INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Games with Noisy Signals About Emotions

Pierpaolo Battigalli and Nicolò Generoso

Working Paper n. 719

This Version: November 1, 2025

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Games with Noisy Signals About Emotions*

Pierpaolo Battigalli† Nicolò Generoso‡

November 1, 2025

Abstract

We formalize a novel framework allowing for the observation of noisy signals about co-players' emotions, or states of mind. Insofar as the latter are belief-dependent, such feedback allows players to draw inferences informing their strategic thinking. We analyze players' strategic reasoning adapting the strong rationalizability solution concept, and we give its epistemic justification in terms of players' rationality and interactive beliefs. The "forward-induction" reasoning entailed by such solution allows players to make inferences about their co-players' beliefs, private information, and future, or past and unobserved behavior based on the behavioral and emotional feedback they obtain as the game unfolds. We illustrate our framework with a signaling-like example, showing that the possibility of betraying lies, e.g., by blushing, may incentivize truth-telling.

1 Introduction

Emotions shape social phenomena and they are often betrayed by some signals, as both common sense and everyday experience suggest. For instance, blushing may reveal embarrassment, and gaze contact may indicate that a person is captivated by a conversation. The relevance of emotional signals is highlighted by a number of experimental studies: emotional leakage occurs when people lie (Porter et al., 2012), nonverbal communication is key in courtship encounters (Givens, 1978), individuals seem to recognize others' predisposition to anger or trustworthiness by observing facial cues (Van Leeuwen et al., 2018; Stirrat and Perrett, 2010), and gesture effectively informs listeners of a speaker's unspoken thoughts (Goldin-Meadow, 1999). Evidence also suggests that states of mind and behavior may be influenced by signals about the emotions of others: individuals tend to mimic others' states of mind, therefore sparking a sort of "emotional

*We thank Carlo Andreatta, Manuel Arnese, Nicola Bariletto, Alessandro Cherubin, Nicodemo De Vito, Shuige Liu, Julien Manili, Silvia Meneghesso, and Nicolas Sourisseau as well as seminar participants at Bocconi, Paris School of Economics, Paris Sorbonne, Bergamo, Naples, Rome La Sapienza, University of Arizona, and University of Southern California for useful comments. This project is funded by the European Union (ERC, TRAITS-GAMES, 101142844). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

†Bocconi University and IGIER. Contact: pierpaolo.battigalli@unibocconi.it.

‡Yale University. Contact: nicolo.generoso@yale.edu.

contagion" (Hatfield et al., 2014; Vásquez and Weretka, 2020). All in all, emotional expressions of others provide useful tools that can be exploited to make social interactions more predictable and manageable (see the survey by Van Kleef and Côté, 2022).

The effect of emotions on decision-making should be of primary interest for economists. Elster (1996, 1998) convincingly argued that a careful study of emotions could help to get a better grip on how decisions are formed, and in this regard psychological game theory – pioneered by Geanakoplos et al. (1989) and substantially developed by Battigalli and Dufwenberg (2009) and Battigalli et al. (2019a) – represents a rich framework to address the issue.[1] To the best of our knowledge, the role of emotional signals has never been formally analyzed. Incorporating such aspect in a formal analysis would yield a more accurate description of reality and new insights when strategic reasoning is studied. Observing such signals allows to make inferences about someone else's state of mind and, insofar as emotions are triggered by beliefs,[2] emotional signals may shed light also on the beliefs of others. Moreover, emotional signals may also depend on actions taken (e.g., lying may cause discomfort and hence emotional leakage), or on personal traits (e.g., a very emotional person may be more likely to betray her state of mind with, say, facial expressions). Therefore, the signals we introduce may allow players to draw conclusions not only on the *beliefs* of others, but also on their past *behavior* and *traits*. Such inferential reasoning can thus fruitfully inform strategic thinking.

In Section 1.1, we sketch out some heuristic examples to clarify the phenomena we aim to model. In Section 1.2, we present our contribution. In Section 1.3, we elaborate on our methodological position, and we briefly discuss the related literature.

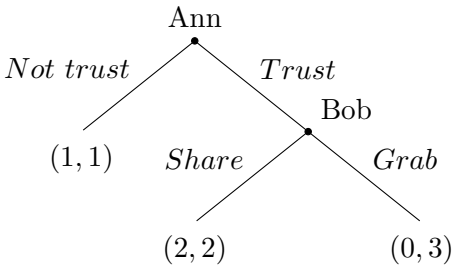
1.1 Heuristic examples

As hinted above, emotional signals can shed light on the emotions of others, on their personal traits, on their future behavior, and on past actions. We sketch some examples where this occurs. (Unless otherwise stated, game trees with payoff vectors at terminal nodes are interpreted as *game forms with monetary payoffs* that do not necessarily represent players' player's utilities.)

Example 1 (Trust mini-game). The following game form with monetary payoffs is widely used in the experimental literature, to assess whether guilt may shape the second mover's (Bob's) behavior.

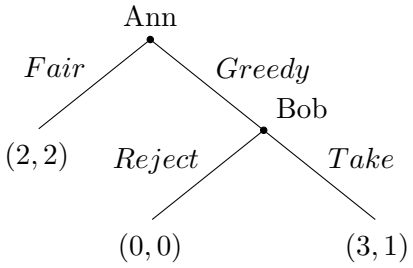
[1] The main innovation introduced by the theoretical apparatus of PGT consists in letting players' utilities depend on (their own and their opponents') beliefs. In this way, a wide array of belief-dependent sentiments and emotions, ranging from reciprocity to self-esteem, can be modeled. See Battigalli and Dufwenberg (2022) for a survey of recent developments in the literature.

[2] For instance, disappointment may be a consequence of unmet expectations about others' behavior and guilt may be generated by the failure to live up to (one's own or others') expectations.



Behrens and Kret (2019) find that face-to-face contact may foster cooperation and pro-social behavior. We can enrich the traditional representation of the Trust Mini-Game by allowing Bob to receive a signal about Ann's emotions before making his choice. A relevant emotional state in this setting is Ann's trustfulness – for instance, smiling may convey her desire to cooperate, and Bob would make inferences about Ann's *emotions* upon observing such signal. In this environment, trustfulness could be thought of as the extent to which Ann expects Bob to share. Upon seeing Ann smile, Bob may infer that she expects him to share, and this would provide him with further incentives to avoid letting her down. ▲

Example 2 (Ultimatum Mini-Game). Consider the following game form.



If Bob gets angry after receiving a greedy offer, he may decide to forego \$1 to punish Ann. Rejections at the second stage can be accounted for by the model of frustration and anger of Battigalli et al. (2019b). Van Leeuwen et al. (2018) suggest that individuals playing such game in the lab can infer how much their opponents are prone to anger by observing facial cues. Such cues cannot concern Bob's frustration, because frustration arises only as a consequence of others' choices. Nonetheless, facial cues may provide hints about a player's *personal trait*, that is, how prone one is to getting angry. ▲

Example 3 (Negotiation). Successful coordination and exchange of information are key in negotiations. Verbal (e.g., statements) and nonverbal (e.g., gesture) emotional expressions allow to infer the counterparts' intentions (Druckman and Olekalns, 2008). Elfenbein et al. (2007) find that individuals with a better emotion recognition accuracy attain better outcomes in negotiation exercises. In a stylized situation, we could imagine two agents engaging in an alternating-offer bargaining procedure. We could also assume that irritation may arise if one party receives an offer that is far from her minimal acceptable outcome, or that impatience may emerge as the negotiation lengthens. In the former case, emotional signal allow to better assess others' reservation values, which may be thought of as a *personal trait*. In the latter, one party may

realize that the other could settle for less advantageous terms to end the costly delay, therefore using emotional signals to make inferences about others' *future behavior*. ▲

Example 4 (Police interrogation). Police manuals recommend to pay attention to stereotypical cues such as gaze aversion and fidgeting to detect lies when questioning suspects. Whether this helps officers or not is unclear – in fact, evidence suggests that doing so may hamper lie detection (DePaulo et al., 2003). Yet, the majority of policemen participating in the experiment of Mann et al. (2004) declared that they look primarily at gaze aversion to detect lies in interrogations. Hence, they (perhaps mistakenly) use emotional cues to infer *past unobserved actions* of others and to infer whether the suspects' stories coincide with their actual past behavior. ▲

1.2 Our contribution

We develop a novel framework to model sequential psychological games where players receive signals, here called "messages", about their opponents' states of mind, in the form, e.g., of facial cues or involuntary behavior. In this regard, our contribution is twofold. First of all, an innovation is represented by the proposed framework, and by the incorporation of emotional feedback in game-theoretic analysis. More specifically, we allow such signals to be generated stochastically, and we take this generative process to be driven by the agents' states of mind. Relatedly, we also discuss how emotions are generated by players' beliefs and behavior as the game unfolds, and how signals about emotions allow to make inferences when reasoning strategically.

Second, we carry out an analysis of the key features that allow to derive behavioral predictions. We first give a definition of players' rationality as the conjunction of several requirements concerning players' cognitive sophistication and optimality of plans and behavior. We provide an explicit formal analysis of players' inferential reasoning and of players' rationality, showing that the set of states corresponding to the event "player i is rational" is a measurable subset of the set of states of the world. While measurability is essentially a mathematical property, it has relevant conceptual meaning, because we interpret measurable sets of states of the world as the events about which players can form their beliefs. Saying that rationality is an event implies that a given player may wonder about her opponents' rationality, incorporating such event into her strategic reasoning.

Finally, we propose a rationalizability-like solution concept to predict behavior, and we justify it in terms of underlying assumptions about players' rationality and interactive beliefs (Theorem 1). Such solution concept is particularly suited to our context, since it entails a form of forward-induction reasoning. That is, players try to make sense of (i.e., they try to *rationalize*) the information they receive as the game unfolds in a way that is consistent with their opponents being rational and strategically sophisticated. This solution concept is particularly compelling in our framework, because it captures the idea that players use emotional signals to infer their opponents' beliefs, private information, future behavior, and past unobserved actions. We apply our solution procedure to a simple situation, showing how the possibility of betraying false statements with emotional messages (e.g., by blushing) may represent a strong enough incentive for the disclosure of private information.

1.3 The bigger picture: methodology and related literature

In this section, we discuss some of our modeling choices, relating them to the existing literature. First, our work builds on the methodological paper of Battigalli et al. (2019a) in the way it models psychological games and belief-dependent motivations. Differently from such paper, states of the world include a description of *how players would behave* also at histories that do not realize, rather than just a description of *how the game unfolds*. In this, our approach mirrors that of Battigalli and De Vito (2021). Like them, we explicitly distinguish between players' plans and descriptions of behavior, requiring that they coincide for rational players, but may differ otherwise. This means that we do not assume that players necessarily know how they would behave at different contingencies: they can plan what to do, but they may fail to stick to their own plans.

Our approach to modeling rationality presents some innovations as well. Rationality is traditionally understood as the conjunction of several features, concerning both behavior and cognition. Some of these assumptions are typically implied by the modeling tools employed. For example, a "correct" updating policy is embedded in the definition of conditional probability systems (cf. Axiom 3 in Battigalli and Siniscalchi, 1999 and the analysis in Battigalli et al., 2023), which are conventionally used to model beliefs in sequential games, and transparency of coherence between beliefs of different orders follows from the choice of positing a type structure (cf. Battigalli and Siniscalchi, 1999 and Dekel and Siniscalchi, 2015). We instead construct a rich state space, and take the desired rationality features to be properties holding *only at some states* – this way, each requirement becomes an explicit assumption, represented by an event in a state space. Like in Battigalli et al. (2020), we do not posit a type structure, and we take instead an infinite hierarchical system of beliefs to be the epistemic/doxastic type of a player: with this, a player's way of thinking is described by a map that associates an infinite hierarchy of beliefs to each history she may observe. In a state of the world, such descriptions of "ways of thinking" will be coupled with descriptions of behavior and with personal traits. For rational players, we impose some cognitive sophistication properties (i.e., that beliefs of different orders be coherent, and that beliefs be updated consistently with evidence and according to the rules of conditional probabilities), as well as the requirement that rational players plan optimally and implement their plans. All in all, our notion of rationality shares similarities with the traditional one, but it is more explicit and more structured.

With this approach, showing the measurability of the event "player i is rational" is non-trivial. Even if this comes at a cost, we believe that our language features enough flexibility to model a wide variety of cognitive failures and behavioral inconsistencies. The richness of our framework also allows to let players entertain the possibility that some of their opponents be in some sense unsophisticated.[3] Such a level of expressiveness seems to be a prerequisite for

[3]In contrast to our approach, in a canonical type structure, types are collectively coherent hierarchies of conditional beliefs (in the words of Dekel and Siniscalchi, 2015). This means that, by construction, the possibility that an opponent features —for example— some incoherence between her first- and second-order beliefs is inconceivable for any player.

the introduction of elements of bounded rationality (or, more generally, of departures from a canonical notion of rationality) in strategic settings, as well as for a rigorous analysis of such phenomena.

Lastly, we build on our analysis of rationality to formalize a solution concept that captures the implications of meaningful hypotheses about players' rationality and strategic reasoning, that we interpret as *common strong belief in rationality*. Our solution concept is a version of strong directed rationalizability (a.k.a. strong Δ-rationalizability, see the textbook of Battigalli et al., 2025, Battigalli, 2003, Battigalli and Siniscalchi, 2003, Battigalli and Tebaldi, 2019 and relevant references therein), which characterize in standard settings the utility-relevant implications of rationality, some belief restrictions, and common strong belief in both (Battigalli and Prestipino, 2013). We prove that the same holds in our framework (Theorem 1). Our epistemic analysis is different from the usual one because of our type-structure-free approach. Our result establishes that a procedure carried out taking into account only beliefs of a finite order captures the implications of epistemic assumptions that are formulated in terms of infinite hierarchies of beliefs. This is in the same spirit of Battigalli et al. (2020), and it leverages technical results proved in Battigalli and Tebaldi (2019).

Roadmap The paper is organized as follows. Section 2 introduces our framework. Section 3 formalizes the inferential reasoning players carry out upon observing messages about their opponents. Section 4 defines rationality. Section 5 introduces the solution concept. Section 6 provides the epistemic justification for the proposed procedure. Section 7 concludes.

2 Formal framework

In the following, for each compact metrizable topological space S, we denote by $\mathcal{B}(S)$ its Borel σ-algebra and by $\Delta(S)$ the space of Borel probability measures on S. Sets of probability measures are endowed with the topology of weak convergence, Cartesian products with the product topology, finite sets with the discrete topology, and subsets of topological spaces with the relative topology. Moreover, we maintain that the (finite) set of players is I, and that the games we model unfold within a single period and last at most $L \in \mathbb{N}$ stages.

For a set X and for each $n \in \mathbb{N}$, we let X^n denote the n-fold product of X, with generic element x^n. Moreover, given $\bar{x}^n \in X^n$ with $n \in \mathbb{N}$, we let $\bar{x}_k$ denote its k-th coordinate (with $k \in \{1, \dots, n\}$). Lastly, we also define $X^0 := \{\varnothing_X\}$, i.e., the singleton containing the empty sequence of elements of X.

The remainder of this section is organized as follows. Section 2.1 describes how emotions shape feedback and utility. Section 2.2 constructively derives the game tree. Section 2.3 describes players' predispositions to act and to believe as the game unfolds, and relates such attitudes to the generation of emotions. Section 2.4 further elaborates on utility functions.

2.1 Emotions, messages, and utility

We start by describing how emotional feedback is generated and how emotions determine utilities. Emotions are understood as broad categories, not necessarily tied to specific situations.[4] Therefore, our focus here will be independent from any game, and we will embed emotions in specific contexts only later.

First, for each $i \in I$, we denote the (nonempty) finite set of *personal traits of agent* i as Θ_i, and the (nonempty) compact metrizable set of *emotional states* (henceforth simply *emotions*) of agent i as E_i. We let $\Theta := \bigtimes_{i\in I} \Theta_i$ and $E := \bigtimes_{i\in I} E_i$ denote the set of profiles of traits and emotions, respectively. Agents experience streams of emotions: for each $\ell \in \{1, \ldots, L+1\}$, E^ℓ is the set of *streams of emotion profiles* of length ℓ. Given that we will model games lasting at most L stages, we consider the set $E^{\leq L+1} := \bigcup_{\ell=1}^{L+1} E^\ell$, which represents the possible streams of emotions experienced by agents in the situation of interest.[5] For each $i \in I$, $E_i^{\leq L+1}$ has an analogous meaning.

We posit, for each $i \in I$, a (nonempty) finite set of *conceivable emotional messages* (or signals), $M_{i,\mathrm{e}}$, and we let $M_\mathrm{e} := \bigtimes_{i\in I} M_{i,\mathrm{e}}$.[6] Furthermore, for each $i \in I$, let Y_i be the finite (nonempty) set of *material outcomes*, and define the set of *collective outcomes* as $Y := \bigtimes_{i\in I} Y_i$.

We now turn to the key elements of our analysis. First, we define a continuous *feedback function about emotions and traits* $\tilde{f}_\mathrm{e} : \tilde{A} \times \Theta \times E^{\leq L+1} \to \Delta(M_\mathrm{e})$, where $\tilde{A} = \bigtimes_{i\in I} \tilde{A}_i$ is a generic finite but "universal" set of action profiles that can be taken by agents. We let messages be generated stochastically because messages about emotions are noisy. Note that we allow the message generation to depend also on actions agents can take,[7] as well as on their traits. Second, we define a profile of continuous *psychological utility functions* $(\tilde{v}_i : Y \times \Theta \times E^{\leq L+1} \to \mathbb{R})_{i\in I}$. Differently from conventional utilities, they do not depend only on outcomes and traits, but also on the streams of emotions experienced by players.

2.2 The game tree

We now move to the description of game-specific aspects. Although the peculiarity of our framework is that players receive messages related to the emotions and traits of others, as the play unfolds they also receive messages about previous moves, or "previous-play messages". We do not assume players necessarily observe their co-players" previous moves – they only receive some "previous-play messages" that contain some information about how the game has been played up to a given point. As a special case, such messages may exactly pin down the actions chosen by others. Whenever a player is called upon to act, her available actions are self-evident, regardless of whether she perfectly recalls how the game unfolded up to that point. Given that

[4] For instance, someone may get angry if his favorite football team loses or if he is disappointed by the behavior of someone – the emotion experienced is arguably the same, but the situations that triggered it may be different.

[5] Players can experience a stream of emotions of length at most $L+1$ because we assume that they hold some initial emotional state, and then they experience a new one after each stage of the game.

[6] It is useful to assume $M_{i,\mathrm{e}} = \bigtimes_{j\in I\setminus\{i\}} M_{i,j,\mathrm{e}}$, where $M_{i,j,\mathrm{e}}$ is interpreted as the set of messages about j's emotions that i can observe. Whenever $I = \{i, j\}$, $M_{i,\mathrm{e}}$ (resp. $M_{j,\mathrm{e}}$) is isomorphic to $M_{i,j,\mathrm{e}}$ (resp., $M_{j,i,\mathrm{e}}$).

[7] In a game, such actions will be the ones agents can play at a given stage.

the game-specific information players receive is encoded in previous play messages, we posit that the last such message received directly informs a player of her feasible actions at the next stage.[8] This way of modeling players' information throughout the game is the one proposed by Battigalli and Generoso (2024) – we refer the reader to such paper for a more detailed discussion of the conceptual and methodological issues involved in our modeling choice.

For each $i \in I$, we let A_i (with $\emptyset \neq A_i \subseteq \tilde{A}_i$) be the finite set of *potentially available actions* of player i in the given game, and $M_{i,\mathrm{p}}$ the finite set of *previous play messages* player i can receive. In our framework, feedback pertains to both the actions previously chosen (and possibly not observed) by players and the emotional states of others. The Roman subscript in our notation is a mnemonic for these domains. We let $A := \bigtimes_{i\in I} A_i$ and $M_{\mathrm{p}} := \bigtimes_{i\in I} M_{i,\mathrm{p}}$ be the sets of profiles of actions and previous play messages, respectively. We also posit a *previous play message generating function*, $\tilde{f}_{\mathrm{p}} : \bigcup_{\ell=1}^{L} A^{\ell} \to M_{\mathrm{p}}$. Note that feedback about previous play – unlike feedback about emotions – is game-dependent, as it is generated according to the rules of the game. Unlike emotional messages, previous play messages are produced deterministically, and letting $\tilde{f}_{\mathrm{p}}$ be the map $a^{\ell} \mapsto (a_i^{\ell})_{i\in I}$ amounts to assuming observed actions. For each $a^{\ell} \in A^{\ell}$, $\ell \in \{1, \ldots, L\}$, and $i \in I$, we let $\tilde{f}_{i,\mathrm{p}}(a^{\ell}) := \mathrm{proj}_{M_{i,\mathrm{p}}} \tilde{f}_{\mathrm{p}}(a^{\ell})$.

We also posit, for each $i \in I$, an *action feasibility correspondence*, $\mathcal{A}_i : M_{i,\mathrm{p}} \cup \{\varnothing_{M_{i,\mathrm{p}}}\} \rightrightarrows A_i$. The interpretation is that $\mathcal{A}_i(m_{i,\mathrm{p}})$ is the set of actions available to her after receiving previous-play message $m_{i,\mathrm{p}}$ (i.e., in the subsequent stage). Moreover, $\varnothing_{M_{i,\mathrm{p}}}$ stands for the situation in which player i has not received any message yet, so that $\mathcal{A}_i(\varnothing_{M_{i,\mathrm{p}}})$ represents the actions player i can choose at the beginning of the game. It is convenient to define $\mathcal{A} : M_{\mathrm{p}} \cup \{\varnothing_{M_{\mathrm{p}}}\} \rightrightarrows A$ to be such that $\mathcal{A}(m_{\mathrm{p}}) := \bigtimes_{i\in I} \mathcal{A}_i(m_{i,\mathrm{p}})$ for each $m_{\mathrm{p}} = (m_{i,\mathrm{p}})_{i\in I}$, and $\mathcal{A}(\varnothing_{M_{\mathrm{p}}}) := \bigtimes_{i\in I} \mathcal{A}_i(\varnothing_{M_{i,\mathrm{p}}})$. To describe the end of the game, we assume that, for each $m_{\mathrm{p}} = (m_{i,\mathrm{p}})_{i\in I}$ and $i \in I$, $\mathcal{A}_i(m_{i,\mathrm{p}}) = \emptyset$ if and only if $\mathcal{A}_j(m_{j,\mathrm{p}}) = \emptyset$ for each $j \in I$. In such case, $\mathcal{A}(m_{\mathrm{p}}) = \emptyset$ as well. In words, as soon as the game is over for one player, it is over for everyone.[9]

We take histories to be sequences of profiles of actions, previous-play messages, and messages about emotions and traits. With $\tilde{f}_{\mathrm{p}}$, $\tilde{f}_{\mathrm{e}}$, and $(\mathcal{A}_i)_{i\in I}$ as primitive elements of our analysis, we can give a constructive definition of the set $\bar{H}$ of *feasible histories*. A history is feasible if, at each stage, (i) the sequence of actions played is allowed by the rules of the game (specifically, by $(\mathcal{A}_i)_{i\in I}$ and $\tilde{f}_{\mathrm{p}}$), and (ii) the previous-play and emotional messages can be generated with positive probability given the feedback functions $\tilde{f}_p$ and $\tilde{f}_{\mathrm{e}}$.[10] For convenience, we assume that the empty history $\varnothing$ belongs to $\bar{H}$.[11] The set of *terminal histories* is $Z := \{h = (a^{\ell}, m_{\mathrm{p}}^{\ell}, m_{\mathrm{e}}^{\ell}) \in$

[8] For instance, the average amateur chess player arguably cannot remember the entire sequence of moves at all the stages of the game. Yet, the disposition of pieces on the chessboard informs him of his feasible moves. For instance, if his king is under check, he can understand which are the legitimate moves he can take (if any) based on such disposition. One can think of previous play messages (e.g. the piece disposition) as summary indicators that (perhaps imperfectly) aggregate past moves and that provide all the information needed to be able to continue the game.

[9] This means that players who are at some stage inactive actually have only one feasible action (say, a dummy action "wait"), which will always be neglected in our notation.

[10] We give a formal definition of feasibility in Appendix C.

[11] The empty history can be thought of as a history of length zero where no action has been played and no

$H : \mathcal{A}(m_{\mathrm{p},\ell}) = \emptyset\}$, and the set of *non-terminal histories* is $H := \bar{H} \setminus Z$.

Note that, at each stage, agents first act, and then observe messages. The previous-play message profile generated at some stage k depends on the entire sequence of action profiles played up to that stage, while emotional feedback depends on actions played, personal traits, and emotions. To ease notation, we let $M := M_{\mathrm{p}} \times M_{\mathrm{e}}$, with generic element $m = (m_{\mathrm{p}}, m_{\mathrm{e}})$ describing in a concise way the (previous-play and emotional) feedback received by all players. The set $M_i := M_{i,\mathrm{p}} \times M_{i,\mathrm{e}}$, with generic element m_i, has analogous meaning.

The assumption that players need not observe others' actions or messages justifies the introduction of the set of *personal histories* of any player i, defined as $\bar{H}_i := \mathrm{proj}_{\bigcup_{\ell=0}^{L} A_i^\ell \times M_i^\ell} \bar{H}$. The set $\bar{H}_i$ collects all the information – in terms of actions played and messages received – player i may have access to as the game unfolds. The sets H_i and Z_i represent the sets of personal non-terminal and terminal histories, respectively. Thus, $\{H_i, Z_i\}$ is a partition of $\bar{H}_i$.

A (weak) "prefix of" relation $\preceq$ can be defined on $\bar{H}$. Given $\hat{h} = (\hat{a}^k, \hat{m}^k), h = (a^\ell, m^\ell) \in \bar{H}$, $\hat{h} \preceq h$ if either $\hat{h} = h$ or $k < \ell$ and $(\hat{a}^k, \hat{m}^k) = (a^k, m^k)$. If $\hat{h} \preceq h$, we say that $\hat{h}$ (weakly) precedes h. Since $\varnothing \in \bar{H}$, it is easy to check that $\bar{H}$, partially ordered by $\preceq$, is a tree, and that the same holds for $\bar{H}_i$.

Lastly, a *consequence function* $\pi : Z \times \Theta \to Y$ specifies how outcomes accrue to players at the end of the game. For each $i \in I$ and $(z, \theta) \in Z \times \Theta$, we let $\pi_i(z, \theta) := \mathrm{proj}_{Y_i} \pi(z, \theta)$. We conclude this section introducing our running example.

Example 5 (Buy me an ice-cream). Child is at home alone and he should do his homework, but he is tempted to play video-games. When Dad gets back from work, Child asks him to buy him an ice-cream. Dad would be happy to reward Child, but he does not know if his son studied. He simply asks him if he has done his homework, and to decide based on the answer. To make the problem more interesting we add two features. First, we assume Child is concerned about his image in Dad's eyes: he dislikes being thought of as a liar, regardless of whether he actually lied or not.[12] Second, we assume that Child may blush when he falsely claims that he has done his homework.

The set of Child's potentially available actions is $A_C := \{w, v, yes, no\}$, where the elements denote doing homework, playing video-games, saying "yes," and saying "no," respectively. As for Dad, we let $A_D := \{buy, not\}$, because he can either buy Child an ice-cream or not. Only Dad observes emotional messages throughout, so let $M_{D,\mathrm{e}} := \{b, \neg b, n\}$, whose elements respectively stand for "blushing," "not blushing," and "uninformative message," and $M_{C,\mathrm{e}} := \{n\}$. Lastly, assume Θ_D is a singleton and let $\theta_C \in \Theta_C \subset \mathbb{R}_+$ denote Child's appreciation for video-games.

We model the situation as follows. Child first privately chooses between homework and video-games, then he answers "yes" or "no" to Dad, and lastly Dad decides whether to buy the ice-cream. Child observes all the actions taken, while Dad observes only the actions taken from the second stage onward. To capture this flow of information, for each $a \in \{w, v\}$, $a' \in \{yes, no\}$,

message has been received yet – i.e., $\varnothing_H = (\varnothing_A, \varnothing_{M_p}, \varnothing_M)$. To simplify notation, we denote it simply as $\varnothing$.

[12] This is a form of image concern. In particular, in our case the concern depends on others' opinions about good actions, i.e., not lying (see Battigalli and Dufwenberg, 2022).

and $a'' \in \{buy, not\}$, we can define function $\tilde{f}_{\mathrm{p}}$ to be such that $(a) \mapsto (a, \bar{a})$, $(a, a') \mapsto \big((a, a'), a'\big)$, $(a, a', a'') \mapsto \big((a, a', a''), (a', a'')\big)$, where the two components are Child's and Dad's previous-play messages, respectively, and $\bar{a}$ is an uninformative message. Action feasibility correspondences are defined in an obvious way.

Recall that we would like to model a situation where Child has image concerns and may blush with positive probability only if he lies after playing video-games. Relevant emotions in this settings are confidence, guilt, and blame. Child is confident if he thinks he can get away with his lie, he might feel guilty for not doing his homework, and he dislikes Dad's blame. A profile of emotional states is an element of the set $E := [0, 1] \times \{0, 1\} \times [0, 1]$, with its three components representing confidence, guilt, and blame, respectively. Confidence and guilt shape emotional feedback in the "second stage" of the situation we have in mind (i.e., when Child decides what to tell Dad). Denote these emotions as $\mathbf{c}_2$ and $\mathbf{g}_2$, where the subscript reminds that these are Child's emotional states during the second stage, and the boldface font is used to distinguish emotions from other objects. Dad's blame instead matters at the end of the game, because Child cares about what Dad eventually thinks of him. Denote such emotion as $\mathbf{b}_3$. To ease notation, we neglect the emotions held at other points of the interaction.

Then, we can assume that Child may blush only if he feels guilty for not doing his homework, and that the probability of not blushing is equal to his confidence: for each $(a, \theta, e^2) \in A \times \Theta \times E^2$,

$$\tilde{f}_{\mathrm{e}}(a, \theta, e^2) = \begin{cases} \mathbf{g}_2\big(\mathbf{c}_2\delta_{\neg b} + (1 - \mathbf{c}_2)\delta_b\big) + (1 - \mathbf{g}_2)\delta_{\neg b} & \text{if } a = yes; \\ \delta_{\neg b} & \text{if } a = no; \end{cases} \tag{1}$$

and equal to δ_n in all other cases.[13] This formulation implies that message b may be generated only after Child's second-stage action and only if he says "yes." Also note that personal trait θ does not affect emotional feedback in this example.

It is natural to try to tie the emotions just discussed to a more structured model. For instance, we hinted at the fact that guilt may arise if Child plays video-games. We will elaborate on this (cf. p. 14) and we will explain how to embed emotions into an interactive situation. Eventually, we will obtain that b may realize only if Child plays video-games and subsequently says "yes." For the moment, we leave the description of emotions and feedback unstructured. This means that Dad can observe a trivial length-one personal history (where he waits and observes uninformative signals about Child's action and emotions) and three length-two personal histories, identified with (yes, b), $(yes, \neg b)$, and $(no, \neg b)$.

Lastly, we describe utility functions. Child's is the most interesting. Recall that he dislikes Dad's blame $\mathbf{b}_3$. In terms of material outcomes, let $Y_C := \{0, 1\}^2$, with generic element $(y_{C,1}, y_{C,2})$, and with the two coordinates indicating whether Child eats the ice-cream and whether he plays video-games, respectively. Then, define

$$\tilde{v}_C(y, \theta, e^4) := y_{C,1} + \theta y_{C,2} - \mathbf{b}_3. \tag{2}$$

[13] We report only Dad's message as subscript, as Child only observes uninformative messages. Moreover, we report the argument e^2 to be consistent with the notation used in the main text, because the relevant feedback is generated in the "second stage' (i.e., after players have experienced a stream of emotions of length 2).

