcal{L}$, enhanced with belief operators B_i, $B^{sh}_{\langle j,i\rangle}$, $B^{sim}_{\langle i,j\rangle}$.

A model $\mathfrak{M}$ consists of a Masba structure plus a valuation function V, where for each propositional variable p, $V(p) \subseteq U$. Truth is evaluated at possible worlds:

1. $w \vDash p$ iff $\mathrm{wst}(w) \in V(p)$.
2. $w \vDash \phi \wedge \psi$ iff $w \vDash \phi$ and $w \vDash \psi$.
3. $w \vDash \neg\phi$ iff $w \nvDash \phi$.
4. $w \vDash B_i\phi$ iff for each $w' \in \mathcal{C}(\mathrm{bst}_i(w))$, $w' \vDash \phi$.
5. $w \vDash B^{sim}_{\langle i,j\rangle}\phi$ iff for each $w' \in \mathcal{C}\left(\mathrm{bst}^{sim}_{\langle i,j\rangle}(w)\right)$, $w' \vDash \phi$.
6. $w \vDash B^{sh}_{\langle i,j\rangle}\phi$ iff for each $w' \in \mathcal{C}\left(\mathrm{bst}^{sh}_{\langle i,j\rangle}(w)\right)$, $w' \vDash \phi$.

The deductive system of MASBA consists of a **KD45** system for the operator B_i, and a $\mathbf{K}, \mathbf{4}, \mathbf{5}$ for $B^{sh}_{\langle j,i \rangle}$; lastly, **K** only for $B^{sim}_{\langle i,j \rangle}$:

1. Tautologies,
2. (K) $B_i(\phi \to \psi) \to (B_i\phi \to B_i\psi)$, similarly for $B^{sh}_{\langle i,j \rangle}$ and $B^{sim}_{\langle i,j \rangle}$,
3. (S) $\neg\big(B_i\phi \ \wedge \ B_i\neg\phi\big)$,
4. (4) $B_i\phi \ \to \ B_iB_i\phi$,
5. (5) $\neg B_i\phi \ \to \ B_i\neg B_i\phi$.

The deductive system of MASBA consists of a **KD45** system for the operator B_i, and a $\mathbf{K}, \mathbf{4}, \mathbf{5}$ for $B^{sh}_{\langle j,i \rangle}$; lastly, **K** only for $B^{sim}_{\langle i,j \rangle}$:

1. Tautologies,
2. (K) $B_i(\phi \to \psi) \to (B_i\phi \to B_i\psi)$, similarly for $B^{sh}_{\langle i,j \rangle}$ and $B^{sim}_{\langle i,j \rangle}$,
3. (S) $\neg\big(B_i\phi \ \wedge \ B_i\neg\phi\big)$,
4. (4) $B_i\phi \ \to \ B_iB_i\phi$,
5. (5) $\neg B_i\phi \ \to \ B_i\neg B_i\phi$.

The framework is *sound* and *complete*[4] showing that MASBA is fully representable in a standard Kripke-Hintikka system.

[4] A proof can be constructed through a canonical model. The complete proof will be appeared on my website.

From now on, we will focus on the *simulative aspects* of MASBA.

From now on, we will focus on the *simulative aspects* of MASBA.

AGM revision operation, denoted by $*$ defined as Levi Identity,

$$(L) \qquad K * \varphi := (K \dot{-} \neg\varphi) + \varphi,$$

From now on, we will focus on the *simulative aspects* of Masba.

AGM revision operation, denoted by $*$ defined as Levi Identity,

$$(L) \qquad K * \varphi := (K \dot{-} \neg\varphi) + \varphi,$$

Three AGM operations will be introduced to suit Masba's need:

From now on, we will focus on the *simulative aspects* of MASBA.

AGM revision operation, denoted by $*$ defined as Levi Identity,

$$(L) \qquad K * \varphi := (K \dot{-} \neg\varphi) + \varphi,$$

Three AGM operations will be introduced to suit MASBA's need:

1. **Expansion**,
2. **Contraction** (by selection),
3. **Revision**.

