Games with Noisy Signals About Emotions*

Pierpaolo Battigalli† Nicolò Generoso‡

February 15, 2022

Abstract

We study games where utilities may depend on emotions, and 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 and economic phenomena, and they are often betrayed by some signals, as both common sense and everyday experience suggest: facial expressions may warn that a student is confused by a lecture, 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. For instance, emotional leakage occurs when people lie (Porter, Ten Brinke, & Wallace, 2012), nonverbal communication is key in courtship encounters (Givens, 1978), individuals seem to recognize others' predisposition to anger or trustworthiness by observing facial cues (Stirrat & Perrett, 2010; Van Leeuwen et al., 2018), and gesture effectively informs listeners of a speaker's unspoken thoughts (Goldin-Meadow, 1999). Moreover, available 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 contagion" (Hatfield, Bensman, Thornton, & Rapson, 2014). All in all, emotional expressions of others provide useful tools that can be exploited to make social interactions more predictable and manageable.

*We thank Manuel Arnese, Nicodemo De Vito, and Julien Manili, as well as seminar audiences at Bocconi, Paris School of Economics, and Paris Sorbonne for useful comments.

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

‡Yale University, New Haven, CT, USA. Contact: nicolo.generoso@yale.edu.

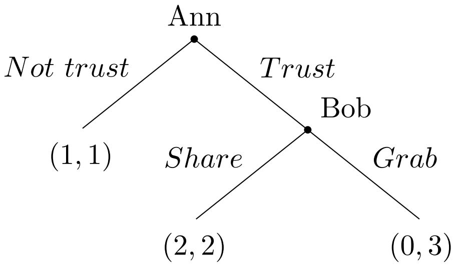
The effect of emotions on decision-making should be of primary interest for economists. Elster (1996, 1998) convincingly argued that a careful study of emotions could help to get a better grip on how decisions are formed, and in this regard Psychological Game Theory (PGT) – pioneered by Geanakoplos, Pearce, and Stacchetti (1989) and substantially developed by Battigalli and Dufwenberg (2009) and Battigalli, Corrao, and Dufwenberg (2019) – represents a rich framework to address the issue.[1] However, to the best of our knowledge, the role of emotional feedback has never been formally analyzed. Incorporating such aspect in a formal analysis would not only result in a more accurate description of reality, but it may also lead to new insights when strategic reasoning is studied. Indeed, 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 feedback 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* – all these aspects are key in interactive environments. As a result, such inferential reasoning can fruitfully inform players' 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 describes our contribution. In Section 1.3, we elaborate on our methodological position, and we briefly discuss the related literature.

1.1 Heuristic examples

As partly mentioned, emotional feedback can shed light on the emotions of others, on their personal traits, on their future behavior, and on past (unobserved) actions. In this section, we briefly describe some examples where this occurs.

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

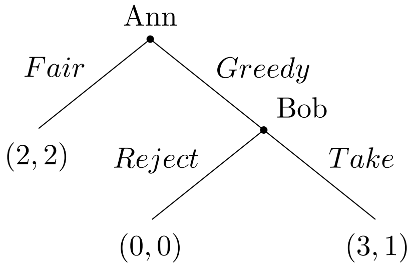


[1] The main innovation introduced by the theoretical apparatus of PGT consists in letting players' utilities depend on (their own and their opponents') beliefs. In this way, a wide array of belief-dependent sentiments and emotions, ranging from reciprocity to self-esteem, can be modeled. See Battigalli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

The findings of Behrens and Kret (2019) suggest that face-to-face contact may foster cooperation and pro-social behavior, and we can indeed enrich the traditional representation of the Trust Mini-Game by allowing Bob to receive a message about Ann's emotions before making his choice. An emotion that may be betrayed by some facial cues when she acts is her trustfulness: for instance, smiling may convey her desire to cooperate, and Bob would make inferences about Ann's *emotions* upon observing such signal. Trustfulness can easily be tied to the present situation – it could be thought of as, for example, the extent to which Ann expects Bob to share. Therefore, 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's hubris, and rejections at the second stage are accounted for by the model of Battigalli, Dufwenberg, and Smith (2019). 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. Importantly, such messages cannot concern Bob's frustration, because frustration arises only as a consequence of opponents' choices – hence, players cannot be frustrated at the root of the game. 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, and emotional expressions – both verbal (e.g., statements) and nonverbal ones (e.g., gesture) – allow to infer the counterparts' intentions (Druckman & Olekalns, 2008). Elfenbein, Der Foo, White, Tan, and Aik (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 a long alternating-offer bargaining procedure. We could also assume that irritation may arise if one party receives an offer that is far from the minimal acceptable outcome, or that impatience may emerge as the negotiation lengthens. In the former case, the proposer could use such signal to fine-tune her beliefs about the counterpart's reservation value, which may be thought of as a *personal trait.* In the latter, one party may realize that the counterpart could settle for less advantageous terms only to end the costly delay. Hence, it could use emotional signals to make inferences about the opponent's propensity to accept an offer in the upcoming rounds of negotiation – hence, on her *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 far from clear – 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, Vrij, and Bull (2004) declared that they look primarily at gaze aversion to detect lies in interrogations. Hence, they (perhaps mistakenly) use emotional messages to infer *past unobserved actions* of others and to infer whether the suspects' stories coincide with their actual past behavior. Our running example in the next sections will share similarities with this setting, and it will investigate the disclosure of private information – a typical economic problem – in face-to-face interactions where emotional leakage may betray lies. ▲

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, as emotional messages can be assumed to be noisy, and we take this generative process to be driven by the agents' states of mind. On a related point, we also discuss how emotions are generated by players' beliefs and behavior as the game unfolds, and how emotional messages allow to make inferences when reasoning strategically.

Secondly, we carry out an analysis of the key features that allow to derive behavioral predictions in the setting of interest. Specifically, 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. Thus, saying that rationality is an event implies that a given player may wonder about her opponents' rationality, incorporating such event into her strategic reasoning. From both a conceptual and a technical point of view, this result is essential to develop a theory of strategic thinking and to derive predictions in the present framework.

Finally, we propose a rationalizability-like solution concept to predict behavior, and we justify it in terms of underlying assumptions about players' rationality and 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. Given that such information includes emotional signals, such procedure provides the adequate tools to formalize the idea that players use such 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 truth-telling.

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 and emphasizing their conceptual implications. First of all, our work builds on the methodological paper of Battigalli, Corrao, and Dufwenberg (2019) in the way it models psychological games and belief-dependent motivations, but it features a major difference. There, states of the world include a description of *how the game unfolds* (i.e., players form first-order beliefs over the set of terminal histories of the game), while in our setting states of the world include a description of *how players would behave* also at histories that do not realize. In this regard, our approach mirrors that of Battigalli and De Vito (2021). Like them, we also explicitly distinguish between players' plans and actual behavior, requiring that they coincide only for rational players. 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 can also fail to stick to their 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: a "correct" belief revision policy is embedded in the definition of Conditional Probability Systems (cf. Axiom 3 in Battigalli & Siniscalchi, 1999), which are conventionally used to model beliefs in sequential games, or coherence between beliefs of different orders follows from the choice of positing a canonical type structure (cf. Battigalli & Siniscalchi, 1999 and Dekel & Siniscalchi, 2015). We instead derive a rich states space, and we take the desired rationality features to be properties holding *only at some states* – this way, each requirement becomes an explicit assumption. Specifically, like in Battigalli, Corrao, and Sanna (2020), we do not postulate a type structure, and we take instead an infinite hierarchical system of beliefs to be the epistemic 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. But not only that: 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

[3] In contrast, for example, types are collectively coherent hierarchies of conditional beliefs (in the words of Dekel & Siniscalchi, 2015) in a canonical type structure. This means that, *by construction*, the possibility that

a prerequisite for 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 theoretic analysis of such phenomena.