As for Dad, he incurs a cost of 1 to buy the ice-cream and he gets a payoff of 2 if he buys the ice-cream when Child did his homework (and 0 otherwise). ▲

2.3 Predispositions to act and believe

We now define "states of the world," which we take to be complete descriptions of players' traits and predispositions to act and believe. The term "predisposition," suggests that we do not want to define only players' behavior and beliefs along the path of the game, but rather how they *would* behave and what they *would* believe conditional on all possible contingencies. A state of the world therefore encompasses all the relevant aspects of a strategic situation.

2.3.1 Behavior

Our first building block is a complete description of a player's behavior conditional on different personal histories. To define such objects, we introduce, for each player $i \in I$, the correspondence $\hat{\mathcal{A}}_i : H_i \rightrightarrows A_i$, where, for each $h_i \in H_i$, $\hat{\mathcal{A}}_i(h_i) := \{a_i \in A_i : \exists\, m_i \in M_i, (h_i, a_i, m_i) \in \bar{H}_i\}$. In words, $\hat{\mathcal{A}}_i(h_i)$ is the set of player i's available actions after personal history h_i. For each $i \in I$, we define the set of i's *personal external states* as:

$$S_i := \bigtimes_{h_i \in H_i} \hat{\mathcal{A}}_i(h_i).$$

The set of personal external state profiles is $S := \bigtimes_{i \in I} S_i$, and we call $s \in S$ an *external state*.

A personal external state is a map from non-terminal personal histories to feasible actions. Elements of S_i can technically be labeled as player i's "strategies", but we refrain from using such terminology because we maintain that strategies are plans *in the minds of players*. In particular, we will allow player i to form beliefs about her own behavior (i.e., over the set S_i): such beliefs are interpreted as the way in which a player expects herself to behave in the future. Importantly, a complete description of player i's contingent behavior, s_i, *may or may not* coincide with what she planned to do before the game started.

2.3.2 Beliefs

We now discuss how to give a complete description of the epistemic features of a player. The mathematical description of a player's way of thinking is a hierarchical system of beliefs, that is, a map from personal histories to hierarchies of beliefs. We define such objects inductively.

First, define the space of *primitive uncertainty* to be $\Omega^0 := S \times \Theta$. This is the basic uncertainty space upon which players form their first-order beliefs.[14] A system of first-order beliefs is any function that maps from $\bar{H}_i$ to the set of Borel probability measures on Ω^0. Therefore, the set of *systems of first-order beliefs* of player i is $\mathcal{T}_{i,1} := [\Delta(\Omega^0)]^{\bar{H}_i}$. We define the sets of profiles of first-order beliefs of players other than i as $\mathcal{T}_{-i,1} := \bigtimes_{j \in I \setminus \{i\}} \mathcal{T}_{j,1}$. Lastly, for each $i \in I$, we let $\Omega^1_{-i} := \Omega^0 \times \mathcal{T}_{-i,1}$.

[14] Note that players form beliefs also over their own traits and personal external states. Later on, we will make the assumption that rational players know their personal traits, while beliefs about one's own behavior are interpreted as players' plans.

Assume now that Ω_{-i}^{k-1}, $\mathcal{T}_{i,k-1}$, and $\mathcal{T}_{-i,k-1}$ have been defined for each $i \in I$ and $k \in \{2,\dots,n\}$. Then, define:

$$\mathcal{T}_{i,n} := \left[\Delta(\Omega_{-i}^{n-1})\right]^{\bar{H}_i}, \quad \mathcal{T}_{-i,n} := \bigtimes_{j\in I\setminus\{i\}} \mathcal{T}_{j,n}, \quad \Omega_{-i}^{n} := \Omega_{-i}^{n-1} \times \mathcal{T}_{-i,n} = \Omega^0 \times \left(\bigtimes_{k=1}^{n} \mathcal{T}_{-i,k}\right).$$

The set of *systems of n-th-order beliefs* of player i is $\mathcal{T}_{i,n}$. We define also $\mathcal{T}_n := \bigtimes_{i\in I} \mathcal{T}_{i,n}$. As a matter of notation, $\mathcal{T}_{i,n}$ denotes the set of systems of n-th order beliefs, while the set of hierarchies of systems of beliefs of order up to n will be denoted as $\mathcal{T}_i^n$, as specified below.

We let the set of *n-th-order hierarchical systems of beliefs* (with $n \in \mathbb{N}$) and the set of *infinite hierarchical systems of beliefs* of player i be, respectively,

$$\mathcal{T}_i^n := \bigtimes_{k=1}^{n} \mathcal{T}_{i,k} = \left[\left(\Delta(\Omega^0)\right)^{\bar{H}_i}\right] \times \bigtimes_{k=2}^{n} \left[\left(\Delta(\Omega_{-i}^{k-1})\right)^{\bar{H}_i}\right], \quad \mathcal{T}_i^\infty := \bigtimes_{n\in\mathbb{N}} \mathcal{T}_{i,n}.$$

Define also $\mathcal{T}_{-i}^n := \bigtimes_{k=1}^n \mathcal{T}_{-i,k}$, $\mathcal{T}^n := \bigtimes_{k=1}^n \mathcal{T}_k$, $\mathcal{T}_{-i}^\infty := \bigtimes_{j\in I\setminus\{i\}} \mathcal{T}_j^\infty$ and $\mathcal{T}^\infty := \bigtimes_{i\in I} \mathcal{T}_i^\infty$.[15]

A generic $\tau_i^\infty \in \mathcal{T}_i^\infty$ is an *epistemic type* of player i. Taking an infinite hierarchical system of beliefs as the epistemic type of a player allows us to conduct an epistemic analysis without resorting to a type structure. The interpretation of such objects is similar to that of personal external states: τ_i^∞ represents a complete description of what player i would believe at different contingencies. Unlike personal external states, we informally assume players know their epistemic types. Finally, note that we have not imposed any requirement in the construction above: cognitive sophistication properties will be modeled as features that hold only at some states.

Remark 1. For each $i \in I$ and $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{T}_i^n$ is compact metrizable.[16]

We conclude with a notational clarification. For each $n \in \mathbb{N}$, $i \in I$, $\tau_{i,n} \in \mathcal{T}_{i,n}$, and $h_i \in \bar{H}_i$, to ease interpretation we denote $\tau_{i,n}(h_i) \in \Delta(\Omega_{-i}^{n-1})$ by $\tau_{i,n}(\cdot\,|h_i)$. Indeed, recall that $\tau_{i,n}$ selects a n-th-order belief for each personal history, and such notation suggests that such belief is the one held by player i conditional on observing personal history h_i. Moreover, given $n \in \mathbb{N}$ and $\tau_i^n \in \mathcal{T}_i^n$, we write $\tau_i^n(\cdot\,|h_i)$ as a shorthand for $(\tau_{i,k}(\cdot\,|h_i))_{k=1}^n$. To ease notation and with a small abuse, given two generic topological spaces X and Y and a measure $\mu \in \Delta(X \times Y)$, for each $A \subseteq X$, we write $\mu(A)$ instead of $\mu(A \times Y)$. Therefore, expressions such as $\tau_{i,n}(\{s_{-i}\}|h_i)$ should be read as $\tau_{i,n}(S_i \times \{s_{-i}\} \times \Theta \times \mathcal{T}_{-i}^{n-1}|h_i)$.

2.3.3 States of the world

We can now define the set of *states of the world* as $\Omega^\infty := \Omega^0 \times \mathcal{T}^\infty$, and measurable subsets of Ω^∞ are *events*. For each $i \in I$, $S_i \times \Theta_i \times \mathcal{T}_i^\infty$ is instead the set of *personal states* of player i.

[15] Note that it is possible to write $\Omega_i^n = \Omega^0 \times \mathcal{T}_i^n$, for each $i \in I$ and $n \in \mathbb{N}$. This explains the presence of superscripts in our notation.

[16] Given that Ω^0 is finite, it is compact metrizable and so is $\Delta(\Omega^0)$ (Aliprantis and Border, 2006, Theorem 15.11). Tychonoff's theorem and Theorem 3.36 of Aliprantis and Border (2006) imply that $\mathcal{T}_{i,1}$, Ω_i^1, and Ω_{-i}^1 are compact metrizable as they are countable products of compact metrizable spaces. An inductive argument shows that $\mathcal{T}_{i,n}$, Ω_i^n, and Ω_{-i}^n are compact metrizable. With this, for each $i \in I$ and $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{T}_i^n$ is a countable product of compact metrizable spaces, and it is therefore compact metrizable as well.

Remark 2. Ω^∞ and $S_i \times \Theta_i \times \mathcal{T}_i^\infty$ are compact metrizable.

A state of the world provides all the relevant game-specific aspects about players, as it encodes their traits and a complete description of their behavior and their beliefs conditional on each possible contingency that may arise as the game unfolds. Throughout, we will interpret measurable sets of states of the world as those events that can be evaluated by players' beliefs of some order. We will show that, under a belief coherence property, it is *as if* players formed their beliefs on Ω^∞ (cf. Lemma 3). Events in Ω^∞ such as "a player is rational" (cf. Lemma 8) can then be assessed by coherent players, and this will be key in defining a theory of strategic reasoning (cf. Section 6; specifically, Lemma 10).

2.3.4 Epistemic types, game unfolding, and emotions

States of the world capture all the game-specific attitudes of players. Yet, we still need to explain how emotions are triggered by players' behavior and beliefs as the game unfolds. It seems reasonable to think that feelings such as surprise, guilt, or frustration arise from the unfolding of the game (e.g., from players' choices) and from endogenous beliefs (e.g., from player's expectations). In our running example we introduced broad concepts such as guilt, distrust, or blame, but the situation at hand also suggested a very natural interpretation of such emotions (e.g., Child feels guilty if he plays video-games instead of studying). Our aim is to tie streams of emotions experienced by players during the game to states of the world.

First, we discuss how players' beliefs are realized as the game unfolds. The realized beliefs of a player at some personal history are the beliefs held by that player at predecessors of such history (i.e., along the "path" that led to such history). We define a profile of *realized-beliefs functions* $\rho := (\rho_h)_{h \in \bar{H}}$, where, for each $h = (h_i)_{i \in I} \in \bar{H}$, ρ_h is the map $\tau^\infty \mapsto ((\tau_i^\infty(\cdot\,|h_i'))_{h_i' \preceq h_i})_{i \in I}$. In words, $\rho_h(\tau^\infty)$ is the stream of belief profiles realized along h.[17]

Then, we define a continuous *emotion-generating function*, $\varepsilon : \bar{H} \times \mathcal{T}^\infty \to \Delta(E^{\leq L+1})$, and we make the following assumptions about it. First, only realized beliefs matter in the generation of emotions: for each $h = (h_i)_{i \in I} \in \bar{H}$, the section of ε at h is given by $\varepsilon_h := \bar{\varepsilon}_h \circ \rho_h$, with $\bar{\varepsilon}_h : \rho_h(\mathcal{T}^\infty) \to \Delta(E^{\leq L+1})$. Second, along histories of a given length $\ell \in \{1, \ldots, L\}$ players experience streams of emotion profiles of length $\ell + 1$:[18] for each $h \in \bar{H}^\ell$, $\operatorname{supp} \varepsilon_h \subseteq E^{\ell+1}$. Third, we posit a belief-order $K \in \mathbb{N}$ such that beliefs of order higher than $K \in \mathbb{N}$ are irrelevant for the generation of emotions: for each $\tau^\infty, \bar{\tau}^\infty \in \mathcal{T}^\infty$, $\tau^K = \bar{\tau}^K$ implies $\varepsilon(h, \tau^\infty) = \varepsilon(h, \bar{\tau}^\infty)$ for each $h \in \bar{H}$. For simplicity, we write the argument of ε directly as elements of $\bar{H} \times \mathcal{T}^K$.

By linking the generation of emotions to game-specific contingencies, function ε completes the definition of a game with feedback about emotions. A *game with feedback about emotions* is a structure $\Gamma := \langle I, \mathcal{A}, \tilde{f}_{\mathrm{p}}, \tilde{f}_{\mathrm{e}}, \pi, \varepsilon, (\Theta_i, A_i, M_{i,\mathrm{p}}, M_{i,\mathrm{e}}, Y_i, E_i, \tilde{v}_i)_{i \in I} \rangle$, with its components as defined in previous sections. It is informally assumed that these elements are commonly known.

[17] A brief comment on notation. Indexing objects by (personal) histories should be read as "at such history." So, for instance, $\rho_h(\tau^\infty)$ are the beliefs realized at h, $u_{i,h_i}(s, \theta, \tau^K)$ the utility i expects at h_i given (s, θ, τ^K), and $\zeta(z|h_i; s, \theta, \tau^K)$ the probability of z occurring given (s, θ, τ^K) and conditional on h_i occurring (cf. Section 4.4).

[18] Indeed, it is reasonable to assume that an emotion profile is generated after each stage. Hence, a given length-ℓ history induces one emotion for each of its $\ell + 1$ weak predecessors.

Next, we retrieve a *profile of feedback functions about beliefs* (or, simply, *emotional feedback*) $f_{\mathrm{e}} := (f_{h,\mathrm{e}} : S \times \Theta \times \mathcal{T}^K \to \Delta(M_{\mathrm{e}}))_{h \in H}$. For each $h \in H$, $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$ and $m_{\mathrm{e}} \in M_{\mathrm{e}}$, let

$$f_{h,\mathrm{e}}(s, \theta, \tau^K)[m_{\mathrm{e}}] := \int_{E^{L(h)+1}} \tilde{f}_{\mathrm{e}}(s(h), \theta, e^{L(h)+1})[m_{\mathrm{e}}] \cdot \varepsilon(h, \tau^K)[\mathrm{d}e^{L(h)+1}], \tag{3}$$

where $L(h)$ denotes the length of each history $h \in \bar{H}$. Continuity of $\tilde{f}_{\mathrm{e}}$ and ε implies the following.

Remark 3. For each $h \in H$, $f_{h,\mathrm{e}}$ is continuous.

Note that the domain of feedback functions is now a set over which players form well-defined beliefs. In Section 3, we present assumptions on f_{e} that allow to prove technical results.

We also express the feedback about previous play in a way that depends on profiles (s, θ, τ^K). To do so, define $f_{\mathrm{p}} := (f_{h,\mathrm{p}} : S \times \Theta \times \mathcal{T}^K \to \Delta(M_{\mathrm{p}}))_{h \in H}$ to be such that, for each $h = (a^t, m^t) \in H$, $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$ and $m_{\mathrm{p}} \in M_{\mathrm{p}}$,

$$f_{h,\mathrm{p}}(s, \theta, \tau^K)[m_{\mathrm{p}}] = \begin{cases} 1 & \text{if } \tilde{f}_{\mathrm{p}}(a^t, s(a^t)) = m_{\mathrm{p}}; \\ 0 & \text{otherwise.} \end{cases}$$

In words, $f_{h,p}(s, \theta, \tau^K)$ is a degenerate probability measure concentrated on the message about previous play that would be generated by $\tilde{f}_{\mathrm{p}}$ if players behave as described by s after history h. Note that f_{p} does not depend on personal traits or on systems of beliefs. This is consistent with the idea that previous-play messages only pertain to (past) behavior, which is entirely summarized by personal external states.

Finally, we let $f := (f_h : S \times \Theta \times \mathcal{T}^K \to \Delta(M))_{h \in H}$ summarize the generation of messages about both previous play and emotions. To this end, for each $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$, define

$$f_h(s, \theta, \tau^K) := f_{h,\mathrm{p}}(s, \theta, \tau^K) \otimes f_{h,\mathrm{e}}(s, \theta, \tau^K), \tag{4}$$

where $\otimes$ denotes the product of measures. Consistently with the notation used so far, we let $f_{i,h} = \mathrm{marg}_{M_i} \circ f_h$ for each $i \in I$.

Example 5 (Buy me an ice-cream, continued). A generic personal external state of Dad is indicated as $a_1.a_2.a_3$, where a_1, a_2, and a_3 are the actions prescribed after histories (yes, b), $(yes, \neg b)$, and $(no, \neg b)$, respectively. A generic personal external state of Child is instead $a_1.a_2$, with a_1 (resp., a_2) denoting the first-stage (resp., second-stage) action he would play.[19]

Defining the generative process for all the streams of emotion profiles is notationally costly. To ease the exposition, we only define how the emotions appearing in equations (1) and (2) (cf. p. 10) are generated. For simplicity, we further assume emotions to be generated deterministically,

[19] Actually, according to our definition, a personal external state of Child should be a map from the set of histories where Child is active, $\{\varnothing, (V), (H)\}$, to A_C. Letting Child's personal external states take the form of $a_1.a_2$ amounts to not specifying the action prescribed at the (personal) history that is not allowed for by the first-stage action. This is inconsequential, and it comes with an advantage in terms of parsimony of notation.

and we let $K = 1$. First, Child is *guilty* if he plays video-games instead of doing his homework. Hence, we simply impose $\mathbf{g}_2 = 1$ after history (v), and $\mathbf{g}_2 = 0$ after (w).[20] Second, Child's *confidence* is his belief of getting away with his lie even if he blushes. Given that we are only interested in such emotion when he lies after playing video-games, we can let $\mathbf{c}_2 = \tau_{C,1}\big(\{s_D : s_D\big((yes, b)\big) = buy\}|v\big)$. As for Dad, *blame* ($\mathbf{b}_3$) is equal to the probability with which he believes Child lied. Let $L := \big\{v.yes, w.no\big\}$ be the set of lies,[21] and z_D the terminal personal history observed by Dad during the game unfolding. Then, $\mathbf{b}_3 = \tau_{D,1}(L|z_D)$. We neglect the generation of emotions at other stages of the game, as they are ultimately inconsequential.

Feedback now takes a tractable form. We obtain, for each $(s, \theta, \tau^1) \in S \times \Theta \times \mathcal{T}^1$,

$$f_{(v),\mathrm{e}}(s, \theta, \tau^1) = \begin{cases} q\delta_{\neg b} + (1-q)\delta_b & \text{if } s_C(v) = Y; \\ \delta_{\neg b} & \text{if } s_C(v) = N; \end{cases} \qquad f_{(w),\mathrm{e}}(s, \theta, \tau^1) = f_{(w),\mathrm{e}}(s, \theta, \tau^1) = \delta_{\neg b}.$$

where $q = \tau_{C,1}\big(\{s_D : s_D\big((yes, b)\big) = buy\}|v\big)$ and subscripts of f_{e} denote the length-one history after which emotional messages are generated. ▲

From now onward, we will base our analysis on f_{e} rather than on $\tilde{f}_{\mathrm{e}}$. In some sense, emotions seems therefore to be bypassed. This raises the question of why we have not started expressing directly feedback functions as dependent on players' beliefs. The reason is essentially pedagogical. The game-independent notion of "emotion" allowed us to give a constructive definition of the game tree. We believe this approach is helpful to understand the double role of emotions. On the one hand, they drive emotional feedback independently of a specific game. On the other hand, they are triggered by players' behavior and beliefs during the game unfolding.

2.4 Utility

We now only need to express utility functions in game-dependent terms. In doing so, we leverage function ε introduced in Section 2.3.4. For each player $i \in I$, a *game-dependent psychological utility function* is a function $v_i : Z \times \Theta \times \mathcal{T}^K \to \mathbb{R}$, defined, for each $(z, \theta, \tau^K) \in Z \times \Theta \times \mathcal{T}^K$:

$$v_i(z, \theta, \tau^K) := \int_{E^{L(z)+1}} \tilde{v}_i(\pi(z, \theta), \theta, e^{L(z)+1}) \cdot \varepsilon(z, \tau^K)[\mathrm{d}e^{L(z)+1}].$$

Conceptually, $v_i(z, \theta, \tau^K)$ can be thought of as i's expected utility, if she knew that the game unfolded according to z, and if she knew her opponents' beliefs and traits. Note that functions $(v_i)_{i \in I}$ depend only on hierarchies of beliefs of order up to K, because higher-order beliefs do not trigger emotions. The assumption that $\tilde{v}_i$ and ε are continuous gives the following.

Remark 4. For each $i \in I$, v_i is continuous.

It is useful to express utility functions as depending on players' personal external states, rather than on terminal histories, as players form beliefs over S and not over Z. In conventional

[20] Recall that boldface letters represent emotional states. E.g., $\mathbf{g}_2$ describes whether Child feels guilty during stage 2.

[21] Recall that v and w stand for video-games and homework, respectively, and *yes* and *no* are the answers Child can give to the question "did you do your homework?".

settings, an external state s induces a unique terminal history, but in the present framework multiple histories can be induced by the same profile (s, θ, τ^K), as players' behavior may depend on the stochastic signals they observe. Hence, we can derive the distribution over terminal histories induced by any profile (s, θ, τ^K). To do so, it is convenient to retrieve from f a profile of functions $g := (g_h : S \times \Theta \times \mathcal{T}^K \to \Delta(A \times M))_{h \in H}$ that specifies how profiles of actions and messages are stochastically generated after each non-terminal history. In other words, g describes the probability measure over the immediate successors of any history h for each underlying profile (s, θ, τ^K). For each $h \in H$, $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$, and $(a, m) \in A \times M$, let

$$g_h(s, \theta, \tau^K)[(a, m)] := \begin{cases} f_h(s, \theta, \tau^K)[m] & \text{if } a = s(h); \\ 0 & \text{otherwise.} \end{cases}$$

In words, once we fix (s, θ, τ^K) and h, the probability that profile (a, m) realizes can be positive if and only if a is consistent with the behavior described by s after h. In such case, the probability of realization of (a, m) is simply the probability of m, as specified by feedback function f_h.

For each history $h \in \bar{H}$ and profile $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$, let $\zeta(h|s, \theta, \tau^K)$ denote the probability that h realizes given (s, θ, τ^K).[22] For a given game-dependent psychological utility function v_i of player i, we let the *external-state-dependent psychological utility function* describe the psychological utility of player i as a function of the external states. For each $i \in I$, we define $u_i : S \times \Theta \times \mathcal{T}^K \to \mathbb{R}$ to be such that, for each $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$,

$$u_i(s, \theta, \tau^K) := \sum_{z \in Z} v_i(z, \theta, \tau^K) \zeta(z|s, \theta, \tau^K).$$

Note that the domain of functions $(u_i)_{i \in I}$ is a set over which players form their beliefs of order $K + 1$, and about whose elements inferences can be made using emotional signals (cf. Section 3). In light of this, we say that $S \times \Theta \times \mathcal{T}^K$ is the set of *utility-relevant states*.

Remark 5. For each $i \in I$, u_i is continuous.[23]

Example 5 (Buy me an ice-cream, continued). Game-dependent psychological utilities are easily retrieved. For each $z = (a_{C,1}, a_{C,2}, m_D, a_D) \in Z$ and $(\theta, \tau^1) \in \Theta \times \mathcal{T}^1$, let $\pi_{C,1}(z)$ and $\pi_{C,2}(z)$ denote the two coordinates of Child's material outcome along z (i.e., whether he gets the ice-cream and whether he plays video-games). Then,

$$v_C(z, \theta, \tau^1) := \pi_{C,1}(z) + \theta \pi_{C,2}(z) - \tau_{D,1}(L|z_D).$$

Also recall that Dad gets a payoff of 2 from buying the ice-cream when Child did his homework and that ice-cream costs 1 to him. Denoting as z_D Dad's terminal personal history induced by z, we can define

$$v_D(z, \theta, \tau^1) := \begin{cases} 2\tau_{D,1}(\{s_C : s_C(\varnothing) = w\}|z_D) - 1 & \text{if } a_D = B; \\ 0 & \text{if } a_D = N. \end{cases}$$

[22] An explicit definition of such probability is in Appendix C.

[23] For each $h \in \bar{H}$, $\zeta(h|\cdot)$ is a continuous function on $S \times \Theta \times \mathcal{T}^K$ (this can be checked using the fact that functions $(f_h)_{h \in H}$ are continuous as per Remark 3). Continuity of functions $(v_i)_{i \in I}$ (Remark 4) then implies the result.

Deriving $(u_i)_{i\in I}$ is straightforward but notationally tedious. We postpone a detailed analysis to Section B.1 in Appendix B. ▲

3 Inferences on opponents' behavior, traits, and beliefs

Observing emotional and previous-play messages provides players with the means to make inferences about others' behavior and realized beliefs. Intuitively, the flow of information available to a player allows her to gradually restrict the set of utility-relevant states that are consistent with the observed evidence (i.e., the realized personal histories). In the following, we formalize such reasoning: Section 3.1 discusses mild assumptions about feedback functions, and Section 3.2 describes the ways in which the game may unfold for each utility-relevant state.

3.1 Properties of feedback

In this section, we discuss properties that make feedback "well-behaved." In particular, Definition 1 gives a condition for the feedback about others a player may observe to be independent from that player's own beliefs, and Definition 2 gives a notion of simplicity for feedback. An additional natural requirement consists in imposing some measurability condition on the set of utility-relevant states that allow a given message to be generated with positive probability after each history, and Definition 3 is in this spirit.[24]

First, we formalize the idea that, at any history, the beliefs of a player should not influence the generation of messages she may observe. This is natural if we stick to our interpretation of the messages a player can observe as messages about the emotions of *others*. In the following, for each $i \in I$ and $h \in H$, we let $f_{i,h,\mathrm{e}} = \mathrm{marg}_{M_{i,\mathrm{e}}} \circ f_{h,\mathrm{e}}$.[25] This map describes the emotional feedback each player may observe at each history.

Definition 1. *Feedback $f_{\mathrm{e}} = (f_{h,\mathrm{e}})_{h\in H}$ is **own-belief independent** if, for each $i \in I$, $h \in H$, $s \in S$, $\theta \in \Theta$, and $\tau_{-i}^K \in \mathcal{T}_{-i}^K$, the section $f_{i,h,\mathrm{e}}(s, \theta, \cdot, \tau_{-i}^K)$ of $f_{i,h,\mathrm{e}}$ is constant on $\mathcal{T}_i^K$.*

Own-belief independence requires that the generation of the messages a player can receive be independent from her own beliefs if we keep fixed a profile (s, θ, τ_{-i}^K). Note that the messages generated by player i's state of mind may shape her opponents' beliefs, and thus the realization of messages player i can observe at later stages. In some sense, then, a player's beliefs may influence the generation of her future messages. Own-belief independence *does not* rule this out, because such effect is incorporated in the realized history and own-belief independence applies when we keep the realized history fixed.

The most elementary feedback structure satisfying own-belief independence has two features: (i) only first-order beliefs (of others) matter, and (ii) the generation of messages about a player's

[24] Note that such assumptions are ultimately assumptions about functions $\tilde{f}_{\mathrm{e}}$ and ε. However, expressing them in terms of f_{e} comes with a substantial advantage in terms of notation and interpretation.

[25] As suggested by notation, marg denotes a marginalization map. For each measure μ on a finite product space $X \times Y$, $\mathrm{marg}_X \mu$ is a measure on X defined, for each $x \in X$ as $(\mathrm{marg}_X \mu)(x) := \sum_{y\in Y} \mu(x, y)$.

emotions (observed by her opponents) at any history depends exclusively on the beliefs she holds at (the personal history induced by) such history. Formally, we give the following.

Definition 2. *Feedback* $f_{\mathrm{e}} = (f_{h,\mathrm{e}})_{h\in H}$ *is* ***simple*** *if* (i) $K = 1$, *and* (ii) *for each* $i \in I$, $h = (h_i)_{i\in I} \in H$, $(s,\theta) \in S\times\Theta$, *and* $\tau_i^1, \bar{\tau}_i^1 \in \mathcal{T}_i^1$, $\tau_i^1(\cdot\,|h_i) = \bar{\tau}_i^1(\cdot\,|h_i)$ *implies that* $\operatorname{marg}_{M_{j,i,\mathrm{e}}} f_{h,\mathrm{e}}(s,\theta,\tau_i^1,\tau_{-i}^1) = \operatorname{marg}_{M_{j,i,\mathrm{e}}} f_{h,\mathrm{e}}(s,\theta,\bar{\tau}_i^1,\tau_{-i}^1)$ *for each* $j \in I\setminus\{i\}$ *and* $\tau_{-i}^1 \in \mathcal{T}_{-i}^1$.

Recall that $M_{j,i,\mathrm{e}}$ in the previous definition is the set of messages about i that j may observe.

Simplicity is a mild requirement. Indeed, the majority of psychological motivations can be modeled resorting to first-order beliefs only (Battigalli and Dufwenberg, 2022), so that point (i) does not seem to be too restrictive. Condition (ii) requires that a player's emotional leakage be independent of the realized beliefs of previous stages, so that only the last realized belief plays a role – this too seems reasonable. Note that feedback is simple in all the examples we mentioned.

Next, we give conditions about feedback that allow players to make inferences.[26]

Definition 3. *Feedback* $f_{\mathrm{e}} = (f_{h,\mathrm{e}})_{h\in H}$ *is:*

1. ***semi-regular*** *if, for each* $h \in H$, *the correspondences* $(\tau^K \mapsto \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^K))_{(s,\theta)\in S\times\Theta}$ *are measurable. That is, if for each* $h \in H$ *and* $m_{\mathrm{e}} \in M_{\mathrm{e}}$ *the lower inverse of* $\{m_{\mathrm{e}}\}$ *of each of the correspondences* $(\tau^K \mapsto \operatorname{supp} f_h(s,\theta,\tau^K))_{(s,\theta)\in S\times\Theta}$ *is measurable;*

2. ***regular*** *if, for each* $h \in H$ *and* $m_{\mathrm{e}} \in M_{\mathrm{e}}$, *the lower inverse of* $\{m_{\mathrm{e}}\}$ *of each of the correspondences* $(\tau^K \mapsto \operatorname{supp} f_h(s,\theta,\tau^K))_{(s,\theta)\in S\times\Theta}$ *is a measurable rectangle.*

Semi-regularity is weaker than regularity, and it arguably represents the minimal assumption needed to allow players to carry out a "well-defined" reasoning about possible ways in which the game may unfold in Section 3.2), as it ensures that eventualities such as "receiving message $m_{i,\mathrm{e}}$ with positive probability at (personal) history h_i" can be assessed by player $i \in I$. This is formalized by the following.

Remark 6. If feedback is semi-regular, sets $\left\{(s,\theta,\tau^K) \in S\times\Theta\times\mathcal{T}^K : m_{\mathrm{e}} \in \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^K)\right\}$ and $\left\{(s,\theta,\tau^K) \in S\times\Theta\times\mathcal{T}^K : m_{i,\mathrm{e}} \in \operatorname{supp} f_{i,h,\mathrm{e}}(s,\theta,\tau^K)\right\}$ are measurable for each $h \in H$, $m_{\mathrm{e}} \in M_{\mathrm{e}}$, $i \in I$, and $m_{i,\mathrm{e}} \in M_{i,\mathrm{e}}$. [27]

Regularity is instead a slightly stronger requirement, but it has a reasonable conceptual justification. With regularity players are able to disentangle the different factors at play in the generation of messages. With this, we mean that each player is able to assess also, for example,

[26] Recall that, for given measurable space $(X,\mathcal{X})$, topological space Y, and correspondence $\gamma : X \rightrightarrows Y$, the *lower inverse* of γ, $\gamma^{-1} : 2^Y \to 2^X$, is defined to be such that $\gamma^{-1}(A) = \{x \in X : \gamma(x)\cap A \neq \emptyset\}$ for each $A \subseteq Y$. Correspondence γ is said to be *measurable* if $\gamma^{-1}(F) \in \mathcal{X}$ for each closed $F \subseteq Y$. Moreover, given a countable sequence of measurable spaces $(X_k,\mathcal{X}_k)_{k\in K}$ and the product measurable space $\left(\bigtimes_{k\in K} X_k, \bigotimes_{k\in K}\mathcal{X}_k\right)$, a *measurable rectangle* is a set $\bigtimes_{k\in K} Y_k \subseteq \bigtimes_{k\in K} X_k$, with $Y_k \in \mathcal{X}_k$ for each $k \in K$.