Expansion. For a multi-agent, multi-compartment setup in MASBA, the expansion $+$ is defined:

$$+^{sim}_{\langle i,j\rangle}(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w',$$

where:

Expansion. For a multi-agent, multi-compartment setup in MASBA, the expansion $+$ is defined:

$$+^{sim}_{\langle i,j\rangle}(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w',$$

where:

$\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w'$, and $\|\varphi\| \sqsubseteq w'$,

Expansion. For a multi-agent, multi-compartment setup in Masba, the expansion $+$ is defined:

$$+^{sim}_{\langle i,j\rangle}(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w',$$

where:

$$\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w', \text{ and } \|\varphi\| \sqsubseteq w',$$

$$\mathrm{bst}^{sim}_{\langle i,j\rangle}(w') = \mathrm{bst}^{sim}_{\langle i,j\rangle}(w) \cup \mathrm{bst}^{sh}_{\langle j,i\rangle}(w),$$

Expansion. For a multi-agent, multi-compartment setup in MASBA, the expansion $+$ is defined:

$$+^{sim}_{\langle i,j\rangle}(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w',$$

where:

$\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w'$, and $\|\varphi\| \sqsubseteq w'$,

$\mathrm{bst}^{sim}_{\langle i,j\rangle}(w') = \mathrm{bst}^{sim}_{\langle i,j\rangle}(w) \cup \mathrm{bst}^{sh}_{\langle j,i\rangle}(w)$,

$\mathrm{wst}(w') = \mathrm{wst}(w)$, and,

Expansion. For a multi-agent, multi-compartment setup in MASBA, the expansion $+$ is defined:

$$+^{sim}_{\langle i,j \rangle}(\mathcal{C}(\text{bst}^{sh}_{\langle j,i \rangle}(w)) = w',$$

where:

$$\mathcal{C}(\text{bst}^{sh}_{\langle j,i \rangle}(w)) = w', \text{ and } \|\varphi\| \sqsubseteq w',$$
$$\text{bst}^{sim}_{\langle i,j \rangle}(w') = \text{bst}^{sim}_{\langle i,j \rangle}(w) \cup \text{bst}^{sh}_{\langle j,i \rangle}(w),$$
$$\text{wst}(w') = \text{wst}(w), \text{ and},$$
$$\text{bst}(w') = \text{bst}(w), \text{ for } k \neq i, j.$$

Expansion. For a multi-agent, multi-compartment setup in MASBA, the expansion $+$ is defined:

$$+^{sim}_{\langle i,j\rangle}(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w',$$

where:

$$\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) = w', \text{ and } \|\varphi\| \sqsubseteq w',$$
$$\mathrm{bst}^{sim}_{\langle i,j\rangle}(w') = \mathrm{bst}^{sim}_{\langle i,j\rangle}(w) \cup \mathrm{bst}^{sh}_{\langle j,i\rangle}(w),$$
$$\mathrm{wst}(w') = \mathrm{wst}(w), \text{ and},$$
$$\mathrm{bst}(w') = \mathrm{bst}(w), \text{ for } k \neq i, j.$$

A simple expansion occurs as

$$\mathcal{C}\big(\mathcal{C}(\mathrm{bst}^{sim}_{\langle i,j\rangle}(w)) + \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))\big) = \Big\{+^{sim}_{\langle i,j\rangle}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) \mid w \sqsubseteq \mathcal{C}(\mathrm{bst}^{sim}_{\langle i,j\rangle}(w))\Big\}$$

Contraction. In MASBA, contraction operation given by:

Contraction. In Masba, contraction operation given by:

$$\dot{-}^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big) = w',$$

is defined by a selection function γ, such that:

$$\gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big) \sqsubseteq \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)),$$

meaning, that from $\|\varphi\| \sqsubseteq \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))$, keep only those worlds consistent with $b^{sh}_{\langle j,i\rangle}$:

Contraction. In Masba, contraction operation given by:

$$\dot{-}^{sim}_{\langle i,j\rangle}(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) = w',$$

is defined by a selection function γ, such that:

$$\gamma_{(b^{sim}_{\langle i,j\rangle})}(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) \sqsubseteq \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)),$$

meaning, that from $\|\varphi\| \sqsubseteq \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w))$, keep only those worlds consistent with $b^{sh}_{\langle j,i\rangle}$:

1. If $\mathcal{C}(\mathsf{bst}^{sim}_{\langle i,j\rangle}(w)) \cup \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) = \emptyset$, then,
2. $\gamma_{(b^{sim}_{\langle i,j\rangle})}(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) = \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) \cup \mathcal{C}(\mathsf{bst}^{sim}_{\langle i,j\rangle}(w))$.