Lastly, a word on our solution concept. 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 essentially a version of strong Δ-rationalizability (see Battigalli & Tebaldi, 2019 and relevant references therein), which captures in standard settings the utility-relevant implications of rationality, some belief restrictions, and common strong belief in both (Battigalli & 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. In light of this, the relevance of our result lies in the fact that it establishes that a procedure carried out taking into account only beliefs of a finite order actually 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 This paper is organized as follows. Section 2 introduces the formal framework. Section 3 formalizes the inferential reasoning players carry out upon observing messages about their opponents. Section 4 defines rationality. Section 5 introduces the strong Δ-rationalizability solution concept. Section 6 provides the epistemic justification for the proposed procedure. Section 7 concludes. Appendix A collects proofs. Appendix B provides a detailed analysis of our running example. Appendix C summarizes our notation, for the reader's convenience.

2 Formal framework

The aim of this section is to define a dynamic game with feedback about emotions. 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 $T \in \mathbb{N}$ 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 $\bar{x}_m$ denote its m-th coordinate (with $m \in \{1, \ldots, n\}$). 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

an opponent features some incoherence between her first- and second-order beliefs is *inconceivable* for any player.

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

Emotions are often betrayed by some signals (e.g., facial cues) and affect well-being: it is then natural to start the definition of our formal framework by describing how emotional feedback is generated and how emotions determine utilities. Moreover, 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 of all, 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 *emotions* of agent i as E_i. With this, 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 $t \in \{1,\ldots,T+1\}$, E^t is the set of *streams of emotion profiles* of length t. Given that we will model games lasting at most T stages, we define for convenience the set $E^{\leq T+1} := \bigcup_{t=1}^{T+1} E^t$, which represents the possible streams of emotions experienced by agents in the situation of interest. Players can experience a stream of emotions of length at most $T+1$ because we assume that they hold some initial emotional state, and then they experience a new one after each stage of the game. For each $i \in I$, $E_i^{\leq T+1}$ has an analogous meaning.

We posit for each $i \in I$ a (nonempty) finite set of *conceivable messages*, M_i, and we let $M := \bigtimes_{i\in I} M_i$.[5] 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* $\tilde{f} : \tilde{A} \times \Theta \times E^{\leq T+1} \to \Delta(M)$, where $\tilde{A}$ is a generic finite set of action profiles that can be taken by agents. As mentioned, we let messages be generated stochastically because messages about emotions are noisy. Moreover, we allow the message generation to depend also on actions agents can take,[6] as well as on their traits (indeed, recall that our feedback is informative also in that respect, as in Examples 2 and 4). Second, we define a profile of continuous *psychological utility functions* $(\tilde{v}_i : Y \times \Theta \times E^{\leq T+1} \to \mathbb{R})_{i\in I}$. Differently from conventional utilities, they do not depend only on outcomes and traits, but also on the streams of emotions experienced by all players. Importantly, such dependence on unobserved variables (e.g., the emotions of others) is understood in a state-dependent sense.

[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] It may be useful to assume $M_i = \bigtimes_{j\in I\setminus\{i\}} M_{i,j}$, where $M_{i,j}$ is interpreted as the set of messages about j's emotions that i can observe – this comes in handy when games with more than two players are studied. On the other hand, whenever $I = \{i,j\}$, M_i (resp. M_j) is isomorphic to $M_{i,j}$ (resp., $M_{j,i}$).

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

2.2 The game tree

The peculiarity of our framework is that players receive messages about emotions and traits of others. Moreover, we do not assume players observe their opponents' 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. We take the position that, whenever a player is called to act, her available actions are self-evident, regardless of whether she perfectly recalls how the game unfolded up to that point. Given that 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. Put alternatively, we model game-specific information as a *stream* rather than as a *stock*. The difference is relevant, as a stock-based approach implicitly requires perfect recall for a game tree to be meaningfully defined, while our approach clearly separates a game form from players' cognitive abilities.[7] This way of modeling players' information throughout the game is the one proposed by Battigalli and Generoso (2021) – we refer the reader to such paper for a more detailed discussion of the conceptual and methodological issues involved in our modeling choice. Also, with respect to Battigalli and Generoso (2021), we restrict attention to multistage games for simplicity.

For each $i \in I$, we let A_i be the (nonempty) finite set of *potentially available actions* of player i, and $M_{i,p}$ the finite set of *previous play messages* player i can receive. As a mnemonic, we write $m_{i,x}$ to denote a message i can observe about x. This could be previous play (in which case $x = p$), or player j's state of mind (in which case $x = j$). We let $A := \bigtimes_{i \in I} A_i$ and $M_p := \bigtimes_{i \in I} M_{i,p}$ be the set of profiles of actions and previous play messages, respectively. We also posit a *previous play message generating function*, $p : \bigcup_{t=1}^{T} A^t \to M_p$. Note that, unlike emotional messages, previous play messages are produced deterministically, and that letting p be the map $a^t \mapsto (a^t)_{i \in I}$ basically amounts to assuming observable actions. For each $a^t \in A^t$, $t \in \{1, \ldots, T\}$, and $i \in I$, we let $p_i(a^t) := \operatorname{proj}_{M_{i,p}} p(a^t)$.

Then, we posit, for each $i \in I$, an *action feasibility correspondence*, $\mathcal{A}_i : M_{i,p} \cup \{\varnothing_{M_{i,p}}\} \rightrightarrows A_i$. The interpretation is straightforward: for a given previous play message profile $m_{i,p} \in M_{i,p}$ received by player i at a given stage, $\mathcal{A}_i(m_{i,p})$ is the set of actions available to her at the subsequent stage. Moreover, $\varnothing_{M_{i,p}}$ stands for the situation in which player i has not received any message yet: thus, $\mathcal{A}_i(\varnothing_{M_{i,p}})$ represents the actions player i can choose at the first stage. It is convenient to define $\mathcal{A} : M_p \cup \{\varnothing_{M_p}\} \rightrightarrows A$ to be such that $\mathcal{A}(m_p) := \bigtimes_{i \in I} \mathcal{A}_i(m_{i,p})$ for each $m_p = (m_{i,p})_{i \in I}$, and $\mathcal{A}(\varnothing_{M_p}) := \bigtimes_{i \in I} \mathcal{A}_i(\varnothing_{M_{i,p}})$. Lastly, we assume that, for each $m_p = (m_{i,p})_{i \in I}$ and $i \in I$, $\mathcal{A}_i(m_{i,p}) = \emptyset$ if and only if $\mathcal{A}_j(m_{j,p}) = \emptyset$ for each $j \in I$. In such case, also $\mathcal{A}(m_p) = \emptyset$.

[7] 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. Building on this example, we 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: the exact sequence of moves that led to a specific piece disposition is *not* strictly needed to figure out the ways in which the game may continue.

In words, as soon as the game is over for one player, it is over for everyone.[8]

We take histories to be sequences of profiles of actions, previous play messages, and messages about emotions and traits. With $\tilde{f}$ 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*.[9] For convenience, we assume that the empty history $\varnothing$ belongs to $\bar{H}$.[10] We say that a history $(a^t, m_p^t, m^t) \in A^t \times M_p^t \times M^t$ (with $t \in \{1, \ldots, T\}$) is *feasible* if:

1. $a_1 \in \mathcal{A}(\varnothing_{M_{i,p}})$, and, for each $k \in \{1, \ldots, t-1\}$, $a_{k+1} \in \mathcal{A}(m_{p,k})$;

2. for each $k \in \{1, \ldots, t\}$, $m_{p,k} = p(a^k)$;

3. for each $k \in \{1, \ldots, t\}$, there exists $(\theta, e^k) \in \Theta \times E^k$ such that $m_k \in \operatorname{supp} \tilde{f}(a_k, \theta, e^k)$.