[27] Fix $h \in H$ and, for each $(s,\theta) \in S\times\Theta$, let $\gamma_{s,\theta}$ be the correspondence $\tau^K \mapsto \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^K)$. Then, the first set is $\bigcup_{(s,\theta)}\{(s,\theta)\}\times\gamma_{s,\theta}^{-1}(m_{\mathrm{e}})$, which is measurable because $\gamma_{s,\theta}$ is measurable. As for the second set, we write it as $\bigcup_{(s,\theta)}\left(\{(s,\theta)\}\times\left\{\bigcup_{m_{-i,\mathrm{e}}}\gamma_{s,\theta}^{-1}(m_{i,\mathrm{e}},m_{-i,\mathrm{e}})\right\}\right)$, which is again easily seen to be measurable.

the hierarchical systems of beliefs of (each of) her opponents that allow her to observe some message with positive probability at some history. Formally, this means that the projection onto $\mathcal{T}_j^K$ of a set of the kind $\{(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K : m_i \in \operatorname{supp} f_{i,h,\mathrm{e}}(s,\theta,\tau^K)\}$ is measurable for each $j \in I \setminus \{i\}$. This does not hold for all measurable subsets of $S \times \Theta \times \mathcal{T}^K$,[28] and it is ensured precisely by the rectangular shape assumed by such set under regularity of feedback.

While semi-regularity is easily acceptable, one may wonder about how restrictive regularity actually is. It turns out that the two conditions coincide whenever feedback is also simple.[29]

Proposition 1. *Let feedback be simple. Then, it is semi-regular if and only if it is regular.*

Example 5 (Buy me an ice-cream, continued). Informative messages are generated after history (v), depending on Child's subsequent action. Feedback is simple because it depends only on Child's first-order beliefs held after (v). To check (semi-)regularity of feedback, focus on message b and history (v). We have:

$$\begin{aligned}&\{(s,\theta,\tau^1) : b \in \operatorname{supp} f_{D,(v),\mathrm{e}}(s,\theta,\tau^1)\}\\ &= \{s_C : s_C(v) = yes\} \times S_D \times \Theta \times \{\tau_C^1 : \tau_{C,1}(\{s_D : s_D((yes,b)) = buy\}|v) < 1\} \times \mathcal{T}_D^1,\end{aligned}$$

which is a measurable rectangle. Similar considerations apply to message $\neg b$ and to history (w). In addition, note that the generation of feedback is independent of Dad's beliefs of any order, and this ensures own-belief independence. ▲

3.2 Making inferences

Recall that multiple (terminal and non-terminal) histories may arise from an underlying utility-relevant state. A crucial part of players' reasoning pertains therefore to the understanding of the possible paths the game can follow given any underlying state.

For each $i \in I$, we let $\mathbf{H}_i : S \times \Theta \times \mathcal{T}^K \rightrightarrows \bar{H}_i$ be the correspondence that collects the set of i's personal histories that are possible given each utility relevant state. Intuitively, a personal history $h_i = (a_i^\ell, m_i^\ell)$ is possible at (s,θ,τ^K) if (i) i's observed behavior (i.e., a_i^ℓ) is consistent with s_i, and (ii) the feedback i observes at each stage (i.e., m_i^ℓ) can be generated with positive probability given (s,θ,τ^K) according to feedback f. Given that we informally assume that players know the rules of interaction, such correspondence can be retrieved by player i, by reasoning about how the game may unfold. The interpretation of $\mathbf{H}_i$ is straightforward, and to ease exposition we defer its formal definition to the proof of Lemma 1 (Appendix A, p. 36).

The following result ensures that, under semi-regularity of feedback, the set of utility-relevant states allowing for any given personal history of any player is measurable.

Lemma 1. *If feedback is semi-regular, $\mathbf{H}_i$ is measurable for each $i \in I$.*

Upon observing a personal history, players can then check whether it is consistent with a given utility-relevant state, leveraging the personal history correspondences just defined. In

[28] Indeed, projections onto Polish spaces of Borel sets are analytic but not Borel, in general (cf. Definition 12.23 and Theorem 12.24 of Aliprantis and Border, 2006).

[29] Proofs are collected in Appendix A.

particular, the set of utility-relevant states consistent with $h_i \in \bar{H}_i$ is $(\mathbf{H}_i)^{-1}(h_i)$. Given that player i is assumed to know her epistemic type $\bar{\tau}_i^\infty$ (hence, the induced hierarchical system of finite-order beliefs $\bar{\tau}_i^K$), we can focus on the section of such set at $\bar{\tau}_i^K$:

$$\Omega^K_{-i,\bar{\tau}_i^K}(h_i) := \big\{(s,\theta,\tau^K_{-i}) \in \Omega^K_{-i} : h_i \in \mathbf{H}_i(s,\theta,\bar{\tau}_i^K,\tau^K_{-i})\big\}.$$

For each $i \in I$ and $\tau_i^K \in \mathcal{T}_i^K$, we call the sequence of sets $\big(\Omega^K_{-i,\tau_i^K}(h_i)\big)_{h_i \in \bar{H}_i}$ the *inference sets* of player i when her hierarchical system of beliefs of order K is τ_i^K. Note that the inferences players can make are in general linked to their beliefs of order up to K. The following is an immediate consequence of Lemma 1.

Remark 7. If feedback is semi-regular, $\Omega^K_{-i,\tau_i^K}(h_i)$ is measurable for each $i \in I$, $h_i \in \bar{H}_i$, and $\tau_i^K \in \mathcal{T}_i^K$.

Example 5 (Buy me an ice-cream, continued). Assume $\tau_{C,1}\big(\{s_D : s_D\big((yes,b)\big) = buy\}|v\big) = \frac{1}{2}$, so that Child blushes with probability $\frac{1}{2}$ after lying. If $s_C = v.yes$, then $\mathbf{H}_D(s,\theta,\tau^1) = \{(yes,b),(yes,\neg b)\}$. If instead $s_C = v.no$ or $s_C = w.no$, then $\mathbf{H}_D(s,\theta,\tau^1) = \{(no,\neg b)\}$. Lastly, if $s_C = w.yes$, then $\mathbf{H}_D(s,\theta,\tau^1) = \{(yes,\neg b)\}$. Other correspondences are derived analogously.

The set of profiles (s,θ,τ^1) consistent with Dad's personal history $(yes,\neg b)$ is

$$\begin{aligned}(\mathbf{H}_D)^{-1}\big((yes,\neg b)\big) =& \big\{(s,\theta,\tau^1) : s_C = w.yes\big\} \\ &\cup \big\{(s,\theta,\tau^1) : s_C = v.yes,\ \tau_{1,C}\big(\{s_D : s_D\big((yes,b)\big) = buy\}|v\big) > 0\big\},\end{aligned}$$

and such set can be seen to be measurable. Given that Dad's beliefs do not play a role in the generation of feedback, Dad's corresponding inference set is just $\Omega^1_{C,\tau_D^1}\big((yes,\neg b)\big) = (\mathbf{H}_D^{-1}\big((yes,\neg b)\big)$ for each $\tau_D^1 \in \mathcal{T}_D^1$. Similar considerations apply to other cases. ▲

4 Rationality

In this section, we describe rationality as the conjunction of several features. First, we analyze cognitive sophistication requirements: rational players' beliefs should satisfy a natural notion of coherence (Section 4.1), they should be consistent with evidence (Section 4.2), and they should be updated according to Bayes rule throughout the game (Section 4.3). Second, the plan of a player is required to satisfy an optimality criterion (Section 4.4), and to coincide with the player's actual behavioral predisposition (Section 4.5). Third, we define rationality of a player as the conjunction of the aforementioned properties, proving that it is an event (Section 4.6).

4.1 Coherence

We say that a hierarchy of beliefs is coherent if lower-order beliefs can be recovered from higher-order ones through marginalization. In some sense, beliefs of different orders along a coherent hierarchy "agree" on relevant events.

Definition 4. *Epistemic type τ_i^∞ of player $i \in I$ is* ***coherent*** *if, for each $n \in \mathbb{N}$ and $h_i \in H_i$,*[30]

$$\text{marg}_{\Omega_{-i}^{n-1}} \tau_{i,n+1}(\cdot|h_i) = \tau_{i,n}(\cdot|h_i).$$

Let $\mathcal{T}_{i,C}^\infty$ denote the set of coherent epistemic types of player i and C_i the set of personal states $(s_i, \theta_i, \tau_i^\infty)$ such that $\tau_i^\infty \in \mathcal{T}_{i,C}^\infty$.

Lemma 2. *For each $i \in I$, C_i is closed.*

The following result is adapted from Brandenburger and Dekel (1993), and it establishes that a coherent epistemic type of a player can be identified with a system of beliefs over the space of primitive uncertainty and (not necessarily coherent) epistemic types of her opponents.

Lemma 3. *For each $i \in I$, there exists an homeomorphism $\varphi_i : \mathcal{T}_{i,C}^\infty \to \left[\Delta(S \times \Theta \times \mathcal{T}_{-i}^\infty)\right]^{\bar{H}_i}$ such that, for each $h_i \in \bar{H}_i$, $\text{marg}_{\Omega_{-i}^{n-1}} \varphi_i(\tau_i^\infty)(\cdot|h_i) = \tau_{i,n}(\cdot|h_i)$.*

4.2 Knowledge-implies-belief

According to the reasoning described in Section 3, upon observing h_i, a player who knows her epistemic type can rule out states that are inconsistent with the occurrence of such history. We now formally require that the $(K+1)$-th-order beliefs held by a player at each personal history be consistent with such inferential reasoning. The expression "knowledge-implies-belief" suggests that knowing that a history has realized must imply believing (i.e., assigning probability one to) the set of utility-relevant states that allow for such history.

Definition 5. *Epistemic type τ_i^∞ of player $i \in I$ satisfies* ***knowledge-implies-belief*** *if, for each $h_i \in \bar{H}_i$,*

$$\tau_{i,K+1}\left(\Omega_{-i,\tau_i^K}^K(h_i)\middle|h_i\right) = 1.$$

Let $\mathcal{T}_{i,KB}^\infty$ be the set of player i's epistemic types satisfying knowledge-implies-belief, and KB_i the set of personal states $(s_i, \theta_i, \tau_i^\infty)$ such that $\tau_i^\infty \in \mathcal{T}_{i,KB}^\infty$.

Lemma 4. *If feedback is regular and own-belief independent, KB_i is measurable for each $i \in I$.*

Note that not assuming coherence makes our notion of knowledge-implies-belief very weak, as it requires that only $(K+1)$-th-order beliefs be updated consistently with evidence. When considering rational (hence, coherent) players, however, beliefs of all order conform to such inferential reasoning under the hypotheses of Lemma 4.[31]

Example 5 (Buy me an ice-cream, continued). Seeing Child blush is the most informative message for Dad, because it perfectly reveals a lie. We already highlighted that $\Omega_{C,\tau_D^1}^1\big((yes, b)\big) =$

[30] In the following, we slightly abuse notation by writing Ω_{-i}^0 instead of Ω^0, to ease the exposition.

[31] Coherence implies that beliefs of order higher than $K+1$ conform to the inferential reasoning we outlined. With regularity of feedback, we conclude that also lower-order beliefs do so: by coherence they assign probability one to the projections of inference sets onto Ω^0 and Ω_{-i}^n (with $1 \leq n < K$), and measurability of such projections is implied by regularity.

$\{v.yes\} \times S_D \times \Theta \times \{\tau_C^1 : \tau_C^1(\{s_D : s_D\big((yes, b)\big) = buy\}|v) > 0\}$. Knowledge-implies-belief then ensures that, for example, Dad's second-order beliefs after personal history (yes, b) are such that $\tau_{D,2}\big(\{s_C\}|(yes, b)\big) = 0$ for each $s_C \in S_C \setminus \{v.yes\}$. With coherence, the same reasoning extends to beliefs of different orders. ▲

4.3 Belief updating

To model how cognitively rational players should update their beliefs it is useful to unpack the mechanisms through which information accrues to players. After a given personal history, a player observes three pieces of information: she first observes the action she plays, and then she observes the realized previous-play and emotional messages.

We want to formalize the idea that player i uses each piece of information independently, and timing is key for this purpose. Specifically, player i should first update her beliefs *about her personal external state* upon seeing the action she chooses.[32] Then, she can take into account the messages she receives to update her beliefs *about others* using Bayes rule. Note that (both previous-play and emotional) messages do not provide novel information about a player's personal external state once the player observes the actions she takes.

Conceptually, it is as if we were endowing a player with a fictitious "interim belief" held at stage "$k+\frac{1}{2}$," that is, after playing at stage k, but before having observed any messages. In such metaphor, we should impose that a player does not change her beliefs about her personal external state after acting. This formalizes a notion of *own-action independence* of beliefs, capturing the idea that own actions and messages should be used to make inferences in "parallel" ways.

Before proceeding, we introduce some notation. Recall that $\zeta(h|s, \theta, \tau^K)$ was defined as the probability that h realizes when the utility-relevant state is (s, θ, τ^K) (cf. Section 2.4). Taking player i's perspective, we denote as $\zeta(h|h_i; s, \theta, \tau^K)$ the probability that h realizes when the utility-relevant state is (s, θ, τ^K), *and* conditional on observing h_i.

Finally, define $(f_{i,h_i} : S \times \Theta \times \mathcal{T}^K \to \Delta(M_i))_{h_i \in H_i}$ to be such that, for each $h_i \in H_i$, $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$ and $(a_i, m_i) \in A_i \times M_i$,

$$f_{i,h_i}(s, \theta, \tau^K)[m_i] := \sum_{h \in \bar{H}(h_i)} f_{i,h}(s, \theta, \tau^K)[m_i] \cdot \zeta(h|h_i; s, \theta, \tau^K),$$

where $\bar{H}(h_i)$ is the set of "complete" histories compatible with h_i. In words, these functions describe the generation of messages of a given player $i \in I$ (as a function of the utility-relevant state) at a given personal history. Recall that the feedback functions $(f_h)_{h \in H}$ derived before conditioned instead on "complete" histories.

At this point, we can formally describe belief updating of any player $i \in I$. As a prerequisite, we require that beliefs about one's self and about others satisfy a form of *independence* – i.e., that the belief held at each given history on the space $(S \times \Theta \times \mathcal{T}_{-i}^K, \mathcal{B}(S \times \Theta \times \mathcal{T}_{-i}^K))$ be obtained as a product measure starting from measures on $(S_i \times \Theta_i, \mathcal{B}(S_i \times \Theta_i))$ and $(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K, \mathcal{B}(S_{-i} \times$

[32] Recall that we do not assume that players know their personal external states (i.e., how they would behave throughout the game).

$\Theta_{-i} \times \mathcal{T}_{-i}^K)$). Formally, for each $h_i \in \bar{H}_i$,

$$\tau_{i,K+1}(\,\cdot\,|h_i) = \text{marg}_{S_i \times \Theta_i}\, \tau_{i,K+1}(\,\cdot\,|h_i) \otimes \text{marg}_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K}\, \tau_{i,K+1}(\,\cdot\,|h_i), \tag{I}$$

where $\otimes$ is used to denote the product of measures.

Next, we require that the chain rule of conditional probability hold for personal external states. For each $i \in I$ and $h_i \in \bar{H}_i$, denote $S_i(h_i)$ be the set of i's personal external states that do not prevent h_i, as $S_i(h_i, a_i)$ set of player i's personal external states that allow h_i and that prescribe a_i at h_i, and as $\bar{H}_i(h_i, a_i)$ the set of immediate successors of h_i where a_i is played. The *chain rule* holds if, for each $h_i \in H_i$, $a_i \in \hat{\mathcal{A}}_i(h_i)$, $h_i' \in \bar{H}_i(h_i, a_i)$ and $s_i \in S_i(h_i, a_i)$,[33]

$$\tau_{i,K+1}(\{s_i\}|h_i') \cdot \tau_{i,K+1}(S_i(h_i, a_i)|h_i) = \tau_{i,K+1}(\{s_i\}|h_i). \tag{CR}$$

Finally, we require that *Bayes rule* hold for anything else, after playing. Let $M_i(h_i, a_i)$ be the set of messages player i can observe after playing a_i at personal history h_i. For each $h_i \in H_i$, $a_i \in \hat{\mathcal{A}}_i(h_i)$, $m_i \in M_i(h_i, a_i)$ and $F \in \mathcal{B}(S_{-i} \times \Theta \times \mathcal{T}_{-i}^K)$, we impose

$$\begin{aligned}
&\tau_{i,K+1}(F|h_i') \cdot \int_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^K} f_{i,h_i}(s_i, s_{-i}, \theta, \tau_i^K, \tau_{-i}^K)[m_i] \cdot \big(\text{marg}_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^K}\, \tau_{i,K+1}\big)\big(\text{d}(s_{-i}, \theta, \tau_{-i}^K)|h_i\big) \\
&= \int_F f_{i,h_i}(s_i, s_{-i}, \theta, \tau_i^K, \tau_{-i}^K)[m_i] \cdot \big(\text{marg}_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^K}\, \tau_{i,K+1}\big)\big(\text{d}(s_{-i}, \theta, \tau_{-i}^K)|h_i\big),
\end{aligned} \tag{BR-a_i}$$

where s_i above is any element of $S_i(h_i, a_i)$, and $h_i' = (h_i, (a_i, m_i))$.

Definition 6. *Epistemic type τ_i^∞ of player $i \in I$ satisfies **correct belief updating** if* (I), (CR), *and* (BR-a_i) *hold. Let $\mathcal{T}_{i,CBU}^\infty$ be the set of epistemic types of player i that satisfy correct belief updating, and CBU_i the set of personal states $(s_i, \theta_i, \tau_i^\infty)$ such that $\tau_i^\infty \in \mathcal{T}_{i,CBU}^\infty$.*

Lemma 5. *For each $i \in I$, $\mathcal{T}_{i,CBU}^\infty$ is measurable.*

Example 5 (Buy me an ice-cream, continued). Dad is inactive and does not receive any informative message throughout the first stage. In the second stage, he is inactive so he does not need to update his beliefs about his personal external state with (CR). However, (BR-a_i) applies. Focus for the sake of the example on Dad's beliefs about $F := \{s_C : s_C(\varnothing) = v\} \times \Theta \times \mathcal{T}_C^1$ (i.e., "Child played video-games"), and say that he observes $(yes, \neg b)$. The probability of observing such personal history at the second stage (i.e., after a "dummy" length-1 personal history h_D^1) as a function of a profile (s, θ, τ_C^1) can be checked to be

$$f_{D,h_D^1,\tau_D^1}(s, \theta, \tau_C^1)[(Y, \neg b)] = \begin{cases} 1 & \text{if } s_C = w.yes; \\ 1-q & \text{if } s_C = v.yes; \\ 0 & \text{if } s_C \in \{v.no, w.no\}; \end{cases}$$

where $q = \tau_{C,1}\big(\{s_D : s_D\big((yes, b)\big) = buy\}|v\big)$. As a result, the probability with which epistemic type τ_D^∞ expects to observe $(yes, \neg b)$ is

$$\underbrace{\tau_{D,2}(\{w.yes\}|h_D^1)}_{=:\alpha(\tau_D^\infty)} + \underbrace{\int_{\{v.yes\} \times \Theta \times \mathcal{T}_C^1} (1-q) \cdot \big(\text{marg}_{S_C \times \Theta \times \mathcal{T}_C^1}\, \tau_{D,2}\big)\big(\text{d}(s_C, \theta, \tau_C^1)|h_D^1\big)}_{=:\beta(\tau_D^\infty)}.$$

[33] Recall from Section 2.3.2 that we use obvious abbreviations for marginal probabilities.

If $\alpha(\tau_D^\infty) + \beta(\tau_D^\infty) > 0$, (BR-$a_i$) implies that:

$$\tau_{D,2}\big(F|(yes, \neg b)\big) = \frac{\beta(\tau_D^\infty)}{\alpha(\tau_D^\infty) + \beta(\tau_D^\infty)}.$$

Note that our belief updating rule implies that players *do not* change their beliefs about others after they act if they do not observe any informative message in the meantime. Hence, Dad's final blame (i.e., the probability with which he believes Child lied), which concerns Child's actions, actually arises at the second stage, since his beliefs about Child's actions do not change after his action. ▲

4.4 Rational planning

As a prerequisite for rational planning, we require that a player know her personal trait. Since realized utilities at the end of the game are affected by players' traits, different trait-types of a player may want to behave differently at some points of the game, and knowing one's own trait is necessary to plan how to behave optimally.

Definition 7. *Player i* ***knows her personal trait*** *at personal state $(s_i, \bar{\theta}_i, \tau_i^\infty) \in S_i \times \Theta_i \times \mathcal{T}_i^\infty$ if, for each $h_i \in \bar{H}_i$, $\tau_{i,K+1}\left(\{\bar{\theta}_i\}|h_i\right) = 1$. Let KT_i be the set of personal states where player i knows her personal trait.*[34]

Next, we retrieve a *plan* of player i (technically, a behavior strategy) from her epistemic type τ_i^∞, denoted as

$$\sigma(\tau_i^\infty) \in \Sigma_i := \bigtimes_{h_i \in H_i} \Delta(\hat{\mathcal{A}}_i(h_i)).$$

It is defined, for each $h_i \in H_i$ and $a_i \in \hat{\mathcal{A}}_i(h_i)$, as[35]

$$\sigma(\tau_i^\infty)(a_i|h_i) := \tau_{i,K+1}(S_i(h_i, a_i)|h_i),$$

where $S_i(h_i, a_i)$ is the set of personal external states of player i consistent with h_i that prescribe a_i at h_i (cf. Section 4.3).

We argue that such an object is what one can legitimately label as a "strategy." Indeed, we take a strategy to be a plan in the mind of a player, and the derivation of $\sigma(\tau_i^\infty)$ follows this intuition. A plan specifies how a player expects herself to behave at each contingency she could observe, and a player's plan coincides with her behavioral predisposition s_i if and only if $\sigma(\tau_i^\infty)(s_i(h_i)|h_i) = 1$ for each $h_i \in H_i$.

Next, we define a player's expected utility conditional on observing a given personal history. For each $i \in I$, define the profile of functions $(u_{i,h_i} : S \times \Theta \times \mathcal{T}^K \to \mathbb{R})_{h_i \in H_i}$ to be such that, for each $h_i \in H_i$ and $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$,

$$u_{i,h_i}(s, \theta, \tau^K) := \sum_{z \in Z} v_i(z, \theta, \tau^K)\zeta(z|h_i; s, \theta, \tau^K). \tag{5}$$

[34] Recall that we are not assuming coherence. Thus, our choice of working with beliefs over utility-relevant events, although reasonable, is ultimately arbitrary. We choose to impose this condition on beliefs of order $K+1$ because such beliefs are the ones used by a player to figure out her optimal plan. Also recall that $\tau_{i,K+1}\left(\{\bar{\theta}_i\}|h_i\right)$ is a shortcut for $\tau_{i,K+1}\left(S \times \{\bar{\theta}_i\} \times \Theta_{-i} \times \mathcal{T}_{-i}^K|h_i\right)$.

[35] Also in this case, we rely on beliefs of order $K+1$ as we did to define knowledge of one's personal trait.

In words, $u_{i,h_i}(s,\theta,\tau^K)$ is the expected utility of player i after h_i when the utility-relevant state is (s,θ,τ^K).

Defining rational planning requires some care because in our setting players may fail to be dynamically consistent. This can happen when a player's utility depends on her own plan, Dynamic inconsistency arises for this reason, for instance, in models of frustration and anger (Battigalli et al., 2019b), anticipatory feelings (Caplin and Leahy, 2001), and reference-dependence (Kőszegi and Rabin, 2006).

In such settings, the appropriate notion of rational planning is a form of "intra-personal equilibrium": each self of a player chooses the "optimal" available action taking as given the actions chosen by other selves.[36] This notion is referred to as *one-step optimality* . To give a formal definition of such property, we need to define the "(expected) utility of taking a given action at a given personal history." As a maintained assumption throughout this section, we take systems of beliefs to satisfy *independence* (see Section 4.3).

Note that choosing $a_i \in \hat{\mathcal{A}}_i(h_i)$ at h_i induces a distribution over personal histories of the form $h_i' = (h_i, a_i, m_i)$. Such personal histories are the instances where the player is going to act next. The distribution can depend on the player's type, as well as on the emotional feedback. We denote the vector of such distributions for a player with hierarchical system of beliefs τ_i^{K+1} as $(\mu(\cdot\,|h_i,a_i,\tau_i^{K+1}) \in \Delta(H_i))_{h_i\in H_i, a_i\in\hat{\mathcal{A}}_i(h_i)}$. It is possible to obtain an explicit expression of such distribution using the same steps as in Section 4.3. There, we used the map $f_{i,h_i} : S\times\Theta\times\mathcal{T}^K \to \Delta(M_i)$ to determine the probability that a message $m_i = (m_{i,\mathrm{p}}, m_{i,\mathrm{e}})$ is generated after personal history h_i (as a function of the utility-relevant state). Taking the section of such map at some s_i with $s_i(h_i) = a_i$ gives the distribution over messages that player i expects to receive after taking action a_i at h_i.[37] To emphasize the role of a_i, we denote such map as f_{i,h_i,a_i}. Then, the probability that personal history $h_i' = (h_i, a_i, m_i)$ realizes after taking action a_i at h_i is

$$\mu(h_i'|h_i,a_i,\tau_i^K) = \int f_{i,h_i,a_i}(s_{-i},\theta,\tau_i^K,\tau_{-i}^K)[m_i]\cdot \mathrm{marg}_{S_{-i}\times\Theta\times\mathcal{T}_{-i}^K}\,\tau_{i,K+1}(\mathrm{d}(s_{-i},\theta,\tau_{-i}^K)|h_i).$$

At this point, we can define player i's "local" *decision utility functions*, $(\bar{u}_{i,h_i} : \hat{\mathcal{A}}_i(h_i)\times\Theta_i\times \mathcal{T}_i^{K+1}\to\mathbb{R})_{h_i\in H_i}$. For each $h_i\in H_i$, $a_i\in\hat{\mathcal{A}_i}(h_i)$, $\theta_i\in\Theta_i$, and $\tau_i^{K+1}\in\mathcal{T}_i^{K+1}$,

$$\bar{u}_{i,h_i}(a_i,\theta_i,\tau_i^{K+1}) = \sum_{h_i'\in\bar{H}_i}\mu_i(h_i'|h_i,a_i,\tau_i^{K+1})\int u_{i,h_i}(s,\theta,\tau_i^K,\tau_{-i}^K)\tau_{i,K+1}(\mathrm{d}s,\theta_i,\mathrm{d}\theta_{-i},\mathrm{d}\tau_{-i}^K|h_i').$$

Note that personal trait θ_i is being held fixed when taking the expectation: this is needed to identify the optimal actions for each given personal trait and belief system.

[36] Note that this presumes that the player understands her dynamic inconsistency and plans accordingly. For this reason, one could legitimately talk about "sophisticated" planning.

[37] Note that insofar as h_i and a_i are fixed, the only relevant difference between personal external states that prescribe a_i at h_i is in the actions prescribed at future reachable contingencies. But those prescriptions are irrelevant to the generation of emotional feedback thanks to our assumptions. In particular, we assumed that only *realized* emotional states mattered in the generation of emotional feedback, and the description of future behavior cannot affect present emotions.

The actions that maximize player i's decision utility at personal history h_i are collected by the best reply correspondence $r_{i,h_i} : \Theta_i \times \mathcal{T}_i^\infty \rightrightarrows \hat{\mathcal{A}}_i(h_i)$, defined as

$$(\theta_i, \tau_i^\infty) \mapsto \arg \max_{a_i \in \hat{\mathcal{A}}_i(h_i)} \bar{u}_{i,h_i}(a_i, \theta_i, \tau_i^{K+1}).$$

We are now ready to give our desired definition. In the following, we refer to a "plan," which is understood as a profile of probability measures over available actions and generically denoted $\sigma_i = (\sigma_i(\cdot\,|h_i) \in \Delta(\hat{\mathcal{A}}_i(h_i))_{h_i \in H_i}$. Recall that $\sigma_i(\tau_i^\infty)$ denotes the plan induced by a type τ_i^∞.

Definition 8. *A plan σ_i is **one-step optimal** for $(\theta_i, \tau_i^\infty)$ if, for each $h_i \in H_i$,*

$$\operatorname{supp} \sigma_i(\cdot\,|h_i) \subseteq r_{i,h_i}(\theta_i, \tau_i^\infty).$$

*Player i **plans rationally** at personal state $(s_i, \theta_i, \tau_i^\infty)$ if she knows her personal trait (i.e., $(s_i, \theta_i, \tau_i^\infty) \in KT_i$), her epistemic type τ_i^∞ satisfies independence, and her plan $\sigma(\tau_i^\infty)$ is one-step optimal. Let RP_i denote the set of states where player i plans rationally.*

Lemma 6. *For each $i \in I$, RP_i is closed.*

As is well-known, deterministic one-step optimal plans may fail to exist. This typically happens when a player's utility depends on her own plan, as is the case, for example, when she is affected by disappointment aversion or anxiety. When no such dependence exist, preferences depend on the psychological states induced by others' beliefs, and on the anticipation of such states. Many interesting models belong in this category, including models with image concerns or guilt aversion (see, e.g., the survey article by Battigalli and Dufwenberg, 2022, or the discussion in Section 6 of Battigalli et al., 2019a). Moreover, under such "independence" assumption, preferences admit an almost standard expected utility formulation.

This motivates the following definition. To state it, we denote as $\tau_{i,-i}^K = (\tau_{i,k,-i})_{k=1}^K$ the system of beliefs *about others* induced by system of beliefs τ_i^K. For each $k \in \{1, \ldots, K\}$, $\tau_{i,k,-i}$ is the k-th order belief about others' personal external states, traits, and others' systems of beliefs of order up to $k-1$. Formally, for each $h_i \in \bar{H}_i$, $\tau_{i,1,-i}(\cdot\,|h_i) = \operatorname{marg}_{S_{-i} \times \Theta} \tau_{i,1}(\cdot\,|h_i)$ and $\tau_{i,k,-i}(\cdot\,|h_i) = \operatorname{marg}_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^{k-1}} \tau_{i,k}(\cdot\,|h_i)$ for each $k \in \{2, \ldots, K\}$. Note that $\tau_{i,-i}^K$ is obtained by system of beliefs τ_i^K essentially by excluding i's beliefs about her own behavior (i.e., the marginal of her beliefs of any order on S_i). Finally, we can state our desired condition as the requirement that a player's utility be the same whenever her system of beliefs differ only in the induced beliefs about her own behavior.

Definition 9. *Player i's preferences are **own-plan independent** if for each $(s, \theta, \tau_{-i}^K) \in S \times \Theta \times \mathcal{T}_{-i}^K$, and $\tau_i^K, \bar{\tau}_i^K \in \mathcal{T}_i^K$,*

$$\tau_{i,-i}^K = \bar{\tau}_{i,-i}^K \Longrightarrow u_i(s, \theta, \tau_i^K, \tau_{-i}^K) = u_i(s, \theta, \bar{\tau}_i^K, \tau_{-i}^K).$$

Under own-plan independence, preferences are dynamically consistent and rational plans can be obtained by dynamic programming methods. To illustrate, we focus on the "(expected) utility of following a given plan from a given personal history onward." Previously, we used the notation

$\zeta(z|h_i; s, \theta, \tau^K)$ to denote the probability of terminal history z conditional on having reached personal history h_i, when the utility-relevant state is (s, θ, τ^K). The notation $\zeta(z|h_i; \sigma_i, s_{-i}, \theta, \tau^K)$ has analogous meaning, but player i's behavior is described by a plan $\sigma_i \in \bigtimes_{h_i \in H_i} \Delta(\hat{\mathcal{A}}_i(h_i))$.[38] Note that such probability is affected only by the behavior prescribed by σ_i after personal history h_i, that is, by the continuation plan $(\sigma_i(\cdot\,|h_i'))_{h_i' \succeq h_i}$. We slightly abuse notation by letting

$$u_{i,h_i}(\sigma_i, s_{-i}, \theta, \tau^K) = \sum_{z \in Z} v_i(z, \theta, \tau^K) \cdot \zeta(z|h_i; \sigma_i, s_{-i}, \theta, \tau^K).$$

Conceptually, this is an intuitive modification of (5). The expected utility of following plan σ_i from h_i onward for type $(\theta_i, \tau_i^\infty)$ is then

$$\hat{u}_{i,h_i}(\sigma_i, \theta_i, \tau_i^\infty) := \int u_{i,h_i}(\sigma_i, s_{-i}, \theta_i, \theta_{-i}, \tau_i^K, \tau_{-i}^K) \cdot \tau_{i,K+1}(s_i, \theta_i, \mathrm{d}(s_{-i}, \theta_{-i}, \tau_{-i}^K)|h_i)).$$

Say that a plan σ_i^* is **sequentially optimal** for $(\theta_i, \tau_i^\infty)$ if, for each $h_i \in H_i$,

$$\sigma_i^* \in \arg\max_{\sigma_i \in \Sigma_i} \hat{u}_{i,h_i}(\sigma_i, \theta_i, \tau_i^\infty).$$

Standard dynamic programming results give the following.[39]

Remark 8. Assume player i has own-plan independent preferences. Then, a plan is sequentially optimal for $(\theta_i, \tau_i^\infty)$ if and only if it is one-step optimal for $(\theta_i, \tau_i^\infty)$. Moreover, a pure sequentially optimal plan exists.[40]

The issue of dynamic (in)consistency with psychological preferences is discussed in detail by (Battigalli and Dufwenberg, 2009, Section 6.3) and Battigalli et al. (2019a), to which we refer the interested reader.