Contraction. In MASBA, contraction operation given by:

$$\dot{-}^{sim}_{\langle i,j\rangle}(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) = w',$$

is defined by a selection function γ, such that:

$$\gamma_{(b^{sim}_{\langle i,j\rangle})}(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) \sqsubseteq \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)),$$

meaning, that from $\|\varphi\| \sqsubseteq \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w))$, keep only those worlds consistent with $b^{sh}_{\langle j,i\rangle}$:

1. If $\mathcal{C}(\mathsf{bst}^{sim}_{\langle i,j\rangle}(w)) \cup \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) = \emptyset$, then,
2. $\gamma_{(b^{sim}_{\langle i,j\rangle})}(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) \ = \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w)) \cup \mathcal{C}(\mathsf{bst}^{sim}_{\langle i,j\rangle}(w))$.

When multiple compartments take part simultaneously, we can modify this selection function accordingly.

Revision. The final step in *simulative belief ascription* is revision,

Revision. The final step in *simulative belief ascription* is revision, $*^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big) = w'$, defined by the Levi Identity:

Revision. The final step in *simulative belief ascription* is revision, $*^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big) = w'$, defined by the Levi Identity:

$$\begin{aligned}\mathcal{C}(\mathrm{bst}^{sim}_{\langle i,j\rangle}(w)) *_{\langle i,j\rangle} \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) \\ := \Big(\gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))\big)\Big) +^{sim}_{\langle i,j\rangle} \mathcal{C}\big(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big),\end{aligned}$$

where,

Revision. The final step in *simulative belief ascription* is revision, $*^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big) = w'$, defined by the Levi Identity:

$$\begin{aligned}\mathcal{C}(\mathrm{bst}^{sim}_{\langle i,j\rangle}(w)) *_{\langle i,j\rangle} \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) \\ := \Big(\gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))\big) +^{sim}_{\langle i,j\rangle} \mathcal{C}\big(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big),\end{aligned}$$

where,

$$\|\varphi\| \sqsubseteq \mathcal{C}\big(\mathrm{bst}^{sh}_{\langle j,u\rangle}(w)\big),$$

Revision. The final step in *simulative belief ascription* is revision, $*^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big) = w'$, defined by the Levi Identity:

$$\begin{aligned}
&\mathcal{C}(\mathrm{bst}^{sim}_{\langle i,j\rangle}(w)) *_{\langle i,j\rangle} \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)) \\
&\qquad := \Big(\gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))\big)\Big) +^{sim}_{\langle i,j\rangle} \mathcal{C}\big(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big),
\end{aligned}$$

where,

$$\begin{aligned}
&\|\varphi\| \sqsubseteq \mathcal{C}\big(\mathrm{bst}^{sh}_{\langle j,u\rangle}(w)\big), \\
&\mathrm{wst}(w) = \mathrm{wst}(w'),
\end{aligned}$$

Revision. The final step in *simulative belief ascription* is revision, $*^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big) = w'$, defined by the Levi Identity:

$$\begin{aligned}\mathcal{C}(\mathrm{bst}^{sim}_{\langle i,j\rangle}(w)) *_{\langle i,j\rangle} \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))\\ := \Big(\gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))\Big) +^{sim}_{\langle i,j\rangle} \mathcal{C}\big(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w)\big),\end{aligned}$$

where,

$$\begin{aligned}&\|\varphi\| \sqsubseteq \mathcal{C}\big(\mathrm{bst}^{sh}_{\langle j,u\rangle}(w)\big),\\ &\mathrm{wst}(w) = \mathrm{wst}(w'),\\ &\mathrm{bst}_k(w) = \mathrm{bst}(w') \text{ for } k \neq i, j,\end{aligned}$$