The set of *terminal histories* is $Z := \{h = (a^t, m_p^t, m^t) \in H : t = T \text{ or } \mathcal{A}(m_{p,t}) = \emptyset\}$, and the set of *non-terminal histories* is $H := \bar{H} \setminus Z$.

It is worth stressing that, at each stage, agents first act, and then observe messages. Indeed, the previous play message profile generated at some stage k depends on the entire sequence of action profiles played up to that stage, including the k-th-stage action profile. Similarly, emotional feedback depends by definition on the actions players may choose.

The assumption that players need not observe their opponents' actions or messages justifies the introduction, for each $i \in I$, of the set of *personal histories* of player i, defined as $\bar{H}_i := \operatorname{proj}_{\bigcup_{t=1}^{T} A_i^t \times M_{i,p}^t \times M_i^t} \bar{H}$. The set $\bar{H}_i$ basically 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.

A (weak) "prefix of" relation $\preceq$ can be defined on $\bar{H}$. Given $\tilde{h} = (\tilde{a}^k, \tilde{m}_p^k, \tilde{m}^k), h = (a^\ell, m_p^\ell, m^\ell) \in \bar{H}$, $\tilde{h} \preceq h$ if either $\tilde{h} = h$ or $k < \ell$ and $(\tilde{a}^k, \tilde{m}_p^k, \tilde{m}^k) = (a^k, m_p^k, m^k)$. If $\tilde{h} \preceq h$, we say that $\tilde{h}$ (weakly) precedes h. Under the assumption that $\varnothing \in \bar{H}$, 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 $\pi_i(z, \theta) := \operatorname{proj}_{Y_i} \pi(z, \theta)$. We conclude this section introducing our running example.

Example 5 (Buy me an ice-cream) Child is at home alone, and he faces a typical childhood dilemma: he should do his homework, but playing video-games is quite tempting. When Mom gets back from work, Child asks her to buy him an ice-cream. Mom would be happy to reward Child, but she cannot know (nor check) if her son studied. She simply decides to ask him if he has done his homework, and to act based on the answer she will receive. To make the problem more interesting we add two features. First, we assume Child is concerned about his image in

[8] 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.

[9] As for $\tilde{f}$, we let its domain be $\tilde{A} \times \Theta \times E^{\leq T+1} = A \times \Theta \times E^{\leq T+1}$.

[10] The empty history can be thought of as a history of length zero where no action has been played and 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 $\varnothing$.

Mom's eyes: he dislikes being though of as a liar, *regardless of whether he actually lied or not.*[11] Second, we assume that Child may blush when he falsely affirms that he has done his homework.

Formally, we define $A_C := \{H, V, Y, N\}$, where the elements denote doing homework, playing video-games, saying "yes", and saying "no", respectively.[12] As for Mom, we let $A_M := \{B, N\}$, because she can either buy Child an ice-cream or not. Only Mom observes emotional messages throughout, so let $M_M := \{b, \neg b, n\}$ – whose elements respectively stand for "blushing", "not blushing", and "uninformative message" –, and $M_C := \{n\}$. Lastly, assume Θ_M is a singleton and let $\Theta_C := \Lambda \times N \subset \mathbb{R}^2_+$, where $\lambda \in \Lambda$ denotes Child's sensitivity for Mom's opinion, and $\nu \in N$ his appreciation for video-games.

We model the situation as follows: Child first decides between homework and video-games without being observed by Mom, then he answers "yes" or "no" to Mom, and lastly Mom decides whether to buy the ice-cream. Formally, for each $a \in \{H, V\}$, $b \in \{Y, N\}$, and $c \in \{B, N\}$, we can define function p to be such that $(a) \mapsto (a, \bar{a})$, $(a, b) \mapsto \big((a, b), b\big)$, $(a, b, c) \mapsto \big((a, b, c), (b, c)\big)$, where the two components are Child's and Mom's previous play messages, respectively, and $\bar{a}$ is an uninformative message. With this, suitable feasibility correspondences are easily defined.

As for emotions, let $E_C := C \times G = [0, 1] \times \{0, 1\}$ and $E_M := B \times D = [0, 1]^2$, where C, G, B, and D respectively stand for confidence, guilt, blame, and distrust. To distinguish them from actions, the elements of such sets will be denoted in bold. Moreover, for each $t \in \{1, \dots, 4\}$, we let $e^t = (e_0, \dots, e_{t-1})$, where e_0 is the initial emotion profile and e_k the profile (here, a pair) of emotions held by players after stage k.

Recall that we would like to model a situation where Child may blush with positive probability only after he lies after having played video-games. In order to do so, we can assume that Child may blush only if he feels guilty for not having done his homework, and that the probability of not blushing is equal to his confidence.[13] Define $\tilde{f}$ to be, for each $(a, \theta, e^2) \in A \times \Theta \times E^2$,

$$\tilde{f}(a, \theta, e^2) = \begin{cases} \mathbf{g}_1\big(\mathbf{c}_1\delta_{\neg b} + (1 - \mathbf{c}_1)\delta_b\big) + (1 - \mathbf{g}_1)\delta_{\neg b} & \text{if } a = Y; \\ \delta_{\neg b} & \text{if } a = N; \end{cases} \tag{1}$$

and equal to δ_n in all other cases.[14] This formulation implies that message b may be generated only after Child's second-stage action and only if he says "yes". Moreover, we hinted at the fact that guilt may arise if Child plays video-games: we will elaborate on this (cf. p. 15), and we will eventually obtain that b may realize only if Child plays video-games and subsequently

[11] 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 & Dufwenberg, forthcoming).

[12] H and Y have been introduced to define the sets of non-terminal histories and collective outcomes. Yet, they are the most natural symbols to denote also Child's actions: we will stick to this notation throughout the example, hoping that this does not cause confusion. Moreover, in the context of this example, actions in uppercase, messages in lowercase, and emotions in bold.

[13] The claim that emotional expressions can effectively help telling truths from lies seems to be backed by relevant evidence (Gadea, Aliño, Espert, & Salvador, 2015; Matsumoto & Hwang, 2018; Warren, Schertler, & Bull, 2009), even though it is disputed. Besides that, *which* feeling is the foremost driver of emotional leakage when lies are told is not clear – hence, our choice of confidence is suggested by common sense only.

[14] We report only Mom's message as subscript, as Child only observes uninformative messages.

says "yes". At that moment, players have experienced a length-two stream of emotions, and this explains the presence of $e^2 \in E^2$ in the definition of $\tilde{f}$. Hence, Mom can observe a trivial length-one personal history (where she waits and observes uninformative signals about Child's action and emotions) and three length-two personal histories, identified with (Y, b), $(Y, \neg b)$, and $(N, \neg b)$.

Lastly, we describe utility functions. All terminal histories in this game have length three, hence players eventually experience a length-four stream of emotions. Let $Y_C := \{0,1\}^2$, with generic element $(y_{C,1}, y_{C,2})$, and $Y_M := \{0,1\}$. Then, define:

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

In words, $y_{C,1}$ (resp., $y_{C,2}$) indicates whether Child eats the ice-cream (resp., plays video-games): he enjoys both, but dislikes Mom's blame. The cost incurred by Mom to buy the ice-cream is instead y_M. Her utility function captures the idea that it is worth buying the ice-cream (in which case, $y_{C,1} = 1$) only if she trusts Child to a relatively high extent (in particular, if $\mathbf{d}_3 \leq \frac{1}{2}$). Again, the rationale of our definition of emotions will become clear when we discuss how emotions arise from players' beliefs in Section 2.3.4 (cf. p. 15). ▲

2.3 Predispositions to act and believe

The aim of this section is to provide a definition of a "state of the world", which we take to be a complete description of players' traits and predispositions to act and believe. Using the term "predisposition", we mean that we do not want to define only players' behavior and beliefs along the path of the game. Rather, we describe how players *would* behave and what they *would* believe conditional on all possible contingencies. Hence, a state of the world encompasses all the relevant aspects of a strategic situation. Finally, we also discuss how the game unfolding and players' game-specific attitudes translate into emotions.