We conclude with an illustration.

Example 5 (Buy me an ice-cream, continued). Suppose that Child's epistemic type τ_C^∞ satisfies independence and knowledge of personal trait, and that his system of second-order beliefs $\tau_{C,2}$ is such that, for each $h_C \in \{\varnothing, (w), (v)\}$,

$$\mathrm{marg}_{S_D}\, \tau_{C,2}(\cdot\,|h_C) = \delta_{not.buy.not}; \tag{6}$$

$$\mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L^c|(yes, \neg b)\big)|h_C\big] = \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes, b)\big)|h_C\big] = \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L^c|(no, \neg b)\big)|h_C\big] = 1. \tag{7}$$

In words, (6) says that, at each history where he is active, Child is sure that Dad would behave according to *not.buy.not* (i.e., that Dad would buy him the ice-cream only if he says "no" without blushing).[41] Equation (7) instead says that Child thinks that Dad would be sure he is a liar

[38] The derivation of objects of this kind is conceptually straightforward. We offer an explicit discussion in Appendix C.

[39] See Kreps (2013) for an overview.

[40] Randomization is superfluous because the expected utility is affine in the probabilities assigned by plans to personal external states.

[41] Note that this implies that he would blush for sure if he says "yes" after having played video-games.

if and only if he blushes. (Recall that L is the set of personal external states where Child lies, and note that L^c denotes its complement.) The expected utilities of saying "yes" or "no" after doing homework and playing videogames can be retrieved as[42]

$$\begin{aligned}
\bar{u}_{C,(w)}(yes, \theta_C, \tau_C^2) &= \tau_{C,2}\big(\{s_D : s_D\big((yes, \neg b)\big) = buy\}|w\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes, \neg b)\big)|w\big] = 1;\\
\bar{u}_{C,(w)}(no, \theta_C, \tau_C^2) &= \tau_{C,2}\big(\{s_D : s_D\big((no, \neg b)\big) = buy\}|w\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(no, \neg b)\big)|w\big] = 0;\\
\bar{u}_{C,(v)}(yes, \theta_C, \tau_C^2) &= \theta + q\big(\tau_{C,2}\big(\{s_D : s_D\big((yes, \neg b)\big) = buy\}|v\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes, \neg b)\big)|v\big]\big) +\\
&\quad + (1-q)\big(\tau_{C,2}\big(\{s_D : s_D\big((yes, b)\big) = buy\}|v\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes, b)\big)|v\big]\big)\\
&= \theta - 1;\\
\bar{u}_{C,(v)}(no, \theta_C, \tau_C^2) &= \theta + \tau_{C,2}\big(\{s_D : s_D\big((no, \neg b)\big) = buy\}|v\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(no, \neg b)\big)|v\big]\big) = \theta;
\end{aligned}$$

where the second equality in each line follows from assumptions (6) and (7) on Child's beliefs.

Under (6) and (7), it is optimal for Child to say *yes* after doing homework, and to say *no* after playing video-games. Knowing this, at the beginning of the game, Child chooses between doing his homework and saying *yes*, and playing video-games and saying *no*. These two courses of actions yield expected utilities of 1 and θ, respectively. The latter is preferred if and only if $\theta \geq 1$. Note that Child's preferences are own-plan independent. ▲

4.5 Consistency

As a final building block for our definition of rationality, we require that rational players effectively carry out their plans – that is, the behavior described by their personal external states coincides with what they plan to do.

Definition 10. *Player i is **consistent** at personal state $(s_i, \theta_i, \tau_i^\infty)$ if, for each $h_i \in H_i$,*

$$\sigma(\tau_i^\infty)(s_i(h_i)|h_i) = 1.$$

Let CON_i be the set of personal states where player i is consistent.

Lemma 7. *For each $i \in I$, CON_i is closed.*

4.6 Rationality

We take rationality to be the conjunction of the requirements listed in Sections 4.1-4.5.

Definition 11. *Player i is **rational** at personal state $(s_i, \theta_i, \tau_i^\infty)$ if $(s_i, \theta_i, \tau_i^\infty) \in C_i \cap KB_i \cap BR_i \cap RP_i \cap CON_i$. Let R_i denote the set of personal states where i is rational.*

By the results of previous sections, the following is straightforward.

Lemma 8. *If feedback is regular and own-belief independent, R_i is measurable for each $i \in I$.*

[42] A detailed derivation of Child's local decision utilities is given in Appendix B.1.

Our notion of rationality deserves some comments. First, it is richer than the one usually adopted in the literature because we distinguish plans from objective behavior (cf. also Battigalli and De Vito, 2021). Moreover, we require a player's plan to assign positive probability, at each personal history, only to optimal actions: *in conjunction with consistency*, this implies that a player's personal external state must prescribe optimal actions at each personal history, and not only at personal histories it allows for. This is motivated by the observation that players do not commit to personal external states (in fact, they need not even know their true ones).

In light of Lemma 3, rational (hence, coherent) players are able to formulate beliefs over the set of personal states of opponents. Measurability of $R_i \subseteq S_i \times \Theta_i \times \mathcal{T}_i^\infty$ $(i \in I)$ ensures that a rational player $j \in I \setminus \{i\}$ can wonder about the rationality of i in a well-defined way.

5 Strong Δ-rationalizability

The aim of this section is to consider some profile Δ of restricted sets of beliefs — suggested by the application and context — and define a strong Δ-rationalizability procedure for the framework developed so far. Such procedure is a version of the strong rationalizability procedure that incorporates some contextual and transparent restrictions to players' beliefs (see Battigalli and Tebaldi, 2019, Battigalli et al., 2019a and relevant references therein). This in turn builds on earlier concepts of rationalizability for sequential games (Pearce, 1984). The epistemic foundations of our solution concept will be discussed in Section 6 – for the moment, it is enough to note that it captures the behavioral implications of rationality and forward-induction reasoning. This way of reasoning posits that players interpret others' moves as purposeful choices: in this way, they try to rationalize such moves, making inferences about opponents' beliefs, traits, and future behavior.

We begin with some terminology. A profile of *belief restrictions* is $\Delta = (\Delta_i)_{i \in I}$, where, for each $i \in I$, $\Delta_i = (\Delta_{\theta_i})_{\theta_i \in \Theta_i}$ and $\Delta_{\theta_i} \in \mathcal{B}(\mathcal{T}_i^{K+1})$. That is, each trait-type of a given player is associated to a measurable subset of the set of hierarchical system of beliefs of order $K+1$ of that player, and such mapping reflects some belief restrictions that are deemed relevant in the applications at hands. For notational convenience, define, for each $i \in I$ and $\theta_i \in \Theta_i$, $\Delta_{\theta_i}^\infty := \Delta_{\theta_i} \times \left(\bigtimes_{k \geq K+2} \mathcal{T}_{i,k}\right)$. Throughout this section and the next one, assume that a game and a profile Δ are fixed.

Given a measure μ defined over the measurable space $(D, \mathcal{B}(D))$ with D Polish, we denote by μ^* the *outer measure* defined over $(D, 2^D)$ defined, for each $F \subseteq D$, as:[43]

$$\mu^*(F) := \inf \left\{ \mu(G) \in [0,1] : G \in \mathcal{B}(D), F \subseteq G \right\}.$$

Then, we say that a $(K+1)$-th-order system of beliefs of player i $\tau_{i,K+1}$, *strongly believes* $F \in 2^{\Omega_{-i}^K}$ if, for each $h_i \in H_i$, $F \cap \Omega^K_{-i,\tau_i^K}(h_i) \neq \emptyset$ implies $\tau^*_{i,K+1}(F|h_i) = 1$, where τ_i^K is the K-th-order hierarchical system of beliefs obtained by taking, for each $h_i \in H_i$ the marginals of $\tau_{i,K+1}(\cdot\,|h_i)$ over the tuple of sets $(\Omega^0, (\Omega^n_{-i})_{n=1}^{K-1})$.

[43] Note that the following definition implies that $\mu^*(F) = \mu(F)$ if F is Borel, and that F differs from a Borel set only by a μ^*-null set if F is analytic but not Borel.

Consider the following procedure.[44]

Definition 12. *First, define* $\mathbf{P}_i^{\Delta}(0) := S_i \times \Theta_i \times \mathcal{T}_i^K$, $\mathbf{P}_{-i}^{\Delta}(0) := S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K$, *and* $\mathbf{P}^{\Delta}(0) := S \times \Theta \times \mathcal{T}^K$. *Then, for each* $n \geq 1$ *and* $i \in I$, $(s_i, \theta_i, \tau_i^K) \in \mathbf{P}_i^{\Delta}(n)$ *if and only if there exists* $\bar{\tau}_{i,K+1} \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ *such that:*

1. $(\tau_i^K, \bar{\tau}_{i,K+1}) \in \text{proj}_{\mathcal{T}_i^{K+1}} \left(\mathcal{T}_{i,C}^{\infty} \cap \Delta_{\theta_i}^{\infty}\right)$;
2. *for each* $h_i \in H_i$, $s_i(h_i) \in r_{i,h_i}(\theta_i, (\tau_i^K, \bar{\tau}_{i,K+1}))$;
3. *for each* $h_i \in H_i$, $\tau_{i,K+1}(S_i(h_i, s_i(h_i))|h_i) = 1$;
4. *for each* $k \in \{1, \ldots, n-1\}$, $\bar{\tau}_{i,K+1}$ *strongly believes* $\mathbf{P}_{-i}^{\Delta}(k)$.

Define $\mathbf{P}_{-i}^{\Delta}(n) := \bigtimes_{j \in I \setminus \{i\}} \mathbf{P}_j^{\Delta}(n)$ *and* $\mathbf{P}^{\Delta}(n) := \bigtimes_{i \in I} \mathbf{P}_i^{\Delta}(n)$.

In Definition 12, utility-relevant states are iteratively deleted if they fail to satisfy some requirements that mirror closely the rationality conditions of Section 4, plus the strong-belief requirement. However, this procedure is carried out on utility-relevant states, rather than on states of the world. Note that for standard games (i.e., when $K = 0$) utility-relevant states have the form (θ, s), and we obtain the strong rationalizability procedure of Battigalli (2003) and Battigalli and Prestipino (2013).

Lemma 9. *Fix a profile of belief restrictions* Δ. *For each* $n \in \mathbb{N}$ *and* $i \in I$, (i) *if feedback is regular and own-belief independent,* $\mathbf{P}_i^{\Delta}(n)$ *is analytic, and* (ii) $\mathbf{P}_i^{\Delta}(n) \subseteq \mathbf{P}_i^{\Delta}(n-1)$.

Thanks to Lemma 9, the limit of the sequence $(\mathbf{P}^{\Delta}(n))_{n \in \mathbb{N}}$ is well-defined: we say that a utility-relevant state (s, θ, τ^K) is *strongly* Δ*-rationalizable* if $(s, \theta, \tau^K) \in \mathbf{P}^{\Delta}(\infty) := \bigcap_{n \in \mathbb{N}} \mathbf{P}^{\Delta}(n)$. Note that, without additional assumptions, the set of strongly Δ-rationalizable states may be empty because Δ may entail restrictions on endogenous beliefs that are ultimately inconsistent with strategic reasoning.

However, nonemptiness obtains in a number of cases of interest. For instance, one can show that the set of strongly Δ-rationalizable states is nonempty when feedback is regular and own-belief independent, players' preferences are own-plan independent, and Δ only restricts initial beliefs about traits, or beliefs about such beliefs.

A slightly simpler and more convenient procedure has been proposed in the literature for standard games. We adapt it to our framework with belief-dependent preferences.

Definition 13. *First, define* $\mathbf{Q}_i^{\Delta}(0) := S_i \times \Theta_i \times \mathcal{T}_i^K$, $\mathbf{Q}_{-i}^{\Delta}(0) := S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K$, *and* $\mathbf{Q}(0)^{\Delta} := S \times \Theta \times \mathcal{T}^K$. *Then, for each* $n \geq 1$ *and* $i \in I$, $(s_i, \theta_i, \tau_i^K) \in \mathbf{Q}_i^{\Delta}(n)$ *if and only if*

0M $(s_i, \theta_i, \tau_i^K) \in \mathbf{Q}_i^{\Delta}(n-1)$;

and there exists $\bar{\tau}_{i,K+1} \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ *such that*

[44] In th following, we denote by $\mathcal{T}_{i,K+1,KB}$ and $\mathcal{T}_{i,K+1,CBU}$ the set of systems of beliefs of order $K+1$ that satisfy knowledge-implies-belief and correct belief updating, respectively.

1M $(\tau_i^K, \bar{\tau}_{i,K+1}) \in \text{proj}_{\mathcal{T}_i^{K+1}} \left(\mathcal{T}_{i,C}^{\infty} \cap \Delta_{\theta_i}^{\infty}\right)$*;*

2M for each $h_i \in H_i$*,* $s_i(h_i) \in r_{i,h_i}(\theta_i, (\tau_i^K, \bar{\tau}_{i,K+1}))$*;*

3M for each $h_i \in H_i$*,* $\bar{\tau}_{i,K+1}(S_i(h_i, s_i(h_i))|h_i) = 1$*;*

4M $\bar{\tau}_{i,K+1}$ *strongly believes* $\mathbf{Q}_{-i}^{\Delta}(n-1)$*.*

Define $\mathbf{Q}_{-i}^{\Delta}(n) := \bigtimes_{j \in I \setminus \{i\}} \mathbf{Q}_j^{\Delta}(n)$ *and* $\mathbf{Q}^{\Delta}(n) := \bigtimes_{i \in I} \mathbf{Q}_i(n)$*.*

Such procedure has been called as "naive" strong Δ-rationalizability (Battigalli and Prestipino, 2013). We could also label it as "memoryless," or "one-step," as each elimination round only relies on the previous step (to appreciate this, compare requirements 0M and 4M of Definition 13 with requirement 4 of Definition 12). Adapting the proof of Lemma 9, one gets the following.

Remark 9. Fix a profile of belief restrictions Δ. For each $n \in \mathbb{N}$ and $i \in I$, (i) if feedback is regular and own-belief independent, $\mathbf{Q}_i^{\Delta}(n)$ is analytic, and (ii) $\mathbf{Q}_i^{\Delta}(n) \subseteq \mathbf{Q}_i^{\Delta}(n-1)$.

Remark 9 implies that $\mathbf{Q}^{\Delta}(\infty) := \bigcap_{n \in \mathbb{N}} \mathbf{Q}^{\Delta}(n)$ is meaningfully defined. It is natural to wonder if the two procedures are equivalent. We provide an affirmative answer for a special case of belief restrictions (see Battigalli and Prestipino, 2013 for a similar result concerning standard games) under the assumption of own-plan independence of preferences. We say that $\Delta = (\Delta_{\theta_i})_{i \in I, \theta_i \in \Theta_i}$ is *rectangular* if, for each $i \in I$ and $\theta_i \in \Theta_i$, Δ_{θ_i} is a measurable rectangle. This means that, for each $i \in I$ and $\theta_i \in \Theta_i$, there exists a profile of measurable sets $((B_{\theta_i,n,h_i})_{h_i \in \bar{H}_i})_{n=1}^{K+1}$ such that $B_{\theta_i,n,h_i} \subseteq \Delta(\Omega_{-i}^{n-1})$ and $\Delta_{\theta_i} = \bigtimes_{n=1}^{K+1} \bigtimes_{h_i \in \bar{H}_i} B_{\theta_i,n,h_i}$.[45] In words, B_{θ_i,n,h_i} is the measurable set of n-th-order beliefs player i is allowed to hold at history h_i when her trait is θ_i.

Proposition 2. *Assume that the profile of belief restrictions* Δ *is rectangular and preferences are own-plan independent for each player. For all* $i \in I$ *and* $n \in \mathbb{N} \cup \{0\}$*,* $\mathbf{P}_i^{\Delta}(n) = \mathbf{Q}_i^{\Delta}(n)$*.*

We conclude with an illustration of the procedure.

Example 5 (Buy me an ice-cream, continued). For simplicity, we do not impose belief restrictions and we assume $\Theta_C = \{\theta', \theta''\}$, with $0 < \theta' < 1 < \theta''$. To keep the exposition simple, we describe the procedure only informally.[46] Moreover, given condition 3 of Definition 12, we can assume players have deterministic plans coinciding with their personal external state: for simplicity, we talk directly of optimal personal external states. Lastly, with own-plan independence (hence, consistency) of preferences, we can assume that Child commits to a plan among $w.yes$, $w.no$, $v.yes$, and $v.no$ at the root of the game.

Step 1 It is possible to check that $v.no$ grants Child a strictly higher expected utility than $w.no$ at the root of the game.[47] Thus, $\text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(1) = \{w.yes, v.yes, v.no\} \times \{\theta', \theta''\}$.

[45] With some abuse, we write Ω_{-i}^0 instead of Ω^0 to ease notation.

[46] A formal analysis is reported in Appendix B.

[47] Intuitively, if he correctly expects to play *no* in the second stage, he would be sure not to blush after his report. Then, his expectation about Dad's behavior (which he knows to depend on the fact that he observes personal history $(no, \neg b)$) will be exactly the same regardless of whether he plays $w.no$ or $v.no$, as they both give rise to Dad's personal history $(no, \neg b)$. Playing video-games thus allows Child to secure a higher expected utility.

Turning to Dad, note that condition 1 of Definition 12 implies that he is sure that Child played video-games in the first stage whenever he observes (yes, b). He is better off not buying him the ice-cream in such case, so that $\text{proj}_{S_D} \mathbf{P}_D(1) = \big\{s_D \in S_D : s_D\big((yes, b)\big) = not\big\}$.

Step 2 Taking into account the first step, Child realizes that Dad will not buy the ice-cream if he sees him blush. This undermine Child confidence, who will blush for sure upon choosing $v.yes$.[48] Moreover, Dad will spot Child's lie for sure, and Child's image concerns then imply that $v.yes$ is strictly worse than $v.no$. Now note that $v.no$ ensures a utility of θ coming from video-games: for trait-type θ'', this is higher than the maximum utility that $w.yes$ may lead to (i.e., the utility of 1 coming from the ice-cream). Hence, $\text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(2) = \big(\{w.yes, v.no\} \times \{\theta'\}\big) \cup \big(\{v.no\} \times \{\theta''\}\big)$. On the other hand, Dad's strong belief in $\mathbf{P}_C(1)$ leads him to conclude that it must be the case that Child played video-games in the first stage whenever he observes $(N, \neg b)$. Thus, upon observing $(no, \neg b)$, he is sure that Child did not do his homework, and he will not buy him the ice-cream in such case. We obtain $\text{proj}_{S_D} \mathbf{P}_D(2) = \big\{s_D : s_D\big((no, \neg b)\big) = s_D\big((yes, b)\big) = not\big\}$. That is, he will not buy Child an ice-cream if he observes $(no, \neg b)$ or (yes, b).

Step 3 This step has no behavioral implications for Child, because trait-type θ' is not sure of Dad's behavior after $(yes, \neg b)$, so both $w.yes$ and $v.no$ can be optimal for some belief (e.g., the latter is optimal if he is sure that Dad would not buy him the ice-cream also if he observes $(yes, \neg b)$). Dad instead concludes, by strong belief in $\mathbf{P}_C(1)$ and $\mathbf{P}_C(2)$, that personal history $(yes, \neg b)$ realizes if and only if Child did his homework. Upon observing such personal history, he should therefore buy him an ice-cream. Thus, $\text{proj}_{S_D} \mathbf{P}_D(3) = \big\{not.buy.not\big\}$.

Step 4 At this point, by strong belief in all previous steps, Child is sure that Dad will believe him and buy him an ice-cream if she observes $(yes, \neg b)$. Therefore, $w.yes$ allows to secure the ice-cream without being blamed. Trait-type θ' finds it optimal to play according to $w.yes$, as his value of video-games (i.e., θ') is lower than that of ice-cream (i.e., 1). Hence, $\text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(4) = \{(w.yes, \theta'), (v, no, \theta'')\}$.

Subsequent steps of the procedure do not yield further behavioral implications. This result shows that the possibility of betraying a lie through an emotional signal provides Child with a strong enough incentive to truthfully reveal the action he privately chose. This "full disclosure" result seems interesting, as we believe that this basic structure of interaction can be applied also to other situations, where (i) player 1 privately chooses an action and makes a declaration about his behavior to player 2, (ii) player 1 dislikes being perceived as a liar, and (iii) player 2 acts after observing player 1's report. Resorting to image concern motivations may be less reasonable in different economic settings, but our insights would still apply if we replace condition (ii) with

[48] Recall that Child's confidence (i.e., his ability to not blush) is exactly the probability with which he believes Dad would buy him an ice-cream despite the blushing.

the possibility for player 2 to enforce a punishment. The possibility of using emotional feedback to assess the truthfulness of a statement makes our framework well-suited for the analysis of information transmission in situations where factors like facial mimicry are crucial (e.g., political speeches, sales pitches, or face-to-face bargaining). ▲

6 Epistemic justification of strong Δ-rationalizability

In this section, we show that the proposed procedure captures the utility-relevant implications of some meaningful epistemic assumptions, that is, players' rationality and strong belief in rationality, as well as common strong (correct) belief in the restrictions described by Δ.[49] The notion of strong belief requires that a player be certain of a given event about her opponents whenever it is not falsified by evidence (cf. the definition of strong belief for hierarchical systems of beliefs given in Section 5). Imposing strong belief in rationality therefore essentially entails an assumption about players' belief-revision policy.

In order to carry out a formal analysis, we introduce two operators, that define sets which formally represent the propositions "player i would believe event F_{-i}, were she to observe personal history h_i" and "player i strongly believes event F_{-i}." To invoke Lemma 3, we restrict attention to coherent epistemic types of a player. Then, we formalize the notion of "degree of strategic sophistication," and we prove the main result of the paper.

For each player $i \in I$, personal history $h_i \in \bar{H}_i$, and event $F_{-i} \in \mathcal{B}(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{\infty})$, we define the *belief operator* of player i at personal history h_i and the *strong belief operator*, as:

$$\mathrm{B}_{i,h_i}(F_{-i}) := \left\{(s_i, \theta_i, \tau_i^{\infty}) \in C_i : \varphi_i(\tau_i^{\infty})(F|h_i) = 1\right\}; \quad \mathrm{SB}_i(F_{-i}) := \bigcap_{h_i \in H_i : \Omega^{\infty}_{-i,\tau_i^K}(h_i) \cap F_{-i} \neq \emptyset} \mathrm{B}_{i,h_i}(F_{-i}).$$

Note that the intersection in the definition of the strong belief operator is taken over personal histories that do not contradict event F_{-i}. This clarifies the interpretation of strong belief as "belief whenever possible." Under the usual technical assumptions, the belief and strong belief operators can be seen as maps from $\mathcal{B}(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{\infty})$ to $\mathcal{B}(C_i)$.

Lemma 10. *If feedback is regular and own-belief independent,* $\mathrm{B}_{i,h_i}(F_{-i})$ *and* $\mathrm{SB}_i(F_{-i})$ *are measurable for all* $i \in I$, $h_i \in \bar{H}_i$, *and* $F_{-i} \in \mathcal{B}(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{\infty})$.

Lastly, the set of personal states of a given player in which given belief restrictions (specified by Δ) are met is denoted $[\Delta_i] = \{(s_i, \theta_i, \tau_i^{\infty}) \in S_i \times \Theta_i \times \mathcal{T}_i^{\infty} : \tau_i^{\infty} \in \Delta_{\theta_i}^{\infty}\}$.

Remark 10. For each $i \in I$, $[\Delta_i]$ is measurable.[50]

[49] Battigalli and Siniscalchi (2002) provide an epistemic justification of strong rationalizability, neglecting restrictions on players' beliefs. For an epistemic foundation of strong directed rationalizability, see Battigalli and Prestipino (2013) and relevant references therein. Battigalli and Tebaldi (2019) and Battigalli et al. (2020) analyze strong directed rationalizability in a class of infinite games and in psychological games, respectively.

[50] The remark follows from the fact that $[\Delta_i]$ can be written as $S_i \times \bigcup_{\theta_i \in \Theta_i}(\{\theta_i\} \times \Delta_{\theta_i}^{\infty})$. Then, measurability of Δ_{θ_i} (which is assumed) yields the desired result.

At this point, we can turn to the description of players' degrees of strategic sophistication. For each $i \in I$, we define the following:

$$\mathbf{R}_i^{\Delta}(1) := R_i \cap [\Delta_i], \quad \mathbf{R}_{-i}^{\Delta}(1) := \bigtimes_{j \in I \setminus \{i\}} \mathbf{R}_j^{\Delta}(1), \quad \mathbf{R}^{\Delta}(1) := \bigtimes_{i \in I} \mathbf{R}_i^{\Delta}(1).$$

Then, for each $n \geq 2$, define:

$$\mathbf{R}_i^{\Delta}(n) := \mathbf{R}_i^{\Delta}(n-1) \cap \mathrm{SB}_i(\mathbf{R}_{-i}^{\Delta}(n-1)), \quad \mathbf{R}_{-i}^{\Delta}(n) := \bigtimes_{j \in I \setminus \{i\}} \mathbf{R}_j^{\Delta}(n), \quad \mathbf{R}^{\Delta}(n) := \bigtimes_{i \in I} \mathbf{R}_i^{\Delta}(n).$$

In words, the first degree of strategic sophistication consists in being rational and holding beliefs that satisfy the relevant restrictions described by profile Δ. A second-order strategically sophisticated player maintains whenever possible that her opponents are first-order strategically sophisticated, on top of being rational herself. A third-order strategically sophisticated player is rational and maintains whenever possible that her opponents are second-order strategically sophisticated. Were the latter hypothesis to be contradicted by evidence, a third-order strategically sophisticated player would "switch" to the assumption that her opponents are only first-order strategically sophisticated, unless also this weaker hypothesis is contradicted. The bottom line is that, under our epistemic assumptions, players ascribe to opponents the highest level of strategic sophistication consistent with evidence, i.e., they comply with the "best rationalization principle" (see, e.g., Battigalli and Prestipino, 2013 and the relevant references therein).

Remark 11. Fix a profile of belief restrictions Δ. If feedback is regular and own-belief independent, $\mathbf{R}_i^{\Delta}(n)$ is measurable and $\mathbf{R}_i^{\Delta}(n+1) \subseteq \mathbf{R}_i^{\Delta}(n)$ for each $i \in I$ and $n \in \mathbb{N}$.[51]

Given that $(\mathbf{R}_i^{\Delta}(n))_{n \in \mathbb{N}}$ is decreasing for each $i \in I$, so is $(\mathbf{R}^{\Delta}(n))_{n \in \mathbb{N}}$. Thus, we can define $\mathbf{R}^{\Delta}(\infty) := \bigcap_{n \in \mathbb{N}} \mathbf{R}^{\Delta}(n)$, which is measurable because of Remark 11. We say that $\mathbf{R}^{\Delta}(\infty)$ is the event in which (i) players are rational, (ii) players' beliefs satisfy restrictions Δ, and (iii) there is common strong belief in (i) and (ii). The following result establishes the epistemic justification of strong Δ-rationalizability.

Theorem 1. *Fix a profile of belief restrictions Δ. If feedback is regular and own-belief independent, $\mathbf{P}_i^{\Delta}(n) = \mathrm{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n)$ for each $i \in I$ and $n \in \mathbb{N}$.*

By Theorem 1, strong Δ-rationalizability characterizes the utility-relevant implications of rationality, the belief restrictions Δ and n-mutual strong belief of the conjunction of rationality and the belief restriction.

[51] That the sequence $(\mathbf{R}_i^{\Delta}(n))_{n \in \mathbb{N}}$ is decreasing is immediate. The first part of the remark follows instead from an induction argument. As for the basis step, note that $\mathbf{R}_i^{\Delta}(1) = R_i \cap [\Delta_i]$, and both R_i and $[\Delta_i]$ are measurable (as per Lemma 8 and Remark 10). Then, assuming that $\mathbf{R}_i^{\Delta}(k)$ is measurable for each $i \in I$ and $k \in \{1, \ldots, n\}$, we write $\mathbf{R}_i^{\Delta}(n+1) = \mathbf{R}_i^{\Delta}(n) \cap \mathrm{SB}_i(\mathbf{R}_{-i}^{\Delta}(n))$: $\mathbf{R}_i^{\Delta}(n)$ and $\mathbf{R}_{-i}^{\Delta}(n)$ are measurable by the inductive hypothesis, and $\mathrm{SB}_i(\mathbf{R}_{-i}^{\Delta}(n))$ is measurable as per Lemma 10.

7 Conclusion

In this paper, we introduced a novel framework (i.e., a theoretical language and some structural assumptions) to incorporate noisy emotional feedback into games, that may effectively be adapted to relevant applications. Our framework can be used to derive testable theoretical predictions about the extent to which the appraisal of others' emotion affects choices (cf. Examples 1, 2, and 3), and to analyze an important economic problem such as information transmission in face-to-face interactions (cf. Examples 4 and 5). Our framework can be naturally applied to situations such as court hearings, presidential debates, political speeches, bargaining, product advertisement by salesmen, and physician-patient interactions. In all these settings, emotional leakage may shape incentives in interesting ways that would not be captured by standard models

In addition to calling for applied models, we believe that our contribution also adds value at a more abstract level. First, our rich description of rationality has the merit of disentangling the different requirements rational players should satisfy, as already emphasized in Section 1.3. In particular, specific failures of rationality both on the cognitive side (e.g., failure to update beliefs consistently with evidence) and the behavioral side (e.g., failure to implement plans) may be analyzed from an analyst's perspective. Even more interestingly, our language is rich enough to model situations in which players may contemplate and reason about cognitive failures of opponents. Such expressiveness is a key step toward a rigorous analysis of the implications of failures of rationality in strategic settings. In this regard, future research may investigate the utility-relevant implications of different sets of assumptions about players' cognitive and behavioral features.

All in all, we believe that the present paper offers an innovative and flexible way to analyze a pervasive phenomenon such as emotional leakage in face-to-face interactions. In this regard, we see our contribution as foundational, in that it provides the tools to model a class of relevant situations and a meaningfully-founded solution procedure to predict behavior. As showed by our running example, it is possible to derive tractable applications and interesting predictions, and further research along this lines would lead to a better understanding of how decisions are formed in a number of social interactions.

A Proofs

Proof of Proposition 1 (p. 19)

Fix $h = (h_i)_{i\in I} \in H$, $(s,\theta) \in S \times \Theta$ and $\bar{m}_{\mathrm{e}} \in M_{\mathrm{e}}$. Recall that we can write $\bar{m}_{\mathrm{e}} = ((\bar{m}_{i,j,\mathrm{e}})_{i\in I})_{j\in I\setminus\{i\}}$ (cf. footnote 6), where $m_{i,j,\mathrm{e}}$ is a message i observes about j. Then, to ease notation, let $\bar{\ell}_i = (\bar{m}_{j,i,\mathrm{e}})_{j\in I\setminus\{i\}}$. In words, $\bar{\ell}_i$ is i's emotional leakage (i.e., the profile of messages about i received by her opponents) implied by $\bar{m}_{\mathrm{e}}$. Note that $\bar{\ell}_i$ belongs to the set $L_i := \bigtimes_{j\in I\setminus\{i\}} M_{j,i,\mathrm{e}}$, and that $\bar{m}_{\mathrm{e}} = (\bar{\ell}_i)_{i\in I}$.