Revision. The final step in *simulative belief ascription* is revision, $*^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\text{bst}^{sh}_{\langle j,i\rangle}(w)\big) = w'$, defined by the Levi Identity:

$$\begin{aligned}\mathcal{C}(\text{bst}^{sim}_{\langle i,j\rangle}(w)) *_{\langle i,j\rangle} \mathcal{C}(\text{bst}^{sh}_{\langle j,i\rangle}(w)) \\ := \Big(\gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\text{bst}^{sh}_{\langle j,i\rangle}(w))\big)\Big) +^{sim}_{\langle i,j\rangle} \mathcal{C}\big(\text{bst}^{sh}_{\langle j,i\rangle}(w)\big),\end{aligned}$$

where,

$$\begin{aligned}&\|\varphi\| \sqsubseteq \mathcal{C}\big(\text{bst}^{sh}_{\langle j,u\rangle}(w)\big),\\ &\text{wst}(w) = \text{wst}(w'),\\ &\text{bst}_k(w) = \text{bst}(w') \text{ for } k \neq i,j,\\ &\text{bst}^{sim}_{\langle i,j\rangle}(w') = \big(\text{bst}^{sim}_{\langle i,j\rangle}(w)\big) * \mathcal{C}(\text{bst}^{sh}_{\langle j,i\rangle}).\end{aligned}$$

Revision continues.

Revision continues.

$*^{sim}_{\langle i,j \rangle}$ in MASBA is a *simulative belief revision* operation, by taking $\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i \rangle})$ with a minimal revision of $\mathrm{bst}^{sim}_{\langle i,j \rangle}(w)$:

Revision continues.

$*^{sim}_{\langle i,j\rangle}$ in Masba is a *simulative belief revision* operation, by taking $\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle})$ with a minimal revision of $\mathsf{bst}^{sim}_{\langle i,j\rangle}(w)$:

$$\mathcal{C}\big(\mathsf{bst}^{sim}_{\langle i,j\rangle}(w) * \mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle})\big) \\ = \Big\{ *^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}) \mid w \in \gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\mathsf{bst}^{sh}_{\langle j,i\rangle}(w))\big)\Big\},$$

Revision continues.

$*^{sim}_{\langle i,j\rangle}$ in MASBA is a *simulative belief revision* operation, by taking $\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle})$ with a minimal revision of $\mathrm{bst}^{sim}_{\langle i,j\rangle}(w)$:

$$\begin{aligned}\mathcal{C}\big(&\mathrm{bst}^{sim}_{\langle i,j\rangle}(w) * \mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle})\big)\\ &= \Big\{ *^{sim}_{\langle i,j\rangle}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}) \mid w \in \gamma_{(b^{sim}_{\langle i,j\rangle})}\big(\mathcal{C}(\mathrm{bst}^{sh}_{\langle j,i\rangle}(w))\big)\Big\},\end{aligned}$$

Here, the agent j revises the simulative belief state $b^{sim}_{\langle i,j\rangle}$ with respect to *shared belief state* of j to i.

Masba, an extension of $\mathcal{F}$ incorporating *simulative* and *shared* belief states, provides a modular internal-worlds semantics for simulative belief ascriptions between agents. By treating a world as

$$w = \langle u,\ b_{i(1\leq i\leq n)},\ b^{sh}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)},\ b^{sim}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)}\rangle,$$

Masba supports:

MASBA, an extension of $\mathcal{F}$ incorporating *simulative* and *shared* belief states, provides a modular internal-worlds semantics for simulative belief ascriptions between agents. By treating a world as

$$w = \left\langle u,\ b_{i(1\leq i\leq n)},\ b^{sh}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)},\ b^{sim}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)}\right\rangle,$$

MASBA supports:

1. Multiple doxastic compartments: $b,\ b^{sh},\ b^{sim}$,

MASBA, an extension of $\mathcal{F}$ incorporating *simulative* and *shared* belief states, provides a modular internal-worlds semantics for simulative belief ascriptions between agents. By treating a world as

$$w = \left\langle u,\ b_{i(1 \leq i \leq n)},\ b^{sh}_{\langle i,j \rangle (1 \leq i,j \leq n \,|\, i \neq j)},\ b^{sim}_{\langle i,j \rangle (1 \leq i,j \leq n \,|\, i \neq j)} \right\rangle,$$

MASBA supports:

1. Multiple doxastic compartments: b, b^{sh}, b^{sim},
2. Local, modular updates rather than global ones,

Masba, an extension of $\mathcal{F}$ incorporating *simulative* and *shared* belief states, provides a modular internal-worlds semantics for simulative belief ascriptions between agents. By treating a world as

$$w = \langle u,\ b_{i(1\leq i\leq n)},\ b^{sh}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)},\ b^{sim}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)}\rangle,$$