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 $h_i \in H_i$, $\hat{\mathcal{A}}_i(h_i) := \{a_i \in A_i : \exists (m_{i,p}, m_i) \in M_{i,p} \times M_i, (h_i, a_i, m_{i,p}, m_i) \in \bar{H}_i\}$.[15] For each $i \in I$, the set of *personal external states* of i is:

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

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. Thus, elements of S_i can technically be labeled as player i's "strategies", but we refrain from using such terminology: we maintain that strategies are plans of actions *in the minds of players.*

[15] The reader may recognize in such correspondences the "conventional" action feasibility correspondences that are customarily used in the literature on dynamic games.

Hence, they are part of the players' ways of thinking, and will be described by different mathematical objects.[16] 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.

Another conceptual point has to be stressed. Our definition of action feasibility correspondences was motivated by the observation that also players who do not perfectly recall the actions they previously played (or the messages they previously received) must be able to play the game, which requires realizing the actions available at any given stage. This may seem at odds with our choice of using correspondences $(\hat{\mathcal{A}}_i)_{i\in I}$ to define external states. However, we interpret elements of S_i as objective *descriptions* of how player i would behave conditional on different contingencies: despite the reliance on $(\hat{\mathcal{A}}_i)_{i\in I}$, such interpretation remains valid even if we do not assume that player i recalls all the actions she played and the messages she received.

2.3.2 Beliefs

We now discuss how to give a complete description of the epistemic[17] 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 of all, the space of *primitive uncertainty* is $\Omega^0 := S \times \Theta$. This is the basic uncertainty space upon which players form their first-order beliefs.[18] 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 and of everyone as $\mathcal{T}_{-i,1} := \bigtimes_{j\in I\setminus\{i\}} \mathcal{T}_{j,1}$ and $\mathcal{T}_1 := \bigtimes_{i\in I} \mathcal{T}_{i,1}$, respectively. Lastly, for each $i \in I$, we let $\Omega^1_{-i} := \Omega^0 \times \left(\bigtimes_{j\in I\setminus\{i\}} \mathcal{T}_{j,1}\right)$.

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

$$\mathcal{T}_{i,n} := \left[\Delta(\Omega^{n-1}_{-i})\right]^{\bar{H}_i}, \quad \Omega^n_i := \Omega^{n-1}_i \times \mathcal{T}_{i,n} = \Omega^0 \times \left(\bigtimes_{k=1}^{n} \mathcal{T}_{i,k}\right);$$

$$\mathcal{T}_{-i,n} := \bigtimes_{j\in I\setminus\{i\}} \mathcal{T}_{j,n}, \quad \Omega^n_{-i} := \Omega^{n-1}_{-i} \times \mathcal{T}_{-i,n} = \Omega^0 \times \left(\bigtimes_{k=1}^{n} \mathcal{T}_{-i,k}\right).$$

$\mathcal{T}_{i,n}$ is the set of *systems of n-th-order beliefs* of player i. 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 $\mathcal{T}^n_i$.

[16] 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.

[17] To be precise, we should use the term "doxastic" to refer to the falsifiable beliefs held by players. Indeed, the word "epistemic" should refer to knowledge, which can be seen as *correct and justified* certainty. On the other hand, the beliefs held by players are falsifiable, in general. Yet, we use the term "epistemic" because it is entrenched in the literature.

[18] Later on, we will make the assumption that rational players know their personal traits, while the rationale of letting player i be uncertain about her future behavior has already been discussed.

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 $\mathcal{T}^\infty := \bigtimes_{i\in I} \mathcal{T}_i^\infty$.[19]

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 (cf. the discussion in Section 1.3). 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, however, 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 then be modeled as features that hold only at some states of the world (cf. Section 4).

Remark 1 For each $i \in I$ and $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{T}_i^n$ is compact metrizable.[20]

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 $(\tau_{i,k}(\,\cdot\,|h_i))_{k=1}^n$. Lastly, 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. The following is obvious in light of Remark 1.

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

As already mentioned, 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

[19] 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.

[20] Given that Ω^0 is finite, it is compact metrizable and so is $\Delta(\Omega^0)$ (Aliprantis & 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.

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 about the behavior of others). Indeed, in our running example we introduced broad concepts such as guilt, distrust, or blame, but the situation at hands also suggested a very natural interpretation of such emotions (e.g., Child feels guilty if he plays video-games instead of studying, or Mom distrusts Child if she thinks that he is lying). Therefore, our aim is now to tie streams of emotions experienced by players during the game to states of the world – i.e., to the relevant aspects of the strategic interaction.

First of all, 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). Hence, 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.[21]

Then, we define a continuous *emotion-generating function*, $\varepsilon : \bar{H} \times \mathcal{T}^\infty \to \Delta(E^{\leq T+1})$, and we make some reasonable 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 $\bar{\varepsilon}_h : \rho_h(\mathcal{T}^\infty) \to \Delta(E^{\leq T+1})$. Second, along histories of a given length $t \in \{1, \ldots, T\}$ players experience streams of emotion profiles of length $t+1$:[22] for each $h \in \bar{H}^t$, $\operatorname{supp} \varepsilon_h \subseteq E^{t+1}$. Third, 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.

Definition 1 *A* ***game with feedback about emotions*** *is a structure*

$$\Gamma := \langle I, \mathcal{A}, \tilde{f}, p, \pi, \varepsilon, (\Theta_i, A_i, M_{i,p}, M_i, Y_i, E_i, \tilde{v}_i)_{i \in I} \rangle.$$

It is informally assumed that the elements of the previous definition are commonly known.

[21] 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 $\zeta_{h_i}(z|s, \theta, \tau^K)$ the probability of z occurring given (s, θ, τ^K) and conditional on h_i occurring (cf. Section 4.4).

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

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

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

where, for each history $h \in \bar{H}$, we let $L(h)$ denote its length.

Remark 3 For each $h \in H$, f_h is continuous.[23]

Note that the domain of feedback functions is now a set over which players form well-defined beliefs. Moreover, it is often convenient to refer to the probabilities of realization of messages received by a given player. Hence, for each $i \in I$ and $h \in H$, let $f_{i,h} = \mathrm{marg}_{M_i} \circ f_h$.[24] In Section 3, we will present reasonable properties of feedback that allow to prove technical results.

Example 5 (Buy me an ice-cream, continued) A generic personal external state of Mom is indicated as $a_1.a_2.a_3$, where a_1, a_2, and a_3 are the actions prescribed after histories (Y, b), $(Y, \neg b)$, and $(N, \neg b)$, 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.[25]

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,[26] and we let $K = 1$. First of all, Child is *guilty* if he plays video-games instead of doing his homework. Hence, we simply impose $\mathbf{g}_1 = 1$ after history (V), and $\mathbf{g}_1 = 0$ after (H).[27] Moreover, 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 $\mathbf{c}_1 = \tau_{C,1}\big(\{s_M : s_M\big((Y, b)\big) = B\}|V\big)$. As for Mom, *blame* ($\mathbf{b}_3$) is equal to the probability with which she believes Child lied, while *distrust* ($\mathbf{d}_3$) is equal to her skepticism about Child's report. Let $L := \big\{V.Y, H.N\big\}$ be the set of lies and $G := \big\{H.Y, H.N\big\}$ the set of "good behaviors", and let z_M denote the terminal personal history observed by Mom during the game unfolding. Then, $\mathbf{b}_3 = \tau_{M,1}(L|z_M)$ and $\mathbf{d}_3 = 1 - \tau_{M,1}(G|z_M)$. How emotions are generated at other stages of the game can be discussed extending these reasoning in a straightforward way.