Consider now $\{(s,\theta)\} \times \{\tau^1 \in \mathcal{T}^1 : \bar{m}_{\mathrm{e}} \in \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^1)\}$, where we let $K = 1$ because

feedback is simple (cf. point (i) of Definition 2). It is possible to check that:

$$\{\tau^1 \in \mathcal{T}^1 : \bar{m}_{\mathrm{e}} \in \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^1)\} = \bigcap_{i\in I} \big\{\tau^1 \in \mathcal{T}^1 : \bar{\ell}_i \in \operatorname{supp}(\operatorname{marg}_{L_i} f_{h,\mathrm{e}}(s,\theta,\tau^1))\big\}. \tag{8}$$

Simplicity of feedback implies that, for each $i \in I$, $\big\{\tau^1 \in \mathcal{T}^1 : \bar{\ell}_i \in \operatorname{supp}(\operatorname{marg}_{L_i} f_{h,\mathrm{e}}(s,\theta,\tau^1))\big\}$ depends exclusively on $\tau_i^1(\,\cdot\,|h_i)$ (cf. point (ii) of Definition 2). Let $B_i \subseteq \Delta(\Omega^0)$ be the set of i's first-order beliefs allowing for $\bar{\ell}_i$ at h_i. Hence, for each $i \in I$,

$$\big\{\tau^1 \in \mathcal{T}^1 : \bar{\ell}_i \in \operatorname{supp}(\operatorname{marg}_{L_i} f_{h,\mathrm{e}}(s,\theta,\tau^1))\big\} = B_i \times \Big(\bigtimes_{h_i' \neq h_i} \Delta(\Omega^0)\Big) \times \Big(\bigtimes_{j\in I\setminus\{i\}} \mathcal{T}_j^1\Big).$$

Then, expression (8) can be rewritten as

$$\{\tau^1 \in \mathcal{T}^1 : \bar{m}_{\mathrm{e}} \in \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^1)\} = \Big(\bigtimes_{i\in I} B_i\Big) \times \Big(\bigtimes_{i\in I}\bigtimes_{h_i' \neq h_i} \Delta(\Omega^0)\Big),$$

which is a rectangle. However, $\{\tau^1 \in \mathcal{T}^1 : \bar{m}_{\mathrm{e}} \in \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^1)\}$ is measurable because of semi-regularity of feedback. Sections of measurable sets in product measurable spaces are measurable by definition, and therefore B_i is measurable for each $i \in I$. Hence, $\{\tau^1 \in \mathcal{T}^1 : \bar{m}_{\mathrm{e}} \in \operatorname{supp} f_{h,\mathrm{e}}(s,\theta,\tau^1)\}$ is a measurable rectangle, proving regularity. ■

Proof of Lemma 1 (p. 19)

We begin by defining the correspondence $\mathbf{H}_i : S \times \Theta \times \mathcal{T}^K \rightrightarrows \bar{H}_i$. We proceed inductively on the length of target personal histories to retrieve a sequence of correspondences $(\mathbf{H}_i^\ell : S \times \Theta \times \mathcal{T}^K \rightrightarrows \bar{H}_i^\ell)_{\ell=0}^T$, where $\mathbf{H}_i^\ell$ specifies i's possible personal histories of length ℓ for each utility-relevant state.

First, $\mathbf{H}_i^0$ is simply the correspondence $(s,\theta,\tau^K) \mapsto \{\varnothing\}$. Then, assume $\mathbf{H}_i^k$ has been defined for $k \in \{0,\dots,\ell-1\}$. Define $\mathbf{H}_i^\ell$ to be such that, for each $(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K$,

$$\begin{aligned}\mathbf{H}_i^\ell(s,\theta,\tau^K) := \big\{(a_i^\ell, m_i^\ell) \in \bar{H}_i^\ell : {} & (a_i^{\ell-1}, m_i^{\ell-1}) \in \mathbf{H}_i^{\ell-1}(s,\theta,\tau^K),\\ & a_{i,\ell} = s_i(a_i^{\ell-1}, m_i^{\ell-1}),\\ & m_{i,\ell} \in \operatorname{supp} f_{i,(a_i^{\ell-1},m_i^{\ell-1}),\mathrm{e}}(s,\theta,\tau^K)\big\}.\end{aligned}$$

In words, $h_i^\ell \in \mathbf{H}_i^\ell(s,\theta,\tau^K)$ if (i) its immediate predecessor $h_i^{\ell-1}$ belongs to $\mathbf{H}_i^{\ell-1}(s,\theta,\tau^K)$, (ii) i's behavior at $h_i^{\ell-1}$ is described by s_i, and (iii), the message i receives after $h_i^{\ell-1}$ is in the support of $f_{i,h_i^{\ell-1},\mathrm{e}}(s,\theta,\tau^K)$.

Finally, for each $(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K$, let

$$\mathbf{H}_i(s,\theta,\tau^K) := \bigcup_{\ell=0}^{L} \mathbf{H}_i^\ell(s,\theta,\tau^K).$$

Next, we move to the proof of the lemma. It is easy to check that the claim follows if we prove that the correspondences in the sequence $(\mathbf{H}_i^\ell : S \times \Theta \times \mathcal{T}^K \rightrightarrows \bar{H}_i^\ell)_{\ell=0}^L$ are measurable. We do so by induction, starting to note that $\mathbf{H}_i^0$ is trivially measurable. As basis step, assume that

$\mathbf{H}_i^k$ is measurable for $k \in \{0, \dots, \ell-1\}$. Consider then $(\mathbf{H}_i^\ell)^{-1}$. For each $h_i^\ell = (a_i^\ell, m_i^\ell) \in \bar{H}_i^\ell$, we have

$$\begin{aligned}(\mathbf{H}_i^\ell)^{-1}(h_i^\ell) = \big\{(s,\theta,\tau^K) : (a_i^{\ell-1}, m_i^{\ell-1}) \in \mathbf{H}_i^{\ell-1}(s,\theta,\tau^K),\\ a_{i,\ell} = s_i(a_i^{\ell-1}, m_i^{\ell-1}),\\ m_{i,\ell} \in \operatorname{supp} f_{i,(a_i^{\ell-1}, m_i^{\ell-1}),\mathrm{e}}(s,\theta,\tau^K)\big\}\\ = \big\{(s,\theta,\tau^K) : (a_i^{\ell-1}, m_i^{\ell-1}) \in \mathbf{H}_i^{\ell-1}(s,\theta,\tau^K)\big\}\\ \cap \big\{(s,\theta,\tau^K) : a_{i,\ell} = s_i(a_i^{\ell-1}, m_i^{\ell-1})\big\}\\ \cap \big\{(s,\theta,\tau^K) : m_{i,\ell} \in \operatorname{supp} f_{i,(a_i^{\ell-1}, m_i^{\ell-1}),\mathrm{e}}(s,\theta,\tau^K)\big\}.\end{aligned}$$

Consider the three intersected sets. The first one is simply $(\mathbf{H}_i^{\ell-1})^{-1}(a_i^{\ell-1}, m_i^{\ell-1})$, and it is measurable by the inductive hypothesis. The second set is measurable because it takes the form $\tilde{S}_i \times S_{-i} \times \Theta \times \mathcal{T}^K$ for some $\tilde{S}_i \subseteq S_i$, and any subset of S_i is measurable (because S_i is finite and equipped with the discrete σ-algebra). Finally, the third set is measurable by semi-regularity of feedback.

To conclude the proof, note that that each $Q \subset \bar{H}_i^\ell$ is trivially closed, and $(\mathbf{H}_i^\ell)^{-1}(Q) = \bigcup_{h_i \in Q} (\mathbf{H}_i^\ell)^{-1}(h_i)$ is measurable. This checks the definition of measurability of correspondence $\mathbf{H}_i^\ell$. The result follows. ■

Proof of Lemma 2 (p. 21)

With some abuse, let $\Omega_{-i}^0 = S \times \Theta$ to simplify notation. Then, we rewrite $\mathcal{T}_{i,C}^\infty$ as follows:

$$\mathcal{T}_{i,C}^\infty = \bigcap_{n \in \mathbb{N}} \bigcap_{h_i \in \bar{H}_i} \left\{ \tau_i^\infty \in \mathcal{T}_i^\infty : \operatorname{marg}_{\Omega_{-i}^{n-1}} \tau_{i,n+1}(\,\cdot\,|h_i) = \tau_{i,n}(\,\cdot\,|h_i) \right\}.$$

Fix generic $\bar{n} \in \mathbb{N}$ and $\bar{h}_i \in \bar{H}_i$, and consider the corresponding set in the intersection above. Take a sequence $(\tau_{i,k}^\infty)_{k \in \mathbb{N}}$ of elements of such set converging to $\bar{\tau}_i^\infty$. This implies that $\tau_{i,\bar{n}+1,k}(\,\cdot\,|\bar{h}_i)$ converges to $\bar{\tau}_{i,\bar{n}+1}(\,\cdot\,|\bar{h}_i)$ in the topology of weak convergence. Then, by continuity of the marginalization map, $\operatorname{marg}_{\Omega_{-i}^{\bar{n}-1}} \bar{\tau}_{i,\bar{n}+1,k}(\,\cdot\,|\bar{h}_i) = \bar{\tau}_{i,\bar{n},k}(\,\cdot\,|\bar{h}_i)$. The same holds for any $n \in \mathbb{N}$ and $h_i \in \bar{H}_i$, as the chosen $\bar{n}$ and $\bar{h}_i$ were generic. Thus, $\mathcal{T}_{i,C}^\infty$ can be written as a countable intersection of closed sets. Arbitrary intersections of closed sets are closed, so we conclude that $\mathcal{T}_{i,C}^\infty$ is closed as well. Then, also C_i is closed, and the same holds for each $i \in I$. ■

Proof of Lemma 3 (p. 21)

The following auxiliary result is Lemma 1 of Brandenburger and Dekel (1993).

Lemma A1. *Let $(Z_n)_{n \in \mathbb{N} \cup \{0\}}$ be a sequence of Polish spaces, and define*

$$\Xi := \left\{ (\xi_n)_{n \in \mathbb{N}} : \forall\, n \geq 1,\ \xi_n \in \Delta\left(\bigtimes_{k=0}^{n-1} Z_k \right),\ \operatorname{marg}_{\times_{k=0}^{n-1} Z_k} \xi_{n+1} = \xi_n \right\}.$$

Then, there exists an homeomorphism $\psi : \Xi \to \Delta\left(\times_{n \in \mathbb{N} \cup \{0\}} Z_n\right)$.

In our setting, fixing $i \in I$, we denote $Z_0 = \Omega^0$, and $Z_n = \mathcal{T}_{-i,n}$ for each $n \in \mathbb{N}$. All such sets are compact metrizable (hence, Polish), as implied by Remark 2.

At this point, for each $h_i \in \bar{H}_i$, define $\gamma_{h_i} : \mathcal{T}_{i,C}^{\infty} \to \Xi$ to be the map $\tau_i^{\infty} \mapsto \tau_i^{\infty}(\cdot|h_i)$. Note that γ_{h_i} is clearly continuous for each $h_i \in \bar{H}_i$. Moreover, by Lemma A1, also the map $\varphi_{h_i} := \psi \circ \gamma_{h_i} : \mathcal{T}_{i,C}^{\infty} \to \Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})$ is continuous. Define now the map $\varphi_i := (\varphi_{h_i})_{h_i \in H_i} : \mathcal{T}_{i,C}^{\infty} \to \left[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\right]^{H_i}$. We want to show that it is indeed an homeomorphism.[52]

It is immediate to see that φ_i is continuous and that it satisfies the condition of Lemma 3. The latter fact implies that (i) φ_i is one-to-one, and (ii) φ_i^{-1} is continuous on $\varphi_i(\mathcal{T}_{i,C}^{\infty})$. Lastly, we show that $\varphi_i(\mathcal{T}_{i,C}^{\infty}) = \left[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\right]^{H_i}$. Indeed, $\varphi_i(\mathcal{T}_{i,C}^{\infty}) \subseteq \left[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\right]^{H_i}$ holds by definition. To see that $\left[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\right]^{H_i} \subseteq \varphi_i(\mathcal{T}_{i,C}^{\infty})$, take $t_i \in \left[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\right]^{H_i}$ and define $\tau_i^{\infty} \in \mathcal{T}_i^{\infty}$ to be such that, for each $n \in \mathbb{N}$ and $h_i \in H_i$, $\tau_{i,n}(\cdot|h_i) = \operatorname{marg}_{\Omega_{-i}^{n-1}} t_i(\cdot|h_i)$: by construction, $\tau_i^{\infty} \in \mathcal{T}_{i,C}^{\infty}$ and $\varphi_i(\tau_i^{\infty}) = t_i$, so that $\tau_i^{\infty} \in \varphi_i(\mathcal{T}_{i,C}^{\infty})$. ■

Proof of Lemma 4 (p. 21)

We first state some preparatory results for the proof of Lemma 4.

Lemma A2. *If feedback is own-belief independent, the collection* $\left\{\Omega^K_{-i,\tau_i^K}(h_i)\right\}_{\tau_i^K \in \mathcal{T}_i^K}$ *is finite for each* $i \in I$ *and* $h_i \in \bar{H}_i$.

Proof of Lemma A2. Fix $i \in I$ and $h_i^{\ell} \in \bar{H}_i$. We start by noting that:

$$\Omega^K_{-i,\tau_i^K}(h_i^{\ell}) = \bigcup_{(s,\theta)\in S\times\Theta} \left(\Omega^K_{-i,\tau_i^K,s,\theta}(h_i^{\ell})\right) = \bigcup_{(s,\theta)\in S\times\Theta} \left(\{s\} \times \{\theta\} \times (\mathbf{H}^{\ell}_{i,\tau_i^K,s,\theta})^{-1}(h_i^{\ell})\right), \tag{9}$$

where subscripts denote sections of the correspondence $(\mathbf{H}_i^{\ell})^{-1}$. Focus on $(\mathbf{H}^{\ell}_{i,\tau_i^K,s,\theta})^{-1}(h_i^{\ell})$. Denoting as h_i^k the k-long predecessor of h_i^{ℓ} (with $k \in \{0,\dots,\ell\}$), it can be written as:

$$\begin{aligned}(\mathbf{H}^{\ell}_{i,\tau_i^K,s,\theta})^{-1}(h_i^{\ell}) := \Big\{ & \tau_{-i}^K \in \mathcal{T}_{-i}^K : \forall\, k \in \{1,\dots,\ell\},\ \underline{(1,k)}\ a_{i,k} = s_i(h_i^{k-1}), \\ & \underline{(2,k)}\ m_{i,k} \in \bigcup_{\substack{h_{-i}^{k-1}:(h_i^{k-1},h_{-i}^{k-1}) \\ \in(\mathbf{H}_j^{k-1}(s,\theta,\tau^K))_{j\in I}}} \operatorname{supp} f_{i,(h_i^{k-1},h_{-i}^{k-1})}(s,\theta,\tau^K) \Big\}.\end{aligned}$$

Note that this expression for $(\mathbf{H}^{\ell}_{i,\tau_i^K,s,\theta})^{-1}(h_i^{\ell})$ features requirements that we labeled $\underline{(1,k)}$ and $\underline{(2,k)}$ (for $k \in \{1,\dots,\ell\}$). Denote as $G_{i,k} \subseteq \mathcal{T}_{-i}^K$ the set where condition (i,k) from the above definition holds. Note that $G_{1,k}$ is independent from players' hierarchical systems of beliefs for each $k \in \{1,\dots,\ell\}$. On the other hand, it is easy to check that, for each $k \in \{1,\dots,\ell\}$, $G_{2,k}$ belongs to following collection:

$$\left\{\tau_{-i} \in \mathcal{T}_{-i}^K : m_{i,k} \in \operatorname{supp} f_{i,(h_i^{k-1},h_{-i}^{k-1})}\big((a_{i,k},a_{-i}),\theta,(\tau_i^K,\tau_{-i}^K)\big)\right\}_{a_{-i}^{k-1}\in A_{-i}^{k-1},h_{-i}^{k-1}\in\bar{H}_{-i}^{k-1}},$$

[52] That is, a continuous one-to-one function with continuous inverse. Moreover, in order to establish that $\mathcal{T}_{i,C}^{\infty}$ and $\left[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\right]^{H_i}$ are actually homeomorphic, we will show that φ_i is also onto.

which is finite (by finiteness of A_{-i} and $\bar{H}_{-i}$) and independent from $\tau_i^K \in \mathcal{T}_i^K$ (by own-belief independence).

Thus, since $(\mathbf{H}^{\ell}_{i,\tau_i^K,s,\theta})^{-1}(h_i^\ell) = \bigcap_{k=1}^{\ell}\bigcap_{i=1}^{2} G_{i,k}$, the foregoing argument allows us to conclude that the collection $\{(\mathbf{H}^{\ell}_{i,\tau_i^K,s,\theta})^{-1}(h_i^\ell)\}_{\tau_i^K \in \mathcal{T}_i^K}$ is finite. With equation (9) and finiteness of set $S \times \Theta$, the desired result follows. ■

The proof is greatly simplified if we can partition the sets $\mathcal{T}_i^K$ $(i \in I)$ into measurable sets such that, all the hierarchical systems of beliefs in each of the cells of the partition lead to the same inference set (for a given personal history $h_i \in \bar{H}_i$). To do so, for each $i \in I$ and $h_i \in \bar{H}_i$, define the relation $\sim_{h_i}$ to be such that

$$\tau_i^K \sim_{h_i} \bar{\tau}_i^K \iff \Omega^K_{-i,\tau_i^K}(h_i) = \Omega^K_{-i,\bar{\tau}_i^K}(h_i).$$

It is routine to check that, for each $h_i \in \bar{H}_i$, $\sim_{h_i}$ is an equivalence relation. We can then define equivalence classes of elements of $\mathcal{T}_i^K$ in a standard way, as $[\tau_i^K]_{h_i} := \{\bar{\tau}_i^K \in \mathcal{T}_i^K : \bar{\tau}_i^K \sim_{h_i} \tau_i^K\}$.

Before checking that such classes are measurable for each $i \in I$ and $h_i \in \bar{H}_i$, we report two auxiliary results. The first is essentially a strengthening of Lemma 1 implied by regularity of feedback. The second is a result on measurable rectangles in product measurable spaces.

Lemma A3. *Let feedback be regular. For each $i \in I$, $(s,\theta) \in S \times \Theta$, and $h_i \in \bar{H}_i$, $(\mathbf{H}_{i,s,\theta})^{\ell}(h_i)$ is a union of measurable rectangles.*

Proof of Lemma A3. The proof is as that of Lemma 1: it is enough to replace semi-regularity with regularity. ■

Lemma A4. *Let $(X,\mathcal{X})$ and $(Y,\mathcal{Y})$ be measurable spaces, A, B, and $C \subseteq A \times B$ finite sets, and $((R_{a,b})_{a\in A})_{b\in C_a}$ a profile of measurable rectangles in $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$.*[53] *Let $R^* := \bigcap_{a\in A}\bigcup_{b\in C_a} R_{a,b}$. Then, for each $\bar{x} \in X$, $\{x \in X : R^*_x = R^*_{\bar{x}}\} \in \mathcal{X}$.*

Proof of Lemma A4. First recall that, by standard results, a finite union of measurable rectangles can be written as a finite union of disjoint measurable rectangles. Hence, for each $a \in A$, $\bigcup_{b\in C_a} R_{a,b} = \bigcup_{d\in D(a)} Q_{a,d}$ for some finite profile of disjoint measurable rectangles $(Q_{a,d})_{d\in D(a)}$ (note that we make the dependence of such profile on a explicit). Consider now the profile $((Q_{a,d})_{d\in D(a)})_{a\in A}$: we show that $\bigcap_{a\in A}\bigcup_{d\in D(a)} Q_{a,d}$ is a union of disjoint measurable rectangles. In particular, it is enough to show that this holds when $|A| = 2$ – then, an easy induction proves that the same holds for any finite A. Let $A = \{\alpha,\beta\}$. We claim that:

$$\bigcap_{a\in A}\bigcup_{d\in D(a)} Q_{a,d} = \bigcup_{i\in D(\alpha)}\bigcup_{j\in D(\beta)} (Q_{\alpha,i} \cap Q_{\beta,j}),$$

where the right hand side is clearly a finite union of (disjoint) measurable rectangles.

Fix $x \in \bigcap_{a\in A}\bigcup_{d\in D(a)} Q_{a,d}$. This implies that, for each $a \in A$, there is $d \in D(a)$ such that $x \in Q_{a,d}$. However, note that, for each $a \in A$, sets of the profile $(Q_{a,d})_{d\in D(a)}$ are disjoint. Hence,

[53] Note that we are allowing C not to have a rectangular shape. This justifies the presence of C_a (that is, the section of C at $a \in A$) in the definition of the profile of measurable rectangles.

for each $a \in A$, there is a unique $d^* \in D(a)$ such that $x \in Q_{a,d^*}$. Note that $A = \{\alpha, \beta\}$ and let $i^* \in D(\alpha)$ and $j^* \in D(\beta)$ be such that $x \in Q_{\alpha,i^*}$ and $x \in Q_{\beta,j^*}$ – that is, $x \in Q_{\alpha,i^*} \cap Q_{\beta,j^*}$. With this, we can conclude that $x \in \bigcup_{i \in D(\alpha)} \bigcup_{j \in D(\beta)} (Q_{\alpha,i} \cap Q_{\beta,j})$.

Now fix $x \in \bigcup_{i \in D(\alpha)} \bigcup_{j \in D(\beta)} (Q_{\alpha,i} \cap Q_{\beta,j})$. This implies that there are $i^* \in D(\alpha)$ and $j^* \in D(\beta)$ such that $x \in Q_{\alpha,i^*} \cap Q_{\beta,j^*}$ (specifically, such i^* and j^* are unique). This means that, for each $a \in A = \{\alpha, \beta\}$, there is $d \in D(a)$ such that $x \in Q_{a,d}$ – that is, $x \in \bigcap_{a \in A} \bigcup_{d \in D(a)} Q_{a,d}$.

At this point, we can conclude that the set of interest R^* is a finite union of (disjoint) measurable rectangles. For simplicity, write it as $R^* = \bigcup_{k \in K} R^*_k$, where K is finite and the measurable rectangles $(R^*_k)_{k \in K}$ are disjoint. Fix a generic $\bar{x} \in X$. If $\bar{x} \in \operatorname{proj}_X R^*$, it means that there is a (unique) $\bar{k} \in K$ such that $\bar{x} \in \operatorname{proj}_X R^*_{\bar{k}}$. Then, $\{x \in X : R^*_x = R^*_{\bar{x}}\} = \operatorname{proj}_X R^*_{\bar{k}}$, which is measurable as $R^*_{\bar{k}}$ is a measurable rectangle.

If instead $\bar{x} \notin \operatorname{proj}_X R^*$, $R^*_{\bar{x}} = \emptyset$ and $\{x \in X : R^*_x = R^*_{\bar{x}}\} = \operatorname{proj}_X (R^*_{\bar{k}})^C$. Now, $(R^*_{\bar{k}})^C$ is the complement of a measurable rectangle, hence it can be written as a (finite) union of disjoint measurable rectangles. The projection onto X of such union is simply the (finite) union of the projections of such measurable rectangles onto X, which are all measurable. Again, we conclude that $\{x \in X : R^*_x = R^*_{\bar{x}}\}$ is measurable, and this gives the desired result. ■

We can now check the measurability of the partition induced by $\sim_{h_i}$ $(i \in I, h_i \in \bar{H}_i)$.

Lemma A5. *If feedback is regular, $[\tau_i^K]_{h_i}$ is measurable for each $i \in I$ and $h_i \in \bar{H}_i$.*

Proof of Lemma A5. Fix generic $i \in I$, $h_i^\ell \in \bar{H}_i^\ell$, and $\bar{\tau}_i^K \in \mathcal{T}_i^K$, and note that, for each $\tau_i^K \in \mathcal{T}_i^K$,

$$\Omega^K_{-i,\tau_i^K}(h_i^\ell) = \bigcup_{(s,\theta) \in S \times \Theta} \left(\Omega^K_{-i,\tau_i^K,s,\theta}(h_i^\ell) \right),$$

where $\Omega^K_{-i,\tau_i^K,s,\theta}(h_i^\ell)$ is the section of $\Omega^K_{-i,\tau_i^K}(h_i^\ell)$ at (s,θ). Thus, it can be checked that, for each $\tau_i^K \in \mathcal{T}_i^K$, $\Omega^K_{-i,\tau_i^K}(h_i^\ell) = \Omega^K_{-i,\bar{\tau}_i^K}(h_i^\ell)$ if and only if, for each $(s,\theta) \in S \times \Theta$, $\Omega^K_{-i,\tau_i^K,s,\theta}(h_i^\ell) = \Omega^K_{-i,\bar{\tau}_i^K,s,\theta}(h_i^\ell)$. Note that, for each $\tau_i^K \in \mathcal{T}_i^K$,

$$\Omega^K_{-i,\tau_i^K,s,\theta}(h_i^\ell) = \{s\} \times \{\theta\} \times (\mathbf{H}^\ell_{i,\tau_i^K,s,\theta})^{-1}(h_i^\ell).$$

Then, we can say that, for each $\tau_i^K \in \mathcal{T}_i^K$, $\Omega^K_{-i,\tau_i^K,s,\theta}(h_i^\ell) = \Omega^K_{-i,\bar{\tau}_i^K,s,\theta}(h_i^\ell)$ if and only if $(\mathbf{H}^\ell_{i,\tau_i^K,s,\theta})^{-1}(h_i^\ell) = (\mathbf{H}^\ell_{i,\bar{\tau}_i^K,s,\theta})^{-1}(h_i^\ell)$.

Note that for each $i \in I$, $\tau_i \in \mathcal{T}_i^K$ and $h_i^\ell \in \bar{H}_i$ we can write $(\mathbf{H}^\ell_{i,s,\theta})^{-1}(h_i^\ell)$ as:

$$(\mathbf{H}^\ell_{i,s,\theta})^{-1}(h_i^\ell) = \Big\{ \tau^K \in \mathcal{T}^K : \forall\, k \in \{1,\dots,\ell\},\ \underline{(1,k)}\ a_{i,k} = s_i(h_i^{k-1}),$$
$$\underline{(2,k)}\ \exists\, h_{-i}^{k-1} \in \mathbf{H}^{k-1}_{-i,s,\theta}(\tau^K_{-i}), m_{i,k} \in \operatorname{supp} f_{i,(h_i^{k-1},h_{-i}^{k-1})}(s,\theta,\tau^K) \Big\}.$$

As before, let $G_{i,k} \subseteq \mathcal{T}^K_{-i}$ denote the set where condition $\underline{(i,k)}$ from the definition above holds, and define $G^* := \bigcap_{k=1}^t \bigcap_{j=1}^2 G_{j,k}$. Let $G^*_{\tau_i^K}$ denote the section of G^* at a generic $\tau_i^K \in \mathcal{T}_i^K$. With this, we observe that, for each $\tau_i^K \in \mathcal{T}_i^K$, $(\mathbf{H}^\ell_{i,\tau_i^K,s,\theta})^{-1}(h_i^\ell) = (\mathbf{H}^\ell_{i,\bar{\tau}_i^K,s,\theta})^{-1}(h_i^\ell)$ if and only

if $G^*_{\tau_i^K} = G^*_{\bar{\tau}_i^K}$. Next, note that $G_{1,k}$ is either empty or equal to $\mathcal{T}^K$ for each $k \in \{1, \dots, \ell\}$. On the other hand, $G_{2,k}$ is a (finite) union of measurable rectangles as per Lemma A3. Hence, $\bigcap_{k=1}^{\ell} \bigcap_{j=1}^{2} G_{j,k}$ and is a (finite) intersection of (finite) unions of measurable rectangles. Then, by Lemma A4, the set $\{\tau_i^K \in \mathcal{T}_i^K : G^*_{\tau_i^K} = G^*_{\bar{\tau}_i^K}\}$ is measurable, and this establishes the result. ■

Lemmas A2 and A5 imply the following convenient result.

Corollary A1. *If feedback is own-belief independent, $\{[\tau_i^K]_{h_i} : \tau_i^K \in \mathcal{T}_i^K\}$ is a finite partition of $\mathcal{T}_i^K$ for each $i \in I$ and $h_i \in \bar{H}_i$. If feedback is also regular, such partition is made of measurable sets.*

Next, we discuss measurability in $\Delta(X)$, where X is a separable topological space. In particular, the following is Proposition 7.25 of Bertsekas and Shreve (1996).

Lemma A6. *let X be a separable topological space, and $\mathcal{F}$ a collection of subsets of X that is closed under finite intersections and for which $\sigma(\mathcal{F}) = \mathcal{B}(X)$. Consider the sequence of functions $(\vartheta_F : \Delta(X) \to [0,1])_{F \in \mathcal{F}}$, where, for each $F \in \mathcal{F}$, ϑ_F is the map $\xi \mapsto \xi(F)$. Then,*

$$\mathcal{B}(\Delta(X)) = \sigma\Bigg(\bigcup_{F \in \mathcal{F}} \bigcup_{B \in \mathcal{B}(\mathbb{R})} \vartheta_F^{-1}(B) \Bigg).$$

When $\mathcal{F}$ is taken to be the collection of Borel sets of X, Lemma A6 gives the following, which is the definition of the Borel σ-algebra of $\Delta(X)$ used, e.g., by Dubins and Freedman (1964).

Remark A1. Let X be separable. $\mathcal{B}(\Delta(X))$ is the smallest σ-algebra that makes the evaluation maps $(\xi \mapsto \xi(B))_{B \in \mathcal{B}(X)}$ measurable.

We are now ready to start the proof of Lemma 4. Fix a generic $i \in I$ and rewrite:

$$\begin{aligned}
\mathcal{T}^{\infty}_{i,KB} &= \left\{ \tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \forall\, h_i \in \bar{H}_i, \tau_{i,K+1} \left(\Omega^K_{-i,\tau_i^K}(h_i) \middle| h_i \right) = 1 \right\} \\
&= \bigcap_{h_i \in \bar{H}_i} \left\{ \tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \exists\, [\bar{\tau}_i^K]_{h_i} \subseteq \mathcal{T}_i^K, \tau_i^K \in [\bar{\tau}_i^K]_{h_i}, \tau_{i,K+1} \left(\Omega^K_{-i,[\bar{\tau}_i^K]_{h_i}}(h_i) \middle| h_i \right) = 1 \right\} \\
&= \bigcap_{h_i \in \bar{H}_i} \bigcup_{[\bar{\tau}_i^K]_{h_i}} \left(\left([\bar{\tau}_i^K]_{h_i} \times \bigtimes_{k \geq K+1} \mathcal{T}_{i,k} \right) \cap \left\{ \tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \tau_{i,K+1} \left(\Omega^K_{-i,[\bar{\tau}_i^K]_{h_i}}(h_i) \middle| h_i \right) = 1 \right\} \right). \qquad (10)
\end{aligned}$$

Consider the expression within parentheses. The first set is measurable as per Lemma A5. As for the second one, it is measurable because the set $\left\{ \tau_{i,K+1}(\,\cdot\,|h_i) \in \Delta(\Omega^K_{-i}) : \tau_{i,K+1}\left(\Omega^K_{-i,\tau_i^K}(h_i) \middle| h_i\right) = 1 \right\}$ is measurable as per Remark A1. Then, the intersection and the union of equation (10) are countable (in particular, Corollary A1 ensures that the union over equivalence classes is finite). All in all, we conclude that $\mathcal{T}^{\infty}_{i,KB}$ can be written as the countable intersection and union of measurable sets, hence it is measurable. $KB_i = S_i \times \Theta_i \times \mathcal{T}^{\infty}_{i,KB}$ is measurable as well, and the same is true for each $i \in I$. ■

Proof of Lemma 5 (p. 23)

Fix a generic $i \in I$ and define the following:

$$\mathcal{T}_{i,CR}^{\infty} := \bigcap_{h_i \in H_i} \bigcap_{a_i \in \hat{\mathcal{A}}_i(h_i)} \bigcap_{h_i' \in \bar{H}_i(h_i)} \bigcap_{s_i \in S_i(h_i,a_i)} \{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \text{(CR) holds}\}; \tag{11}$$

$$\mathcal{T}_{i,BR}^{\infty} := \bigcap_{h_i \in H_i} \bigcap_{a_i \in \hat{\mathcal{A}}_i(h_i)} \bigcap_{m_i \in M_i(h_i,a_i)} \bigcap_{F \in \mathcal{B}(S_{-i} \times \Theta \times \mathcal{T}_{-i}^K)} \{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \text{(BR-}a_i\text{) holds}\},$$

$$\mathcal{T}_{i,I}^{\infty} := \bigcap_{h_i \in H_i} \{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \text{(I) holds}\}.$$

Note that $\mathcal{T}_{i,CBU}^{\infty} = \mathcal{T}_{i,CR}^{\infty} \cap \mathcal{T}_{i,BR}^{\infty}$. To establish the desired result, we prove that both $\mathcal{T}_{i,CR}^{\infty}$ and $\mathcal{T}_{i,BR}^{\infty}$ are measurable.

