Games with noisy signals about emotions

Pierpaolo Battigalli
Bocconi and IGIER

Nicolò Generoso
Yale

Bocconi University
24 May 2024

Introduction: Motivation

- Starting point: well-being depends on **experienced emotions**.
- Key idea: **noisy signals** (e.g., facial cues) may betray emotions.
  - Everyday experience: blushing = embarrassment, gaze contact = interest, smiling = happiness, etc.
  - Literature:
    - Emotional leakage is associated with **lies and deception** (Porter et al. 2012; Matsumoto and Hwang 2018);
    - Nonverbal communication expresses **(dis)liking** (Givens 1978);
    - Gesture informs an audience of a speaker's **(unspoken) thoughts** (Goldin-Meadow 1999);
    - Facial cues allow to recognize others' **trustworthiness** or **predisposition to anger** (Van Leeuwen et al. 2018; Stirrat and Perrett, 2010)
    - **Emotional contagion** occurs when people recognize and mimic others' emotions (Hatfield et al. 2014; see Vasquez and Weretka 2020 for an economic analysis).
- Question: can emotional signals **shape behavior** when individuals **reason strategically**?

Introduction: Motivation

- Starting point: well-being depends on **experienced emotions**.
- Key idea: **noisy signals** (e.g., facial cues) may betray emotions.
  - Everyday experience: blushing = embarrassment, gaze contact = interest, smiling = happiness, etc.
  - Literature:
    - Emotional leakage is associated with **lies and deception** (Porter et al. 2012; Matsumoto and Hwang 2018);
    - Nonverbal communication expresses **(dis)liking** (Givens 1978);
    - Gesture informs an audience of a speaker's **(unspoken) thoughts** (Goldin-Meadow 1999);
    - Facial cues allow to recognize others' **trustworthiness** or **predisposition to anger** (Van Leeuwen et al. 2018; Stirrat and Perrett, 2010)
    - **Emotional contagion** occurs when people recognize and mimic others' emotions (Hatfield et al. 2014; see Vasquez and Weretka 2020 for an economic analysis).
- Question: can emotional signals **shape behavior** when individuals **reason strategically**?

Introduction: Motivation

- Starting point: well-being depends on **experienced emotions**.
- Key idea: **noisy signals** (e.g., facial cues) may betray emotions.
  - Everyday experience: blushing = embarrassment, gaze contact = interest, smiling = happiness, etc.
  - Literature:
    - Emotional leakage is associated with **lies and deception** (Porter et al. 2012; Matsumoto and Hwang 2018);
    - Nonverbal communication expresses **(dis)liking** (Givens 1978);
    - Gesture informs an audience of a speaker's **(unspoken) thoughts** (Goldin-Meadow 1999);
    - Facial cues allow to recognize others' **trustworthiness** or **predisposition to anger** (Van Leeuwen et al. 2018; Stirrat and Perrett, 2010)
    - **Emotional contagion** occurs when people recognize and mimic others' emotions (Hatfield et al. 2014; see Vasquez and Weretka 2020 for an economic analysis).
- Question: can emotional signals **shape behavior** when individuals **reason strategically**?

Introduction: Motivation

- Starting point: well-being depends on **experienced emotions**.
- Key idea: **noisy signals** (e.g., facial cues) may betray emotions.
    - Everyday experience: blushing = embarrassment, gaze contact = interest, smiling = happiness, etc.
    - Literature:
        - Emotional leakage is associated with **lies and deception** (Porter et al. 2012; Matsumoto and Hwang 2018);
        - Nonverbal communication expresses **(dis)liking** (Givens 1978);
        - Gesture informs an audience of a speaker's **(unspoken) thoughts** (Goldin-Meadow 1999);
        - Facial cues allow to recognize others' **trustworthiness** or **predisposition to anger** (Van Leeuwen et al. 2018; Stirrat and Perrett, 2010)
        - **Emotional contagion** occurs when people recognize and mimic others' emotions (Hatfield et al. 2014; see Vasquez and Weretka 2020 for an economic analysis).
- Question: can emotional signals **shape behavior** when individuals **reason strategically**?

Introduction: Contribution

- We formalize a **general framework** to model emotional feedback.
- We show how to derive behavioral predictions:
  - Definition of **rationality features**: both *cognitive* (coherence of beliefs, belief updating consistent with evidence and according to the rules of conditional probabilities), and *behavioral* (rational planning, consistent implementation of plans).
  - These features hold *only at some states of the world*: each requirement is an explicit assumption, and players can entertain the possibility of cognitive or behavioral failures of their opponents.
  - We prove that **rationality** is a well-defined **event**.
  - We define a version of the **strong rationalizability** solution procedure: we show it captures the implications of rationality and common strong belief in rationality.

Introduction: Contribution

- We formalize a **general framework** to model emotional feedback.
- We show how to derive behavioral predictions:
  - Definition of **rationality features**: both *cognitive* (coherence of beliefs, belief updating consistent with evidence and according to the rules of conditional probabilities), and *behavioral* (rational planning, consistent implementation of plans).
  - These features hold *only at some states of the world*: each requirement is an explicit assumption, and players can entertain the possibility of cognitive or behavioral failures of their opponents.
  - We prove that **rationality** is a well-defined **event**.
  - We define a version of the **strong rationalizability** solution procedure: we show it captures the implications of rationality and common strong belief in rationality.

Introduction: Contribution

- We formalize a **general framework** to model emotional feedback.
- We show how to derive behavioral predictions:
    - Definition of **rationality features**: both *cognitive* (coherence of beliefs, belief updating consistent with evidence and according to the rules of conditional probabilities), and *behavioral* (rational planning, consistent implementation of plans).
    - These features hold *only at some states of the world*: each requirement is an explicit assumption, and players can entertain the possibility of cognitive or behavioral failures of their opponents.
    - We prove that **rationality** is a well-defined **event**.
    - We define a version of the **strong rationalizability** solution procedure: we show it captures the implications of rationality and common strong belief in rationality.

Introduction: Contribution

- We formalize a **general framework** to model emotional feedback.
- We show how to derive behavioral predictions:
  - Definition of **rationality features**: both *cognitive* (coherence of beliefs, belief updating consistent with evidence and according to the rules of conditional probabilities), and *behavioral* (rational planning, consistent implementation of plans).
  - These features hold *only at some states of the world*: each requirement is an explicit assumption, and players can entertain the possibility of cognitive or behavioral failures of their opponents.
  - We prove that **rationality** is a well-defined **event**.
  - We define a version of the **strong rationalizability** solution procedure: we show it captures the implications of rationality and common strong belief in rationality.

Introduction: Related literature (non-exhaustive)

Psychological games and belief-dependent preferences

- Articles: Geanakoplos, Pearce, and Stacchetti (1989); Battigalli and Dufwenberg (2009); Battigalli, Corrao and Dufwenberg (2019).
- Survey: Battigalli and Dufwenberg (2022).

Epistemic game theory

- Epistemic analysis without type structures: Battigalli, Corrao and Sanna (2020).
- Consistency between behavior and intentions: Battigalli and De Vito (2021).
- Strong rationalizability (and ancestors): Pearce (1984); Battigalli (1997); Battigalli and Siniscalchi (2002); Battigalli and Prestipino (2013).

What's next?

1. Definition of a **general framework**
2. Description of **inferences**
3. Formalization of a notion of **rationality**
4. Definition of a **solution concept** to derive predictions
5. Concluding remarks

What's next?

1. Definition of a **general framework**
2. Description of **inferences**
3. Formalization of a notion of **rationality**
4. Definition of a **solution concept** to derive predictions
5. Concluding remarks

Framework: Warm-up

To give an idea of the phenomena we model, here is a heuristic example (called buy me an ice-cream).

Situation:

- Child is at home alone: can choose *Homework* or *Video-games*.
- Mom gets back home. Child: "Mom, can you please buy me an ice-cream?" Mom: "Did you do your homework?".
  Child can choose to answer *Yes* or *No* (note: *Yes* = "I did my homework"). *But* Child may *blush* if he lies.
- Mom decides whether to *Buy* the ice-cream or *Not*.

More general and relevant problem: *disclosure of information*.
Is lying worth it if lies can be spotted?

Framework: Warm-up

To give an idea of the phenomena we model, here is a heuristic example (called buy me an ice-cream).

Situation:

- Child is at home alone: can choose *Homework* or *Video-games*.
- Mom gets back home. Child: "Mom, can you please buy me an ice-cream?" Mom: "Did you do your homework?".
  Child can choose to answer *Yes* or *No* (note: *Yes* = "I did my homework"). *But* Child may *blush* if he lies.
- Mom decides whether to *Buy* the ice-cream or *Not*.

More general and relevant problem: *disclosure of information*.
Is lying worth it if lies can be spotted?

Framework: Warm-up

To give an idea of the phenomena we model, here is a heuristic example (called buy me an ice-cream).

Situation:

- Child is at home alone: can choose *Homework* or *Video-games*.
- Mom gets back home. Child: "Mom, can you please buy me an ice-cream?" Mom: "Did you do your homework?".
  Child can choose to answer *Yes* or *No* (note: *Yes* = "I did my homework"). *But* Child may *blush* if he lies.
- Mom decides whether to *Buy* the ice-cream or *Not*.

More general and relevant problem: *disclosure of information*.
Is lying worth it if lies can be spotted?

Framework: Masterplan

Mechanisms at play:

1. Emotions determine **utilities** and **emotional feedback**: these mechanisms are embedded in an interactive setting.
2. Emotions are triggered by the **game unfolding** and **endogenous beliefs**.
3. Observed emotional feedback further **informs beliefs**.

To fix ideas (in green, the game-specific dynamics):

Framework: Masterplan

Mechanisms at play:

1. Emotions determine **utilities** and **emotional feedback**: these mechanisms are embedded in an interactive setting.
2. Emotions are triggered by the **game unfolding** and **endogenous beliefs**.
3. Observed emotional feedback further **informs beliefs**.

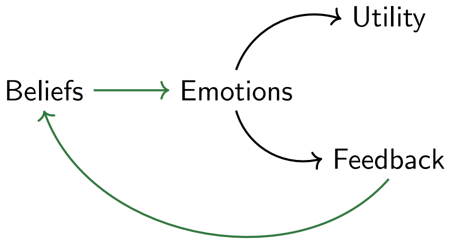
To fix ideas (in green, the game-specific dynamics):

Framework: Masterplan

Mechanisms at play:

1. Emotions determine **utilities** and **emotional feedback**: these mechanisms are embedded in an interactive setting.
2. Emotions are triggered by the **game unfolding** and **endogenous beliefs**.
3. Observed emotional feedback further **informs beliefs**.

To fix ideas (in green, the game-specific dynamics):

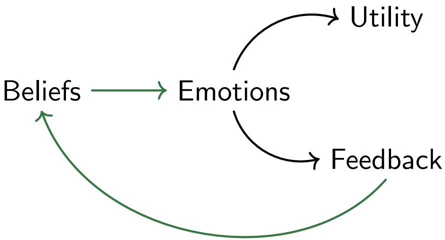


Framework: Masterplan

In buy me an ice-cream:

- **Feedback**: blushing or not.
- **Utility**: we suppose Child dislikes being seen as a liar.
- Relevant **"emotions"**: confidence (feedback), image concern (utility).
- Tie with **beliefs**:
    - Confidence: more confident if he thinks *that Mom would still buy him the ice-cream, even if he blushes.*
    - Image concern: feels blame if he thinks *that Mom thinks he lied.*

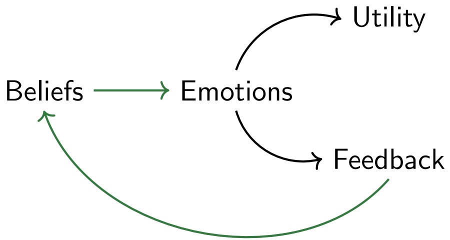
Framework: Masterplan



In buy me an ice-cream:

- **Feedback**: blushing or not.
- **Utility**: we suppose Child dislikes being seen as a liar.
- Relevant **"emotions"**: confidence (feedback), image concern (utility).
- Tie with **beliefs**:
  - Confidence: more confident if he thinks *that Mom would still buy him the ice-cream, even if he blushes.*
  - Image concern: feels blame if he thinks *that Mom thinks he lied.*

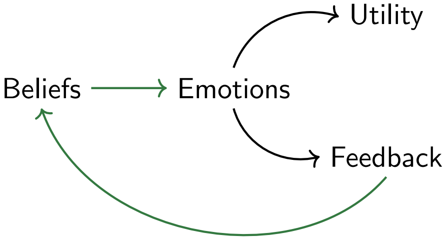
Framework: Masterplan



In buy me an ice-cream:

- **Feedback**: blushing or not.
- **Utility**: we suppose Child dislikes being seen as a liar.
- Relevant **"emotions"**: confidence (feedback), image concern (utility).
- Tie with **beliefs**:
  - Confidence: more confident if he thinks *that Mom would still buy him the ice-cream, even if he blushes.*
  - Image concern: feels blame if he thinks *that Mom thinks he lied.*

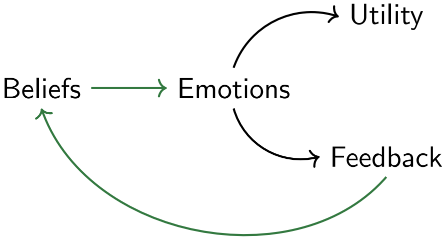
Framework: Masterplan



In buy me an ice-cream:

- **Feedback**: blushing or not.
- **Utility**: we suppose Child dislikes being seen as a liar.
- Relevant **"emotions"**: confidence (feedback), image concern (utility).
- Tie with **beliefs**:
  - Confidence: more confident if he thinks *that Mom would still buy him the ice-cream, even if he blushes.*
  - Image concern: feels blame if he thinks *that Mom thinks he lied.*

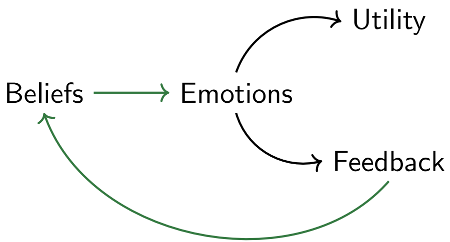
Framework: Masterplan



In buy me an ice-cream:

- **Feedback**: blushing or not.
- **Utility**: we suppose Child dislikes being seen as a liar.
- Relevant **"emotions"**: confidence (feedback), image concern (utility).
- Tie with **beliefs**:
  - Confidence: more confident if he thinks *that Mom would still buy him the ice-cream, even if he blushes.*
  - Image concern: feels blame if he thinks *that Mom thinks he lied.*

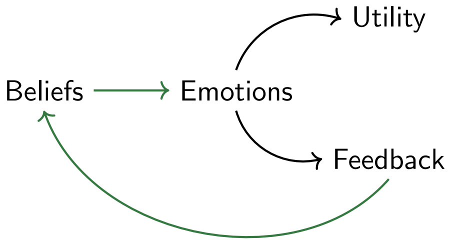
Framework: Masterplan



In buy me an ice-cream:

- **Feedback**: blushing or not.
- **Utility**: we suppose Child dislikes being seen as a liar.
- Relevant **“emotions”**: confidence (feedback), image concern (utility).
- Tie with **beliefs**:
  - Confidence: more confident if he thinks *that Mom would still buy him the ice-cream, even if he blushes.*
  - Image concern: feels blame if he thinks *that Mom thinks he lied.*

Framework: Masterplan



In buy me an ice-cream:

- **Feedback**: blushing or not.
- **Utility**: we suppose Child dislikes being seen as a liar.
- Relevant **"emotions"**: confidence (feedback), image concern (utility).
- Tie with **beliefs**:
  - Confidence: more confident if he thinks *that Mom would still buy him the ice-cream, even if he blushes*.
  - Image concern: feels blame if he thinks *that Mom thinks he lied*.