[23] Taking a sequence $(s_n, \theta_n, \tau_n^K)_{n \in \mathbb{N}}$ converging to $(\bar{s}, \bar{\theta}, \bar{\tau}^K)$, it can be showed that, for each $m \in M$ and $h \in H$, $f_h(s_n, \theta_n, \tau_n^K)[m]$ converges to $f_h(\bar{s}, \bar{\theta}, \bar{\tau}^K)[m]$, by continuity of $\tilde{f}$ and by our assumptions about ε. This implies that $f_h(s_n, \theta_n, \tau_n^K)$ converges to $f_h(\bar{s}, \bar{\theta}, \bar{\tau}^K)$, proving continuity of f_h.

[24] 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 $(\mathrm{marg}_X \, \mu)(x) := \sum_{y \in Y} \mu(x, y)$.

[25] 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.

[26] That is, we will take (with some abuse) the range of function ε to be $E^{\leq T+1}$ instead of $\Delta(E^{\leq T+1})$. This choice is innocuous, as the former set can obviously be embedded in the latter.

[27] Recall that boldface letters represent emotional states. E.g., $\mathbf{g}_1$ describes whether Child feels guilty or not. Subscripts instead stand for the stage at which emotions are considered.

With this, feedback takes a very tractable form: we obtain, for each $(s, \theta, \tau^1) \in S \times \Theta \times \mathcal{T}^1$,

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

where $q = \tau_{C,1}\big(\{s_M : s_M\big((Y,b)\big) = B\}|V\big)$ and subscripts of f denote the length-one history after which emotional messages are generated. ▲

We conclude with a methodological consideration. Emotions are ultimately bypassed, as it is convenient to base the analysis on f rather than on $\tilde{f}$. This begs the question of why we have not started expressing directly feedback functions as dependent on players' beliefs. The reason is essentially pedagogical: starting from the game-independent notion of "emotion" allowed us to give a *constructive* definition of the game tree. We believe this approach to be 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. An axiomatic approach would obscure this distinction.

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 they encode the only relevant factors needed to generate emotions.

Remark 4 For each $i \in I$, v_i is continuous.[28]

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 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 $g := (g_h : S \times \Theta \times \mathcal{T}^K \to \Delta(A \times M_p \times M))_{h \in H}$ that specifies how profiles of actions and messages generate after each non-terminal history. In other words, g allows to describe with which probability the game moves from a non-terminal history to any of its immediate

[28] The remark follows from continuity of functions $\tilde{v}_i$ and ε, both of which are assumed.

successors once we fix an underlying profile (s,θ,τ^K). For each $h \in H$, $(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K$, and $(a, m_p, m) \in A \times M_p \times M$, let

$$g_h(s,\theta,\tau^K)[(a,m_p,m)] := \begin{cases} f_h(s,\theta,\tau^K)[m] & \text{if } a = s(h), m_p = p\big(\operatorname{proj}_{\bigcup_{t=0}^T A^t} h, s(h)\big); \\ 0 & \text{otherwise.} \end{cases}$$

In words, once we fix (s,θ,τ^K) and h, the probability that profile (a, m_p, m) realizes can be positive if and only if a is consistent with the behavior described by s after h and m_p is the previous play message that would be generated by function p after the sequence of action profiles induced by s and h. If both this conditions hold, then the probability of realization of (a, m_p, m) is simply the probability of m, as specified by feedback functions.

Now define function $\zeta : S \times \Theta \times \mathcal{T}^K \to \Delta(Z)$, to be such that, for each $(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K$ and $z \in Z$:

$$\zeta(z|s,\theta,\tau^K) := \prod_{t=0}^{L(z)} g_{h^t(z)}(s,\theta,\tau^K)[(a_{t+1}(z), m_{p,t+1}(z), m_{t+1}(z))], \tag{4}$$

where $h^t(z)$ is the truncation of z at stage t, and $a_t(z)$, $m_{p,t}(z)$, and $m_t(z)$ are the t-th-stage action, previous play message, and message components of z, respectively.

Then, 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 state, rather than on the terminal history reached. 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, it is convenient to define $S \times \Theta \times \mathcal{T}^K$ as the set of *utility-relevant states.*

Remark 5 For each $i \in I$, u_i is continuous.[29]

Example 5 (Buy me an ice-cream, continued) Game-dependent psychological utilities are easily retrieved: for each $z = (a_{C,1}, a_{C,2}, m_M, a_M) \in Z$ and $(\theta, \tau^1) \in \Theta \times \mathcal{T}^1$,

$$v_C(z,\theta,\tau^1) := \pi_{C,1}(z) + \nu\pi_{C,2}(z) - \lambda\tau_{M,1}(L|z_M); \quad v_M(z,\theta,\tau^1) := \begin{cases} 2\tau_{M,1}(G|z_M) - 1 & \text{if } a_M = B; \\ 0 & \text{if } a_M = N; \end{cases}$$

where π is the consequence function,[30] z_M is the terminal personal history observed by Mom, and $\pi_{C,1}(z)$ and $\pi_{C,2}(z)$ are the two coordinates of Child's outcome after z.

Deriving $(u_i)_{i\in I}$ is straightforward, yet costly in terms of notation: we postpone a detailed analysis to Section B.1 in Appendix B. To fix ideas, consider a utility-relevant state (s,θ,τ^1) with

[29] The remark follows from continuity of functions $(v_i)_{i\in I}$ (Remark 4) and of ζ (which is straightforward once we observe that functions $(f_h)_{h\in H}$ are continuous as per Remark 3).

[30] In the present case, outcomes do not depend on traits, so we avoid specifying the dependence of π on them.

$s_C = H.Y$. This implies that Mom's personal history $(Y, \neg b)$ will be generated with certainty, and the resulting terminal history depends on the action prescribed by s_M afterwards. We have:

$$u_C(s, \theta, \tau^1) = \begin{cases} 1 - \lambda \tau_{M,1}\big(L|(Y, \neg b)\big) & \text{if } s_M\big((Y, \neg b)\big) = B; \\ -\lambda \tau_{M,1}\big(L|(Y, \neg b)\big) & \text{if } s_M\big((Y, \neg b)\big) = N. \end{cases}$$

▲

3 Inferences on opponents' behavior, traits, and beliefs

In the present framework, players have the means to make inferences about their opponents' behavior and realized beliefs as they observe emotional messages. Thus, 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 personal histories realized). The focus here is on utility-relevant states, because players cannot make inferences about beliefs of others of order higher than K. In the following, we formalize such reasoning: Section 3.1 discusses mild assumptions about feedback functions that allow to prove meaningful results, 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 2 gives a condition for the feedback about others a player may observe to be independent from that player's own beliefs, and Definition 3 gives a notion of simplicity for feedback. Moreover, as one may expect, a 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 some history. This is indeed necessary for a player to be able to reason about which messages she (or her opponents) may observe at different points of the game – Definition 4 is in this spirit. It is worth specifying that these are *not* maintained assumptions.[31]

First of all, we formalize the idea that, at any history, the beliefs of a player should not play any role on 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*.

Definition 2 *Feedback $f = (f_h)_{h \in H}$ is **own-belief independent** if, for each $i \in I$, $h \in H$, $s \in S$, $\theta \in \Theta$, and $\tau^K_{-i} \in \mathcal{T}^K_{-i}$, the section $f_{i,h,(s,\theta,\tau^K_{-i})}$ of $f_{i,h}$ is constant on $\mathcal{T}^K_i$.*

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^K_{-i}). 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

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

out. Indeed, such effect is incorporated in the realized history, that is crucial in determining feedback, 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,[32] and (ii) the generation of messages about a player's 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 3 *Feedback* $f = (f_h)_{h\in H}$ *is* ***simple*** *if* (i) $K = 1$, *and* (ii) *for each* $i \in I$, $h = (h_i)_{i\in I} \in H$, *and* $(s,\theta) \in S \times \Theta$, $\tau_i^1(\,\cdot\,|h_i) = \bar{\tau}_i^1(\,\cdot\,|h_i)$ *implies that, for each* $j \in I \setminus \{i\}$ *and* $\tau_{-i}^1 \in \mathcal{T}_{-i}^1$, $\mathrm{marg}_{M_{j,i}} f_h(s,\theta,\tau_i^1,\tau_{-i}^1) = \mathrm{marg}_{M_{j,i}} f_h(s,\theta,\bar{\tau}_i^1,\tau_{-i}^1)$.