Masba supports:

1. Multiple doxastic compartments: b, b^{sh}, b^{sim},
2. Local, modular updates rather than global ones,
3. Distinguishing between common beliefs and simulative beliefs,

Masba, an extension of $\mathcal{F}$ incorporating *simulative* and *shared* belief states, provides a modular internal-worlds semantics for simulative belief ascriptions between agents. By treating a world as

$$w = \langle u,\ b_{i(1\leq i\leq n)},\ b^{sh}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)},\ b^{sim}_{\langle i,j\rangle(1\leq i,j\leq n\,|\,i\neq j)}\rangle,$$

Masba supports:

1. Multiple doxastic compartments: $b,\ b^{sh},\ b^{sim}$,
2. Local, modular updates rather than global ones,
3. Distinguishing between common beliefs and simulative beliefs,
4. Incorporating AGM-style revision for simulative belief ascriptions, better suited to dominating view in *mental simulation*.

Thank you!

[1] Peter Aczel. *Non-Well-Founded Sets*. Number no. 14 in CSLI Lecture Notes. Center for the Study of Language and Information, 1988. ISBN 978-0-937073-21-6 978-0-937073-22-3.

[2] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. In Horacio Arló-Costa, Vincent F. Hendricks, and Johan Van Benthem, editors, *Readings in Formal Epistemology*, pages 195–217. Springer International Publishing, 2016. ISBN 978-3-319-20450-5 978-3-319-20451-2. doi: 10.1007/978-3-319-20451-2_13.

[3] G. Aucher. Generalizing AGM to a multi-agent setting. *Logic Journal of IGPL*, 18(4):530–558, 2010-08-01. ISSN 1367-0751, 1368-9894. doi: 10.1093/jigpal/jzp037.

[4] John Cantwell. A Formal Model of Multi-Agent Belief-Interaction. *Journal of Logic, Language and Information*, 14:397–422, 2005. doi: 10.1007/s10849-005-4019-8.

[5] John Cantwell. A Model for Updates in a Multi-Agent Setting. *Journal of Applied Non-Classical Logics*, 17(2): 183–196, 2007-01. ISSN 1166-3081, 1958-5780. doi: 10.3166/jancl.17.183-196.

[6] Zhuang Chen, Jincenzi Wu, Jinfeng Zhou, Bosi Wen, Guanqun Bi, Gongyao Jiang, Yaru Cao, Mengting Hu, Yunghwei Lai, Zexuan Xiong, and Minlie Huang. ToMBench: Benchmarking Theory of Mind in Large Language Models. 2024-12-08. doi: 10.48550/arXiv.2402.15052.

[7] Jelle Gerbrandy. *Bisimulations on Planet Kripke*. Institute for Logic, Language and Computation, Universiteit van Amsterdam, 1999. ISBN 978-90-5776-019-8.

[8] Jelle Gerbrandy and Willem Groeneveld. Reasoning about Information Change. *Journal of Logic, Language and Information*, 6:147–169, 1997.

[9] Alvin I. Goldman. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, 2006. ISBN 978-0195369830.

[10] Robert M. Gordon. Folk Psychology as Simulation. *Mind & Language*, 1(2):158–171, 1986-06. ISSN 0268-1064, 1468-0017. doi: 10.1111/j.1468-0017.1986.tb00324.x.

[11] Robert M. Gordon. Simulation Without Introspection or Inference From Me to You. In Martin Davies and Tony Stone, editors, *Mental Simulation: Evaluations and Applications - Reading in Mind and Language*, Readings in Mind and Language, pages 53–67. Wiley-Blackwell, 1995-10. ISBN 978-0-631-19873-4.

[12] James W. A. Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, Michael S. A. Graziano, and Cristina Becchio. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 8(7):1285–1295, 2024-05-20. ISSN 2397-3374. doi: 10.1038/s41562-024-01882-z.

[13] Alex. Wilf, Sihyun Shawn. Lee, Paul Pu. Liang, and Louis-Philippe Morency. Think twice: Perspective-taking improves large language models' theory-of-mind capabilities. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceddings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Bangkok, Thailand, 2024. Association for Computational Linguistics. URL `https://arxiv.org/abs/2311.10227`.