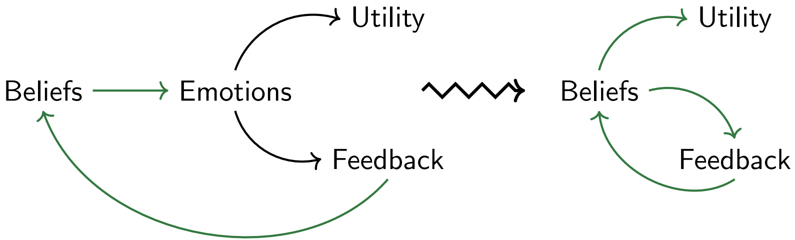
Framework: Masterplan

But it is convenient to retrieve a "reduced form" representation (entirely game-dependent):

Why? Because we work with states of the world that specify how players would behave/what they would think at all game-specific contingencies.

Framework: Masterplan

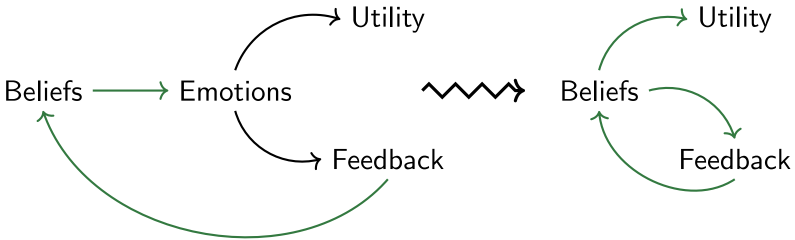
But it is convenient to retrieve a "reduced form" representation (entirely game-dependent):



Why? Because we work with states of the world that specify how players would behave/what they would think at all game-specific contingencies.

Framework: Masterplan

But it is convenient to retrieve a "reduced form" representation (entirely game-dependent):



Why? Because we work with states of the world that specify how players would behave/what they would think at all game-specific contingencies.

Framework: Basic ingredients

Notation: for a generic indexed profile of sets $(X_i)_{i \in I}$, $X := \times_{i \in I} X_i$. For a generic set X, X^n ($n \in \mathbb{N}$) is the set of sequences of elements of X of length n.

Standard ingredients (all finite):

- Set of **players** I.
- Set of **actions** of i: A_i.
- Set of **personal traits** of i: Θ_i. Player i knows θ_i (informal assumption).
- Set of **outcomes** for i: Y_i.
- Set of **messages** i may observe: M_i.

Non-standard ingredients (compact metrizable):

- Set of **emotions** of i: E_i.
- Set of **streams of emotions** of i: $E_i^{\leq T+1} := \bigcup_{t=1}^{T+1} E_i^t$ ($T \in \mathbb{N}$).

Framework: Basic ingredients

Notation: for a generic indexed profile of sets $(X_i)_{i \in I}$, $X := \times_{i \in I} X_i$. For a generic set X, X^n ($n \in \mathbb{N}$) is the set of sequences of elements of X of length n.

Standard ingredients (all finite):

- Set of **players** I.
- Set of **actions** of i: A_i.
- Set of **personal traits** of i: Θ_i. Player i knows θ_i (informal assumption).
- Set of **outcomes** for i: Y_i.
- Set of **messages** i may observe: M_i.

Non-standard ingredients (compact metrizable):

- Set of **emotions** of i: E_i.
- Set of **streams of emotions** of i: $E_i^{\leq T+1} := \bigcup_{t=1}^{T+1} E_i^t$ ($T \in \mathbb{N}$).

Framework: Basic ingredients

Notation: for a generic indexed profile of sets $(X_i)_{i\in I}$, $X := \times_{i\in I} X_i$. For a generic set X, X^n ($n \in \mathbb{N}$) is the set of sequences of elements of X of length n.

Standard ingredients (all finite):

- Set of **players** I.
- Set of **actions** of i: A_i.
- Set of **personal traits** of i: Θ_i. Player i knows θ_i (informal assumption).
- Set of **outcomes** for i: Y_i.
- Set of **messages** i may observe: M_i.

Non-standard ingredients (compact metrizable):

- Set of **emotions** of i: E_i.
- Set of **streams of emotions** of i: $E_i^{\leq T+1} := \bigcup_{t=1}^{T+1} E_i^t$ ($T \in \mathbb{N}$).

Framework: Basic ingredients

Notation: for a generic indexed profile of sets $(X_i)_{i \in I}$, $X := \times_{i \in I} X_i$. For a generic set X, X^n ($n \in \mathbb{N}$) is the set of sequences of elements of X of length n.

Standard ingredients (all finite):

- Set of **players** I.
- Set of **actions** of i: A_i.
- Set of **personal traits** of i: Θ_i. Player i knows θ_i (informal assumption).
- Set of **outcomes** for i: Y_i.
- Set of **messages** i may observe: M_i.

Non-standard ingredients (compact metrizable):

- Set of **emotions** of i: E_i.
- Set of **streams of emotions** of i: $E_i^{\leq T+1} := \bigcup_{t=1}^{T+1} E_i^t$ ($T \in \mathbb{N}$).

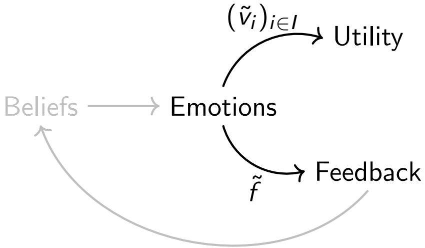
Framework: Emotions, utilities, feedback

We start with:

- Continuous **feedback function**: $\tilde{f} : A \times \Theta \times E^{\leq T+1} \to \Delta(M)$.
  Interpretation: messages are *stochastic* because messages about emotions are *noisy*.
- Profile of continuous **psychological utility functions**: $(\tilde{v}_i : Y \times \Theta \times E^{\leq T+1} \to \mathbb{R})_{i \in I}$.

Framework: Emotions, utilities, feedback

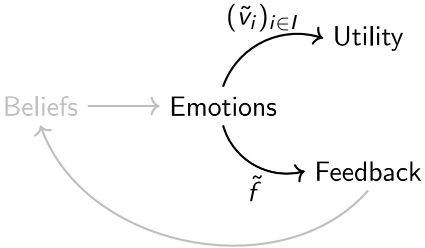
We start with:



- Continuous **feedback function**: $\tilde{f} : A \times \Theta \times E^{\leq T+1} \to \Delta(M)$.
  Interpretation: messages are *stochastic* because messages about emotions are *noisy*.
- Profile of continuous **psychological utility functions**: $(\tilde{v}_i : Y \times \Theta \times E^{\leq T+1} \to \mathbb{R})_{i\in I}$.

Framework: Emotions, utilities, feedback

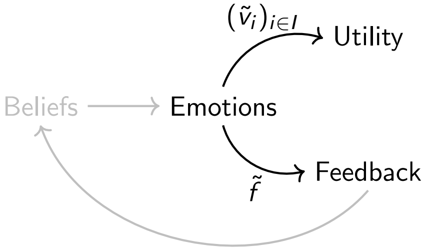
We start with:



- Continuous **feedback function**: $\tilde{f} : A \times \Theta \times E^{\leq T+1} \to \Delta(M)$.
  Interpretation: messages are *stochastic* because messages about emotions are *noisy*.
- Profile of continuous **psychological utility functions**: $(\tilde{v}_i : Y \times \Theta \times E^{\leq T+1} \to \mathbb{R})_{i\in I}$.

Framework: Emotions, utilities, feedback

We start with:



- Continuous **feedback function**: $\tilde{f} : A \times \Theta \times E^{\leq T+1} \to \Delta(M)$.
  Interpretation: messages are *stochastic* because messages about emotions are *noisy*.
- Profile of continuous **psychological utility functions**: $(\tilde{v}_i : Y \times \Theta \times E^{\leq T+1} \to \mathbb{R})_{i\in I}$.

Framework: Emotions and beliefs

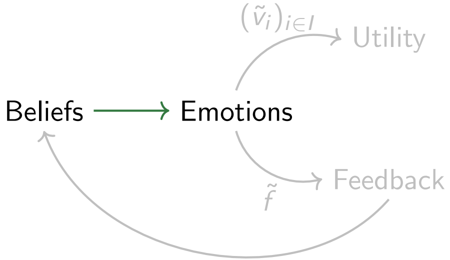
Next, focus on:

We need:

- A description of the rules of interaction (i.e., a **game form**).
- A description of how players would **behave** and **think**.
- A function mapping these **attitudes into emotions**.

Framework: Emotions and beliefs

Next, focus on:

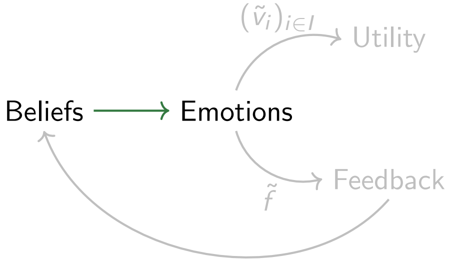


We need:

- A description of the rules of interaction (i.e., a **game form**).
- A description of how players would **behave** and **think**.
- A function mapping these **attitudes into emotions**.

Framework: Emotions and beliefs

Next, focus on:

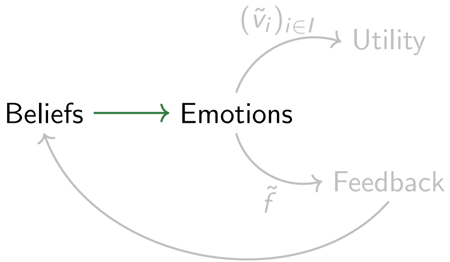


We need:

- A description of the rules of interaction (i.e., a **game form**).
- A description of how players would **behave** and **think**.
- A function mapping these **attitudes into emotions**.

Framework: Emotions and beliefs

Next, focus on:

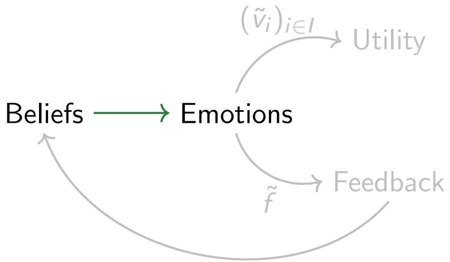


We need:

- A description of the rules of interaction (i.e., a **game form**).
- A description of how players would **behave** and **think**.
- A function mapping these **attitudes into emotions**.

Framework: Emotions and beliefs

Next, focus on:

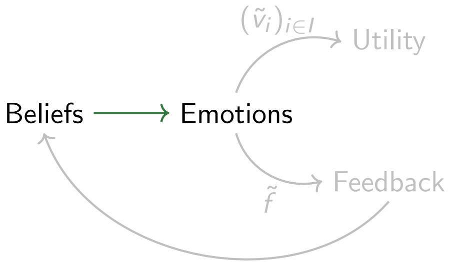


We need:

- A description of the rules of interaction (i.e., a **game form**).
- A description of how players would **behave** and **think**.
- A function mapping these **attitudes into emotions**.

Framework: Emotions and beliefs

Next, focus on:



We need:

- A description of the rules of interaction (i.e., a **game form**).
- A description of how players would **behave** and **think**.
- A function mapping these **attitudes into emotions**.

Framework: Game form

- Assumption: players **need not observe** others' moves; they only get imperfect information about how the game is being played.
- Innovation: we model game-specific information as a **stream**; after each stage, players receive messages about co-players' moves.
- Set of **previous play messages** (PPM) of player i: $M_{i,p}$ (finite). PPMs generated based on the game unfolding (through a function $P : \bigcup_{t=0}^{T} A^t \to M_p$).
- Assumption: Players realize which actions are **feasible** at next stage only by looking at their last PPM (through correspondence $\mathcal{A}_i : M_{i,p} \rightrightarrows A_i$).

Framework: Game form

- Assumption: players **need not observe** others' moves; they only get imperfect information about how the game is being played.
- Innovation: we model game-specific information as a **stream**; after each stage, players receive messages about co-players' moves.
- Set of **previous play messages** (PPM) of player i: $M_{i,p}$ (finite). PPMs generated based on the game unfolding (through a function $P : \bigcup_{t=0}^{T} A^t \to M_p$).
- Assumption: Players realize which actions are **feasible** at next stage only by looking at their last PPM (through correspondence $\mathcal{A}_i : M_{i,p} \rightrightarrows A_i$).

Framework: Game form

- Assumption: players **need not observe** others' moves; they only get imperfect information about how the game is being played.
- Innovation: we model game-specific information as a **stream**; after each stage, players receive messages about co-players' moves.
- Set of **previous play messages** (PPM) of player i: $M_{i,p}$ (finite). PPMs generated based on the game unfolding (through a function $P : \bigcup_{t=0}^{T} A^t \to M_p$).
- Assumption: Players realize which actions are **feasible** at next stage only by looking at their last PPM (through correspondence $\mathcal{A}_i : M_{i,p} \rightrightarrows A_i$).

Framework: Game form

- Assumption: players **need not observe** others' moves; they only get imperfect information about how the game is being played.
- Innovation: we model game-specific information as a **stream**; after each stage, players receive messages about co-players' moves.
- Set of **previous play messages** (PPM) of player i: $M_{i,p}$ (finite). PPMs generated based on the game unfolding (through a function $P : \bigcup_{t=0}^{T} A^t \to M_p$).
- Assumption: Players realize which actions are **feasible** at next stage only by looking at their last PPM (through correspondence $\mathcal{A}_i : M_{i,p} \rightrightarrows A_i$).

Framework: Game form

- Assumption: players **need not observe** others' moves; they only get imperfect information about how the game is being played.
- Innovation: we model game-specific information as a **stream**; after each stage, players receive messages about co-players' moves.
- Set of **previous play messages** (PPM) of player i: $M_{i,p}$ (finite). PPMs generated based on the game unfolding (through a function $P : \bigcup_{t=0}^{T} A^t \to M_p$).
- Assumption: Players realize which actions are **feasible** at next stage only by looking at their last PPM (through correspondence $\mathcal{A}_i : M_{i,p} \rightrightarrows A_i$).

Framework: Game form

- Retrieve the set $\bar{H}$ of **feasible histories** (sequences of profiles of actions, PPMs, and emotional messages). (a^t, m_p^t, m^t) is feasible if:
  1. action profiles are feasible given the last PPM profile;
  2. PPM profiles are generated according to the game unfolding;
  3. emotional message profiles are generated with positive probability for some profile of traits and streams of emotions, given the last action profile played. More details
- $\bar{H}$ is partitioned into the set of **terminal histories** Z and the set of **non-terminal histories** H.
- Note: a given player has information only about the *actions she took* and the *messages she received*. If the history is $(a_j^t, m_{j,p}^t, m_j^t)_{j\in I}$, i "knows"/observes $(a_i^t, m_{i,p}^t, m_i^t)$.
- Set of **personal histories** of player i, $\bar{H}_i$: (feasible) sequences of actions played and messages received by i. Non-terminal personal histories of i: H_i.
- **Material outcomes** accrue to players at the end of the game.