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

This specification is a good compromise between richness and tractability. Indeed, the vast majority of psychological motivations can be modeled resorting to first-order beliefs only (Battigalli & Dufwenberg, forthcoming), so that point (i) does not seem to be too restrictive. As for (ii), it basically 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. Specifically, feedback would be simple in all the examples we mentioned. For instance, in Example 1, the only informative message is generated after the first stage based on Ann's initial beliefs. In Example 2, feedback depends only on traits and it is entirely belief-independent – hence, trivially simple. In Examples 3 and 4, we can model message generation relying on the parties' and the suspect's first-order beliefs held after the last offer received or at an appropriate or "interrogation stage", respectively. Finally, feedback in Example 5 is also simple.

Next, we give conditions about feedback that allow players to make inferences.[33]

Definition 4 *Feedback* $f = (f_h)_{h\in H}$ *is:*

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

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

Remark 6 The following are true:

1. if feedback is regular, it is semi-regular;

[32] Recall that the highest order of utility-relevant beliefs was denoted by $K \in \mathbb{N}$.

[33] Recall that, for each measurable space $(X, \mathcal{X})$, topological space Y, and correspondence $\gamma : X \rightrightarrows Y$, the *lower inverse* of γ, $\gamma^\ell : 2^Y \to 2^X$, is defined to be such that $\gamma^\ell(A) = \{x \in X : \gamma(x) \cap A \neq \emptyset\}$ for each $A \subseteq Y$. Correspondence γ is said to be *measurable* if $\gamma^\ell(F) \in \mathcal{X}$ 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 $k \in K$.

2. if feedback is semi-regular, sets $\big\{(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K : m \in \operatorname{supp} f_h(s,\theta,\tau^K)\big\}$ and $\big\{(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K : m_i \in \operatorname{supp} f_{i,h}(s,\theta,\tau^K)\big\}$ are measurable for each $h \in H$, $m \in M$, $i \in I$, and $m_i \in M_i$.[34]

Semi-regularity is arguably the minimal assumption needed to allow players to carry out a "well-defined" reasoning about possible ways in which the game may unfold (cf. Lemma 1 in Section 3.2), as it ensures that eventualities such as "receiving message m_i with positive probability at (personal) history h_i" can be assessed by player $i \in I$ (cf. point 2 of Remark 6). Regularity is instead a slightly stronger requirement, but it has a reasonable conceptual justification. While semi-regularity only establishes that the set of utility-relevant states allowing for any message of any player to realize with positive probability at any history is measurable, with regularity players are also 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, 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 $\mathcal{T}^K_j$ of a set of the kind $\big\{(s,\theta,\tau^K) \in S \times \Theta \times \mathcal{T}^K : m_i \in \operatorname{supp} f_{i,h}(s,\theta,\tau^K)\big\}$ is measurable for each $j \in I \setminus \{i\}$ – this *does not* hold for all measurable subsets of $S \times \Theta \times \mathcal{T}^K$,[35] 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.[36]

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). Hence, regularity and semi-regularity coincide. To check (semi-)regularity of feedback, focus on message b and history (V). We have:

$$\begin{aligned}&\big\{(s,\theta,\tau^1) : b \in \operatorname{supp} f_{M,V}(s,\theta,\tau^1)\big\} =\\ =&\{s_C : s_C(V) = Y\} \times S_M \times \Theta \times \big\{\tau^1_C : \tau_{C,1}\big(\{s_M : s_M\big((Y,b)\big) = B\}|V\big) < 1\big\} \times \mathcal{T}^1_M,\end{aligned}$$

which can be checked to be a measurable rectangle. Similar considerations apply to message $\neg b$ and to history (H). In addition, note that the generation of feedback is entirely independent of Mom'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

[34] Point 1 is obvious. As for point 2, fix $h \in H$ and, for each $(s,\theta) \in S \times \Theta$, let $\gamma_{s,\theta}$ be the correspondence $\tau^K \mapsto \operatorname{supp} f_h(s,\theta,\tau^K)$. Then, the first set is $\bigcup_{(s,\theta)} \big\{(s,\theta)\big\} \times \gamma^\ell_{s,\theta}(m)$, which is measurable because $\gamma_{s,\theta}$ is measurable. As for the second set, we write it as $\bigcup_{(s,\theta)} \big(\{(s,\theta)\} \times \big\{\bigcup_{m_{-i}} \gamma^\ell_{s,\theta}(m_i,m_{-i})\big\}\big)$, which is again easily seen to be measurable.

[35] 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 & Border, 2006).

[36] Proofs are collected in Appendix A.

the possible paths the game can follow given any underlying state. For instance, the occurrence of a personal history may be inconsistent with some utility-relevant states: by realizing this, one can make inferences about the true utility-relevant state (cf. Section **??**).

For each $t \in \{1, \dots, T\}$ and $i \in I$, we inductively define a length-t *action sequence correspondence* $\mathbf{A}^t : S \times \Theta \times \mathcal{T}^K \rightrightarrows \operatorname{proj}_{A^t} \bar{H}$, and a length-t *personal history correspondence of player i* $\mathbf{H}_i^t : S \times \Theta \times \mathcal{T}^K \rightrightarrows \bar{H}_i^t$, which collect, respectively, all the t-long sequences of action profiles that can materialize and all the t-long personal histories player i can observe given a utility-relevant state. We denote by $\bar{H}_i^t$ the set of personal histories of player $i \in I$ of length t, but in general we will drop such superscripts when no direct reference to length is made. Given that we informally assume that players know the rules of interaction, such correspondences can be retrieved by players, by reasoning about how the game may unfold.

Fixing $i \in I$ and $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$, define:

$$\begin{aligned}
\mathbf{A}^1(s,\theta,\tau^K) &:= \big\{a^1 \in \operatorname{proj}_{A^1} \bar{H} : a = (s_i(\varnothing))_{i \in I}\big\}; \\
\mathbf{H}_i^1(s,\theta,\tau^K) &:= \big\{(a_i, m_{i,p}, m_i) \in \bar{H}_i^1 : (1)\ a_i = s_i(\varnothing),\ (2)\ m_{i,p} \in p_i(\mathbf{A}^1(s,\theta,\tau^K)), \\
&\qquad (3)\ m_i \in \operatorname{supp} f_{i,\varnothing}(s,\theta,\tau^K)\big\}.
\end{aligned}$$