Step 1: $\mathcal{T}_{i,CR}^{\infty}$ is measurable. Fix $h_i \in H_i$, $a_i \in \hat{\mathcal{A}}_i(h_i)$, $h_i' \in \bar{H}_i(h_i)$, and $s_i \in S_i(h_i, a_i)$, and consider the corresponding set in equation (11):

$$\left\{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \tau_{i,K+1}(\{s_i\}|h_i') \cdot \tau_{i,K+1}(S_i(h_i,a_i)|h_i) = \tau_{i,K+1}(\{s_i\}|h_i)\right\}.$$

Note that the intersections in (11) are finite. Thus, it is enough to prove that the above set is measurable to conclude that $\mathcal{T}_{i,CR}^{\infty}$ is also measurable. We will actually do more: we will prove that the above set is closed – hence the intersection of (11) will also be closed.

Consider a sequence $(\tau_{i,n}^{\infty})_{n \in \mathbb{N}}$ of elements of $\mathcal{T}_{i,CR}^{\infty}$ converging to $\bar{\tau}_i^{\infty}$. Note that $\mathcal{T}_{i,CR}^{\infty}$ is a product space, and recall that convergence in product spaces occurs coordinate-wise under the assumed product topology. Thus, $\tau_{i,K+1,n}(\cdot\,|h_i') \to \bar{\tau}_{i,K+1}(\cdot\,|h_i')$ and $\tau_{i,K+1,n}(\cdot\,|h_i) \to \bar{\tau}_{i,K+1}(\cdot\,|h_i)$. Moreover, by the properties of the weak convergence topology, if $\tau_{i,K+1,n}(\cdot\,|h_i') \to \bar{\tau}_{i,K+1}(\cdot\,|h_i')$, then it must be the case that $\tau_{i,K+1,n}(C|h_i') \to \bar{\tau}_{i,K+1}(C|h_i')$ for every Borel set C with empty boundary (see Theorem 15.3 in Aliprantis and Border, 2006). Now notice that $\{s_i\}$, which is a shorthand for $\{s_i\} \times S_{-i} \times \Theta \times \mathcal{T}_{-i}^K$, is a clopen set because it is the product of clopen sets: $\{s_i\}$ is a subset of a finite space (hence it is clopen), and $S_{-i} \times \Theta \times \mathcal{T}_{-i}^K$ is a compact metrizable space (and for each compact metrizable space X, both X and $\emptyset$ are clopen). Clopen sets have empty boundaries, so we conclude that $\tau_{i,K+1,n}(\{s_i\}|h_i')$ converges to $\bar{\tau}_{i,K+1}(\{s_i\}|h_i')$. An entirely analogous point applies to show that $\tau_{i,K+1,n}(\{s_i\}|h_i) \to \bar{\tau}_{i,K+1}(\{s_i\}|h_i)$ and $\tau_{i,K+1,n}(S_i(h_i,a_i)|h_i) \to \bar{\tau}_{i,K+1}(S_i(h_i,a_i)|h_i)$. Wrapping up, we obtain

$$\bar{\tau}_{i,K+1}(\{s_i\}|h_i') \cdot \bar{\tau}_{i,K+1}(S_i(h_i,a_i)|h_i) = \bar{\tau}_{i,K+1}(\{s_i\}|h_i),$$

so that $\bar{\tau}_i^{\infty} \in \mathcal{T}_{i,CR}^{\infty}$, as desired. We conclude that $\mathcal{T}_{i,CR}^{\infty}$ is closed, hence measurable.

Step 2: $\mathcal{T}_{i,BR}^{\infty}$ is measurable. Consider now

$$\mathcal{T}_{i,BR}^{\infty} := \bigcap_{h_i \in H_i} \bigcap_{a_i \in \hat{\mathcal{A}}_i(h_i)} \bigcap_{m_i \in M_i(h_i,a_i)} \bigcap_{F \in \mathcal{B}(S_{-i} \times \Theta \times \mathcal{T}_{-i}^K)} \{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \text{(BR-}a_i\text{) holds}\}.$$

Note that in the expression above the intersection over $\mathcal{B}(S_{-i} \times \Theta \times \mathcal{T}_{-i}^K)$ is uncountable. Yet, $S_{-i} \times \Theta \times \mathcal{T}_{-i}^K$ is a compact metrizable space (it is the product of two finite spaces, S_{-i} and Θ, and of $\mathcal{T}_{-i}^K$, which is compact metrizable as per Remark 2), hence it is second countable – i.e., it

admits a countable base $\mathscr{B}$. Therefore, each Borel set in $S_{-i} \times \Theta \times \mathcal{T}_{-i}^K$ can be obtained through countable unions or intersections of elements of $\mathscr{B}$. We can then write:

$$\mathcal{T}_{i,BR}^\infty := \bigcap_{h_i \in H_i} \bigcap_{a_i \in \hat{\mathcal{A}}_i(h_i)} \bigcap_{m_i \in M_i(h_i,a_i)} \bigcap_{B \in \mathscr{B}} \{\tau_i^\infty \in \mathcal{T}_i^\infty : \text{(BR-}a_i\text{) holds}\}.$$

Note that now the intersections are countable: proving measurability of the intersected sets would then imply the desired result. Therefore, fix $h_i \in H_i$, $a_i \in \hat{\mathcal{A}}_i(h_i)$, $m_i \in M_i(h_i, a_i)$, and $B \in \mathscr{B}$, and consider the corresponding set in the above intersection:

$$\Big\{\tau_i^\infty \in \mathcal{T}_i^\infty : \tau_{i,K+1}(B|h_i') \cdot \int_{S_{-i}\times\Theta\times\mathcal{T}_{-i}^K} f_{i,h_i,s_i^*,\tau_i^K}(\,\cdot\,)[m_i] \cdot \big(\operatorname{marg} \tau_{i,K+1}\big)\big(\mathrm{d}(s_{-i},\theta,\tau_{-i}^K)|h_i\big)$$
$$= \int_B f_{i,h_i,s_i^*,\tau_i^K}(\,\cdot\,)[m_i] \cdot \big(\operatorname{marg} \tau_{i,K+1}\big)\big(\mathrm{d}(s_{-i},\theta,\tau_{-i}^K)|h_i\big)\Big\}, \tag{12}$$

where we write simply "marg" instead of "$\operatorname{marg}_{S_{-i}\times\Theta\times\mathcal{T}_{-i}^K}$" to ease notation.

In order to show its measurability, we show that the above set is the inverse image of a measurable set in $\mathbb{R}$ through a measurable function $\psi : \mathcal{T}_i^\infty \to \mathbb{R}$. To retrieve such function, we proceed in three steps:

1. Let ψ_1 be the map $\tau_i^\infty \mapsto \tau_{i,K+1}(B|h_i')$. Such map is measurable. Indeed, it is the composition of the two maps $\tau_i^\infty \mapsto \tau_{i,K+1}(\,\cdot\,|h_i')$ and $\tau_{i,K+1}(\,\cdot\,|h_i') \mapsto \tau_{i,K+1}(B|h_i')$: the former is continuous (hence, measurable), and the latter is measurable (by the properties of the Borel σ-algebras of sets of probability measures and by the fact that B is measurable, cf. Remark A1). Compositions of measurable maps are measurable, hence ψ_1 is measurable.

2. Let ψ_2 be the map

$$\tau_i^\infty \mapsto \int_{S_{-i}\times\Theta\times\mathcal{T}_{-i}^K} f_{i,h_i,s_i^*,\tau_i^K}(\,\cdot\,)[m_i] \cdot \big(\operatorname{marg} \tau_{i,K+1}\big)\big(\mathrm{d}(s_{-i},\theta,\tau_{-i}^K)|h_i\big).$$

   Such map is continuous. To see it, consider a sequence $(\tau_{i,n}^\infty)_{n\in\mathbb{N}}$ of elements of $\mathcal{T}_i^\infty$ converging to $\bar{\tau}_i^\infty$. This implies that $\tau_{i,K+1,n}(\,\cdot\,|h_i)$ converges to $\bar{\tau}_{i,K+1}(\,\cdot\,|h_i)$. Now note that, since the marginalization map is continuous, $\operatorname{marg} \tau_{i,K+1,n}(\,\cdot\,|h_i)$ converges to $\operatorname{marg} \bar{\tau}_{i,K+1}(\,\cdot\,|h_i)$. Since $f_{i,h_i,s_i^*,\tau_i^K}(\,\cdot\,)[m_i] : S_{-i} \times \Theta \times \mathcal{T}_{-i}^K \to [0,1]$ is continuous and bounded, by the very definition of the topology of weak convergence $\psi_2(\tau_{i,n}^\infty)$ converges to $\psi_2(\bar{\tau}_i^\infty)$. This proves continuity (hence, measurability) of ψ_2.

3. Let ψ_3 be the map

$$\tau_i^\infty \mapsto \int_B f_{i,h_i,s_i^*,\tau_i^K}(\,\cdot\,)[m_i] \cdot \big(\operatorname{marg} \tau_{i,K+1}\big)\big(\mathrm{d}(s_{-i},\theta,\tau_{-i}^K)|h_i\big).$$

   By arguments analogous to those of the previous point, ψ_3 is continuous.

Now define function $\psi : \mathcal{T}_i^\infty \to \mathbb{R}$ as $\psi := \psi_1 \cdot \psi_2 - \psi_3$, and note that the set in (12) can be written as $\{\tau_i^\infty \in \mathcal{T}_i^\infty : \psi(\tau_i^\infty) = 0\} = \psi^{-1}(\{0\})$. As a final step note that $\{0\} \in \mathcal{B}(\mathbb{R})$ and that ψ is measurable (as sums and products of measurable maps are measurable). We conclude that the set of interest is measurable, and this establishes measurability of $\mathcal{T}_{i,BR}^\infty$.

Step 3: $\mathcal{T}_{i,I}^{\infty}$ is measurable. We are actually going to prove that $\mathcal{T}_{i,I}^{\infty}$ is closed. To this end, write the set as

$$\mathcal{T}_{i,I}^{\infty} = \bigcap_{h_i \in \bar{H}_i} \{\tau_i^{\infty} : \tau_{i,K+1}(\cdot\,|h_i) = \text{marg}_{S_i \times \Theta_i}\, \tau_{i,K+1}(\cdot\,|h_i) \otimes \text{marg}_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{K}}\, \tau_{i,n,K+1}(\cdot\,|h_i)\}.$$

To show that each set in the above intersection is closed, fix $h_i \in \bar{H}_i$ and consider a sequence $(\tau_{i,n}^{\infty})_{n\in\mathbb{N}}$ converging to $\bar{\tau}_i^{\infty}$. This sequence induces a sequence of conditional beliefs of order $K+1$, $(\tau_{i,n,K+1}(\cdot\,|h_i))_{n\in\mathbb{N}}$ that (under the assumed product topology) converges to $\bar{\tau}_{i,K+1}(\cdot\,|h_i)$. By construction, we have that the following holds for each n:

$$\tau_{i,n,K+1}(\cdot\,|h_i) = \text{marg}_{S_i \times \Theta_i}\, \tau_{i,n,K+1}(\cdot\,|h_i) \otimes \text{marg}_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{K}}\, \tau_{i,n,K+1}(\cdot\,|h_i)\}.$$

Moreover, convergence of $(\tau_{i,n,K+1}(\cdot\,|h_i))_{n\in\mathbb{N}}$ and continuity of the marginalization maps implies

$$\text{marg}_{S_i \times \Theta_i}\, \tau_{i,n,K+1}(\cdot\,|h_i) \to \bar{\tau}_{i,K+1}(\cdot\,|h_i), \qquad \text{marg}_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{K}}\, \tau_{i,n,K+1}(\cdot\,|h_i) \to \bar{\tau}_{i,K+1}(\cdot\,|h_i),$$

in the topology of weak convergence. To conclude that

$$\bar{\tau}_{i,K+1}(\cdot\,|h_i) = \text{marg}_{S_i \times \Theta_i}\, \bar{\tau}_{i,K+1}(\cdot\,|h_i) \otimes \text{marg}_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{K}}\, \bar{\tau}_{i,K+1}(\cdot\,|h_i),$$

we state without proof the following fact. Consider standard Borel spaces $(X, \mathcal{B}(X))$ and $(X', \mathcal{B}(X'))$ and sequences of measures $(\mu_n)_{n\in\mathbb{N}}$ and $(\mu'_n)_{n\in\mathbb{N}}$ defined on them; if $X \times X'$ is separable, then $\mu_n \otimes \mu'_n \to \mu \otimes \mu'$ if and only if $\mu_n \to \mu$ and $\mu'_n \to \mu'$. This fact, together with the convergence of marginals, implies (A).

Conclusion. Given the measurability of $\mathcal{T}_{i,CR}^{\infty}$, $\mathcal{T}_{i,BR}^{\infty}$, and $\mathcal{T}_{i,I}^{\infty}$, $\mathcal{T}_{i,CBU}^{\infty}$ is measurable. ■

Proof of Lemma 6 (p. 26)

We begin with two preliminary observations. First, note that $RP_i = S_i \times \text{proj}_{\Theta_i \times \mathcal{T}_i^{\infty}} RP_i$ because personal external states are irrelevant in the definition of rational planning. Hence, it is enough to prove that $\widetilde{RP}_i := \text{proj}_{\Theta_i \times \mathcal{T}_i^{\infty}} RP_i$ is closed.

Second, Berge's maximum theorem, together with the observation that the objective function $u_{i,h_i} : \hat{\mathcal{A}}_i(h_i) \times \Theta_i \times \mathcal{T}_i^{\infty} \to \mathbb{R}$ is continuous (which is easily checked), implies that $r_{i,h_i} : \Theta_i \times \mathcal{T}_i^{\infty} \rightrightarrows \hat{\mathcal{A}}_i(h_i)$ is a upper hemicontinuous correspondence. For each sequences $(\theta_{i,n}, \tau_{i,n}^{\infty})_{n\in\mathbb{N}} \in \bigtimes_{n\in\mathbb{N}} (\Theta_i \times \mathcal{T}_i^{\infty})$ and $(a_{i,n})_{n\in\mathbb{N}} \in \bigtimes_{n\in\mathbb{N}} r_{i,h_i}(\theta_{i,n}, \tau_{i,n}^{\infty})$, and $a_i^* \in \hat{\mathcal{A}}_i(h_i)$, $(\theta_{i,n}, \tau_{i,n}^{\infty}) \to (\bar{\theta}_i, \bar{\tau}_i^{\infty})$ and $a_{i,n} \to a_i^*$ only if $a_i^* \in r_{i,h_i}(\bar{\theta}_i, \bar{\tau}_i^{\infty})$.

Next, we move to the main part of the proof, where we show that $\widetilde{RP}_i$ is closed. Consider a sequence $(\theta_{i,n}, \tau_{i,n}^{\infty})_{n\in\mathbb{N}}$ of elements of $\widetilde{RP}_i$ converging to $(\bar{\theta}_i, \bar{\tau}_i^{\infty})$. It is straightforward to check that $(\bar{\theta}_i, \bar{\tau}_i^{\infty}) \in \text{proj}_{\Theta_i \times \mathcal{T}_i^{\infty}} KT_i$ (i.e., $\bar{\tau}_i^{\infty}$ knows θ_i), and that $\bar{\tau}_i^{\infty}$ satisfies independence.[54] Proving that $\text{supp}\, \sigma_i(\bar{\tau}_i^{\infty})(\cdot\,|h_i) \subseteq r_{i,h_i}(\bar{\theta}_i, \bar{\tau}_i^{\infty})$ for each $h_i \in H_i$ then gives the desired result.

[54] The latter claim follows from relatively straightforward results in measure theory. In particular, consider measurable spaces $(X, \mathcal{X})$ and $(X', \mathcal{X}')$, and sequences of measures $(\mu_n)_{n\in\mathbb{N}}$ and $(\mu'_n)_{n\in\mathbb{N}}$ defined on them. If $X \times X'$ is separable, $\mu_n \times \mu'_n \to \mu \times \mu'$ if and only if $\mu_n \to \mu$ and $\mu'_n \to \mu'$, where convergence is assumed to occur in the topology of weak convergence of measures. This fact, together with the product structure and separability of $\mathcal{T}_i^{\infty}$, can be employed to show that independence is preserved in the limit.

To this end, fix $h_i \in H_i$ and note that $\sigma(\tau_{i,n}^\infty)(\cdot|h_i)$ converges to $\sigma(\bar{\tau}_i^\infty)(\cdot|h_i)$ because $\tau_{i,K+1,n}(\cdot|h_i)$ converges to $\bar{\tau}_{i,K+1}(\cdot|h_i)$. Given that $\sigma(\tau_{i,n}^\infty)(\cdot|h_i)$ and $\sigma(\bar{\tau}_i^\infty)(\cdot|h_i)$ are probability measures defined over the finite set $\hat{\mathcal{A}}_i(h_i)$, we have that $\sigma(\tau_{i,n}^\infty)(a_i|h_i) \to \sigma(\bar{\tau}_i^\infty)(a_i|h_i)$ for each $a_i \in \hat{\mathcal{A}}_i(h_i)$. Consider then $a_i^* \in \operatorname{supp}\sigma_i(\bar{\tau}_i^\infty)(\cdot|h_i)$. By the aforementioned convergence, it must be that there is $n_1 \in \mathbb{N}$ such that $a_i^* \in \operatorname{supp}\sigma_i(\tau_{i,n}^\infty)(\cdot|h_i) \subseteq r_{i,h_i}(\theta_{i,n}, \tau_{i,n}^\infty)$ for all $n \geq n_1$. Construct therefore the sequence $(a_{i,n})_{n\in\mathbb{N}}$ to be such that $a_{i,n} = a_i^*$ for all $n \geq n_1$ and $a_{i,n}$ is picked arbitrarily from $r_{i,h_i}(\theta_{i,n}, \tau_{i,n}^\infty)$ for all $n < n_1$. This sequence by construction satisfies the properties that $a_{i,n} \in r_{i,h_i}(\theta_{i,n}, \tau_{i,n}^\infty)$ and $a_{i,n} \to a_i^*$. Upper hemicontinuity of r_{i,h_i} implies that $a_i^* \in r_{i,h_i}(\lim_{n\to\infty}(\theta_{i,n}, \tau_{i,n}^\infty)) = r_{i,h_i}(\bar{\theta}_i, \bar{\tau}_i^\infty)$. This concludes the proof that $(\bar{\theta}_i, \bar{\tau}_i^\infty) \in \widetilde{RP}_i$. The desired result then follows. ■

Proof of Lemma 7 (p. 28)

We start by fixing a generic $i \in I$ and by rewriting:

$$CON_i = \bigcap_{h_i \in H_i} \left\{ (s_i, \theta_i, \tau_i^\infty) \in \mathcal{T}_i^\infty : \sigma(\tau_i^\infty)(s_i(h_i)|h_i) = 1 \right\}.$$

Then, fix $\bar{h}_i \in H_i$ and focus on the corresponding set in the above intersection. Consider a sequence of elements of such set, $(s_{i,n}, \theta_{i,n}, \tau_{i,n}^\infty)_{n\in\mathbb{N}}$, converging to $(\bar{s}_i, \bar{\theta}_i, \bar{\tau}_i^\infty)$. Convergence implies that there is $\bar{n} \in \mathbb{N}$ such that, for each $n \geq \bar{n}$, $s_{i,n} = \bar{s}_i$ (this follows from finiteness of S_i). Therefore, $(s_{i,n}, \theta_{i,n}, \tau_{i,n}^\infty) = (\bar{s}_i, \theta_{i,n}, \tau_{i,n}^\infty) \in CON_i$ and $\tau_{i,K+1,n}(S_i(\bar{h}_i, \bar{s}_i(\bar{h}_i))|\bar{h}_i) = 1$ for each $n \geq \bar{n}$. Moreover, by convergence of $\tau_{i,n}^\infty$ to $\bar{\tau}_i^\infty$, $\tau_{i,K+1,n}(\cdot|\bar{h}_i)$ converges to $\bar{\tau}_{i,K+1}(\cdot|\bar{h}_i)$. As mentioned in earlier proofs (see the proofs of Lemmas 5 and 6), this implies that $\tau_{i,K+1,n}(\{s_i\}|\bar{h}_i)$ converges to $\bar{\tau}_{i,K+1}(\{s_i\}|\bar{h}_i)$ for each $s_i \in S_i$. We conclude that also $\bar{\tau}_i^\infty$ is such that $\bar{\tau}_{i,K+1}(S_i(\bar{h}_i, \bar{s}_i(\bar{h}_i))|\bar{h}_i) = 1$, proving that the set of interest is closed. Hence, CON_i is a finite intersection of closed sets, hence it is closed, and the same holds for each $i \in I$. ■

Proof of Lemma 8 (p. 28)

The result follows from Lemmas 2, 4, 5, 6, and 7, because, for each player $i \in I$, R_i is a finite intersection of measurable sets. ■

Proof of Lemma 9 (p. 30)

We first state and prove an auxiliary result.

Lemma A7. *Fix $i \in I$ and analytic $F \subseteq \Omega_{-i}^K$. The set $\left\{\tau_i^{K+1} \in \mathcal{T}_i^{K+1} : \tau_{i,K+1}$ strongly believes $F\right\}$ is measurable.*

Proof. We rewrite the set of interest as $\mathcal{T}_i^K \times \left\{\tau_{i,K+1} : \tau_{i,K+1}$ strongly believes $F\right\}$. Then,

$$\begin{aligned}
&\left\{\tau_{i,K+1} : \tau_{i,K+1} \text{ strongly believes } F\right\} \\
=&\left\{\tau_{i,K+1} : \forall\, h_i \in H_i, \left(F \cap \Omega_{-i}^K(h_i) \neq \emptyset\right) \implies \left(\forall\, G \in \mathcal{B}(\Omega_{-i}^K), F \subset G \implies \tau_{i,K+1}(G|h_i) \geq 1\right)\right\}
\end{aligned}$$

$$= \bigcap_{h_i: F \cap \Omega_{-i}^K(h_i) \neq \emptyset} \; \bigcap_{G \in \mathcal{B}(\Omega_{-i}^K): F \subset G} \big\{\tau_{i,K+1} : \tau_{i,K+1}(G|h_i) \geq 1\big\}$$
$$= \bigcap_{h_i: F \cap \Omega_{-i}^K(h_i) \neq \emptyset} \; \bigcap_{G \in \mathscr{B}: F \subset G} \big\{\tau_{i,K+1} : \tau_{i,K+1}(G|h_i) \geq 1\big\},$$

where the first equality holds by definition of strong belief, the second is obvious, and the third follows once we note that Ω_{-i}^K is Polish (hence, separable), hence second countable (we let $\mathscr{B}$ denote its countable base). With Remark A1, it is easy to see that all the intersected sets above are measurable. Given that the intersections are countable, our result follows. ■

We proceed by induction to prove Lemma 9. As for part (i), we start by noting that $\mathbf{P}_i^{\Delta}(0) = S_i \times \Theta_i \times \mathcal{T}_i^{\infty}$ is trivially measurable (hence, analytic) for each $i \in I$. Now assume by induction that $\mathbf{P}_i^{\Delta}(k)$ is analytic for $k \in \{1, \ldots, n\}$, with $n \in \mathbb{N}$: we show that $\mathbf{P}_i^{\Delta}(n+1)$ is analytic. Define $\mathcal{T}_{i,KB}^{K+1}$, $\mathcal{T}_{i,C}^{K+1}$, and $\mathcal{T}_{i,CBU}^{K+1}$ as the set of $(K+1)$-th-order hierarchical systems of beliefs where knowledge-implies-belief, coherence, and the Bayes rule hold, respectively. By inspection of the proofs of Lemmas 2, 4, and 5, such sets can be checked to be measurable.

Next, consider the following sets.

$$\begin{aligned}
P_1 :=& \big\{(s_i, \theta_i, \tau_i^{K+1}) : \tau_i^{K+1} \in \mathcal{T}_{i,KB}^{K+1} \cap \mathcal{T}_{i,C}^{K+1} \cap \mathcal{T}_{i,CBU}^{K+1} \cap \Delta_{\theta_i}\big\};\\
P_2 :=& \big\{(s_i, \theta_i, \tau_i^{K+1}) : \forall\, h_i \in H_i, s_i(h_i) \in r_{i,h_i}(\theta_i, \tau_i^{K+1})\big\};\\
P_3 :=& \Theta_i \times \big\{(s_i, \tau_i^{K+1}) : \forall\, h_i \in H_i, \tau_{i,1}(S_i(h_i, s_i(h_i))|h_i) = 1\big\};\\
P_4 :=& S_i \times \Theta_i \times \big\{\tau_i^{K+1} : \forall\, k \in \{1, \ldots, n\}, \tau_{i,K+1} \text{ strongly believes } \mathbf{P}_{-i}^{\Delta}(k)\big\}.
\end{aligned}$$

P_1 measurable, by our foregoing observation about $\mathcal{T}_{i,CBU}^{K+1}$, $\mathcal{T}_{i,C}^{K+1}$, and $\mathcal{T}_{i,CBU}^{K+1}$, and because Δ_{θ_i} is assumed to be measurable for each $i \in I$ and $\theta_i \in \Theta_i$. P_3 can be showed to be measurable by an argument similar to that of the proof of Lemma 7. P_4 is measurable as per Lemma A7, once we note that sets $(\mathbf{P}_{-i}^{\Delta}(k))_{k=1}^n$ are analytic by the inductive hypothesis. As for P_2, note that we can rewrite the first intersected set as follows:

$$\begin{aligned}
&\bigcap_{h_i \in H_i} \big\{(s_i, \theta_i, \tau_i^{K+1}) : \forall\, s_i' \in S_i, \bar{u}_{i,h_i}(s_i, \theta_i, \tau_i^{K+1}) \geq \bar{u}_{i,h_i}(s_i', \theta_i, \tau_i^{K+1})\big\}\\
&= \bigcap_{h_i \in H_i} \bigcap_{s_i' \in S_i} \bigcup_{(s_i, \theta_i) \in S_i \times \Theta_i} \Big(\{(s_i, \theta_i)\} \times \big\{\tau_i^{K+1} \in \mathcal{T}_i^{K+1} : \bar{u}_{i,h_i}(s_i, \theta_i, \tau_i^{K+1}) \geq \bar{u}_{i,h_i}(s_i', \theta_i, \tau_i^{K+1})\big\}\Big).
\end{aligned}$$

In the expression above, the sets within parentheses are measurable – this holds because the map $\tau_i^{K+1} \mapsto \bar{u}_{i,h_i}(s_i, \theta_i, \tau_i^{K+1})$ is continuous for each $i \in I$, $h_i \in H_i$, $s_i \in S_i$, and $\theta_i \in \Theta_i$, and thus the set $\big\{\tau_i^{K+1} \in \mathcal{T}_i^{K+1} : \bar{u}_{i,h_i}(s_i, \theta_i, \tau_i^{K+1}) \geq \bar{u}_{i,h_i}(s_i', \theta_i, \tau_i^{K+1})\big\}$ is measurable for each $s_i' \in S_i$. Then, P_2 is measurable because it is given by finite intersections and unions of measurable sets.

Thus, $\bigcap_{k=1}^4 P_k =: P^*$ is measurable. Note that $\mathbf{P}_i^{\Delta}(n+1) = \text{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} P^*$: since it is the projection over a Polish space of a measurable set, it is analytic. The same holds for each $i \in I$.

Part (ii) is immediate. Obviously, $\mathbf{P}_i^{\Delta}(1) \subseteq \mathbf{P}_i^{\Delta}(0) = S_i \times \Theta_i \times \mathcal{T}_i^K$ trivially holds for each $i \in I$. Assume by induction that, for each $k \in \{1, \ldots, n\}$ and $i \in I$, $\mathbf{P}_i^{\Delta}(k) \subseteq \mathbf{P}_i^{\Delta}(k-1) = S_i \times \Theta_i \times \mathcal{T}_i^K$. We want to show that $\mathbf{P}_i^{\Delta}(n+1) \subseteq \mathbf{P}_i^{\Delta}(n)$. Then, for each $q \in \mathbb{N}$,

let $P_4(q) = S_i \times \Theta_i \times \left\{\tau_i^{K+1} \in \mathcal{T}_{i,C}^{K+1} : \forall k \in \{1,\dots,q-1\}, \tau_{i,K+1} \text{ strongly believes } \mathbf{P}_{-i}^{\Delta}(k)\right\}$. Note that, for each $k \in \mathbb{N}$, we can write $\mathbf{P}_i^{\Delta}(k) = \text{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K}(P_1 \cap P_2 \cap P_3 \cap P_4(k-1))$, and that, for each $k \in \mathbb{N}$, $P_4(k) \subseteq P_4(k-1)$. With this, we conclude that $\mathbf{P}_i^{\Delta}(n+1) = \text{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K}(P_1 \cap P_2 \cap P_3 \cap P_4(n)) \subseteq \text{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K}(P_1 \cap P_2 \cap P_3 \cap P_4(n-1)) = \mathbf{P}_i^{\Delta}(n)$, which yields the desired result. ■

Proof of Proposition 2 (p. 31)

We begin this proof by introducing some terminology and by proving auxiliary results. To ease notation, we denote generic elements of $\mathcal{T}_i^K$ and $\mathcal{T}_{i,K+1}$ ($i \in I$) as τ_i and μ_i, respectively.

Fix a generic $i \in I$. Consider $\mu_i^1, \mu_i^2 \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ and $F^1, F^2 \subseteq \Omega_{-i}^K$. The profile $(\mu_i^k, F^k)_{k\in\{1,2\}}$ is *admissible* if $F^2 \subseteq F^1$ and μ_i^n strongly believes F^n ($n \in \{1,2\}$). As a matter of terminology, for each $F \subseteq \Omega_{-i}^K$ and $\mu_i \in \mathcal{T}_{i,K+1}$, we say that F is *compatible* with μ_i and h_i if

$$F \cap \Omega_{-i,\text{marg}\,\mu_i}^K(h_i) \neq \emptyset,$$

where $\text{marg}\,\mu_i$ is a shorthand to denote the hierarchical system of beliefs of order K obtained by taking the marginals of μ_i over the sets $(\Omega^0, (\Omega_{-i}^n)_{n=1}^{K-1})$. The (F^1, F^2)-*composition* of μ_i^1 and μ_i^2 is $\bar{\mu}_i \in \mathcal{T}_{i,K+1}$ such that $\bar{\mu}_i(\cdot\,|h_i) = \mu_i^2(\cdot\,|h_i)$ whenever F^2 is compatible with μ_i^2 and h_i, and $\bar{\mu}_i(\cdot\,|h_i) = \mu_i^1(\cdot\,|h_i)$ otherwise. For each sequence $(\mu_i^k, F^k)_{k=1}^n$ where $(F^k)_{k=1}^n$ is a decreasing sequence of subsets of $S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K$ and $\mu_i^k \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ for each $k \in \{1,\dots,n\}$, the $(F^k)_{k=1}^n$-*composition* (or, simply, *composition*) of $(\mu_i^k)_{k=1}^n$ can be defined in a natural way.