Framework: Game form

- Retrieve the set $\bar{H}$ of **feasible histories** (sequences of profiles of actions, PPMs, and emotional messages). (a^t, m_p^t, m^t) is feasible if:
  1. action profiles are feasible given the last PPM profile;
  2. PPM profiles are generated according to the game unfolding;
  3. emotional message profiles are generated with positive probability for some profile of traits and streams of emotions, given the last action profile played. More details
- $\bar{H}$ is partitioned into the set of **terminal histories** Z and the set of **non-terminal histories** H.
- Note: a given player has information only about the *actions she took* and the *messages she received*. If the history is $(a_j^t, m_{j,p}^t, m_j^t)_{j \in I}$, i "knows"/observes $(a_i^t, m_{i,p}^t, m_i^t)$.
- Set of **personal histories** of player i, $\bar{H}_i$: (feasible) sequences of actions played and messages received by i. Non-terminal personal histories of i: H_i.
- **Material outcomes** accrue to players at the end of the game.

Framework: Game form

- Retrieve the set $\bar{H}$ of **feasible histories** (sequences of profiles of actions, PPMs, and emotional messages). (a^t, m_p^t, m^t) is feasible if:
  1. action profiles are feasible given the last PPM profile;
  2. PPM profiles are generated according to the game unfolding;
  3. emotional message profiles are generated with positive probability for some profile of traits and streams of emotions, given the last action profile played. More details
- $\bar{H}$ is partitioned into the set of **terminal histories** Z and the set of **non-terminal histories** H.
- Note: a given player has information only about the *actions she took* and the *messages she received*. If the history is $(a_j^t, m_{j,p}^t, m_j^t)_{j \in I}$, i "knows"/observes $(a_i^t, m_{i,p}^t, m_i^t)$.
- Set of **personal histories** of player i, $\bar{H}_i$: (feasible) sequences of actions played and messages received by i. Non-terminal personal histories of i: H_i.
- **Material outcomes** accrue to players at the end of the game.

Framework: Game form

- Retrieve the set $\bar{H}$ of **feasible histories** (sequences of profiles of actions, PPMs, and emotional messages). (a^t, m_p^t, m^t) is feasible if:
  1. action profiles are feasible given the last PPM profile;
  2. PPM profiles are generated according to the game unfolding;
  3. emotional message profiles are generated with positive probability for some profile of traits and streams of emotions, given the last action profile played. More details
- $\bar{H}$ is partitioned into the set of **terminal histories** Z and the set of **non-terminal histories** H.
- Note: a given player has information only about the *actions she took* and the *messages she received*. If the history is $(a_j^t, m_{j,p}^t, m_j^t)_{j\in I}$, i "knows"/observes $(a_i^t, m_{i,p}^t, m_i^t)$.
- Set of **personal histories** of player i, $\bar{H}_i$: (feasible) sequences of actions played and messages received by i. Non-terminal personal histories of i: H_i.
- **Material outcomes** accrue to players at the end of the game.

Framework: Game form

- Retrieve the set $\bar{H}$ of **feasible histories** (sequences of profiles of actions, PPMs, and emotional messages). (a^t, m_p^t, m^t) is feasible if:
    1. action profiles are feasible given the last PPM profile;
    2. PPM profiles are generated according to the game unfolding;
    3. emotional message profiles are generated with positive probability for some profile of traits and streams of emotions, given the last action profile played. More details
- $\bar{H}$ is partitioned into the set of **terminal histories** Z and the set of **non-terminal histories** H.
- Note: a given player has information only about the *actions she took* and the *messages she received*. If the history is $(a_j^t, m_{j,p}^t, m_j^t)_{j \in I}$, i "knows"/observes $(a_i^t, m_{i,p}^t, m_i^t)$.
- Set of **personal histories** of player i, $\bar{H}_i$: (feasible) sequences of actions played and messages received by i. Non-terminal personal histories of i: H_i.
- **Material outcomes** accrue to players at the end of the game.

Framework: Game form

- Retrieve the set $\bar{H}$ of **feasible histories** (sequences of profiles of actions, PPMs, and emotional messages). (a^t, m_p^t, m^t) is feasible if:
  1. action profiles are feasible given the last PPM profile;
  2. PPM profiles are generated according to the game unfolding;
  3. emotional message profiles are generated with positive probability for some profile of traits and streams of emotions, given the last action profile played. More details
- $\bar{H}$ is partitioned into the set of **terminal histories** Z and the set of **non-terminal histories** H.
- Note: a given player has information only about the *actions she took* and the *messages she received*. If the history is $(a_j^t, m_{j,p}^t, m_j^t)_{j \in I}$, i "knows"/observes $(a_i^t, m_{i,p}^t, m_i^t)$.
- Set of **personal histories** of player i, $\bar{H}_i$: (feasible) sequences of actions played and messages received by i. Non-terminal personal histories of i: H_i.
- **Material outcomes** accrue to players at the end of the game.

Framework: Application to “buy me an ice-cream”

Feedback: Child blushes (b) only if he says *Yes* after *Video-games*, with probability equal to the probability with which he believes Mom would not get him the ice-cream if he blushes.
Intuition: if he is confident that he can get away with his lie even if spotted by Mom, more likely to keep a poker face.

Histories: the timeline is as follows.

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom does not observe it.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom observes $a_{C,2}$ and m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child observes a_M.

So, a terminal history is $z = (a_{C,1}, a_{C,2}, m, a_M)$, but Mom observes only $z_M = (a_{C,2}, m, a_M)$.

More details

Framework: Application to "buy me an ice-cream"

Feedback: Child blushes (b) only if he says *Yes* after *Video-games*, with probability equal to the probability with which he believes Mom would not get him the ice-cream if he blushes.
Intuition: if he is confident that he can get away with his lie even if spotted by Mom, more likely to keep a poker face.

Histories: the timeline is as follows.

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom does not observe it.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom observes $a_{C,2}$ and m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child observes a_M.

So, a terminal history is $z = (a_{C,1}, a_{C,2}, m, a_M)$, but Mom observes only $z_M = (a_{C,2}, m, a_M)$.

More details

Framework: Application to "buy me an ice-cream"

Feedback: Child blushes (b) only if he says *Yes* after *Video-games*, with probability equal to the probability with which he believes Mom would not get him the ice-cream if he blushes.
Intuition: if he is confident that he can get away with his lie even if spotted by Mom, more likely to keep a poker face.

Histories: the timeline is as follows.

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom does not observe it.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom observes $a_{C,2}$ and m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child observes a_M.

So, a terminal history is $z = (a_{C,1}, a_{C,2}, m, a_M)$, but Mom observes only $z_M = (a_{C,2}, m, a_M)$.

More details

Framework: Application to “buy me an ice-cream”

Feedback: Child blushes (b) only if he says *Yes* after *Video-games*, with probability equal to the probability with which he believes Mom would not get him the ice-cream if he blushes.
Intuition: if he is confident that he can get away with his lie even if spotted by Mom, more likely to keep a poker face.

Histories: the timeline is as follows.

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom does not observe it.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom observes $a_{C,2}$ and m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child observes a_M.

So, a terminal history is $z = (a_{C,1}, a_{C,2}, m, a_M)$, but Mom observes only $z_M = (a_{C,2}, m, a_M)$.

More details

Framework: Application to “buy me an ice-cream”

Feedback: Child blushes (b) only if he says *Yes* after *Video-games*, with probability equal to the probability with which he believes Mom would not get him the ice-cream if he blushes.
Intuition: if he is confident that he can get away with his lie even if spotted by Mom, more likely to keep a poker face.

Histories: the timeline is as follows.

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom does not observe it.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom observes $a_{C,2}$ and m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child observes a_M.

So, a terminal history is $z = (a_{C,1}, a_{C,2}, m, a_M)$, but Mom observes only $z_M = (a_{C,2}, m, a_M)$.

More details

Framework: Application to "buy me an ice-cream"

Feedback: Child blushes (b) only if he says *Yes* after *Video-games*, with probability equal to the probability with which he believes Mom would not get him the ice-cream if he blushes.
Intuition: if he is confident that he can get away with his lie even if spotted by Mom, more likely to keep a poker face.

Histories: the timeline is as follows.

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom does not observe it.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom observes $a_{C,2}$ and m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child observes a_M.

So, a terminal history is $z = (a_{C,1}, a_{C,2}, m, a_M)$, but Mom observes only $z_M = (a_{C,2}, m, a_M)$.

More details

Framework: Application to “buy me an ice-cream”

Feedback: Child blushes (b) only if he says *Yes* after *Video-games*, with probability equal to the probability with which he believes Mom would not get him the ice-cream if he blushes.
Intuition: if he is confident that he can get away with his lie even if spotted by Mom, more likely to keep a poker face.

Histories: the timeline is as follows.

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom does not observe it.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom observes $a_{C,2}$ and m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child observes a_M.

So, a terminal history is $z = (a_{C,1}, a_{C,2}, m, a_M)$, but Mom observes only $z_M = (a_{C,2}, m, a_M)$.

More details

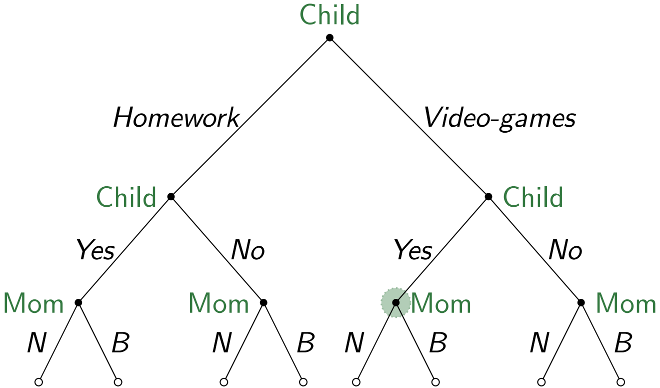
Framework: Application to “buy me an ice-cream”

We obtain the following set of “plays” (sequences of actions):

The unique play after which Child *may* blush is highlighted in green.

Framework: Application to "buy me an ice-cream"

We obtain the following set of "plays" (sequences of actions):



The unique play after which Child *may* blush is highlighted in green.

Framework: Describing behavior

- Set of **personal external states** of i: $S_i := \times_{h_i \in H_i} A_i(h_i)$, with $A_i(h_i)$ set of actions available at h_i.
- Interpretation: $s_i \in S_i$ is a *complete objective description of how player i would behave* at each possible game-specific contingency (i.e., personal history she may observe).
- Note: $s_i \in S_i$ is technically a *strategy* of player i. We avoid this terminology because we will take strategies to be *plans in the minds of players* (hence, they will be part of their ways of thinking).

Framework: Describing behavior

- Set of **personal external states** of i: $S_i := \times_{h_i \in H_i} A_i(h_i)$, with $A_i(h_i)$ set of actions available at h_i.
- Interpretation: $s_i \in S_i$ is a *complete objective description of how player i would behave* at each possible game-specific contingency (i.e., personal history she may observe).
- Note: $s_i \in S_i$ is technically a *strategy* of player i. We avoid this terminology because we will take strategies to be *plans in the minds of players* (hence, they will be part of their ways of thinking).

Framework: Describing behavior

- Set of **personal external states** of i: $S_i := \times_{h_i \in H_i} A_i(h_i)$, with $A_i(h_i)$ set of actions available at h_i.
- Interpretation: $s_i \in S_i$ is a *complete objective description of how player i would behave* at each possible game-specific contingency (i.e., personal history she may observe).
- Note: $s_i \in S_i$ is technically a *strategy* of player i. We avoid this terminology because we will take strategies to be *plans in the minds of players* (hence, they will be part of their ways of thinking).

Framework: Describing behavior

- Set of **personal external states** of i: $S_i := \times_{h_i \in H_i} A_i(h_i)$, with $A_i(h_i)$ set of actions available at h_i.
- Interpretation: $s_i \in S_i$ is a *complete objective description of how player i would behave* at each possible game-specific contingency (i.e., personal history she may observe).
- Note: $s_i \in S_i$ is technically a *strategy* of player i. We avoid this terminology because we will take strategies to be *plans in the minds of players* (hence, they will be part of their ways of thinking).

Framework: Describing ways of thinking

- We want to build **hierarchical systems of beliefs**: maps from personal histories to infinite hierarchies of beliefs.
- Interpretation: one such system is a *complete objective description* of what a player *would think* at each possible contingency.
- Space of primitive uncertainty: $\Omega^0_{-i} := S \times \Theta_{-i}$ (behavior of *everyone* and traits of others). A system of first-order beliefs of i ($\tau_{i,1}$) is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i) \in \Delta(\Omega^0_{-i})$.
- Note: we allow players to form beliefs on *their own behavior*. Why? Players cannot know how they will behave in the future (because they are not committed), and they may also be uncertain about it. Beliefs about how they would act form a **plan**.
- Hierarchies of beliefs defined inductively. $\mathcal{T}_i^n$ is the set of hierarchical systems of beliefs of i of order $n \in \mathbb{N}$. τ_i^n is a map $h_i \mapsto \left(\tau_{i,k}(\cdot|h_i)\right)_{k=1}^n$, with $\tau_{i,k}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{k-1})$.
- $\mathcal{T}_i^\infty$ is the set of **epistemic types** of i.

More details

Framework: Describing ways of thinking

- We want to build **hierarchical systems of beliefs**: maps from personal histories to infinite hierarchies of beliefs.
- Interpretation: one such system is a *complete objective description* of what a player *would think* at each possible contingency.
- Space of primitive uncertainty: $\Omega^0_{-i} := S \times \Theta_{-i}$ (behavior of *everyone* and traits of others). A system of first-order beliefs of i ($\tau_{i,1}$) is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i) \in \Delta(\Omega^0_{-i})$.
- Note: we allow players to form beliefs on *their own behavior*. Why? Players cannot know how they will behave in the future (because they are not committed), and they may also be uncertain about it. Beliefs about how they would act form a **plan**.
- Hierarchies of beliefs defined inductively. $\mathcal{T}_i^n$ is the set of hierarchical systems of beliefs of i of order $n \in \mathbb{N}$. τ_i^n is a map $h_i \mapsto \left(\tau_{i,k}(\cdot|h_i)\right)_{k=1}^n$, with $\tau_{i,k}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{k-1})$.
- $\mathcal{T}_i^\infty$ is the set of **epistemic types** of i.

More details

Framework: Describing ways of thinking

- We want to build **hierarchical systems of beliefs**: maps from personal histories to infinite hierarchies of beliefs.
- Interpretation: one such system is a *complete objective description* of what a player *would think* at each possible contingency.
- Space of primitive uncertainty: $\Omega^0_{-i} := S \times \Theta_{-i}$ (behavior of *everyone* and traits of others). A system of first-order beliefs of i ($\tau_{i,1}$) is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i) \in \Delta(\Omega^0_{-i})$.
- Note: we allow players to form beliefs on *their own behavior*. Why? Players cannot know how they will behave in the future (because they are not committed), and they may also be uncertain about it. Beliefs about how they would act form a **plan**.
- Hierarchies of beliefs defined inductively. $\mathcal{T}_i^n$ is the set of hierarchical systems of beliefs of i of order $n \in \mathbb{N}$. τ_i^n is a map $h_i \mapsto \left(\tau_{i,k}(\cdot|h_i)\right)_{k=1}^n$, with $\tau_{i,k}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{k-1})$.
- $\mathcal{T}_i^\infty$ is the set of **epistemic types** of i.