Suppose that $\mathbf{A}^k(s,\theta,\tau^K)$ and $\mathbf{H}_i^k(s,\theta,\tau^K)$ have been defined for each $k \in \{1, \dots, t-1\}$, with $1 < t \leq T$. For each $k \in \{1, \dots, t\}$, let $\mathbf{H}^k := (\mathbf{H}_i^k)_{i \in I}$. Define:

$$\begin{aligned}
\mathbf{A}^t(s,\theta,\tau^K) := \Bigg\{ &(a^{t-1}, (a_i)_{i \in I}) \in \operatorname{proj}_{A^t} \bar{H} : (1)\ a^{t-1} \in \mathbf{A}^{t-1}(s,\theta,\tau^K), \\
&(2)\ \forall\, i \in I, a_i \in \bigcup_{h_i^{t-1} \in \mathbf{H}_i^{t-1}(s,\theta,\tau^K)} \{s_i(h_i^{t-1})\} \Bigg\}; \\
\mathbf{H}_i^t(s,\theta,\tau^K) := \Bigg\{ &h_i^t = (a_i^t, m_{i,p}^t, m_i^t) \in \bar{H}_i^t : (1)\ (a_i^{t-1}, m_{i,p}^{t-1}, m_i^{t-1}) \in \mathbf{H}_i^{t-1}(s,\theta,\tau^K), \\
&(2)\ a_{i,t} = s_i(a_i^{t-1}, m_{i,p}^{t-1}, m_i^{t-1}),\ (3)\ m_{i,p,t} \in \bigcup_{a_{-i}^t : (a_i^t, a_{-i}^t) \in \mathbf{A}^t(s,\theta,\tau^K)} \{p_i(a_i^t, a_{-i}^t)\}, \\
&(4)\ m_{i,t} \in \bigcup_{\substack{h_{-i}^{t-1} : (h_i^{t-1}, h_{-i}^{t-1}) \\ \in \mathbf{H}^{t-1}(s,\theta,\tau^K)}} \operatorname{supp} f_{i,(h_i^{t-1}, h_{-i}^{t-1})}(s,\theta,\tau^K) \Bigg\}.
\end{aligned}$$

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,* $\mathbf{A}^t$ *and* $\mathbf{H}_i^t$ *are measurable for each* $i \in I$ *and* $t \in \{1, \dots, T\}$*.*

Therefore, 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 particular, the set of utility-relevant states consistent with $h_i \in \bar{H}_i$ is $(\mathbf{H}_i)^\ell(h_i)$. 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 $\bar{\tau}_i^K$:

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

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. Importantly, the definition of inference sets clarifies that the inferences players can make are in general linked to their beliefs of order up to K. The following remark 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 $\tau_i^K \in \mathcal{T}_i^K$.[37]

Example 5 (Buy me an ice-cream, continued) Consider (s, θ, τ^1) with $\tau_{C,1}\big(\{s_M : s_M\big((Y,b)\big) = B\}|V\big) = \frac{1}{2}$. That is, Child would blush with probability $\frac{1}{2}$ after lying. If $s_C = V.Y$, then $\mathbf{H}^2_M(s,\theta,\tau^1) = \{(Y,b),(Y,\neg b)\}$. If instead $s_C = V.N$ or $s_C = H.N$, then $\mathbf{H}^2_M(s,\theta,\tau^1) = \{(N,\neg b)\}$. Lastly, if $s_C = H.Y$, then $\mathbf{H}^2_M(s,\theta,\tau^1) = \{(Y,\neg b)\}$. All other correspondences are analogously derived. To check measurability of $\mathbf{H}^2_M$, focus on Mom's personal history $(Y,\neg b)$:

$$\begin{aligned}(\mathbf{H}^2_M)^\ell\big((Y,\neg b)\big) = &\big\{(s,\theta,\tau^1) : s_C = H.Y\big\} \cup \\ &\cup \big\{(s,\theta,\tau^1) : s_C = V.Y,\ \tau_{1,C}\big(\{s_M : s_M\big((Y,b)\big) = B\}|V\big) > 0\big\},\end{aligned}$$

and such set can be seen to be measurable.[38] Similar considerations apply to other cases.

Finally, note that feedback is own-belief independent for Mom, so that her inference set corresponding to, e.g., $(Y,\neg b)$ is easily derived:

$$\begin{aligned}\Omega^1_{M,\tau^1_M}\big((Y,\neg b)\big) = &\big\{(s,\theta,\tau^1_C) : s_C = H.Y\big\} \cup \\ &\cup \big\{(s,\theta,\tau^1_C) : s_C = V.Y,\ \tau_{1,C}\big(\{s_M : s_M\big((Y,b)\big) = B\}|V\big) > 0\big\},\end{aligned}$$

where $\tau^1_M \in \mathcal{T}^1_M$ is any of Mom's system of first-order beliefs. ▲

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

It seems natural to require rational players to hold coherent beliefs: with this, we mean that we should to be able to recover lower-order beliefs from higher-order ones through marginalization. In the following, we slightly abuse notation by writing Ω^0_{-i} instead of Ω^0, to ease the exposition.

[37] The result follows because $\Omega^K_{-i,\tau_i^K}(h_i)$ is the section at τ_i^K of $\mathbf{H}^\ell_i(h_i) \subseteq S \times \Theta \times \mathcal{T}_K$, which is measurable as per Lemma 1. Measurable sets in measurable product spaces have measurable sections.

[38] Indeed, it is a union of measurable rectangles. Incidentally, we note that the lower inverse of histories correspondences when feedback is regular takes this shape, (cf. Lemma A3 in Appendix A).

Definition 5 *Epistemic type τ_i^∞ of player $i \in I$ is* ***coherent*** *if, for each $n \in \mathbb{N}$ and $h_i \in H_i$,*

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

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 Battigalli and Siniscalchi (1999), and we state it for future reference. It essentially 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 $h_i \in \bar{H}_i$, $\mathrm{marg}_{\Omega_{-i}^{n-1}} \varphi_i(\tau_i^\infty)(\,\cdot\,|h_i) = \tau_{i,n}(\,\cdot\,|h_i)$.*

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. Indeed, 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 6 *Epistemic type τ_i^∞ of player $i \in I$ satisfies* ***knowledge-implies-belief*** *if, for each $h_i \in \bar{H}_i$,*

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

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 we are not assuming coherence: our notion of knowledge-implies-belief is therefore very weak, as it requires that only $(K+1)$-th-order beliefs be updated consistently with evidence. In principle, we may impose a stronger version by requiring that belief of *all* orders conform to such reasoning and measurability of $\mathcal{T}_{i,KB}^\infty$ would hold under the hypotheses of Lemma 4. However, we are going to consider rational (hence, coherent) players later on. Coherence implies that beliefs of order higher than $K+1$ conform to the inferential reasoning we outlined. With regularity of feedback, we can 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.[39]

[39] While Ω^0 is a finite set (and hence the projection onto it of any inference set is trivially measurable), measurability of sets of the kind $\mathrm{proj}_{\Omega_{-i}^n} \Omega_{-i,\tau_i^K}^K(h_i)$ (with $1 \leq n < K$) is non-trivial: indeed, without regularity they can only checked to be analytic. Clearly, this observation applies only when $K > 1$.

Example 5 (Buy me an ice-cream, continued) In this setting, seeing Child blush is obviously the most informative message Mom may wish for. We already highlighted that $\Omega^1_{C,\tau^1_M}\big((Y,b)\big) = \{V.Y\} \times S_M \times \Theta \times \{\tau^1_C : \tau^1_C(\{s_M \in S_M : s_M\big((Y,b)\big) = B\}|V) > 0\}$. Knowledge-implies-belief then ensures that, for example, Mom's second-order beliefs after personal history (Y,b) are such that $\tau_{M,2}\big(\{s_C\}|(Y,b)\big) = 0$ for each $s_C \in S_C \setminus \{V.Y\}$. The same reasoning is easily extended to beliefs of different orders by coherence. ▲

4.3 Belief updating

We now describe how cognitively rational players should update their beliefs upon observing some pieces of evidence, be it the action they play at some stage or the messages they receive. To model such process it is useful to first unpack the mechanisms through which information accrues to players. Fix a stage k, and assume player i has observed personal history h_i. Her beliefs about utility-relevant states will then be described by $\tau_{i,K+1}(\,\cdot\,|h_i)$. After h_i, player i receives three pieces of information: the action she plays, the previous play message and the emotional message she receives. Yet, such pieces of information accrue to her asynchronously: she first observes the action she takes, and then she receives 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.[40] Then, she can take into account the messages she receives to update her beliefs *about others* in a Bayesian way.