We first prove an auxiliary fact.

Lemma A8. *Fix a $i \in I$, an admissible $(\mu_i^k, F^k)_{k\in\{1,2\}}$, and let $\bar{\mu}_i$ be the composition of μ_i^1 and μ_i^2. Then, $\bar{\mu}_i \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ and $\bar{\mu}_i$ strongly believes F^1 and F^2.*

Proof of Lemma A8. That $\bar{\mu}_i \in \mathcal{T}_{i,K+1,KB}$ follows from inspection of the definition of composition. We need to show that $\bar{\mu}_i \in \mathcal{T}_{i,K+1,CBU}$ – that is, we need to show that the chain rule, Bayes rule, and independence are satisfied by $\bar{\mu}_i$. Independence is easily checked, so we are left to prove that the other two properties hold.

Step 1: the chain rule holds. Fix $h_i \in H_i$, $a_i \in \hat{\mathcal{A}}_i(h_i)$, $h_i' \in \bar{H}_i(h_i, a_i)$, and $s_i \in S_i(h_i, a_i)$. We want to show that

$$\bar{\mu}_i(s_i|h_i') \cdot \bar{\mu}_i(S_i(h_i,a_i)|h_i) = \bar{\mu}_i(s_i|h_i) \tag{CR}$$

Notice that, if F^2 is not μ_i^2-compatible with h_i, (CR) boils down to $\mu_i^1(s_i|h_i')\mu_i^1(S_i(h_i,a_i)|h_i) = \mu_i^1(s_i|h_i)$, which is verified as $\mu_i^1 \in \mathcal{T}_{i,K+1,CBU}$.

Suppose then that F^2 is μ_i^2-compatible with h_i. We further distinguish two cases: either F^2 is μ_i^2-compatible with h_i' or not. In the former case, (CR) boils down to $\mu_i^2(s_i|h_i')\mu_i^2(S_i(h_i,a_i)|h_i) = \mu_i^2(s_i|h_i)$, which holds because $\mu_i^2 \in \mathcal{T}_{i,K+1,CBU}$. Focus then on the latter case and notice the following. First, since $F^2 \cap \Omega_{-i,\mu_i^2}^K(h_i') = \emptyset$ and $F^2 \cap \Omega_{-i,\mu_i^2}^K(h_i) \neq \emptyset$, $(\mu_i^2)^*(F^2|h_i) = 1$ and $(\mu_i^2)^*(\Omega_{-i,\mu_i^2}^K(h_i')|h_i) = 0$.[55] Second, since $h_i' \in \bar{H}_i(h_i,a_i)$, each $s_i' \in \Omega_{-i,\mu_i^2}^K(h_i')$ must also

[55] Recall that $(\mu_i^2)^*$ is the outer measure induced by μ_i^2.

belong to $S_i(h_i, a_i)$. Taken together, these observations yield $\mu_i^2(s_i|h_i) = \mu_i^2(S_i(h_i, a_i)|h_i) = 0$.[56] Therefore

$$\begin{aligned}\bar{\mu}_i(s_i|h_i') \cdot \bar{\mu}_i(S_i(h_i, a_i)|h_i) &= \mu_i^1(s_i|h_i') \cdot \mu_i^2(S_i(h_i, a_i)|h_i) \\ &= \mu_i^1(s_i|h_i') \cdot 0 = 0 \\ &= \mu_i^2(s_i|h_i) = \bar{\mu}_i(s_i|h_i),\end{aligned}$$

where the first equality follows from the definition of $\bar{\mu}_i$, under the assumption that F^2 is μ_i^2-compatible with h_i but not with h_i', the second one from the foregoing observations, and the remaining ones are obvious.

We established that the chain rule holds for $\bar{\mu}_i$, and this concludes the first step of the proof.

Step 2: Bayes rule holds. To simplify the notation, let ν_i^1, ν_i^2, and $\bar{\nu}_i$ denote the marginals over $S_{-i} \times \Theta \times \mathcal{T}_{-i}^K$ of μ_i^1, μ_i^2, and $\bar{\mu}_i$, respectively. Fix generic $h_i \in H_i$, $a_i \in \mathcal{A}_i$, $m_i \in M_i(h_i, a_i)$, $G \in \mathcal{B}(S_{-i} \times \Theta \times \mathcal{T}_{-i}^K)$. Let $h_i' = (h_i, (a_i, m_i))$, and denote $f_{i,h_i,s_i^*,\operatorname{marg}\mu_i} : S_{-i} \times \Theta \times \mathcal{T}_{-i}^K \to \Delta(M_i)$ as f for simplicity (s_i^* is a generic element of $S_i(h_i, a_i)$). We want to show that

$$\bar{\nu}_i(G|h_i') \cdot \int_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^K} f(\,\cdot\,)[m_i] \mathrm{d}\bar{\nu}_i(\,\cdot\,|h_i) = \int_G f(\,\cdot\,)[m_i] \mathrm{d}\bar{\nu}_i(\,\cdot\,|h_i). \tag{BR-a_i}$$

We proceed in a way similar to that followed to prove Step 1. Specifically, note the following. First, if F^2 is not μ_i^2-compatible with h_i, then it is not compatible with h_i' either: then, $\bar{\nu}_i(\,\cdot\,|h_i) = \nu_i^1(\,\cdot\,|h_i)$ and $\bar{\nu}_i(\,\cdot\,|h_i') = \nu_i^1(\,\cdot\,|h_i')$, and this yields (BR-$a_i$), as $\mu_i^1 \in \mathcal{T}_{i,K+1,CBU}$. Second, if F^2 is μ_i^2-compatible with both h_i and h_i', $\bar{\nu}_i(\,\cdot\,|h_i) = \nu_i^2(\,\cdot\,|h_i)$ and $\bar{\nu}_i(\,\cdot\,|h_i') = \nu_i^2(\,\cdot\,|h_i')$, and this again yields (BR-a_i), as $\mu_i^2 \in \mathcal{T}_{i,K+1,CBU}$.

Suppose now that F^2 is μ_i^2-compatible with h_i but not with h_i'. We want to show that

$$\nu_i^1(G|h_i') \cdot \int_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^K} f(\,\cdot\,)[m_i] \mathrm{d}\nu_i^2(\,\cdot\,|h_i) = \int_G f(\,\cdot\,)[m_i] \mathrm{d}\nu_i^2(\,\cdot\,|h_i).$$

By assumption, F^2 is such that $F^2 \cap \Omega_{-i,\mu_i^2}^K(h_i') = \emptyset$ and $F^2 \cap \Omega_{-i,\mu_i^2}^K(h_i) \neq \emptyset$, and this implies $(\mu_i^2)^*(F^2|h_i) = 1$ and $(\mu_i^2)^*\big(\Omega_{-i,\mu_i^2}^K(h_i')|h_i\big) = 0$. At this point, it is possible to check that $f(s_{-i}, \theta, \tau_{-i})[m_i] > 0$ only if $(s_{-i}, \theta, \tau_{-i}) \in \operatorname{proj}_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^K} \Omega_{-i,\operatorname{marg}\mu_i^2}(h_i') =: X$.[57] Moreover, $(\nu_i^2)^*(X|h_i) = 0$ by the foregoing observations concerning $(\mu_i^2)^*$. This means that there exists a measurable $Y \subseteq S_{-i} \times \Theta \times \mathcal{T}_{-i}^K$ such that $X \subseteq Y$ and $\nu_i^2(Y|h_i) = (\nu_i^2)^*(X|h_i) = 0$. Clearly, $f(s_{-i}, \theta, \tau_{-i})[m_i] > 0$ only if $(s_{-i}, \theta, \tau_{-i}) \in Y$.

At this point, it is easy to check that

$$\begin{aligned}&\int_{S_{-i} \times \Theta \times \mathcal{T}_{-i}^K} f(\,\cdot\,)[m_i] \mathrm{d}\nu_i^2(\,\cdot\,|h_i) = \int_Y f(\,\cdot\,)[m_i] \mathrm{d}\nu_i^2(\,\cdot\,|h_i) = 0 \\ \geq &\int_G f(\,\cdot\,)[m_i] \mathrm{d}\nu_i^2(\,\cdot\,|h_i) \geq 0,\end{aligned}$$

where the first equality follows from the consideration that $f(\,\cdot\,)[m_i]$ takes positive values only on Y, the second one follows because $\nu_i^2(Y|h_i) = 0$, the first inequality is implied by the fact

[56] There is no need to use outer measures here, as all subsets of S_i (which is finite) are measurable.

[57] Note that X is analytic.

that $G \subseteq S_{-i} \times \Theta \times \mathcal{T}_{-i}^K$ and $f(\cdot)[m_i] : S_{-i} \times \Theta \times \mathcal{T}_{-i}^K \to [0,1]$ is non-negative, and the last inequality holds because $g^*(\cdot)[m_i^*] : S_{-i} \times \Theta \times \mathcal{T}_{-i}^K \to [0,1]$ is non-negative.

Hence, (BR-a_i) hold, and this establishes that $\bar{\mu}_i \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$. Finally, notice that by construction $\bar{\mu}_i$ strongly believes both F^1 and F^2. ■

An easy induction yields the following.

Corollary A2. *Fix a $i \in I$, an admissible $(\mu_i^k, F^k)_{k=1}^n$, and let $\bar{\mu}_i$ be the composition of $(\mu_i^k)_{k=1}^n$. Then, $\bar{\mu}_i \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ and, for each $k \in \{1,\dots,n\}$, μ_i strongly believes F^k.*

At this point, we prove Proposition 2 by induction. As a basis step, note that the statement trivially holds for $n = 0$. Assume by means of induction that it holds for $n \in \mathbb{N}$. We show that, for each $i \in I$, $\mathbf{P}_i^\Delta(n+1) = \mathbf{Q}_i^\Delta(n+1)$.

Step 1: $\mathbf{P}_i^\Delta(n+1) \subseteq \mathbf{Q}_i^\Delta(n+1)$. Take $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i^\Delta(n+1)$. Note that, by Remark 9, $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i^\Delta(n) = \mathbf{Q}_i^\Delta(n)$, where the equality holds by the inductive hypothesis. Therefore, $\mathbf{P}_i^\Delta(n+1) \subseteq \mathbf{Q}_i^\Delta(n)$, and this verifies requirement 0M of Definition 13. We are now left to show that there is $\bar{\mu}_i \in \mathcal{T}_{i,K+1}$ such that conditions 1M-4M of Definition 13 hold. Since $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i^\Delta(n+1)$, conditions 1-4 of Definition 12 are satisfied by some $\bar{\mu}_i \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$. It is readily verified that $\bar{\mu}_i$ satisfies conditions 1M-4M of Definition 13. Hence, $(s_i, \theta_i, \tau_i) \in \mathbf{Q}_i^\Delta(n+1)$.

Step 2: $\mathbf{P}_i^\Delta(n+1) \supseteq \mathbf{Q}_i^\Delta(n+1)$. Pick $(s_i, \theta_i, \tau_i) \in \mathbf{Q}_i^\Delta(n+1)$. This implies that $(s_i, \theta_i, \tau_i) \in \mathbf{Q}_i^\Delta(k)$ for each $k \in \{1,\dots,n\}$. Therefore, for each $k \in \{1,\dots,n\}$, there is $\mu_i^k \in \mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ strongly believing $\mathbf{Q}_i^\Delta(k-1)$ and satisfying conditions 1M-3M of Definition 13. It is easy to check that the sequence $(\mu_i^k, \mathbf{Q}_i^\Delta(k-1))_{k=1}^n$ is admissible. Consider then its composition $\bar{\mu}_i$, which also belongs to $\mathcal{T}_{i,K+1,KB} \cap \mathcal{T}_{i,K+1,CBU}$ as per Corollary A2. Now note the following:

1. Given that, for each $k \in \{1,\dots,n\}$, $(\tau_i, \mu_i^k) \in \mathrm{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$, then $(\tau_i, \bar{\mu}_i) \in \mathrm{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$. To see why this holds, consider that, for each $h_i \in \bar{H}_i$, there is $\bar{k} \in \{1,\dots,n\}$ such that $\bar{\mu}_i(\cdot|h_i) = \mu_i^{\bar{k}}(\cdot|h_i)$. Given that $(\tau_i, \mu_i^{\bar{k}}) \in \mathrm{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$, then $\mathrm{marg}_{\Omega_{-i}^{K-1}} \mu_i^{\bar{k}}(\cdot|h_i) = \tau_{i,K}(\cdot|h_i)$ for each $h_i \in \bar{H}_i$, and the same holds for each $k \in \{1,\dots,n\}$ (cf. condition 1M of Definition 13). Therefore, we can conclude that, for each $h_i \in \bar{H}_i$, $\mathrm{marg}_{\Omega_{-i}^{K-1}} \bar{\mu}_i(\cdot|h_i) = \tau_{i,K}(\cdot|h_i)$. Coherence of lower-order beliefs is independent of $\bar{\mu}_i$ (it is a feature of τ_i), so that the foregoing observations are enough to conclude that $(\tau_i, \bar{\mu}_i^k) \in \mathrm{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$.

   Similarly, we have that, for each $k \in \{1,\dots,n\}$, $(\tau_i, \mu_i^k) \in \Delta_{\theta_i}$. Recall that Δ is rectangular, so that we can write $\Delta_{\theta_i} = \bigtimes_{n=1}^{K+1} \bigtimes_{h_i \in \bar{H}_i} B_{\theta_i,n,h_i}$ for a suitable profile of measurable sets. Thanks to this, we can conclude that $\tau_i \in \mathrm{proj}_{\mathcal{T}_i^K} \Delta_{\theta_i} = \bigtimes_{n=1}^{K} \bigtimes_{h_i \in \bar{H}_i} B_{\theta_i,n,h_i}$. Moreover, note that, for each $h_i \in \bar{H}_i$, there is $\bar{k} \in \{1,\dots,n\}$ such that $\bar{\mu}_i(\cdot|h_i) = \mu_i^{\bar{k}}(\cdot|h_i)$, and that, for each $k \in \{1,\dots,n\}$, $\mu_i^k \in \mathrm{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i} = \bigtimes_{h_i \in \bar{H}_i} B_{\theta_i,K+1,h_i}$. As a consequence, we have that, for each $h_i \in \bar{H}_i$, $\bar{\mu}_i(\cdot|h_i) \in B_{\theta_i,K+1,h_i}$, and this yields $\bar{\mu}_i \in \mathrm{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i}$. Wrapping up, $\tau_i \in \mathrm{proj}_{\mathcal{T}_i^K} \Delta_{\theta_i}$ and $\bar{\mu}_i \in \mathrm{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i}$ imply $(\tau_i, \bar{\mu}_i) \in \mathrm{proj}_{\mathcal{T}_i^K} \Delta_{\theta_i} \times \mathrm{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i} = \Delta_{\theta_i}$, where the last equality holds because of the rectangularity of Δ_{θ_i}.

   Hence, condition 1 of Definition 12 is met.

2. Recall that, for each $h_i \in H_i$ and $(s_i, \theta_i, (\tau_i, \mu_i)) \in S_i \times \Theta_i \times \mathcal{T}_i^{K+1}$, $\hat{u}_{i,h_i}(s_i, \theta_i, (\tau_i, \mu_i)) = \int_{S_{-i}\times\Theta_{-i}\times\mathcal{T}_{-i}^K} u_{i,h_i,s_i,\theta_i} \mathrm{d}\nu_i(\cdot|h_i)$, where $\nu_i(\cdot|h_i)$ is the marginal over $S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K$ of $\mu_i(\cdot|h_i)$. For each $i \in I$ and $h_i \in H_i$, let $\hat{r}_{i,h_i} : \Theta_i \times \mathcal{T}_i^{K+1} \rightrightarrows S_i$ be the correspondence $(\theta_i, (\tau_i, \mu_i)) \mapsto \arg\max_{s_i'} \hat{u}_{i,h_i}(s_i', \theta_i, (\tau_i, \mu_i))$.

   Since $(s_i, \theta_i, \tau_i) \in \bigcap_{k=1}^n \mathbf{Q}_i^\Delta(k)$ and preferences are own-plan independent, Remark 8 implies that $s_i \in \bigcap_{h_i \in H_i} \hat{r}_{i,h_i}(\theta_i, (\tau_i, \mu_i^k\ =))$ for $k \in \{1, \ldots, n\}$. Fix $h_i \in H_i$. We know that there is $k \in \{1, \ldots, n\}$ such that $\bar{\mu}_i(\cdot|h_i) = \mu_i^k(\cdot|h_i)$. This implies that $\hat{r}_{i,h_i}(\theta_i, (\tau_i, \mu_i^k)) = \hat{r}_{i,h_i}(\theta_i, (\tau_i, \bar{\mu}_i))$, because $\mu_i^k(\cdot|h_i)$ and $\bar{\mu}_i(\cdot|h_i)$ have the same marginal over $S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K$. We conclude that $s_i \in \hat{r}_{i,h_i}(\theta_i, (\tau_i, \bar{\mu}_i))$ for each $h_i \in H_i$, proving that $\bar{\mu}_i$ satisfies condition 2 of Definition 12.

3. Consider that $\mu_i^k(S_i(h_i, s_i(h_i))|h_i) = 1$ for each $k \in \{1, \ldots, n\}$ and $h_i \in H_i$, as μ_i^k satisfies condition 3M of Definition 13. Then note that, for each $h_i \in H_i$, there is $\bar{k} \in \{1, \ldots, n\}$ such that $\bar{\mu}_i(\cdot|h_i) = \mu_i^{\bar{k}}(\cdot|h_i)$. Therefore, for each $h_i \in H_i$, $\bar{\mu}_i(S_i(h_i, s_i(h_i))|h_i) = 1$. This proves that $\bar{\mu}_i$ satisfies condition 3 of Definition 12.

4. By Corollary A2, for each $k \in \{1, \ldots, n\}$, $\bar{\mu}_i$ strongly believes $\mathbf{Q}_i^\Delta(k-1) = \mathbf{P}_i^\Delta(k-1)$, with the equality following from the inductive hypothesis. This implies requirement 4 of Definition 12.

In light of the foregoing remarks, we conclude that $\bar{\mu}_i$ (as obtained above) satisfies conditions 1-4 of Definition 12, proving that $(s_i, \theta_i, \tau_i) \in \mathbf{P}_i^\Delta(n+1)$. This concludes the proof. ■

Proof of Lemma 10 (p. 33)

Fix $i \in I$. Define, for each $\tau_i^K \in \mathcal{T}_i^K$:

$$[\tau_i^K] := \left\{\bar{\tau}_i^K \in \mathcal{T}_i^K : \forall h_i \in \bar{H}_i, \bar{\tau}_i^K \sim_{h_i} \tau_i^K\right\} = \bigcap_{h_i \in \bar{H}_i} [\tau_i^K]_{h_i}.$$

Each such set is nonempty (for each $\tau_i^K \in \mathcal{T}_i^K$, $\tau_i^K \in [\tau_i^K]$ trivially holds). Moreover, by finiteness of $\bar{H}_i$ and by Lemma A5, each such set is measurable.

Now fix $F_{-i} \in \mathcal{B}(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^\infty)$. Recall that:

$$\mathrm{B}_{i,h_i}(F_{-i}) = \left\{(s_i, \theta_i, \tau_i^\infty) \in C_i : \varphi_i(\tau_i^\infty)(F|h_i) = 1\right\}.$$

By continuity of φ_i and by Remark A1, such set is measurable. Then, write:

$$\begin{aligned}
\mathrm{SB}_i(F_{-i}) =& \left\{(s_i, \theta_i, \tau_i^\infty) \in C_i : (\exists\, \bar{\tau}_i^K \in \mathcal{T}_i^K, \tau_i^K \in [\bar{\tau}_i^K]),\right. \\
& \left.(\forall\, h_i \in H_i, \Omega_{-i,[\bar{\tau}_i^K]}^\infty(h_i) \cap F_{-i} \neq \emptyset \Longrightarrow \varphi_i(\tau_i^\infty)(F_{-i}|h_i) = 1)\right\} \\
=& \bigcup_{[\bar{\tau}_i^K]} \left(\left(S_i \times \Theta_i \times [\bar{\tau}_i^K]_{h_i} \times \bigtimes_{k \geq K+1} \mathcal{T}_{i,k} \right) \right. \\
& \left. \cap \left(\bigcap_{h_i : \Omega_{-i,[\bar{\tau}_i^K]}^\infty(h_i) \cap F_{-i} \neq \emptyset} \left\{(s_i, \theta_i, \tau_i^\infty) \in C_i : \varphi_i(\tau_i^\infty)(F_{-i}|h_i) = 1\right\} \right) \right)
\end{aligned}$$

$$= \bigcup_{[\bar{\tau}_i^K]} \left(\left(S_i \times \Theta_i \times [\bar{\tau}_i^K]_{h_i} \times \bigtimes_{k \geq K+1} \mathcal{T}_{i,k} \right) \cap \left(\bigcap_{h_i : \Omega^\infty_{-i,[\bar{\tau}_i^K]}(h_i) \cap F_{-i} \neq \emptyset} \mathrm{B}_{i,h_i}(F_{-i}) \right) \right). \quad (13)$$

In (13), the first set within parentheses is measurable as per Lemma A5, and the second one is a finite intersection of measurable sets, by the foregoing reasoning. Then, the union over equivalence classes is finite, as per Corollary A1. We conclude that the expression in (13) is measurable. Thus, $\mathrm{B}_{i,h_i}(F_i)$ and $\mathrm{SB}_i(F_{-i})$ are measurable. The same clearly holds for each $i \in I$, $h_i \in \bar{H}_i$, and $F_{-i} \in \mathcal{B}(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^\infty)$. ■

Proof of Theorem 1 (p. 34)

We first report an auxiliary result, which is an adaptation of Lemma 3 in Battigalli and Tebaldi (2019). For a Polish set X and a countable collection $\mathcal{C}$ of Borel subsets of X, we call a *conditional probability system* (CPS) on $(X, \mathcal{C})$, any $\mu = (\mu(\cdot\,|C))_{C \in \mathcal{C}} \in [\Delta(X)]^{\mathcal{C}}$ such that:

1. for each $C \in \mathcal{C}$, $\mu(C|C) = 1$;
2. for each $E \in \mathcal{B}(C)$ and $C, D \in \mathcal{C}$, $E \subseteq D \subseteq C$ implies $\mu(E|C) = \mu(E|D)\mu(D|C)$.

Moreover, for each X, Y Polish and for each countable collection $\mathcal{C}$ of Borel subsets of X, a CPS on $(X \times Y, \mathcal{C})$ is a CPS on $X \times Y$ with $\{C \times Y : C \in \mathcal{C}\}$ as collection of conditioning events. If μ is a CPS on $(X, \mathcal{C})$ and ν is a CPS on $(X \times Y, \mathcal{C})$, we write $\mathrm{marg}_X \nu$ as a shorthand for $(\mathrm{marg}_X \nu(\cdot\,|C))_{C \in \mathcal{C}}$. With this, we can state the following.

Lemma A9. *Let X, Y be Polish spaces, $\mathcal{C}$ a countable collection of Borel subsets of C, and $(D_k)_{k=1}^n$ a finite decreasing sequence of Borel subsets of $X \times Y$. If μ is a CPS on $(C, \mathcal{C})$ that strongly believes $(\mathrm{proj}_C D_k)_{k=1}^n$, then there exists a CPS ν on $(C \times X, \mathcal{C})$ that strongly believes $(D_k)_{k=1}^n$ and such that $\mathrm{marg}_C \nu = \mu$.*

We now proceed with the proof of Theorem 1.

For each $i \in I$, $\mathbf{P}_i^\Delta(0) = S_i \times \Theta_i \times \mathcal{T}_i^K = \mathrm{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \left(S_i \times \Theta_i \times \mathcal{T}_i^\infty \right) = \mathrm{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^\Delta(0)$. Assume by induction that, for each $i \in I$ and $k \in \{1, \ldots, n-1\}$, $\mathbf{P}_i^\Delta(k) = \mathrm{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^\Delta(k)$. We want to show that $\mathbf{P}_i^\Delta(n) = \mathrm{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^\Delta(n)$.

First, we show $\mathbf{P}_i^\Delta(n) \subseteq \mathrm{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^\Delta(n)$. Take $(s_i, \theta_i, \tau_i^K) \in \mathbf{P}_i^\Delta(n)$: by definition, there exists $\tau_{i,K+1} \in \mathcal{T}_{i,K+1}$ such that the conditions of Definition 12 are satisfied. Specifically, $\tau_{i,K+1}$ is a CPS on $\left(\Omega_{-i}^K, \{\Omega^K_{-i,\tau_i^K}(h_i)\}_{h_i \in H_i}\right)$, according to the terminology we introduced, where τ_i^K is the K-th-order hierarchy of systems of beliefs induced by $\tau_{i,K+1}$ by taking the marginals over $(\Omega^0, \Omega^1_{-i}, \ldots, \Omega^{K-1}_{-i})$. Moreover, $\tau_{i,K+1}$ strongly believes $(\mathbf{P}_{-i}^\Delta(1), \ldots, \mathbf{P}_{-i}^\Delta(n-1)) = (\mathrm{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^\Delta(1), \ldots, \mathrm{proj}_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K} \mathbf{R}_{-i}^\Delta(n-1))$, with the equality holding by our inductive hypothesis. Then, by Lemma A9, there is a CPS μ on $\left(S \times \Theta \times \mathcal{T}_{-i}^\infty, \{\Omega^K_{-i,\tau_i^K}(h_i)\}_{h_i \in H_i}\right)$ strongly believing $(\mathbf{R}_{-i}^\Delta(1), \ldots, \mathbf{R}_{-i}^\Delta(n-1))$ such that $\mathrm{marg}_{\Omega^K_{-i}} \mu = \tau_{i,K+1}$. Note that we can take the inverse through φ_i of μ (cf. Lemma 3). Let $\bar{\tau}_i^\infty = \varphi_i^{-1}(\mu)$, and note that it induces a $(K+1)$-th-order hierarchy of systems of beliefs $\bar{\tau}_i^{K+1}$ satisfying $\bar{\tau}_i^{K+1} = (\tau_i^K, \tau_{i,K+1})$, since $\mathrm{marg}_{\Omega^K_{-i}} \mu = \tau_{i,K+1}$. Hence, if conditions 2 and 3 of Definition 12 hold for $(\tau_i^K, \tau_{i,K+1})$, they

must hold for $\bar{\tau}_i^{K+1}$. This proves that $(s_i, \theta_i, \bar{\tau}_i^{\infty})$ satisfies both rational planning and coherence. Moreover, $\bar{\tau}_i^{\infty}$ satisfies coherence because φ_i^{-1} maps to $\mathcal{T}_{i,C}^{\infty}$, and it satisfies knowledge-implies-belief and the chain rule because $(\tau_i^K, \tau_{i,K+1})$ satisfies condition 1 of Definition 12. Lastly, it strongly believes $(\mathbf{R}_{-i}^{\Delta}(1), \ldots, \mathbf{R}_{-i}^{\Delta}(n-1))$ as already mentioned. Hence, $(s_i, \theta_i, \bar{\tau}_i^{\infty}) \in \mathbf{R}_i^{\Delta}(n)$, so that $(s_i, \theta_i, \tau_i^K) \in \text{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n)$.

Second, we show $\text{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n) \subseteq \mathbf{P}_i^{\Delta}(n)$. Take $(s_i, \theta_i, \tau_i^K) \in \text{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n)$. Then, by definition there exists $\mu = (\mu_k)_{k \geq K+1} \in \bigtimes_{k \geq K+1} \mathcal{T}_{i,k}$ such that $(s_i, \theta_i, \tau_i^K, \mu) \in \mathbf{R}_i^{\Delta}(n) = R_i \cap \big(\bigcap_{k=1}^{n-1} \text{SB}_i(\mathbf{R}_{-i}^{\Delta}(k))\big)$. Conditions 1, 2, 3 of Definition 12 are satisfied by (τ_i^K, μ_{K+1}), because $(s_i, \theta_i, \tau_i^K, \mu) \in R_i$. At this point, we just need to show that μ_{K+1} strongly believes $(\mathbf{P}_{-i}^{\Delta}(1), \ldots, \mathbf{P}_{-i}^{\Delta}(n-1))$. Note that, by coherence, the K-th-order hierarchy of systems of beliefs induced by μ_{K+1} is exactly τ_i^K. Hence, pick $k \in \{1, \ldots, n-1\}$ and $h_i \in H_i$ such that $\mathbf{P}_i^{\Delta}(k) \cap \Omega_{-i,\tau_i}^K(h_i) \neq \emptyset$. By the inductive hypothesis, the coherence of (τ_i^K, μ), and the definition of inference sets, this is equivalent to writing $\mathbf{R}_{-i}^{\Delta}(k) \cap \Omega_{-i,\tau_i^K}^{\infty}(h_i) \neq \emptyset$. However, if such condition holds, we have that $\varphi_i\big((\tau_i^K, \mu)\big)\big(\mathbf{R}_{-i}^{\Delta}(k) | h_i\big) = 1$, because (τ_i^K, μ) strongly believes $(\mathbf{R}_{-i}^{\Delta}(1), \ldots, \mathbf{R}_{-i}^{\Delta}(n-1))$. At this point, we can write:

$$\mu_{K+1}^*(\mathbf{P}_{-i}^{\Delta}(k)|h_i) = \text{marg}_{\Omega_{-i}^K} \varphi_i\big((\tau_i^K, \mu)\big)\big(\mathbf{P}_{-i}^{\Delta}(k)|h_i\big) = \text{marg}_{\Omega_{-i}^K} \varphi_i\big((\tau_i^K, \mu)\big)\big(\text{proj}_{\Omega_{-i}^K} \mathbf{R}_{-i}^{\Delta}(k)|h_i\big) =$$
$$= \varphi_i\big((\tau_i^K, \mu)\big)\big(\text{proj}_{\Omega_{-i}^K}^{-1}(\text{proj}_{\Omega_{-i}^K} \mathbf{P}_{-i}^{\Delta}(k))\big) \geq \varphi_i\big((\tau_i^K, \mu)\big)(\mathbf{R}_{-i}^{\Delta}(q)) = 1.$$

The same holds for each $k \in \{1, \ldots, n-1\}$ and $h_i \in H_i$, proving that μ_{K+1} strongly believes $(\mathbf{P}_{-i}^{\Delta}(1), \ldots, \mathbf{P}_{-i}^{\Delta}(n-1))$. Hence, $(s_i, \theta_i, \tau_i^K) \in \mathbf{P}_i^{\Delta}(n)$, which yields the desired result. ■

B Strong rationalizability analysis of Example 5

B.1 Utility functions

External-state-dependent utility For convenience, we let z^{not} (resp., z^{buy}) denote a generic terminal history where Dad plays *not* (resp., *buy*). That is, an element of $\{(a_{C,1}, a_{C,2}, m_D, a_D) \in Z : a_D = not\}$ (resp., $\{(a_{C,1}, a_{C,2}, m_D, a_D) \in Z : a_D = buy\}$), and let z_D be the length-two personal history of Dad induced by a generic terminal history z. In the following, consider a generic $\theta_C \in \Theta_C$, and $\tau^1 \in \mathcal{T}^1$. Also, recall that $L = \{w.no, v.yes\} \subset S_C$ and $G = \{w.yes, w.no\} \subset S_C$ are the sets of personal external states where Child lies and does homework, respectively. (The labels L and G are mnemonics for "lies" and "good behavior.") Then:

1. A terminal history z^{not} occurs with certainty if: (i) $s_C = w.no$ and $s_D\big((no, \neg b)\big) = not$; (ii) $s_C = w.yes$ and $s_D\big((yes, \neg b)\big) = not$; (iii) $s_C = v.no$ and $s_D\big((no, \neg b)\big) = not$; (iv) $s_C = v.yes$ and $s_D\big((yes, b)\big) = s_D\big((yes, \neg b)\big) = not$. In such case, $u_D(s, \theta, \tau^1) = 0$, and

$$u_C(s, \theta, \tau^1) = \begin{cases} -\tau_{D,1}(L|z_D^{not}) & \text{if } s_C(\varnothing) = w; \\ \theta - \tau_{D,1}(L|z_D^{not}) & \text{if } s_C(\varnothing) = v. \end{cases}$$

2. A terminal history z^{buy} occurs with certainty if: (i) $s_C = w.no$ and $s_D\big((no, \neg b)\big) = buy$; (ii) $s_C = w.yes$ and $s_D\big((yes, \neg b)\big) = buy$; (iii) $s_C = v.no$ and $s_D\big((no, \neg b)\big) = buy$; (iv)

$s_C = v.yes$ and $s_D\big((yes, b)\big) = s_D\big((yes, \neg b)\big) = buy$. Then,[58]

$$u_D(s, \theta, \tau^1) = 2 \cdot \mathbf{1}[s_C(\varnothing) = w] - 1;$$
$$u_C(s, \theta, \tau^1) = \begin{cases} 1 - \tau_{D,1}(L|z_D^{buy}) & \text{if } s_C(\varnothing) = w; \\ 1 + \theta - \tau_{D,1}(L|z_D^{buy}) & \text{if } s_C(\varnothing) = v. \end{cases}$$

3. A terminal history z^{buy} (resp., z^{not}) occurs with probability q (resp., $1-q$) if $s_C = v.yes$, $s_D\big((yes, b)\big) = buy$, and $s_D\big((yes, \neg b)\big) = not$. Hence,

$$\begin{aligned} u_C(s, \theta, \tau^1) =& q\big(1 + \theta - \tau_{D,1}(L|z_D^{buy})\big) + (1-q)\big(\theta - \tau_{D,1}(L|z_D^{not})\big) = \\ &= q + \theta - q\big(\tau_{D,1}(L|(yes, \neg b))\big) - (1-q)\big(\tau_{D,1}(L|(yes, b))\big); \\ u_D(s, \theta, \tau^1) =& \, q\big(2 \cdot \mathbf{1}[s_C(\varnothing) = w] - 1\big). \end{aligned}$$

4. A terminal history z^{buy} (resp., z^{not}) occurs with probability $1-q$ (resp., q) if $s_C = v.yes$, $s_D\big((yes, b)\big) = not$, and $s_D\big((yes, \neg b)\big) = buy$. Hence,

$$\begin{aligned} u_C(s, \theta, \tau^1) =& (1-q)\big(1 + \theta - \tau_{D,1}(L|z_D^{buy})\big) + q\big(\theta - \tau_{D,1}(L|z_D^{not})\big) = \\ &= (1-q) + \theta - (1-q)\big(\tau_{D,1}(L|(yes, \neg b))\big) - q\big(\tau_{D,1}(L|(yes, b))\big); \\ u_D(s, \theta, \tau^1) =& \, (1-q)\big(2 \cdot \mathbf{1}[s_C(\varnothing) = w] - 1\big). \end{aligned}$$

In words, Child's personal external state unambiguously defines the first two actions of a terminal history. Then, multiple terminal histories may arise only if when Child plays according to $v.yes$ and Dad according to a personal external state that prescribes different actions after observing (yes, b) and $(yes, \neg b)$.