More details

Framework: Describing ways of thinking

- We want to build **hierarchical systems of beliefs**: maps from personal histories to infinite hierarchies of beliefs.
- Interpretation: one such system is a *complete objective description* of what a player *would think* at each possible contingency.
- Space of primitive uncertainty: $\Omega^0_{-i} := S \times \Theta_{-i}$ (behavior of *everyone* and traits of others). A system of first-order beliefs of i ($\tau_{i,1}$) is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i) \in \Delta(\Omega^0_{-i})$.
- Note: we allow players to form beliefs on *their own behavior*. Why? Players cannot know how they will behave in the future (because they are not committed), and they may also be uncertain about it. Beliefs about how they would act form a **plan**.
- Hierarchies of beliefs defined inductively. $\mathcal{T}_i^n$ is the set of hierarchical systems of beliefs of i of order $n \in \mathbb{N}$. τ_i^n is a map $h_i \mapsto \left(\tau_{i,k}(\cdot|h_i)\right)_{k=1}^n$, with $\tau_{i,k}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{k-1})$.
- $\mathcal{T}_i^\infty$ is the set of **epistemic types** of i.

More details

Framework: Describing ways of thinking

- We want to build **hierarchical systems of beliefs**: maps from personal histories to infinite hierarchies of beliefs.
- Interpretation: one such system is a *complete objective description* of what a player *would think* at each possible contingency.
- Space of primitive uncertainty: $\Omega^0_{-i} := S \times \Theta_{-i}$ (behavior of *everyone* and traits of others). A system of first-order beliefs of i ($\tau_{i,1}$) is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i) \in \Delta(\Omega^0_{-i})$.
- Note: we allow players to form beliefs on *their own behavior*. Why? Players cannot know how they will behave in the future (because they are not committed), and they may also be uncertain about it. Beliefs about how they would act form a **plan**.
- Hierarchies of beliefs defined inductively. $\mathcal{T}_i^n$ is the set of hierarchical systems of beliefs of i of order $n \in \mathbb{N}$. τ_i^n is a map $h_i \mapsto \left(\tau_{i,k}(\cdot|h_i)\right)_{k=1}^n$, with $\tau_{i,k}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{k-1})$.
- $\mathcal{T}_i^\infty$ is the set of **epistemic types** of i.

More details

Framework: Describing ways of thinking

- We want to build **hierarchical systems of beliefs**: maps from personal histories to infinite hierarchies of beliefs.
- Interpretation: one such system is a *complete objective description* of what a player *would think* at each possible contingency.
- Space of primitive uncertainty: $\Omega^0_{-i} := S \times \Theta_{-i}$ (behavior of *everyone* and traits of others). A system of first-order beliefs of i ($\tau_{i,1}$) is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i) \in \Delta(\Omega^0_{-i})$.
- Note: we allow players to form beliefs on *their own behavior*. Why? Players cannot know how they will behave in the future (because they are not committed), and they may also be uncertain about it. Beliefs about how they would act form a **plan**.
- Hierarchies of beliefs defined inductively. $\mathcal{T}_i^n$ is the set of hierarchical systems of beliefs of i of order $n \in \mathbb{N}$. τ_i^n is a map $h_i \mapsto \left(\tau_{i,k}(\cdot|h_i)\right)_{k=1}^n$, with $\tau_{i,k}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{k-1})$.
- $\mathcal{T}_i^\infty$ is the set of **epistemic types** of i.

More details

Framework: Describing ways of thinking

- We want to build **hierarchical systems of beliefs**: maps from personal histories to infinite hierarchies of beliefs.
- Interpretation: one such system is a *complete objective description* of what a player *would think* at each possible contingency.
- Space of primitive uncertainty: $\Omega^0_{-i} := S \times \Theta_{-i}$ (behavior of *everyone* and traits of others). A system of first-order beliefs of i ($\tau_{i,1}$) is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i) \in \Delta(\Omega^0_{-i})$.
- Note: we allow players to form beliefs on *their own behavior*. Why? Players cannot know how they will behave in the future (because they are not committed), and they may also be uncertain about it. Beliefs about how they would act form a **plan**.
- Hierarchies of beliefs defined inductively. $\mathcal{T}_i^n$ is the set of hierarchical systems of beliefs of i of order $n \in \mathbb{N}$. τ_i^n is a map $h_i \mapsto \left(\tau_{i,k}(\cdot|h_i)\right)_{k=1}^n$, with $\tau_{i,k}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{k-1})$.
- $\mathcal{T}_i^\infty$ is the set of **epistemic types** of i.

More details

Framework: States of the world

- Set of **states of the world**: $\Omega^\infty := \times_{i\in I}\left(S_i \times \Theta_i \times \mathcal{T}_i^\infty\right)$.
- Interpretation: a profile $(s_i, \theta_i, \tau_i^\infty)_{i\in I}$ describes *every relevant aspect of strategic interaction*.
- **Events** are Borel measurable subsets of Ω^∞.
- Interpretation of measurability: measurable sets are those players can *conceive*, *assess*, and *form beliefs about*. Hence, *players can reason only about events*.

Framework: States of the world

- Set of **states of the world**: $\Omega^\infty := \bigtimes_{i\in I} \left(S_i \times \Theta_i \times \mathcal{T}_i^\infty \right)$.
- Interpretation: a profile $(s_i, \theta_i, \tau_i^\infty)_{i\in I}$ describes *every relevant aspect of strategic interaction*.
- **Events** are Borel measurable subsets of Ω^∞.
- Interpretation of measurability: measurable sets are those players can *conceive*, *assess*, and *form beliefs about*. Hence, *players can reason only about events*.

Framework: States of the world

- Set of **states of the world**: $\Omega^\infty := \bigtimes_{i\in I}\left(S_i \times \Theta_i \times \mathcal{T}_i^\infty\right)$.
- Interpretation: a profile $(s_i, \theta_i, \tau_i^\infty)_{i\in I}$ describes *every relevant aspect of strategic interaction*.
- **Events** are Borel measurable subsets of Ω^∞.
- Interpretation of measurability: measurable sets are those players can *conceive*, *assess*, and *form beliefs about*. Hence, *players can reason only about events*.

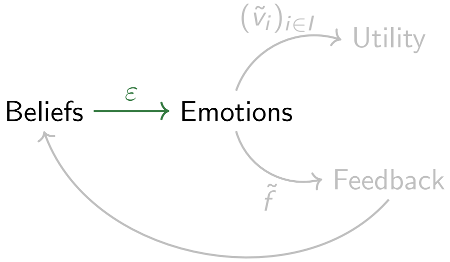
Framework: Emotions and game unfolding

Continuous **emotion-generating function** $\varepsilon : \bar{H} \times \mathcal{T}^\infty \to \Delta(E^{\leq T+1})$.

- Interpretation: *what happened*, and players' *ways of thinking* determine the emotions they experience (e.g., surprise, frustration).
- Assumption: only *realized* beliefs matter (not counterfactual ones).
- Assumption: only beliefs of order *up to* $K \in \mathbb{N}$ matter. So, we can focus on $S \times \Theta \times \mathcal{T}^K$ (**utility-relevant states**, drop K for simplicity).

More details

Framework: Emotions and game unfolding

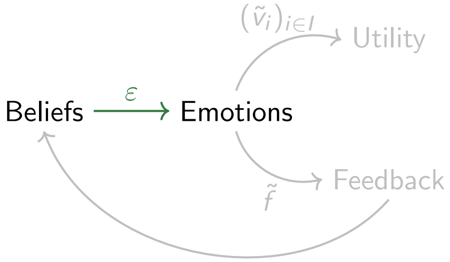


Continuous **emotion-generating function** $\varepsilon : \bar{H} \times \mathcal{T}^\infty \to \Delta(E^{\leq T+1})$.

- Interpretation: *what happened*, and players' *ways of thinking* determine the emotions they experience (e.g., surprise, frustration).
- Assumption: only *realized* beliefs matter (not counterfactual ones).
- Assumption: only beliefs of order *up to* $K \in \mathbb{N}$ matter. So, we can focus on $S \times \Theta \times \mathcal{T}^K$ (**utility-relevant states**, drop K for simplicity).

More details

Framework: Emotions and game unfolding

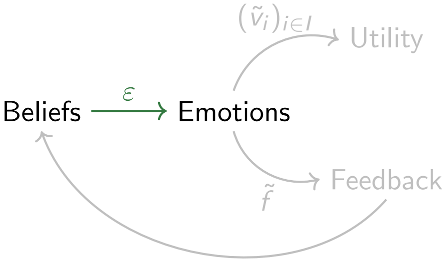


Continuous **emotion-generating function** $\varepsilon : \bar{H} \times \mathcal{T}^\infty \to \Delta(E^{\leq T+1})$.

- Interpretation: *what happened*, and players' *ways of thinking* determine the emotions they experience (e.g., surprise, frustration).
- Assumption: only *realized* beliefs matter (not counterfactual ones).
- Assumption: only beliefs of order *up to* $K \in \mathbb{N}$ matter. So, we can focus on $S \times \Theta \times \mathcal{T}^K$ (**utility-relevant states**, drop K for simplicity).

More details

Framework: Emotions and game unfolding

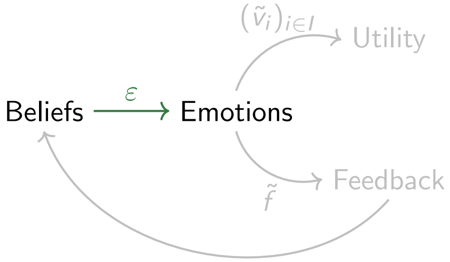


Continuous **emotion-generating function** $\varepsilon : \bar{H} \times \mathcal{T}^{\infty} \to \Delta(E^{\leq T+1})$.

- Interpretation: *what happened*, and players' *ways of thinking* determine the emotions they experience (e.g., surprise, frustration).
- Assumption: only *realized* beliefs matter (not counterfactual ones).
- Assumption: only beliefs of order *up to* $K \in \mathbb{N}$ matter. So, we can focus on $S \times \Theta \times \mathcal{T}^K$ (**utility-relevant states**, drop K for simplicity).

More details

Framework: Emotions and game unfolding

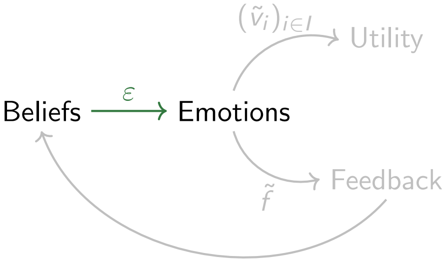


Continuous **emotion-generating function** $\varepsilon : \bar{H} \times \mathcal{T}^\infty \to \Delta(E^{\leq T+1})$.

- Interpretation: *what happened*, and players' *ways of thinking* determine the emotions they experience (e.g., surprise, frustration).
- Assumption: only *realized* beliefs matter (not counterfactual ones).
- Assumption: only beliefs of order *up to* $K \in \mathbb{N}$ matter. So, we can focus on $S \times \Theta \times \mathcal{T}^K$ (**utility-relevant states**, drop K for simplicity).

More details

Framework: Emotions and game unfolding



Continuous **emotion-generating function** $\varepsilon : \bar{H} \times \mathcal{T}^{\infty} \to \Delta(E^{\leq T+1})$.

- Interpretation: *what happened*, and players' *ways of thinking* determine the emotions they experience (e.g., surprise, frustration).
- Assumption: only *realized* beliefs matter (not counterfactual ones).
- Assumption: only beliefs of order *up to* $K \in \mathbb{N}$ matter. So, we can focus on $S \times \Theta \times \mathcal{T}^K$ (**utility-relevant states**, drop K for simplicity).

More details

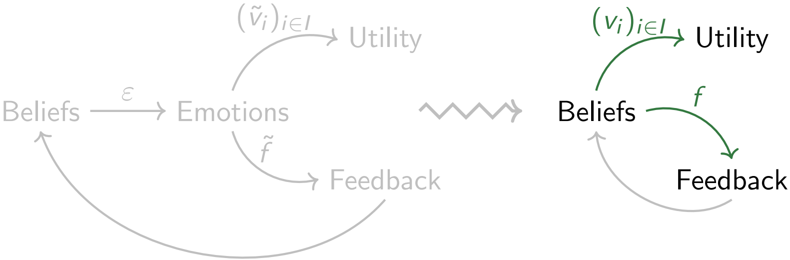
Framework: Reduced form representation

We can derive a "reduced form" representation of feedback and utilities:

- (Continuous) **game-dependent feedback functions** $f := (f_h : S \times \Theta \times \mathcal{T} \to \Delta(M))_{h\in H}$. Interpretation: $f_h(s,\theta,\tau)[m]$ is the probability of m conditional on (s,θ,τ) and given h.
  More details
- (Continuous) **game-dependent psychological utility functions** $(v_i : Z \times \Theta \times \mathcal{T} \to \mathbb{R})_{i\in I}$. More details

Framework: Reduced form representation

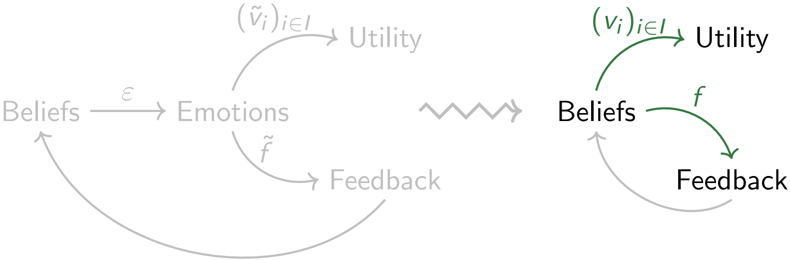
We can derive a "reduced form" representation of feedback and utilities:



- (Continuous) **game-dependent feedback functions** $f := (f_h : S \times \Theta \times \mathcal{T} \to \Delta(M))_{h\in H}$.
  Interpretation: $f_h(s, \theta, \tau)[m]$ is the probability of m conditional on (s, θ, τ) and given h.
  More details
- (Continuous) **game-dependent psychological utility functions** $(v_i : Z \times \Theta \times \mathcal{T} \to \mathbb{R})_{i\in I}$. More details

Framework: Reduced form representation

We can derive a “reduced form” representation of feedback and utilities:



- (Continuous) **game-dependent feedback functions** $f := (f_h : S \times \Theta \times \mathcal{T} \to \Delta(M))_{h \in H}$.
  Interpretation: $f_h(s, \theta, \tau)[m]$ is the probability of m conditional on (s, θ, τ) and given h.
  More details
- (Continuous) **game-dependent psychological utility functions** $(v_i : Z \times \Theta \times \mathcal{T} \to \mathbb{R})_{i \in I}$. More details

Framework: Application to “buy me an ice-cream”

Child’s utility: $\mathbf{1}(I, z)$ and $\mathbf{1}(V, z)$ indicator functions for getting the ice-cream and playing video-games during terminal history z. $L = \{Homework.No, Video\text{-}games.Yes\}$ set of “lies”.

$$v_C(z, \theta, \tau) = \underbrace{\mathbf{1}(I, z)}_{\text{likes ice-cream}} + \overbrace{\theta_C}^{\text{video-games appreciation}} \cdot \underbrace{\mathbf{1}(V, z)}_{\text{likes video-games}} - \overbrace{\tau_M(L|z_M)}^{\text{image concern}}.$$

Mom’s utility: $\mathbf{1}(H, z)$ indicator function for doing *Homework* during z.

$$v_M(z, \theta, \tau) = \begin{cases} \underbrace{2 \cdot \mathbf{1}(H, z)}_{\text{reward Child}} - \underbrace{1}_{\text{ice-cream cost}} & \text{if she } Buys; \\ 0 & \text{if } Not. \end{cases}$$

Framework: Application to “buy me an ice-cream”

Child’s utility: $\mathbf{1}(I,z)$ and $\mathbf{1}(V,z)$ indicator functions for getting the ice-cream and playing video-games during terminal history z. $L = \{Homework.No, Video\text{-}games.Yes\}$ set of “lies”.

$$v_{\mathrm{C}}(z,\theta,\tau) = \underbrace{\mathbf{1}(I,z)}_{\text{likes ice-cream}} + \overbrace{\theta_{\mathrm{C}}}^{\text{video-games appreciation}} \cdot \underbrace{\mathbf{1}(V,z)}_{\text{likes video-games}} - \overbrace{\tau_{\mathrm{M}}(L|z_{\mathrm{M}})}^{\text{image concern}}.$$

Mom’s utility: $\mathbf{1}(H,z)$ indicator function for doing *Homework* during z.

$$v_{\mathrm{M}}(z,\theta,\tau) = \begin{cases} \underbrace{2\cdot\mathbf{1}(H,z)}_{\text{reward Child}} - \underbrace{1}_{\text{ice-cream cost}} & \text{if she } Buys; \\ 0 & \text{if } Not. \end{cases}$$

Framework: Application to "buy me an ice-cream"

Child's utility: $\mathbf{1}(I, z)$ and $\mathbf{1}(V, z)$ indicator functions for getting the ice-cream and playing video-games during terminal history z. $L = \{Homework.No, Video\text{-}games.Yes\}$ set of "lies".

$$v_C(z, \theta, \tau) = \underbrace{\mathbf{1}(I, z)}_{\text{likes ice-cream}} + \overbrace{\theta_C}^{\text{video-games appreciation}} \cdot \underbrace{\mathbf{1}(V, z)}_{\text{likes video-games}} - \overbrace{\tau_M(L|z_M)}^{\text{image concern}}.$$

Mom's utility: $\mathbf{1}(H, z)$ indicator function for doing *Homework* during z.

$$v_M(z, \theta, \tau) = \begin{cases} \underbrace{2 \cdot \mathbf{1}(H, z)}_{\text{reward Child}} - \underbrace{1}_{\text{ice-cream cost}} & \text{if she } Buys; \\ 0 & \text{if } Not. \end{cases}$$

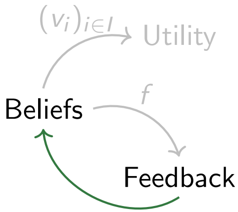
What's next?

1. Definition of a **general framework**
2. Description of **inferences**
3. Formalization of a notion of **rationality**
4. Definition of a **solution concept** to derive predictions
5. Concluding remarks

Inferences: Warm-up

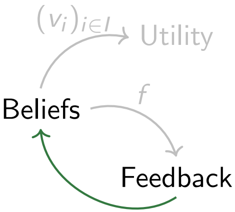
- Players should use of the signals they observe to make inferences about others' behavior or ways of thinking.
- In buy me an ice-cream, Child may *blush* only when he lies saying *Yes*.
- Mom wants to reward Child only if he has done his homework. Should she use the emotional signal to decide what to do?
- Yes! If Child tells her *Yes* and he blushes, Mom should conclude he is lying: better not to buy him the ice-cream in this case.

Inferences: Warm-up



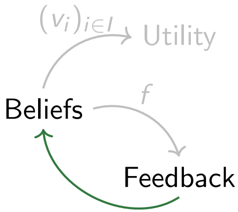
- Players should use of the signals they observe to make inferences about others' behavior or ways of thinking.
- In buy me an ice-cream, Child may *blush* only when he lies saying *Yes*.
- Mom wants to reward Child only if he has done his homework. Should she use the emotional signal to decide what to do?
- Yes! If Child tells her *Yes* and he blushes, Mom should conclude he is lying: better not to buy him the ice-cream in this case.

Inferences: Warm-up



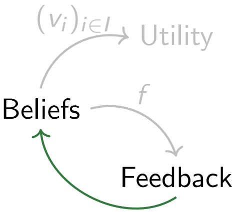
- Players should use of the signals they observe to make inferences about others' behavior or ways of thinking.
- In buy me an ice-cream, Child may *blush* only when he lies saying *Yes*.
- Mom wants to reward Child only if he has done his homework. Should she use the emotional signal to decide what to do?
- Yes! If Child tells her *Yes* and he blushes, Mom should conclude he is lying: better not to buy him the ice-cream in this case.

Inferences: Warm-up



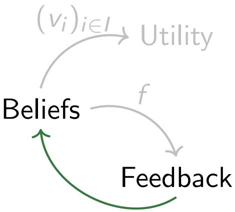
- Players should use of the signals they observe to make inferences about others' behavior or ways of thinking.
- In buy me an ice-cream, Child may *blush* only when he lies saying *Yes*.
- Mom wants to reward Child only if he has done his homework. Should she use the emotional signal to decide what to do?
- Yes! If Child tells her *Yes* and he blushes, Mom should conclude he is lying: better not to buy him the ice-cream in this case.

Inferences: Warm-up



- Players should use of the signals they observe to make inferences about others' behavior or ways of thinking.
- In buy me an ice-cream, Child may *blush* only when he lies saying *Yes*.
- Mom wants to reward Child only if he has done his homework. Should she use the emotional signal to decide what to do?
- Yes! If Child tells her *Yes* and he blushes, Mom should conclude he is lying: better not to buy him the ice-cream in this case.

Inferences: Warm-up



- Players should use of the signals they observe to make inferences about others' behavior or ways of thinking.
- In buy me an ice-cream, Child may *blush* only when he lies saying *Yes*.
- Mom wants to reward Child only if he has done his homework. Should she use the emotional signal to decide what to do?
- Yes! If Child tells her *Yes* and he blushes, Mom should conclude he is lying: better not to buy him the ice-cream in this case.

Inferences: Reasoning about the game unfolding

- Fixing a utility-relevant state (s, θ, τ), several histories may occur: emotional messages are stochastic, and behavior may depend on their realizations. $\mathcal{H}_i(s, \theta, \tau) \subseteq \bar{H}_i$ is the set of personal histories of i that may realize given (s, θ, τ).
- We informally assume players know their beliefs and traits (i.e., i "knows" θ_i and τ_i): so, player i will make inferences about elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$.
- When player i observes h_i (recall, she knows θ_i and τ_i), she **infers** that the set of elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$ consistent with evidence is: $\mathcal{I}_{(\theta_i, \tau_i)}(h_i) := \left\{(s, \theta_{-i}, \tau_{-i}) : h_i \in \mathcal{H}_i(s, (\theta_i, \theta_{-i}), (\tau_i, \tau_{-i}))\right\}$.
- Under regularity conditions, inferences are "well-behaved" (i.e., for each i, h_i, τ_i, θ_i, $\mathcal{I}_{(\theta_i, \tau_i)}(h_i)$ is measurable).

More details

Inferences: Reasoning about the game unfolding

- Fixing a utility-relevant state (s, θ, τ), several histories may occur: emotional messages are stochastic, and behavior may depend on their realizations. $\mathcal{H}_i(s, \theta, \tau) \subseteq \bar{H}_i$ is the set of personal histories of i that may realize given (s, θ, τ).
- We informally assume players know their beliefs and traits (i.e., i "knows" θ_i and τ_i): so, player i will make inferences about elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$.
- When player i observes h_i (recall, she knows θ_i and τ_i), she **infers** that the set of elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$ consistent with evidence is: $\mathcal{I}_{(\theta_i, \tau_i)}(h_i) := \left\{(s, \theta_{-i}, \tau_{-i}) : h_i \in \mathcal{H}_i(s, (\theta_i, \theta_{-i}), (\tau_i, \tau_{-i}))\right\}$.
- Under regularity conditions, inferences are "well-behaved" (i.e., for each i, h_i, τ_i, θ_i, $\mathcal{I}_{(\theta_i, \tau_i)}(h_i)$ is measurable).

More details

Inferences: Reasoning about the game unfolding

- Fixing a utility-relevant state (s, θ, τ), several histories may occur: emotional messages are stochastic, and behavior may depend on their realizations. $\mathcal{H}_i(s, \theta, \tau) \subseteq \bar{H}_i$ is the set of personal histories of i that may realize given (s, θ, τ).
- We informally assume players know their beliefs and traits (i.e., i "knows" θ_i and τ_i): so, player i will make inferences about elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$.
- When player i observes h_i (recall, she knows θ_i and τ_i), she **infers** that the set of elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$ consistent with evidence is: $\mathcal{I}_{(\theta_i, \tau_i)}(h_i) := \left\{(s, \theta_{-i}, \tau_{-i}) : h_i \in \mathcal{H}_i(s, (\theta_i, \theta_{-i}), (\tau_i, \tau_{-i}))\right\}$.
- Under regularity conditions, inferences are "well-behaved" (i.e., for each i, h_i, τ_i, θ_i, $\mathcal{I}_{(\theta_i, \tau_i)}(h_i)$ is measurable).

More details

Inferences: Reasoning about the game unfolding

- Fixing a utility-relevant state (s, θ, τ), several histories may occur: emotional messages are stochastic, and behavior may depend on their realizations. $\mathcal{H}_i(s, \theta, \tau) \subseteq \bar{H}_i$ is the set of personal histories of i that may realize given (s, θ, τ).
- We informally assume players know their beliefs and traits (i.e., i "knows" θ_i and τ_i): so, player i will make inferences about elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$.
- When player i observes h_i (recall, she knows θ_i and τ_i), she **infers** that the set of elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$ consistent with evidence is: $\mathcal{I}_{(\theta_i, \tau_i)}(h_i) := \big\{(s, \theta_{-i}, \tau_{-i}) : h_i \in \mathcal{H}_i(s, (\theta_i, \theta_{-i}), (\tau_i, \tau_{-i}))\big\}$.
- Under regularity conditions, inferences are "well-behaved" (i.e., for each i, h_i, τ_i, θ_i, $\mathcal{I}_{(\theta_i, \tau_i)}(h_i)$ is measurable).

More details

Inferences: Reasoning about the game unfolding

- Fixing a utility-relevant state (s, θ, τ), several histories may occur: emotional messages are stochastic, and behavior may depend on their realizations. $\mathcal{H}_i(s, \theta, \tau) \subseteq \bar{H}_i$ is the set of personal histories of i that may realize given (s, θ, τ).
- We informally assume players know their beliefs and traits (i.e., i "knows" θ_i and τ_i): so, player i will make inferences about elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$.
- When player i observes h_i (recall, she knows θ_i and τ_i), she **infers** that the set of elements of $S \times \Theta_{-i} \times \mathcal{T}_{-i}$ consistent with evidence is: $\mathcal{I}_{(\theta_i, \tau_i)}(h_i) := \left\{(s, \theta_{-i}, \tau_{-i}) : h_i \in \mathcal{H}_i(s, (\theta_i, \theta_{-i}), (\tau_i, \tau_{-i}))\right\}$.
- Under regularity conditions, inferences are "well-behaved" (i.e., for each i, h_i, τ_i, θ_i, $\mathcal{I}_{(\theta_i, \tau_i)}(h_i)$ is measurable).

More details

What's next?

1. Definition of a **general framework**
2. Description of **inferences**
3. Formalization of a notion of **rationality**
4. Definition of a **solution concept** to derive predictions
5. Concluding remarks

Rationality: Warm-up

Studying rationality we want to answer questions like the following ones (in the context of buy me an ice-cream):

- When Mom sees Child's blushing she should infer that he is lying. But assume he says *Yes* without blushing: is he telling the truth or did he manage to keep a poker face? Is there a reasonable way for Mom to update her beliefs?
- If Child did his *Homework*, he can answer *Yes* without blushing afterwards, because he would be telling the truth. Then, does it make sense for him to plan do say *No* after *Homework*?

Rationality: Warm-up

Studying rationality we want to answer questions like the following ones (in the context of buy me an ice-cream):

- When Mom sees Child's blushing she should infer that he is lying. But assume he says *Yes* without blushing: is he telling the truth or did he manage to keep a poker face? Is there a reasonable way for Mom to update her beliefs?
- If Child did his *Homework*, he can answer *Yes* without blushing afterwards, because he would be telling the truth. Then, does it make sense for him to plan do say *No* after *Homework*?

Rationality: Warm-up

Studying rationality we want to answer questions like the following ones (in the context of buy me an ice-cream):

- When Mom sees Child's blushing she should infer that he is lying. But assume he says *Yes* without blushing: is he telling the truth or did he manage to keep a poker face? Is there a reasonable way for Mom to update her beliefs?
- If Child did his *Homework*, he can answer *Yes* without blushing afterwards, because he would be telling the truth. Then, does it make sense for him to plan do say *No* after *Homework*?

Rationality: Warm-up

Studying rationality we want to answer questions like the following ones (in the context of buy me an ice-cream):

- When Mom sees Child's blushing she should infer that he is lying. But assume he says *Yes* without blushing: is he telling the truth or did he manage to keep a poker face? Is there a reasonable way for Mom to update her beliefs?
- If Child did his *Homework*, he can answer *Yes* without blushing afterwards, because he would be telling the truth. Then, does it make sense for him to plan do say *No* after *Homework*?

Rationality: Cognitive side

Rationality is defined as the conjunction of several cognitive and behavioral features.

Cognitive rationality

- Coherence (C): beliefs of different orders are coherent (recall that we work with infinite hierarchies of beliefs, τ_i^∞).
- Believe-what-you-observe (BO): beliefs over utility-relevant states (s, θ, τ) assign probability 1 to states consistent with evidence.
- Correct belief updating (CBU): at each stage, players receive two pieces of information; a_i about their *own behavior*, and $(m_{i,p}, m_i)$ about *others*' behavior and ways of thinking. CBU holds if these pieces of information are used to update beliefs about self and others "correctly" (following the rules of conditional probabilities).

More details

Rationality: Cognitive side

Rationality is defined as the conjunction of several cognitive and behavioral features.

Cognitive rationality

- Coherence (C): beliefs of different orders are coherent (recall that we work with infinite hierarchies of beliefs, τ_i^∞).
- Believe-what-you-observe (BO): beliefs over utility-relevant states (s, θ, τ) assign probability 1 to states consistent with evidence.
- Correct belief updating (CBU): at each stage, players receive two pieces of information; a_i about their *own behavior*, and $(m_{i,p}, m_i)$ about *others*' behavior and ways of thinking. CBU holds if these pieces of information are used to update beliefs about self and others "correctly" (following the rules of conditional probabilities).

More details

Rationality: Cognitive side

Rationality is defined as the conjunction of several cognitive and behavioral features.