Local decision utilities Child's preferences are own-plan independent and hence dynamically consistent. Dad's preferences are also trivially own-plan independent. For ease of exposition, we use the local utilities $(\hat{u}_{i,h_i} : \Delta(S_i) \times \Theta_i \times \mathcal{T}_i^{\infty})_{i \in \{C,D\}, h_i \in H_i}$ defined in Section 4.4, which is inconsequential thanks to Remark 8. The remark also ensures that we can focus on pure plans. Finally, we only consider beliefs of orders up to 2 (note that second-order beliefs are necessary, e.g., for Child to form expectations about Dad's blame). As a result, $\hat{u}_{i,h_i}(s_i, \theta_i, \tau_i^2)$ is interpreted as player i's expected utility of following the deterministic plan s_i from personal history h_i onward, when his trait is θ_i and he holds beliefs described by τ_i^2.

We start from Child. Note that he acts twice in a row, and he is the only active player in the first two stages. Dynamic consistency of his preferences, in conjunction with the rationality requirements embodied in our solution procedure, then implies that we can simply look at his choice between pure plans at the root of the game. To save on notation, we therefore derive functions $\hat{u}_{C,h_C}$ only for $h_C = \varnothing$. Given that the set of possible trait-types for Dad is a singleton, we identify a profile of trait types $\theta = (\theta_C, \theta_D)$ with θ_C, therefore dropping the subscript. We have:

$$\hat{u}_{C,\varnothing}(w.yes, \theta, \tau_C^2) = \tau_{C,2}\big(\{s_D : s_D\big((yes, \neg b)\big) = buy\}|\varnothing\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes, \neg b)\big)|\varnothing\big];$$

[58] We denote as $\mathbf{1}[\,\cdot\,]$ the indicator function. The domain of such function is $\mathcal{B}(S \times \Theta \times \mathcal{T}^1)$.

$$\hat{u}_{C,\varnothing}(w.no,\theta,\tau_C^2) = \tau_{C,2}\big(\{s_D : s_D\big((no,\neg b)\big) = buy\}|\varnothing\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(no,\neg b)\big)|\varnothing\big];$$

$$\hat{u}_{C,\varnothing}(v.yes,\theta,\tau_C^2) = \theta + q\Big(\tau_{C,2}\big(\{s_D : s_D\big((yes,\neg b)\big) = buy\}|\varnothing\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes,\neg b)\big)|\varnothing\big]\Big)$$
$$+ (1-q)\Big(\tau_{C,2}\big(\{s_D : s_D\big((yes,b)\big) = buy\}|\varnothing\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes,b)\big)|\varnothing\big]\Big);$$

$$\hat{u}_{C,\varnothing}(v.no,\theta,\tau_C^2) = \theta + \tau_{C,2}\big(\{s_D : s_D\big((no,\neg b)\big) = buy\}|\varnothing\big) - \mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(no,\neg b)\big)|\varnothing\big]\big).$$

As for Dad, for each $h_D \in \big\{(yes,\neg b),(yes,b),(no,\neg b)\big\}$, $s_D \in S_D$ and $\tau_D^2 \in \mathcal{T}_D^2$,

$$\hat{u}_{D,h_D}(s_D,\tau_D^2) = \begin{cases} 2\tau_{D,2}\big(G|h_D)\big) - 1 & \text{if } s_D\big(h_D\big) = buy; \\ 0 & \text{if } s_D\big(h_D\big) = not; \end{cases}$$

where we avoid specifying the dependence of $\hat{u}_{D,h_D}$ on Dad's personal trait, as Θ_D was assumed to be a singleton.

B.2 Solution procedure

For simplicity, we carry out the "memoryless" solution procedure outlined in Definition 13. This is equivalent to the strong rationalizability procedure (Definition 12) because here preferences are own-plan independence and belief restrictions are absent (cf. Proposition 2).

First step By inspection of the decision utilities defined in the previous section, it is easy to check that $\hat{u}_{C,\varnothing}(w.no,\theta,\tau_C^2) < \hat{u}_{C,\varnothing}(v.no,\theta,\tau_C^2)$ for all θ_C and τ_C^2 (recall that $\theta > 0$). Moreover, if Child's beliefs satisfy (6) and (7) but with $\mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(no,\neg b)\big)|\varnothing\big]\big) = 1$, $w.yes$ is optimal for both Child's trait-types. Such second-order system of beliefs trivially strongly believes $S_D \times \Theta_D \times \mathcal{T}_D^1$, and it can be checked that condition 1 is met by $(\tau_C^1, \bar{\tau}_{C,2})$ for some $\tau_C^1 \in \mathcal{T}_C^1$. Lastly, note that $v.yes$ maximizes $\hat{u}_{C,\varnothing}\big(\,\cdot\,,\theta,(\tau_C^1,\tau_{C,2})\big)$ for both trait-types if $\tau_{C,2}$ is such that $\tau_{C,2}\big(\{s_D : s_D\big((yes,b)\big) = buy\}|\varnothing\big) = 1$ and $\mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}(L|(yes,\neg b))|\varnothing\big] = 0$. Again, such $\tau_{C,2}$ strongly believes $S_D \times \Theta_D \times \mathcal{T}_D^1$, and it can be checked that there is $\tau_C^1 \in \mathcal{T}_C^1$ such that condition 1 is met by $(\tau_C^1, \bar{\tau}_{C,2})$. We conclude that $\text{proj}_{S_C \times \Theta_C}\mathbf{P}_C(1) = \{w.yes, v.yes, v.no\} \times \{\theta', \theta''\}$.

As for Dad, it is immediate to notice that condition 1 of Definition 12 implies that, to survive this deletion step, a profile (s_D,τ_D^1) has to be such that $\tau_{D,1}\big(\{s_C : s_C(\varnothing) = v\}|(yes,b)\big) = 1$. This in turn implies that $\tau_{D,1}\big(G|(yes,b)\big) = 0$ and $\tau_{D,1}\big(L|(yes,b)\big) = 1$. Then, any $\tau_{D,2}$ that we may look for to carry out the procedure has to conform to such features. Therefore, $\hat{u}_{D,(yes,b)}\big(\,\cdot\,,(\tau_D^1,\tau_{D,2})\big)$ is maximized by any s_D such that $s_D\big((yes,b)\big) = not$. On the other hand, if $\tau_{D,2}\big(G|(yes,\neg b)\big) = \tau_{D,2}\big(G|(no,\neg b)\big) = \frac{1}{2}$, any s_D maximizes both $\hat{u}_{D,(yes,\neg b)}\big(\,\cdot\,,(\tau_D^1,\tau_{D,2})\big)$ and $\hat{u}_{D,(no,\neg b)}\big(\,\cdot\,,(\tau_D^1,\tau_{D,2})\big)$. Thus, $\text{proj}_{S_D}\mathbf{P}_D(1) = \big\{s_D : s_D\big((yes,b)\big) = not\big\}$.

Second step We now have to restrict attention to $\tau_{C,2}$ such that, for each non-terminal personal history $h_C \in H_C$, $\tau_{C,2}\big(\{s_D : s_D\big((yes,b)\big) = buy\}|h_C\big\} = 0$ and $\mathbb{E}_{\tau_{C,2}}\big[\tau_{D,1}\big(L|(yes,b)\big)|h_C\big] = 1$. By construction of emotional feedback, we can conclude that lying after playing video-games makes Child blush with certainty. It is easy to check that $\hat{u}_{C,\varnothing}\big(v.yes,\theta,(\tau_C^1,\tau_{C,2})\big) = \theta - 1 < \theta = \hat{u}_{C,\varnothing}\big(v.no,\theta,(\tau_C^1,\tau_{C,2})\big)$ for each $\tau_{C,2}$ satisfying the aforementioned restrictions and for each

$\theta \in \Theta_C$. Thus, any $(s_C, \theta, \tau_C^1) \in \mathbf{P}_C(1)$ with $s_C = v.yes$ fails condition 2 of Definition 12. Moreover, it can be checked that playing according to $h.yes$ yields a utility of at most 1. It follows that for trait-type θ'', such personal external state is never optimal, as $v.no$ yields a utility of $\theta'' > 1$. It follows that $\text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(2) = \big(\{w.yes, v.no\} \times \{\theta'\}\big) \cup \big(\{v.no\} \times \{\theta''\}\big)$.

As for Dad, any $\tau_{D,2}$ strongly believing $\mathbf{P}_C(1)$ must be such that $\tau_{D,2}\big(\{v.no\}|(no, \neg b)\big) = 1$. That is, he is now sure that Child must have played video-games whenever he answers "no." This implies that the aforementioned $\tau_{D,2}$ must be such that $\tau_{D,2}\big(G|(no, \neg b)\big) = 0$. It is easy to see that $\hat{u}_{D,(no,\neg b)}\big(\,\cdot\,, (\tau_D^1, \tau_{D,2})\big)$ is maximized by any s_D with $s_D\big((no, \neg b)\big) = not$, for each $\tau_{D,2}$ strongly believing $\mathbf{P}_C(1)$ and $(\tau_D^1, \tau_{D,2})\big)$ satisfying condition 1 of Definition 12. Hence, $\text{proj}_{S_D} \mathbf{P}_D(2) = \big\{s_D : s_D\big((yes, b)\big) = s_D\big((no, \neg b)\big) = not\big\}$.

Third step We now have to consider only $\tau_{C,2}$ strongly believing $\mathbf{P}_D(2)$, and we can focus on trait-type θ'. This means that $\tau_{C,2}\big(\big\{s_D : s_D\big((no, \neg b)\big) = buy\big\}|h_C\big) = 0$ for each non-terminal $h_C \in H_C$. This implies that $\hat{u}_{C,\varnothing}\big(v.yes, \theta', (\tau_C^1, \tau_{C,2})\big) = \theta'$ for each $\tau_{C,2}$ strongly believing $\mathbf{P}_D(2)$ and for each $(\tau_C^1, \tau_{C,2})$ meeting requirement 1 of Definition 12. Moreover, note that, if $\tau_{C,2}$ has to strongly believe $\mathbf{P}_D(2)$, we obtain $\hat{u}_{C,\varnothing}\big(w.yes, \theta', (\tau_C^1, \tau_{C,2})\big) = \tau_{C,1}(\{not.buy.not\}|\varnothing)$. Thus, both $w.yes$ and $v.no$ can be optimal for trait-type θ', and this leads us to conclude that $\text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(3) = \text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(2) = \big(\{w.yes, v.no\} \times \{\theta'\}\big) \cup \big(\{v.no\} \times \{\theta''\}\big)$.

On the other hand, any $\tau_{D,2}$ strongly believing $\mathbf{P}_C(2)$ is such that $\tau_{D,2}\big(\{w.yes\}|(yes, \neg b)\big) = 1$. Therefore, $\tau_{D,2}\big(L|(yes, \neg b)\big) = 0$ and $\tau_{D,2}\big(G|(yes, \neg b)\big) = 1$. With this, $\hat{u}_{D,(no,\neg b)}\big(\,\cdot\,, (\tau_D^1, \tau_{D,2})\big)$ is maximized by any s_D with $s_D\big((no, \neg b)\big) = not$, for each $\tau_{D,2}$ strongly believing $\mathbf{P}_C(1)$ and $(\tau_D^1, \tau_{D,2})\big)$ satisfying condition 1 of Definition 12. Hence, $\text{proj}_{S_D} \mathbf{R}_D^{\Delta}(3) = \{not.buy.not\}$.

Fourth step At this point, any $\tau_{C,2}$ strongly believing $\mathbf{P}_D(3)$ must assign probability one to $not.buy.not$ at each non-terminal personal history. Hence, $\hat{u}_{C,\varnothing}\big(w.yes, \theta', (\tau_C^1, \tau_{C,2})\big) = 1 > \theta' = \hat{u}_{C,\varnothing}\big(v.no, \theta', (\tau_C^1, \tau_{C,2})\big)$ for each $\tau_{C,2}$ satisfying the above mentioned restrictions and for each $\theta \in \Theta_C$. Therefore, we conclude that $\text{proj}_{S_C \times \Theta_C} \mathbf{R}_C^{\Delta}(4) = \big\{(w.yes, \theta'), (v.no, \theta'')\big\}$.

C A recap on notation

The following table summarizes the pieces of notation we introduced throughout the paper. For sets, we report on the left column the chosen notation, as well as a generic element. The Cartesian product of indexed sets is defined in an intuitive way, and we avoid mentioning it explicitly below.

Notation	Meaning
$I,\ i$	Players
$A_i,\ a_i$	Actions of i
$\Theta,\ \theta_i$	Personal traits of i
$Y_i,\ y_i$	Outcomes of i
$E_i,\ e_i$	Emotions of i
$E^{\leq L}$	Sequences of emotion profiles of length up to L

Notation	Meaning
$M_{i,\mathrm{e}},\ m_{i,\mathrm{e}}$	Emotional messages receivable by i
$M_{i,\mathrm{p}},\ m_{i,\mathrm{p}}$	previous-play messages receivable by i
$M_i = M_{i,p} \times M_i,\ m_i$	Message pairs receivable by i
$\tilde{f}_{\mathrm{e}} : A \times \Theta \times E^{\leq T} \to \Delta(M_{\mathrm{e}})$	Game-independent emotional feedback function
$\tilde{f}_{\mathrm{p}} : \bigcup_{t=1}^{T} A^t \to M_{\mathrm{p}}$	previous-play messages generating function
$\tilde{v}_i : Y \times \Theta \times E^{\leq T} \to \mathbb{R}$	Game-independent psychological utility of i
$\mathcal{A}_i : M_{i,\mathrm{p}} \cup \{\varnothing_{M_{i,\mathrm{p}}}\} \rightrightarrows A_i$	Feasibility correspondence of i
$\bar{H},\ H,\ Z$	Feasible, non-terminal, and terminal histories
$\bar{H}_i,\ H_i,\ Z_i$	Feasible, non-terminal, and terminal personal histories of i
$\bar{H}(h_i)$	Histories compatible with h_i
$Z(h_i)$	Terminal histories possible after h_i
$\bar{H}_i(h_i, a_i)$	Immediate successors of h_i where a_i is played
$M_i(h_i, a_i)$	Message pairs receivable by i after h_i and a_i
$\pi : Z \times \Theta \to Y$	Outcome function
$\hat{\mathcal{A}}_i : H_i \rightrightarrows A_i$	History-dependent feasibility correspondence of i
$S_i = \bigtimes_{h_i \in H_i} \hat{\mathcal{A}}_i(h_i), s_i$	Personal external states of i
$\mathcal{T}_i^{\infty},\ \tau_i^{\infty}$	Epistemic types of i
$\mathcal{T}_i^{K},\ \tau_i^{K}$	Hierarchical systems of beliefs of i of order K
$\mathcal{T}_{i,K+1},\ \tau_{i,K+1}$	Systems of beliefs of i of order $K+1$
$\tau_{i,K+1}(\cdot\,\vert h_i)$	Belief of i of order $K+1$ held at h_i
$\bigtimes_{h_i \in H_i} \Delta(\hat{\mathcal{A}}_i(h_i)),\ \sigma(\tau_i^{\infty})$	Plans of i
$\varepsilon : H \times \mathcal{T}^{\infty} \to \Delta(E^{\leq T})$	Emotion-generating function
$\Omega^{\infty} = \bigtimes_{i \in I}(S_i \times \Theta_i \times \mathcal{T}_i^{\infty})$	States of the world
$S_i \times \Theta_i \times \mathcal{T}_i^{\infty}$	Personal states of i
$S \times \Theta \times \mathcal{T}^{K}$	Utility-relevant states
$(f_{h,\mathrm{e}} : S \times \Theta \times \mathcal{T}^{K} \to \Delta(M_{\mathrm{e}}))_{h \in H}$	Game-dependent emotional feedback functions
$(f_{h,\mathrm{p}} : S \times \Theta \times \mathcal{T}^{K} \to \Delta(M_{\mathrm{p}}))_{h \in H}$	Game-dependent previous-play feedback functions
$(f_h : S \times \Theta \times \mathcal{T}^{K} \to \Delta(M))_{h \in H}$	Game-dependent emotional and previous-play feedback functions
$(g_h : S \times \Theta \times \mathcal{T}^{K} \to \Delta(A \times M))_{h \in H}$	State-history-dependent distribution of action-message profiles
$\zeta(h\vert s, \theta, \tau^{K})$	Probability of realization of history h given utility-relevant state (s, θ, τ^{K})
$\zeta(h\vert h_i; s, \theta, \tau^{K})$	Probability of realization of history h given utility-relevant state (s, θ, τ^{K}) and personal history h_i
$f_{i,h},\ g_{i,h}$	State-dependent distributions over M_i and $A_i \times M_i$ derived from f_h and g_h
$f_{i,h_i},\ g_{i,h_i},\ g_{i,h_i}^{*}$	Expected state-dependent distributions over M_i and $A_i \times M_i$, after h_i, derived from $f_{i,h}$ and $g_{i,h}$
$v_i : Z \times \Theta \times \mathcal{T}^{K} \to \mathbb{R}$	Game-dependent psychological utility of i

Notation	Meaning
$u_i : S \times \Theta \times \mathcal{T}^K \to \mathbb{R}$	State-dependent psychological utility of i
$u_{i,h_i} : S \times \Theta \times \mathcal{T}^K \to \mathbb{R}$	Psychological expected utility of i at h_i
$\bar{u}_{i,h_i} : \hat{\mathcal{A}}_i(h_i) \times \Theta_i \times \mathcal{T}_i^{K+1} \to \mathbb{R}$	Decision utility of i at h_i
$r_{i,h_i} : \Theta \times \mathcal{T}_i^{\infty} \rightrightarrows \hat{\mathcal{A}}_i(h_i)$	Optimality correspondence of i at h_i
$\mathbf{H}_i : S \times \Theta \times \mathcal{T}^K \rightrightarrows \bar{H}_i$	State-dependent personal history correspondence
$\Omega^K_{i,\tau_i^K}(h_i)$	Inference about states of i given beliefs τ_i^K at h_i

We now give an explicit definition of some functions and sets introduced in the main text. First, We say that a history $(a^\ell, m_{\mathrm{p}}^\ell, m_{\mathrm{e}}^\ell) \in A^\ell \times M_{\mathrm{p}}^\ell \times M_{\mathrm{e}}^\ell$ (with $\ell \in \{1, \ldots, L\}$) is *feasible* if:

1. $a_1 \in \mathcal{A}(\varnothing_{M_{\mathrm{p}}})$, and, for each $k \in \{1, \ldots, \ell - 1\}$, $a_{k+1} \in \mathcal{A}(m_{\mathrm{p},k})$;
2. for each $k \in \{1, \ldots, \ell\}$, $m_{\mathrm{p},k} = \tilde{f}_{\mathrm{p}}(a^k)$;
3. for each $k \in \{1, \ldots, \ell\}$, there exists $(\theta, e^k) \in \Theta \times E^k$ such that $m_k \in \operatorname{supp} \tilde{f}_{\mathrm{e}}(a_k, \theta, e^k)$.

Second, in Section 2.4, we introduced the notation $\zeta(h|s, \theta, \tau)$ to denote the probability that history h realizes when the utility-relevant state is (s, θ, τ^K). This probability is obtained as

$$\zeta(h|s, \theta, \tau^K) := \prod_{t=0}^{L(h)} g_{h^\ell}(s, \theta, \tau^K)[(\mathrm{a}_{\ell+1}(z), \mathrm{m}_{\ell+1}(z))], \tag{14}$$

where h^ℓ is the truncation of h at stage $\ell \leq L(h)$, and $\mathrm{a}_\ell(z)$ and $\mathrm{m}_\ell(z)$ are the ℓ-th-stage action and message components of h, respectively. Recall that $g_h(s, \theta, \tau^K) \in \Delta(A \times M)$ specifies the probability that a give profile of actions and (previous-play and emotional) messages is generated at history h when the utility-relevant state is (s, θ, τ^K).

Third, we used the notation $\zeta(h|h'^K)$ to denote the probability that h realizes when the utility-relevant state is (s, θ, τ^K), conditional on having reached h'. This probability is positive only if $h' \preceq h$, and in such case it is simply

$$\zeta(h|h'^K) := \frac{\zeta(h|s, \theta, \tau^K)}{\zeta(h'^K)}.$$

We also used the notation $\zeta(h|h_i; s, \theta, \tau^K)$. The interpretation is similar to the one just discussed, but in this case one conditions on the realization of a personal history of a given player. For each $i \in I$ and $h_i \in \bar{H}_i$, define as $\bar{H}(h_i) := \{h \in \bar{H} : \exists\, h_{-i} \in \bar{H}_{-i},\ h = (h_i, h_{-i})\}$ the set of complete histories compatible with h_i. That is, player i infers that the complete history must belong to $\bar{H}(h_i)$ if she observes h_i. Then,

$$\zeta(h|h_i; s, \theta, \tau^K) := \begin{cases} \dfrac{\zeta(h|s, \theta, \tau^K)}{\sum_{h' \in \bar{H}(h_i)} \zeta(h'|s, \theta, \tau^K)} & \text{if } h \in \bar{H}(h_i); \\ 0 & \text{if } h \notin \bar{H}(h_i). \end{cases}$$

In Section 4.4, we allowed a player's behavior to be defined in terms of plans rather than personal external states. The derivation of the probability of realization of each history is

conceptually similar to the procedure just described. In the following, $\sigma_i \in \bigtimes_{h_i \in H_i} \Delta(\hat{\mathcal{A}}i(h_i))$ denotes a generic plan for player i. We define

$$\zeta(h|\sigma_i, s_{-i}, \theta, \tau^K) := \prod_{\ell=0}^{L(z)} \sigma_i(\mathrm{a}_{i,\ell+1}(z)|\mathrm{h}_i^\ell(z)) \cdot \mathbf{1}\{\mathrm{a}_{-i,\ell+1}(z) = s_{-i}(\mathrm{h}_{-i}^\ell(z))\} \cdot f_{\mathrm{h}^\ell(z)}(s_i^*, s_{-i}, \theta, \tau^K)[\mathrm{m}_{\ell+1}(z)].$$

The functions in Roman font extract action and message profiles realized during terminal history z, as well as suitable predecessors of given lengths, and s_i^* is any personal external state of player i that prescribes $\mathrm{a}_{i,\ell+1}(z)$ at $\mathrm{h}_i^\ell(z)$.[59] The conditional version of such distribution would be

$$\zeta(h|\sigma_i, s_{-i}, \theta, \tau^K) = \begin{cases} \dfrac{\zeta(h|\sigma_i, s_{-i}, \theta, \tau^K)}{\sum_{h' \in \bar{H}(h_i)} \zeta(h'|\sigma_i, s_{-i}, \theta, \tau^K)} & \text{if } h \in \bar{H}_i(h_i); \\ 0 & \text{if } h \notin \bar{H}_i(h_i). \end{cases}$$

References

Aliprantis, C. D. and K. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Berlin: Springer-Verlag. 12, 19, 42

Battigalli, P. (2003): "Rationalizability in infinite, dynamic games with incomplete information," *Research in Economics*, 57, 1–38. 6, 30

Battigalli, P., E. Catonini, and N. De Vito (2025): *Game Theory: Analysis of Strategic Thinking*, Bocconi University: typescript. 6

Battigalli, P., E. Catonini, and J. Manili (2023): "Belief Change, Rationality, and Strategic Reasoning in Sequential Games," *Games and Economic Behavior*, 142, 527–551. 5

Battigalli, P., R. Corrao, and M. Dufwenberg (2019a): "Incorporating belief-dependent motivation in games," *Journal of Economic Behavior & Organization*, 167, 185–218. 2, 5, 26, 27, 29

Battigalli, P., R. Corrao, and F. Sanna (2020): "Epistemic game theory without types structures: An application to psychological games," *Games and Economic Behavior*, 120, 28–57. 5, 6, 33

Battigalli, P. and N. De Vito (2021): "Beliefs, Plans, and Perceived Intentions in Dynamic Games," *Journal of Economic Theory*, 195, 105283. 5, 29

Battigalli, P. and M. Dufwenberg (2009): "Dynamic psychological games," *Journal of Economic Theory*, 144, 1–35. 2, 27

[59] This works because feedback is expressed conditional on a given collective history, and players' personal external states matter in the generation of feedback only through the action they prescribe at the induced personal history.

——— (2022): "Belief-Dependent Motivations and Psychological Game Theory," *Journal of Economic Literature*, 60. 2, 9, 18, 26

Battigalli, P., M. Dufwenberg, and A. Smith (2019b): "Frustration, aggression, and anger in leader-follower games," *Games and Economic Behavior*, 117, 15–39. 3, 25

Battigalli, P. and N. Generoso (2024): "Information Flows and Memory in Games," *Games and Economic Behavior*, 145, 356–376. 8

Battigalli, P. and A. Prestipino (2013): "Transparent restrictions on beliefs and forward-induction reasoning in games with asymmetric information," *The BE Journal of Theoretical Economics*, 13, 79–130. 6, 30, 31, 33, 34

Battigalli, P. and M. Siniscalchi (1999): "Hierarchies of conditional beliefs and interactive epistemology in dynamic games," *Journal of Economic Theory*, 88, 188–230. 5

——— (2002): "Strong belief and forward induction reasoning," *Journal of Economic Theory*, 106, 356–391. 33

——— (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 1, Article 3. 6

Battigalli, P. and P. Tebaldi (2019): "Interactive epistemology in simple dynamic games with a continuum of strategies," *Economic Theory*, 68, 737–763. 6, 29, 33, 51

Behrens, F. and M. E. Kret (2019): "The interplay between face-to-face contact and feedback on cooperation during real-life interactions," *Journal of Nonverbal Behavior*, 43, 513–528. 3

Bertsekas, D. P. and S. Shreve (1996): *Stochastic Optimal Control: The Discrete-Time Case*, Belmont, MA: Athena Scientific. 41

Brandenburger, A. and E. Dekel (1993): "Hierarchies of beliefs and common knowledge," *Journal of Economic Theory*, 59, 189–198. 21, 37

Caplin, A. and J. Leahy (2001): "Psychological expected utility theory and anticipatory feelings," *The Quarterly Journal of Economics*, 116, 55–79. 25

Dekel, E. and M. Siniscalchi (2015): "Epistemic game theory," in *Handbook of Game Theory with Economic Applications*, Elsevier, vol. 4, 619–702. 5

DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper (2003): "Cues to Deception," *Psychological Bulletin*, 129, 74–118. 4

Druckman, D. and M. Olekalns (2008): "Emotions in negotiation," *Group Decision and Negotiation*, 17, 1–11. 3

Dubins, L. and D. Freedman (1964): "Measurable sets of measures," *Pacific Journal of Mathematics*, 14, 1211–1222. 41

Elfenbein, H. A., M. Der Foo, J. White, H. H. Tan, and V. C. Aik (2007): "Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation," *Journal of Nonverbal Behavior*, 31, 205–223. 3

Elster, J. (1996): "Rationality and the emotions," *The Economic Journal*, 106, 1386–1397. 2

——— (1998): "Emotions and economic theory," *Journal of Economic Literature*, 36, 47–74. 2

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological games and sequential rationality," *Games and Economic Behavior*, 1, 60–79. 2

Givens, D. B. (1978): "The nonverbal basis of attraction: Flirtation, courtship, and seduction," *Psychiatry*, 41, 346–359. 1

Goldin-Meadow, S. (1999): "The role of gesture in communication and thinking," *Trends in Cognitive Sciences*, 3, 419–429. 1

Hatfield, E., L. Bensman, P. D. Thornton, and R. L. Rapson (2014): "New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion," *Interpersona*. 2

Kőszegi, B. and M. Rabin (2006): "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, 121, 1133–1165. 25

Kreps, D. M. (2013): *Microeconomic Foundations I: Choice and Competitive Markets*, vol. 1, Princeton, NJ: Princeton University Press. 27

Mann, S., A. Vrij, and R. Bull (2004): "Detecting True Lies: Police Officers' Ability to Detect Suspects' Lies." *Journal of Applied Psychology*, 89, 137. 4

Pearce, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 1029–1050. 29

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

Stirrat, M. and D. I. Perrett (2010): "Valid facial cues to cooperation and trust: Male facial width and trustworthiness," *Psychological Science*, 21, 349–354. 1

Van Kleef, G. A. and S. Côté (2022): "The social effects of emotion," *Annual Review of Psychology*, 73, 1–30. 2

Van Leeuwen, B., C. N. Noussair, T. Offerman, S. Suetens, M. Van Veelen, and J. Van De Ven (2018): "Predictably angry—facial cues provide a credible signal of destructive behavior," *Management Science*, 64, 3352–3364. 1, 3

Vásquez, J. and M. Weretka (2020): "Affective empathy in non-cooperative games," *Games and Economic Behavior*, 121, 548–564. 2