Cognitive rationality

- Coherence (C): beliefs of different orders are coherent (recall that we work with infinite hierarchies of beliefs, τ_i^∞).
- Believe-what-you-observe (BO): beliefs over utility-relevant states (s, θ, τ) assign probability 1 to states consistent with evidence.
- Correct belief updating (CBU): at each stage, players receive two pieces of information; a_i about their *own behavior*, and $(m_{i,p}, m_i)$ about *others*' behavior and ways of thinking. CBU holds if these pieces of information are used to update beliefs about self and others "correctly" (following the rules of conditional probabilities).

More details

Rationality: Cognitive side

Rationality is defined as the conjunction of several cognitive and behavioral features.

Cognitive rationality

- Coherence (C): beliefs of different orders are coherent (recall that we work with infinite hierarchies of beliefs, τ_i^∞).
- Believe-what-you-observe (BO): beliefs over utility-relevant states (s, θ, τ) assign probability 1 to states consistent with evidence.
- Correct belief updating (CBU): at each stage, players receive two pieces of information; a_i about their *own behavior*, and $(m_{i,p}, m_i)$ about *others*' behavior and ways of thinking. CBU holds if these pieces of information are used to update beliefs about self and others "correctly" (following the rules of conditional probabilities).

More details

Rationality: Cognitive side

Rationality is defined as the conjunction of several cognitive and behavioral features.

Cognitive rationality

- Coherence (C): beliefs of different orders are coherent (recall that we work with infinite hierarchies of beliefs, τ_i^∞).
- Believe-what-you-observe (BO): beliefs over utility-relevant states (s, θ, τ) assign probability 1 to states consistent with evidence.
- Correct belief updating (CBU): at each stage, players receive two pieces of information; a_i about their *own behavior*, and $(m_{i,p}, m_i)$ about *others*' behavior and ways of thinking. CBU holds if these pieces of information are used to update beliefs about self and others "correctly" (following the rules of conditional probabilities).

More details

Rationality: Cognitive side

Rationality is defined as the conjunction of several cognitive and behavioral features.

Cognitive rationality

- Coherence (C): beliefs of different orders are coherent (recall that we work with infinite hierarchies of beliefs, τ_i^∞).
- Believe-what-you-observe (BO): beliefs over utility-relevant states (s, θ, τ) assign probability 1 to states consistent with evidence.
- Correct belief updating (CBU): at each stage, players receive two pieces of information; a_i about their *own behavior*, and $(m_{i,p}, m_i)$ about *others*' behavior and ways of thinking. CBU holds if these pieces of information are used to update beliefs about self and others "correctly" (following the rules of conditional probabilities).

More details

Rationality: Behavioral side

Plans and behavior

- Rational planning (RP): player *plans* to choose "optimal actions" at each contingency. Here, a plan is a player's belief on how she would behave: RP means that a player expects herself to act in a way that satisfies an "intra-personal equilibrium" condition.
- Consistency (Con): player implements her plans (i.e., planned and actual behavior coincide).

The set of states where i is **rational** is $R_i = C_i \cap BO_i \cap CBU_i \cap RP_i \cap Con_i \subseteq \Omega^\infty$.
Under regularity conditions, it R_i is Borel (hence, an event that can be assessed and expressed by players).

More details

Rationality: Behavioral side

Plans and behavior

- Rational planning (RP): player *plans* to choose "optimal actions" at each contingency. Here, a plan is a player's belief on how she would behave: RP means that a player expects herself to act in a way that satisfies an "intra-personal equilibrium" condition.
- Consistency (Con): player implements her plans (i.e., planned and actual behavior coincide).

The set of states where i is **rational** is $R_i = C_i \cap BO_i \cap CBU_i \cap RP_i \cap Con_i \subseteq \Omega^\infty$.
Under regularity conditions, it R_i is Borel (hence, an event that can be assessed and expressed by players).

More details

Rationality: Behavioral side

Plans and behavior

- Rational planning (*RP*): player *plans* to choose "optimal actions" at each contingency. Here, a plan is a player's belief on how she would behave: RP means that a player expects herself to act in a way that satisfies an "intra-personal equilibrium" condition.
- Consistency (*Con*): player implements her plans (i.e., planned and actual behavior coincide).

The set of states where i is **rational** is $R_i = C_i \cap BO_i \cap CBU_i \cap RP_i \cap Con_i \subseteq \Omega^\infty$.
Under regularity conditions, it R_i is Borel (hence, an event that can be assessed and expressed by players).

More details

Rationality: Behavioral side

Plans and behavior

- Rational planning (RP): player *plans* to choose "optimal actions" at each contingency. Here, a plan is a player's belief on how she would behave: RP means that a player expects herself to act in a way that satisfies an "intra-personal equilibrium" condition.
- Consistency (Con): player implements her plans (i.e., planned and actual behavior coincide).

The set of states where i is **rational** is $R_i = C_i \cap BO_i \cap CBU_i \cap RP_i \cap Con_i \subseteq \Omega^\infty$.
Under regularity conditions, it R_i is Borel (hence, an event that can be assessed and expressed by players).

More details

Rationality: Behavioral side

Plans and behavior

- Rational planning (RP): player *plans* to choose "optimal actions" at each contingency. Here, a plan is a player's belief on how she would behave: RP means that a player expects herself to act in a way that satisfies an "intra-personal equilibrium" condition.
- Consistency (Con): player implements her plans (i.e., planned and actual behavior coincide).

The set of states where i is **rational** is $R_i = C_i \cap BO_i \cap CBU_i \cap RP_i \cap Con_i \subseteq \Omega^\infty$.
Under regularity conditions, it R_i is Borel (hence, an event that can be assessed and expressed by players).

More details

What's next?

1. Definition of a **general framework**
2. Description of **inferences**
3. Formalization of a notion of **rationality**
4. Definition of a **solution concept** to derive predictions
5. Concluding remarks

Strong rationalizability: Warm-up

With our solution concept, we aim to capture **strategic thinking**. For example (based on buy me an ice-cream):

- If Mom is rational, she believes what she observes: if she sees Child blush, she will conclude he lied. If Child thinks that Mom is sophisticated enough, he understands that blushing means (1) no ice-cream, and (2) being labeled as a liar.
  Then, does it make sense to lie?

Strong rationalizability: Warm-up

With our solution concept, we aim to capture **strategic thinking**. For example (based on buy me an ice-cream):

- If Mom is rational, she believes what she observes: if she sees Child blush, she will conclude he lied. If Child thinks that Mom is sophisticated enough, he understands that blushing means (1) no ice-cream, and (2) being labeled as a liar.
  Then, does it make sense to lie?

Strong rationalizability: Warm-up

With our solution concept, we aim to capture **strategic thinking**. For example (based on buy me an ice-cream):

- If Mom is rational, she believes what she observes: if she sees Child blush, she will conclude he lied. If Child thinks that Mom is sophisticated enough, he understands that blushing means (1) no ice-cream, and (2) being labeled as a liar.
  Then, does it make sense to lie?

Strong rationalizability: Preliminaries

- We are going to propose a procedure of iterated deletion of utility-relevant states (s_i, θ_i, τ_i) for each player $i \in I$.
- In the following, ρ_i is a system of beliefs of order $K+1$ of i: so, ρ_i is a map $h_i \mapsto \rho_i(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i})$.
- Key concept: **strong belief**. Strongly believing an event = assigning probability 1 to it as long as it is consistent with evidence.
- $(\theta_i, \tau_i, \rho_i)$ strongly believes $F \subseteq S \times \Theta_{-i} \times \mathcal{T}_{-i}$ if $\rho_i(F|h_i) = 1$ for each h_i such that $F \cap \mathcal{I}_{(\theta_i, \tau_i)}(h_i) \neq \emptyset$.

Strong rationalizability: Preliminaries

- We are going to propose a procedure of iterated deletion of utility-relevant states (s_i, θ_i, τ_i) for each player $i \in I$.
- In the following, ρ_i is a system of beliefs of order $K+1$ of i: so, ρ_i is a map $h_i \mapsto \rho_i(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i})$.
- Key concept: **strong belief**. Strongly believing an event = assigning probability 1 to it as long as it is consistent with evidence.
- $(\theta_i, \tau_i, \rho_i)$ strongly believes $F \subseteq S \times \Theta_{-i} \times \mathcal{T}_{-i}$ if $\rho_i(F|h_i) = 1$ for each h_i such that $F \cap \mathcal{I}_{(\theta_i,\tau_i)}(h_i) \neq \emptyset$.

Strong rationalizability: Preliminaries

- We are going to propose a procedure of iterated deletion of utility-relevant states (s_i, θ_i, τ_i) for each player $i \in I$.
- In the following, ρ_i is a system of beliefs of order $K+1$ of i: so, ρ_i is a map $h_i \mapsto \rho_i(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i})$.
- Key concept: **strong belief**. Strongly believing an event = assigning probability 1 to it as long as it is consistent with evidence.
- $(\theta_i, \tau_i, \rho_i)$ strongly believes $F \subseteq S \times \Theta_{-i} \times \mathcal{T}_{-i}$ if $\rho_i(F|h_i) = 1$ for each h_i such that $F \cap \mathcal{I}_{(\theta_i,\tau_i)}(h_i) \neq \emptyset$.

Strong rationalizability: Preliminaries

- We are going to propose a procedure of iterated deletion of utility-relevant states (s_i, θ_i, τ_i) for each player $i \in I$.
- In the following, ρ_i is a system of beliefs of order $K+1$ of i: so, ρ_i is a map $h_i \mapsto \rho_i(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i})$.
- Key concept: **strong belief**. Strongly believing an event = assigning probability 1 to it as long as it is consistent with evidence.
- $(\theta_i, \tau_i, \rho_i)$ strongly believes $F \subseteq S \times \Theta_{-i} \times \mathcal{T}_{-i}$ if $\rho_i(F|h_i) = 1$ for each h_i such that $F \cap \mathcal{I}_{(\theta_i,\tau_i)}(h_i) \neq \emptyset$.

Strong rationalizability: Preliminaries

- We are going to propose a procedure of iterated deletion of utility-relevant states (s_i, θ_i, τ_i) for each player $i \in I$.
- In the following, ρ_i is a system of beliefs of order $K+1$ of i: so, ρ_i is a map $h_i \mapsto \rho_i(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i})$.
- Key concept: **strong belief**. Strongly believing an event = assigning probability 1 to it as long as it is consistent with evidence.
- $(\theta_i, \tau_i, \rho_i)$ strongly believes $F \subseteq S \times \Theta_{-i} \times \mathcal{T}_{-i}$ if $\rho_i(F|h_i) = 1$ for each h_i such that $F \cap \mathcal{I}_{(\theta_i,\tau_i)}(h_i) \neq \emptyset$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ ($i \in I$), $\mathbf{P}_{-i}(0) := \times_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\times_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{P}_i(n) \right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ ($i \in I$), $\mathbf{P}_{-i}(0) := \times_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\times_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{P}_i(n) \right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ ($i \in I$), $\mathbf{P}_{-i}(0) := \times_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\times_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{P}_i(n) \right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ ($i \in I$), $\mathbf{P}_{-i}(0) := \times_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\times_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{P}_i(n)\right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ ($i \in I$), $\mathbf{P}_{-i}(0) := \times_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\times_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{P}_i(n) \right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ ($i \in I$), $\mathbf{P}_{-i}(0) := \times_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\times_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{P}_i(n) \right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ $(i \in I)$, $\mathbf{P}_{-i}(0) := \times_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\times_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{P}_i(n) \right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Procedure

- Start with $\mathbf{P}_i(0) := S_i \times \Theta_i \times \mathcal{T}_i$ $(i \in I)$, $\mathbf{P}_{-i}(0) := \bigtimes_{j \neq i} \mathbf{P}_j(0)$.
- $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n)$ if $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i(n-1)$ and there is ρ_i such that
  1. (τ_i, ρ_i) satisfies cognitive rationality (coherence, believe-what-you-observe, correct belief updating);
  2. the plan entailed by (τ_i, ρ_i) satisfies rational planning (intra-personal equilibrium);
  3. s_i is consistent with the plan entailed by (τ_i, ρ_i);
  4. $(\theta_i, \tau_i, \rho_i)$ strongly believes $\mathbf{P}_{-i}(n-1)$.
- The set of **strongly rationalizable utility-relevant states** is $\bigtimes_{i \in I} \mathbf{P}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\bigtimes_{i \in I} \mathbf{P}_i(n) \right) \subseteq S \times \Theta \times \mathcal{T}$.

Strong rationalizability: Justification (hints)

- Recall that rationality of a player, R_i, is a subset of Ω^∞.
- Can define events in Ω^∞ like:
  1. $\mathbf{R}_i(1)$ = player i is rational.
  2. $\mathbf{R}_i(2)$ = player i is rational and strongly believes (believes whenever possible) $\mathbf{R}_{-i}(1)$.

  ...

  n $\mathbf{R}_i(n)$ = player i is rational and strongly believes $\mathbf{R}_{-i}(1), \ldots, \mathbf{R}_{-i}(n-1)$.
- States of the world where there is **rationality and common strong belief in rationality**: $\times_{i\in I} \mathbf{R}_i(\infty) := \bigcap_{n\in\mathbb{N}} \left(\times_{i\in I} \mathbf{R}_i(n)\right) \subseteq \Omega^\infty$.

Theorem

For each $i \in I$ and $n \in \mathbb{N}_0$: $\mathbf{P}_i(n) = \text{proj}_{S_i\times\Theta_i\times T_i} \mathbf{R}_i(n)$.

Note: we can work with finite hierarchical systems of beliefs (in many cases, $K = 1$ is enough) to derive the utility-relevant implications of assumptions that are formulated in terms of infinite hierarchical systems of beliefs.

Strong rationalizability: Justification (hints)

- Recall that rationality of a player, R_i, is a subset of Ω^∞.
- Can define events in Ω^∞ like:
  1. $\mathbf{R}_i(1)$ = player i is rational.
  2. $\mathbf{R}_i(2)$ = player i is rational and strongly believes (believes whenever possible) $\mathbf{R}_{-i}(1)$.

  ...

  n. $\mathbf{R}_i(n)$ = player i is rational and strongly believes $\mathbf{R}_{-i}(1), \ldots, \mathbf{R}_{-i}(n-1)$.
- States of the world where there is **rationality and common strong belief in rationality**: $\times_{i \in I} \mathbf{R}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{R}_i(n) \right) \subseteq \Omega^\infty$.

Theorem

For each $i \in I$ and $n \in \mathbb{N}_0$: $\mathbf{P}_i(n) = \mathrm{proj}_{S_i \times \Theta_i \times T_i} \mathbf{R}_i(n)$.

Note: we can work with finite hierarchical systems of beliefs (in many cases, $K = 1$ is enough) to derive the utility-relevant implications of assumptions that are formulated in terms of infinite hierarchical systems of beliefs.

Strong rationalizability: Justification (hints)

- Recall that rationality of a player, R_i, is a subset of Ω^∞.
- Can define events in Ω^∞ like:
  1. $\mathbf{R}_i(1)$ = player i is rational.
  2. $\mathbf{R}_i(2)$ = player i is rational and strongly believes (believes whenever possible) $\mathbf{R}_{-i}(1)$.

  ...

  n $\mathbf{R}_i(n)$ = player i is rational and strongly believes $\mathbf{R}_{-i}(1), \ldots, \mathbf{R}_{-i}(n-1)$.
- States of the world where there is **rationality and common strong belief in rationality**: $\times_{i\in I} \mathbf{R}_i(\infty) := \bigcap_{n\in\mathbb{N}} \left(\times_{i\in I} \mathbf{R}_i(n)\right) \subseteq \Omega^\infty$.

Theorem

For each $i \in I$ and $n \in \mathbb{N}_0$: $\mathbf{P}_i(n) = \mathrm{proj}_{S_i\times\Theta_i\times T_i} \mathbf{R}_i(n)$.

Note: we can work with finite hierarchical systems of beliefs (in many cases, $K = 1$ is enough) to derive the utility-relevant implications of assumptions that are formulated in terms of infinite hierarchical systems of beliefs.

Strong rationalizability: Justification (hints)

- Recall that rationality of a player, R_i, is a subset of Ω^∞.
- Can define events in Ω^∞ like:
  1. $\mathbf{R}_i(1)$ = player i is rational.
  2. $\mathbf{R}_i(2)$ = player i is rational and strongly believes (believes whenever possible) $\mathbf{R}_{-i}(1)$.

  ...

  n $\mathbf{R}_i(n)$ = player i is rational and strongly believes $\mathbf{R}_{-i}(1), \ldots, \mathbf{R}_{-i}(n-1)$.
- States of the world where there is **rationality and common strong belief in rationality**: $\times_{i \in I} \mathbf{R}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{R}_i(n) \right) \subseteq \Omega^\infty$.

Theorem

For each $i \in I$ and $n \in \mathbb{N}_0$: $\mathbf{P}_i(n) = \text{proj}_{S_i \times \Theta_i \times T_i} \mathbf{R}_i(n)$.

Note: we can work with finite hierarchical systems of beliefs (in many cases, $K = 1$ is enough) to derive the utility-relevant implications of assumptions that are formulated in terms of infinite hierarchical systems of beliefs.

Strong rationalizability: Justification (hints)

- Recall that rationality of a player, R_i, is a subset of Ω^∞.
- Can define events in Ω^∞ like:
  1. $\mathbf{R}_i(1)$ = player i is rational.
  2. $\mathbf{R}_i(2)$ = player i is rational and strongly believes (believes whenever possible) $\mathbf{R}_{-i}(1)$.

  ...

  n $\mathbf{R}_i(n)$ = player i is rational and strongly believes $\mathbf{R}_{-i}(1), \ldots, \mathbf{R}_{-i}(n-1)$.
- States of the world where there is **rationality and common strong belief in rationality**: $\times_{i\in I} \mathbf{R}_i(\infty) := \bigcap_{n\in\mathbb{N}} \left(\times_{i\in I} \mathbf{R}_i(n)\right) \subseteq \Omega^\infty$.

Theorem

For each $i \in I$ and $n \in \mathbb{N}_0$: $\mathbf{P}_i(n) = \mathrm{proj}_{S_i\times\Theta_i\times T_i} \mathbf{R}_i(n)$.

Note: we can work with finite hierarchical systems of beliefs (in many cases, $K = 1$ is enough) to derive the utility-relevant implications of assumptions that are formulated in terms of infinite hierarchical systems of beliefs.

Strong rationalizability: Justification (hints)

- Recall that rationality of a player, R_i, is a subset of Ω^∞.
- Can define events in Ω^∞ like:
  1. $\mathbf{R}_i(1)$ = player i is rational.
  2. $\mathbf{R}_i(2)$ = player i is rational and strongly believes (believes whenever possible) $\mathbf{R}_{-i}(1)$.

  ...

  n $\mathbf{R}_i(n)$ = player i is rational and strongly believes $\mathbf{R}_{-i}(1), \ldots, \mathbf{R}_{-i}(n-1)$.
- States of the world where there is **rationality and common strong belief in rationality**: $\times_{i \in I} \mathbf{R}_i(\infty) := \bigcap_{n \in \mathbb{N}} \left(\times_{i \in I} \mathbf{R}_i(n) \right) \subseteq \Omega^\infty$.

Theorem

For each $i \in I$ and $n \in \mathbb{N}_0$: $\mathbf{P}_i(n) = \text{proj}_{S_i \times \Theta_i \times \mathcal{T}_i} \mathbf{R}_i(n)$.

Note: we can work with finite hierarchical systems of beliefs (in many cases, $K = 1$ is enough) to derive the utility-relevant implications of assumptions that are formulated in terms of infinite hierarchical systems of beliefs.

Strong rationalizability: Application to “buy me an ice-cream”

Assume Child’s appreciation for video-games belongs to $\Theta_C = \{\theta', \theta''\}$ with $0 < \theta' < 1 < \theta''$.

Solution procedure:

1. **Child** Does not make sense to (plan to) say *No* after *Homework*.
   Mom if Child blushes, he must have played *Video-games* $\rightarrow$ she does *Not* buy the ice-cream if (Yes, b).
2. **Child** Consider step 1: if he blushes, Mom will conclude he is a liar $\rightarrow$ this undermines confidence $\rightarrow$ optimal action after *Video-games* is to say *No* $\rightarrow$ it is always optimal to tell the truth.
   Mom Consider step 1: if Child says *No*, he must have played *Video-games* $\rightarrow$ Mom does *Not* buy the ice-cream if $(No, \neg b)$.

Strong rationalizability: Application to "buy me an ice-cream"

Assume Child's appreciation for video-games belongs to $\Theta_C = \{\theta', \theta''\}$ with $0 < \theta' < 1 < \theta''$.

Solution procedure:

1. **Child** Does not make sense to (plan to) say *No* after *Homework*.
   Mom if Child blushes, he must have played *Video-games* $\rightarrow$ she does *Not* buy the ice-cream if (Yes, b).
2. **Child** Consider step 1: if he blushes, Mom will conclude he is a liar $\rightarrow$ this undermines confidence $\rightarrow$ optimal action after *Video-games* is to say *No* $\rightarrow$ it is always optimal to tell the truth.
   Mom Consider step 1: if Child says *No*, he must have played *Video-games* $\rightarrow$ Mom does *Not* buy the ice-cream if $(No, \neg b)$.

Strong rationalizability: Application to “buy me an ice-cream”

Assume Child’s appreciation for video-games belongs to $\Theta_C = \{\theta', \theta''\}$ with $0 < \theta' < 1 < \theta''$.

Solution procedure:

1. **Child** Does not make sense to (plan to) say *No* after *Homework*.
   Mom if Child blushes, he must have played *Video-games* $\rightarrow$ she does *Not* buy the ice-cream if (Yes, b).
2. **Child** Consider step 1: if he blushes, Mom will conclude he is a liar $\rightarrow$ this undermines confidence $\rightarrow$ optimal action after *Video-games* is to say *No* $\rightarrow$ it is always optimal to tell the truth.
   Mom Consider step 1: if Child says *No*, he must have played *Video-games* $\rightarrow$ Mom does *Not* buy the ice-cream if $(No, \neg b)$.

Strong rationalizability: Application to "buy me an ice-cream"

Assume Child's appreciation for video-games belongs to $\Theta_C = \{\theta', \theta''\}$ with $0 < \theta' < 1 < \theta''$.

Solution procedure:

1. **Child** Does not make sense to (plan to) say *No* after *Homework*.
   Mom if Child blushes, he must have played *Video-games* $\rightarrow$ she does *Not* buy the ice-cream if (Yes, b).
2. **Child** Consider step 1: if he blushes, Mom will conclude he is a liar $\rightarrow$ this undermines confidence $\rightarrow$ optimal action after *Video-games* is to say *No* $\rightarrow$ it is always optimal to tell the truth.
   Mom Consider step 1: if Child says *No*, he must have played *Video-games* $\rightarrow$ Mom does *Not* buy the ice-cream if $(No, \neg b)$.

Strong rationalizability: Application to "buy me an ice-cream"

Assume Child's appreciation for video-games belongs to $\Theta_C = \{\theta', \theta''\}$ with $0 < \theta' < 1 < \theta''$.

Solution procedure:

1. **Child** Does not make sense to (plan to) say *No* after *Homework*.
   Mom if Child blushes, he must have played *Video-games* $\rightarrow$ she does *Not* buy the ice-cream if (*Yes*, *b*).
2. **Child** Consider step 1: if he blushes, Mom will conclude he is a liar $\rightarrow$ this undermines confidence $\rightarrow$ optimal action after *Video-games* is to say *No* $\rightarrow$ it is always optimal to tell the truth.
   Mom Consider step 1: if Child says *No*, he must have played *Video-games* $\rightarrow$ Mom does *Not* buy the ice-cream if ($No, \neg b$).

Strong rationalizability: Application to "buy me an ice-cream"

Assume Child's appreciation for video-games belongs to $\Theta_C = \{\theta', \theta''\}$ with $0 < \theta' < 1 < \theta''$.

Solution procedure:

1. **Child** Does not make sense to (plan to) say *No* after *Homework*.
   Mom if Child blushes, he must have played *Video-games* $\rightarrow$ she does *Not* buy the ice-cream if (Yes, b).
2. **Child** Consider step **1**: if he blushes, Mom will conclude he is a liar $\rightarrow$ this undermines confidence $\rightarrow$ optimal action after *Video-games* is to say *No* $\rightarrow$ it is always optimal to tell the truth.
   Mom Consider step **1**: if Child says *No*, he must have played *Video-games* $\rightarrow$ Mom does *Not* buy the ice-cream if $(No, \neg b)$.

Strong rationalizability: Application to “buy me an ice-cream”

Assume Child’s appreciation for video-games belongs to $\Theta_C = \{\theta', \theta''\}$ with $0 < \theta' < 1 < \theta''$.

Solution procedure:

1. **Child** Does not make sense to (plan to) say *No* after *Homework*.
   Mom if Child blushes, he must have played *Video-games* $\rightarrow$ she does *Not* buy the ice-cream if (Yes, b).
2. **Child** Consider step 1: if he blushes, Mom will conclude he is a liar $\rightarrow$ this undermines confidence $\rightarrow$ optimal action after *Video-games* is to say *No* $\rightarrow$ it is always optimal to tell the truth.
   Mom Consider step 1: if Child says *No*, he must have played *Video-games* $\rightarrow$ Mom does *Not* buy the ice-cream if $(No, \neg b)$.

Strong rationalizability: Application to "buy me an ice-cream"

3. **Mom** Consider step 2: Child always tells the truth $\rightarrow$ Mom *Buys* him the ice-cream if $(Yes, \neg b)$.
4. **Child** Consider step 3: Child has to choose between *Homework* plus ice-cream and *Video-games* without ice-cream $\rightarrow$ if θ' he prefers *Homework* plus ice-cream, if θ'' he prefers *Video-games* without ice-cream.

Takeaway: signals about emotions and image concerns yield "full disclosure" (Child honestly reports his action, and Mom believes him).

Strong rationalizability: Application to “buy me an ice-cream”

3. **Mom** Consider step 2: Child always tells the truth $\rightarrow$ Mom *Buys* him the ice-cream if $(Yes, \neg b)$.
4. **Child** Consider step 3: Child has to choose between *Homework* plus ice-cream and *Video-games* without ice-cream $\rightarrow$ if θ' he prefers *Homework* plus ice-cream, if θ'' he prefers *Video-games* without ice-cream.

Takeaway: signals about emotions and image concerns yield “full disclosure” (Child honestly reports his action, and Mom believes him).

Strong rationalizability: Application to "buy me an ice-cream"

3. **Mom** Consider step 2: Child always tells the truth $\rightarrow$ Mom *Buys* him the ice-cream if $(\mathit{Yes}, \neg b)$.
4. **Child** Consider step 3: Child has to choose between *Homework* plus ice-cream and *Video-games* without ice-cream $\rightarrow$ if θ' he prefers *Homework* plus ice-cream, if θ'' he prefers *Video-games* without ice-cream.

Takeaway: signals about emotions and image concerns yield "full disclosure" (Child honestly reports his action, and Mom believes him).

Strong rationalizability: Application to “buy me an ice-cream”

3. **Mom** Consider step 2: Child always tells the truth $\rightarrow$ Mom *Buys* him the ice-cream if $(Yes, \neg b)$.
4. **Child** Consider step 3: Child has to choose between *Homework* plus ice-cream and *Video-games* without ice-cream $\rightarrow$ if θ' he prefers *Homework* plus ice-cream, if θ'' he prefers *Video-games* without ice-cream.

Takeaway: signals about emotions and image concerns yield “full disclosure” (Child honestly reports his action, and Mom believes him).

What's next?

1. Definition of a **general framework**
2. Description of **inferences**
3. Formalization of a notion of **rationality**
4. Definition of a **solution concept** to derive predictions
5. Concluding remarks

Concluding remarks: Future research

Applications

- Emotional leakage and disclosure in face-to-face interactions? Think of negotiations, political speeches, court hearings.
- More likely to accept unfair offers if the proposer smiles?
- More aggressive (conciliatory) when negotiating with happy (angry) counterparts?

Bounded rationality and strategic thinking

- We introduced a rich and expressive language to analyze failures of rationality on both the cognitive and the behavioral side. Here, focus on rationality and common strong belief in rationality.
- Interesting to allow agents to reason strategically about failures of rationality, and to capture behavioral implications of different epistemic assumptions.

Concluding remarks: Future research

Applications

- Emotional leakage and disclosure in face-to-face interactions? Think of negotiations, political speeches, court hearings.
- More likely to accept unfair offers if the proposer smiles?
- More aggressive (conciliatory) when negotiating with happy (angry) counterparts?

Bounded rationality and strategic thinking

- We introduced a rich and expressive language to analyze failures of rationality on both the cognitive and the behavioral side. Here, focus on rationality and common strong belief in rationality.
- Interesting to allow agents to reason strategically about failures of rationality, and to capture behavioral implications of different epistemic assumptions.

Concluding remarks: Future research

Applications

- Emotional leakage and disclosure in face-to-face interactions? Think of negotiations, political speeches, court hearings.
- More likely to accept unfair offers if the proposer smiles?
- More aggressive (conciliatory) when negotiating with happy (angry) counterparts?

Bounded rationality and strategic thinking

- We introduced a rich and expressive language to analyze failures of rationality on both the cognitive and the behavioral side. Here, focus on rationality and common strong belief in rationality.
- Interesting to allow agents to reason strategically about failures of rationality, and to capture behavioral implications of different epistemic assumptions.

Thank you!

References

Battigalli, P. (1997). On rationalizability in extensive games. *Journal of Economic Theory*, 74(1), 40-61.

Battigalli, P., Corrao, R., and Dufwenberg, M. (2019). Incorporating belief-dependent motivation in games. *Journal of Economic Behavior & Organization*, 167 , 185–218.

Battigalli, P., Corrao, R., and Sanna, F. (2020). Epistemic game theory without types structures: An application to psychological games. *Games and Economic Behavior*, 120 , 28–57.

Battigalli, P., and De Vito, N. (2021). Beliefs, plans, and perceived intentions in dynamic games. *Journal of Economic Theory*, 195 , 105283.

Battigalli, P., and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144 (1), 1–35.

Battigalli, P., and Dufwenberg, M. (2022). Belief-dependent motivations and psychological game theory. *Journal of Economic Literature*, 60(3), 833-882.

References

Battigalli, P., and Prestipino, A. (2013). Transparent restrictions on beliefs and forward-induction reasoning in games with asymmetric information. *The BE Journal of Theoretical Economics*, 13 (1), 79–130.

Battigalli, P., and Siniscalchi, M. (2002). Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106 (2), 356–391.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1 (1), 60–79.

Givens, D. B. (1978). The nonverbal basis of attraction: Flirtation, courtship, and seduction. *Psychiatry*, 41 (4), 346–359.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3 (11), 419–429.

References

Hatfield, E., Bensman, L., Thornton, P. D., and Rapson, R. L. (2014). New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion. *Interpersona*.

Matsumoto, D., and Hwang, H. C. (2018). Microexpressions differentiate truths from lies about future malicious intent. *Frontiers in Psychology*, 9 , 2545.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica*, 1029–1050.

Porter, S., Ten Brinke, L., and Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity. *Journal of Nonverbal Behavior*, 36 (1), 23–37.

Stirrat, M., and Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, 21 (3), 349–354.

References

Van Leeuwen, B., Noussair, C. N., Offerman, T., Suetens, S., Van Veelen, M., and Van De Ven, J. (2018). Predictably angry—facial cues provide a credible signal of destructive behavior. *Management Science*, 64 (7), 3352–3364.

Vásquez, J., and Weretka, M. (2021). Co-worker altruism and unemployment. *Games and Economic Behavior*, 130, 224-239.

Game tree: Derivation

- **Action feasibility correspondence** of i: $\mathcal{A}_i : M_{i,p} \rightrightarrows A_i$. Define $\mathcal{A}((m_{i,p})_{i\in I}) := \bigtimes_{i\in I} \mathcal{A}_i(m_{i,p})$.
- First PPM profile is $P(a^0) = P(\varnothing_A) =: m_{p,0}$ (neglected in notation).
- Histories are sequences of profiles of actions and messages: $(a^t, m_p^t, m^t) = (a_k, m_{p,k}, m_k)_{k=1}^t$ is **feasible** if for each $k \in \{1, \ldots, t\}$:
  1. $a_k \in \mathcal{A}(m_{p,k-1})$;
  2. $m_{p,k} = P(a^k)$;
  3. there exists (θ, e^{k+1}) such that $m_k \in \operatorname{supp} \tilde{f}(a_k, \theta, e^{k+1})$.

Back

Game tree: Application to "buy me an ice-cream"

Stage 1 Child chooses $a_{C,1} \in \{Homework, Video\text{-}games\}$.
Mom's PPM is uninformative: $(a_{C,1}) \mapsto \bar{m}_{M,p}$.
No emotional messages.

Stage 2 Child chooses $a_{C,2} \in \{Yes, No\}$, and $m \in \{b, \neg b\}$ realizes.
Mom's PPM reveals stage-2 action: $(a_{C,1}, a_{C,2}) \mapsto a_{C,2}$.
Mom's emotional message is m.

Stage 3 Mom chooses $a_M \in \{Buy, Not\}$.
Child's PPM reveals Mom's action: $(a_{C,1}, a_{C,2}, a_M) \mapsto a_M$.
No emotional messages.

A terminal history as the form $(a_{C,1}, \bar{m}_{M,p}, a_{C,2}, m, a_M)$.

Mom's only length-1 personal history is $(\bar{m}_{M,p})$ (henceforth neglected in notation).
Length-2 personal histories: (Yes, b), $(Yes, \neg b)$, $(No, \neg b)$.
Terminal personal histories: $\{(Yes, b), (Yes, \neg b), (No, \neg b)\} \times \{Buy, Not\}$.

Back

Hierarchical systems of beliefs: Construction

- **Basic space of uncertainty**: $\Omega^0_{-i} := S \times \Theta_{-i}$.
- Set of **first-order systems of beliefs** of i: $\mathcal{T}_{i,1} := \left[\Delta(\Omega^0_{-i})\right]^{\bar{H}_i}$. Generic element $\tau_{i,1}$ is a map $h_i \mapsto \tau_{i,1}(\cdot|h_i)$.
- Define $\Omega^1_{-i} := \Omega^0_{-i} \times \left(\times_{j\neq i} \mathcal{T}_{j,1}\right)$. Set of **second-order systems of beliefs** of i: $\mathcal{T}_{i,2} := \left[\Delta(\Omega^1_{-i})\right]^{\bar{H}_i}$. Generic element $\tau_{i,2}$ is a map $h_i \mapsto \tau_{i,2}(\cdot|h_i)$ and $\tau_{i,2}(\cdot|h_i)$ is a belief over Ω^0_{-i} *and* others' first-order system of beliefs.
- Proceed by induction to retrieve $\mathcal{T}_{i,n} := \left[\Delta(\Omega^{n-1}_{-i})\right]^{\bar{H}_i}$ $(n \in \mathbb{N})$.
- Set of ***n*-th-order hierarchical systems of beliefs** of i: $\mathcal{T}^n_i := \times_{k=1}^n \mathcal{T}_{i,k}$. Generic element τ^n_i is a map $h_i \mapsto (\tau_{i,k}(\cdot|h_i))_{k=1}^n$.
- Set of **epistemic types** of i: $\mathcal{T}^\infty_i = \times_{k\in\mathbb{N}} \mathcal{T}_{i,k}$.

Back

Emotion-generating function: Assumptions

Assumptions about ε:

- Counterfactual beliefs do not matter. Define realized-beliefs map: for each $h = (h_i)_{i\in I}$, β_h is the map $(\tau_i^\infty)_{i\in I} \mapsto ((\tau_i^\infty(\cdot|h_i'))_{h_i' \preceq h_i})_{i\in I}$. Then, for each $h \in H$, ε_h can be written as $\bar{\varepsilon}_h \circ \beta_h$ for some $\bar{\varepsilon}_h$.
- Only beliefs of order up to K matter. For each h, τ^∞, and $\bar{\tau}^\infty$, $\tau^K = \bar{\tau}^K \Longrightarrow \varepsilon(h, \tau^\infty) = \varepsilon(h, \bar{\tau}^\infty)$.

Back

Game-dependent feedback: Derivation

Let $\mathbf{E} = E^{\leq T+1}$ and note that $\varepsilon(h, \tau) \in \Delta(\mathbf{E})$.

Recall that inputs of $\tilde{f}$ are actions, traits, streams of emotions.

Then, the game-dependent representation of feedback is (in green the game-specific ingredients):

$$f_h(s, \theta, \tau)[m] := \int_{\mathbf{E}} \tilde{f}(s(h), \theta, \mathbf{e})[m] \cdot \mathrm{d}\varepsilon(h, \tau).$$

If ε is deterministic, then simply:

$$f_h(s, \theta, \tau)[m] = 1 \text{ iff } m = \tilde{f}(s(h), \theta, \varepsilon(h, \tau)).$$

Back

Psychological utility functions: Hints

Some examples of psychological motivations. For simplicity, consider $I = \{i, j\}$ and emotions profiles in $E = E_i \times E_j$.

1. **Image concerns**: a player dislikes being thought of as "bad". Assume $\Theta_i = \Theta_i^S \times \{Nice_i, Rude_i\}$.
   Game-independent utility is:

$$\tilde{v}_i(y, \theta, e) = \underbrace{y_i}_{\text{own outcome}} - \overbrace{\theta_i^S}^{\text{sensitivity to } j\text{'s opinion}} \cdot \underbrace{e_i}_{\text{image concern}} .$$

   Game-dependent utility is:

$$v_i(z, \theta, \tau) = \underbrace{\pi_i(z)}_{\text{own outcome after } z} - \overbrace{\theta_i^S}^{\text{sensitivity to } j\text{'s opinion}} \cdot \underbrace{\tau_j(Rude_i | z)}_{j\text{'s opinion after } z} .$$

   Note: here, opinions about unobserved traits of others, but there are also opinions about unobserved actions (as in the running example).

Psychological utility functions: Hints

❷ **Guilt aversion**: a player dislikes failing others' expectations. *Game-independent* utility is:

$$\tilde{v}_i(y,\theta,e) = \underbrace{y_i}_{\text{outcome}} - \overbrace{\theta_i^S}^{\text{sensitivity to } j\text{'s disappointment}} \cdot \underbrace{e_j}_{j\text{'s disapp.}} .$$

Game-dependent utility is:

$$v_i(z,\theta,\tau) = \underbrace{\pi_i(z)}_{\text{outcome after } z} - \overbrace{\theta_i^S}^{\text{sensitivity to } j\text{'s disappointment}} \cdot \underbrace{\left[\mathbb{E}_{\tau_j}(\pi_j|\varnothing) - \pi_j(z)\right]^+}_{j\text{'s disapp. after } z} .$$

Back

Inferences: Properties of feedback

- We need conditions that ensure that players' inferential reasoning is well-defined.
- Assumption 1 (Own-belief independence, OBI): *feedback is about others*; at each history, the probabilities of realization of messages about i's co-players do not depend on i's beliefs.
- Assumption 2 (Regularity, Reg): *upon observing some message, players can always "discern" the set of states that did not prevent such message*; the set of utility-relevant states that allow for some message at some history is a measurable rectangle.

Lemma

Assume OBI and Reg hold. Then, $\mathcal{I}_{(\theta_i,\tau_i^K)}(h_i)$ is Borel for each i, h_i, θ_i and τ_i^K.

Back

Rationality: Coherence, believe-what-you-observe

Rationality is defined as the conjunction of several cognitive and behavioral features.

1. **Coherence**: beliefs of different orders are be coherent.
   - Epistemic type $\tau_i^\infty = (\tau_{i,n})_{n\in\mathbb{N}}$, with $\tau_{i,n}$ system of n-th-order beliefs.
   - $\tau_{i,n}$ is a map $h_i \mapsto \tau_{i,n}(\cdot|h_i) \in \Delta(S \times \Theta_{-i} \times \mathcal{T}_{-i}^{n-1})$. Call $S \times \Theta_{-i} \times \mathcal{T}_{-i}^{n-1} = \Omega_{-i}^{n-1}$.
   - τ_i^∞ is **coherent** if, for each h_i and n, $\tau_{i,n}(\cdot|h_i) = \mathrm{marg}_{\Omega_{-i}^{n-1}}\tau_{i,n+1}(\cdot|h_i)$.
   - $C_i \subseteq \Omega^\infty$ is the set of states where i's epistemic type is coherent.

2. **Believe-what-you-observe**: beliefs over utility-relevant states assign probability 1 to the set of states consistent with evidence.
   - τ_i^∞ satisfies **believe-what-you-observe** (BO) if, for each h_i, $\tau_{i,K+1}\big(\mathcal{I}_{\tau_i}(h_i)|h_i\big) = 1$.
   - $BO_i \subseteq \Omega^\infty$ is the set of states where i's epistemic type satisfies BO.

Back

Rationality: Belief updating

3. **Correct belief updating**: a player updates her beliefs about herself and her opponents according to the rules of conditional probabilities.
   - At each stage, i observes two pieces of information: first, a_i is chosen; then, $(m_{i,p}, m_i)$ realizes.
   - a_i is used to update beliefs on *own external states* (S_i); $(m_{i,p}, m_i)$ is used to update beliefs about *others* $(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i})$.
   - Call $S_i(h_i, a_i)$ the set of s_i that allow for h_i and that are such that $s_i(h_i) = a_i$. **Chain rule**: for each h_i, $a_i \in A_i(h_i)$, $s_i \in S_i(h_i, a_i)$,

$$\underbrace{\tau_{i,K+1}(s_i|h_i')}_{\text{prob. of } s_i \text{ after } a_i} \cdot \underbrace{\tau_{i,K+1}(S_i(h_i, a_i)|h_i)}_{\text{prob. of } a_i} = \underbrace{\tau_{i,K+1}(s_i|h_i)}_{\text{prob. of } s_i \text{ before } a_i},$$

   where h_i' is any immediate successor of h_i where a_i is played.

Back

❸ **Correct belief updating** (continued)

- Now assume a_i has been played at h_i. Call $\mu(\cdot|h_i)$ the marginal of $\tau_{i,K+1}(\cdot|h_i)$ on $S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i} =: X$.
  Bayes rule: for each h_i, a_i, $(m_{i,p}, m_i)$, and Borel $F \subseteq X$,

$$\overbrace{\tau_{i,K+1}(F|h_i')}^{\text{prob. of } F \text{ after } (m_{i,p},m_i)} \cdot \underbrace{\int_X g_{h_i,a_i}(m_{i,p}, m_i|x)\mu(\mathrm{d}x|h_i)}_{\text{prob. of } (m_{i,p},m_i)} = \underbrace{\int_F g_{h_i,a_i}(m_{i,p}, m_i|x)\mu(\mathrm{d}x|h_i)}_{\text{prob. of } F \text{ before } (m_{i,p},m_i)},$$

  where $h_i' = (h_i, (a_i, m_{i,p}, m_i))$, and $g_{h_i,a_i}\big((m_{i,p}, m_i)|x\big)$ is the probability of $(m_{i,p}, m_i)$ given x after a_i was played at h_i (retrieved from p and f).
- Note: two "parallel" belief updating procedures. This way, if player i is surprised by her behavior, she still updates her beliefs about others correctly.
- τ_i^∞ satisfies **correct belief updating** if the chain rule and the Bayes rule hold.
- $CBU_i \subseteq \Omega^\infty$ is the set of states where i's epistemic type satisfies correct belief updating.

Back

Rationality: Rational planning

4. **Rational planning**: a player plans to choose only optimal actions, at each possible contingency.
   - Derive a *plan* of epistemic type τ_i^∞, $\sigma(\tau_i^\infty) \in \times_{h_i \in H_i} \Delta(A_i(h_i))$, based on i's beliefs about herself.
   - Derive the *decision utility* of i, $u_{i,h_i} : A_i \times \Theta_i \times \mathcal{T}_i^{K+1} \to \mathbb{R}$. Interpretation: $u_{i,h_i}(a_i, \theta_i, \tau_i^{K+1})$ is the expected utility i from choosing a_i at h_i when her trait is θ_i. This is derived from the fact that beliefs τ_i^{K+1} determine a continuation plan after (h_i, a_i).
   - Let $A_i^*(h_i, \theta_i, \tau_i^{K+1}) = \arg\max_{a_i \in A_i(h_i)} u_{i,h_i}(\cdot, \theta_I, \tau_i^{K+1})$ be the set of optimal actions at h_i when player i's trait and belief system are θ_i and τ_i^{K+1}.
   - Player i **plans rationally** at $(s_i, \theta_i, \tau_i^\infty)$ if, for each h_i, $\text{supp}\sigma(\tau_i^\infty)(\cdot|h_i) \subseteq A_i^*(h_i, \theta_i, \tau_i^{K+1})$.
   - $RP_i \subseteq \Omega^\infty$ is the set of states where i plans rationally.

Back

Rationality: Consistency

5. **Consistency**: planned and actual behavior coincide.
   - Player i is **consistent** at $(s_i, \theta_i, \tau_i^\infty)$ if, for each h_i, $\sigma(\tau_i^\infty)(s_i(h_i)|h_i) > 0$.
   - $Con_i \subseteq \Omega^\infty$ is the set of states where i is consistent.

The set of states where i is **rational** is $R_i = C_i \cap BO_i \cap CBU_i \cap RP_i \cap Con_i$.

Theorem

Assume OBI and Reg hold. Then, R_i is Borel for each i.

Back