Why should a player not update her beliefs about her personal external state upon seeing (previous play and emotional) messages? Intuitively, because *she has already observed the last action she played*, and thus messages cannot shed further light on her personal external state. Specifically, previous play messages depend on the sequence of action profiles played: given that each player observes each of the actions she takes throughout the game, k-th-stage previous play messages allow, if anything, to make inferences about opponents' personal external states.[41] Analogously, once we fix a player's k-th-stage action, also emotional feedback reduces to feedback about others. To see this, consider the definition of feedback functions in (3): at each history, feedback depends on personal external states only through the action players choose at such history. Thus, if a player has already observed her k-th-stage action, emotional feedback provides her with no novel information about her personal external state.

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 having played at stage k, but before having observed any message. In such metaphor, we would have to require that a player does not change her beliefs about anything but her personal external state after acting. This would formalize a notion of *own-action independence* of beliefs, and it captures precisely the aforementioned idea that own actions and messages should be used to make inferences in "parallel" ways.

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

[41] This is true *a fortiori* if we consider that each player has also just observed her k-th-stage action, right before receiving her k-th-stage previous play message.

Before proceeding, we introduce some notation. For each $i \in I$, let $M_i^* := M_{i,p} \times M_i$. Moreover, for each $i \in I$, $h_i \in H_i$, $a_i \in A_i$, we define the sets of immediate successors of h_i where a_i is played as $\bar{H}_i(h_i, a_i) := \{h_i' \in \bar{H}_i : \exists\, m_i^* \in M_i^*, h_i' = (h_i, (a_i, m_i^*))\}$. Lastly, for each $i \in I$, $h_i \in H_i$, and $a_i \in \hat{\mathcal{A}}_i(h_i)$, define the set of message profiles player i can receive after playing a_i at h_i as $M_i^*(h_i, a_i) := \{m_i^* \in M_i^* : (h_i, (a_i, m_i^*)) \in \bar{H}_i\}$.

The underlying utility-relevant state influences the probabilities of realization of different profiles of pieces of information, according to the already defined function profile g (cf. Section 2.4). We generalize the definition of ζ to retrieve a function $\eta : S \times \Theta \times \mathcal{T}^K \to [0,1]^H$ that specifies, for each utility-relevant state, the probability of reaching each non-terminal history conditional on such state.[42] Specifically, for each $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$ and $h \in H$, let:

$$\eta(h|s, \theta, \tau^K) := \prod_{k=0}^{L(h)} g_{h^k}(s, \theta, \tau^K)[(a_{k+1}(h), m_{p,k+1}(h), m_{k+1}(h))],$$

where for each $t \in \{0, \ldots, L(h)\}$, h^k is the k-long truncation of h, and $a_{k+1}(h)$, $m_{p,k+1}(h)$, and $m_{k+1}(h)$ are the $(k+1)$-th-stage, action, previous play message, and message of i along h.

However, in general a player does not know what the realized history is, as she only observes personal histories. For each $i \in I$ and $h_i \in \bar{H}_i$, the set of histories *compatible* with the occurrence of h_i is $\bar{H}(h_i) := \{h \in \bar{H} : \exists\, h_{-i} \in \bar{H}_{-i}, h = (h_i, h_{-i})\}$. Define then $(\eta_{h_i} : S \times \Theta \times \mathcal{T}^K \to \Delta(H))_{h_i \in H_i}$ to be such that, for each $h_i \in H_i$, $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$, and $h \in H$:

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

In words, $\eta_{h_i}(h|s, \theta, \tau^K)$ is the probability player i would assign to the "complete" history being h conditional on having observed h_i, if she knew that the underlying utility-relevant state was (s, θ, τ^K). Clearly, such probability may be positive if and only if h is compatible with h_i.

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

$$g_{i,h_i}(s, \theta, \tau^K)[(a_i, m_{i,p}, m_i)] := \sum_{h \in \bar{H}(h_i)} g_{i,h}(s, \theta, \tau^K)[(a_i, m_{i,p}, m_i)] \cdot \eta_{h_i}(h|s, \theta, \tau^K).$$

Given that we will be interested in the probabilities of realization of messages in $m_i^* \in M_i^*$ *after* some action has been played, for each $i \in I$ and $h_i \in H_i$ define $g_{i,h_i}^* : S \times \Theta \times \mathcal{T}^K \to \Delta(M_i^*)$ as $g_{i,h_i}^* := \mathrm{marg}_{M_i^*} \circ g_{i,h_i}$.

For each $i \in I$ and $h_i \in \bar{H}_i$, we define the set of personal external states of i that do not prevent h_i as $S_i(h_i) := \{s_i \in S_i : \forall\, k \in \{0, \ldots, t-1\}, s_i(h_i^k) = a_{i,k+1}\}$, where h_i^k is the k-long truncation of h_i. Moreover, for $a_i \in A_i$, let $S_i(h_i, a_i) := \{s_i \in S_i(h_i) : s_i(h_i) = a_i\}$ be the set of player i's personal external states that allow h_i and that prescribe a_i at h_i.

[42] We try to employ a suggestive notation. Just like ζ was assigning realization probabilities to elements of Z (capital ζ in Greek), η will assign realization probabilities to elements of H (capital η in Greek).

At this point, we can formally describe belief updating of generic player $i \in I$. First, we require that the *chain rule* hold for personal external states: 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 $s_i \in S_i(h_i, a_i)$,[43]

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

Second, we require that the *Bayes rule* hold for anything else, after playing: 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 $F \in \mathcal{B}(S_{-i} \times \Theta \times \mathcal{T}^K_{-i})$,

$$\begin{aligned} &\tau_{i,K+1}(F|h'_i) \cdot \int_{S_{-i}\times\Theta\times\mathcal{T}^K_{-i}} g^*_{i,h_i,s^*_i,\tau^K_i}(\,\cdot\,)[m^*_i] \cdot \big(\mathrm{marg}_{S_{-i}\times\Theta\times\mathcal{T}^K_{-i}}\, \tau_{i,K+1}\big)\big(\mathrm{d}(s_{-i}, \theta, \tau^K_{-i})|h_i\big) \\ &= \int_F g^*_{i,h_i,s^*_i,\tau^K_i}(\,\cdot\,)[m^*_i] \cdot \big(\mathrm{marg}_{S_{-i}\times\Theta\times\mathcal{T}^K_{-i}}\, \tau_{i,K+1}\big)\big(\mathrm{d}(s_{-i}, \theta, \tau^K_{-i})|h_i\big), \end{aligned} \tag{BR-a_i}$$

where s^*_i above is any element of $S_i(h_i, a_i)$, and $h'_i = (h_i, (a_i, m^{)}_i))$.

Definition 7 *Epistemic type τ^∞_i of player $i \in I$ satisfies **correct belief updating** if,* (i) *for for each $h_i \in H_i$, $a_i \in \hat{\mathcal{A}}_i(h_i)$, $h'_i \in \bar{H}_i(h_i)$ and $s_i \in S_i(h_i, a_i)$,* (CR) *holds, and* (ii) *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 $F \in \mathcal{B}(S_{-i} \times \Theta \times \mathcal{T}^K_{-i})$,* (BR-$a_i$) *holds.*

Let $\mathcal{T}^\infty_{i,CBU}$ be the set of epistemic types of player i that satisfy (i) *and* (ii)*, and CBU_i the set of personal states $(s_i, \theta_i, \tau^\infty_i)$ such that $\tau^\infty_i \in \mathcal{T}^\infty_{i,CBU}$.*

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

Example 5 (Buy me an ice-cream, continued) Mom can observe only one personal history of length one, that is, one in which she is inactive and receives uninformative messages – call it h^1_M. Afterwards, Mom observes $(m_{M,p}, m_M) \in \big\{(Y, b), (Y, \neg b), (N, \neg b)\big\}$ – that is, she observes Child's second-stage action and the emotional feedback. Note that we neglect her own action because she is inactive. The histories compatible with h^1_M are (V) and (H), which occur with probability one at (s, θ, τ^K) if and only if $s_c(\varnothing) = V$ and $s_C(\varnothing) = H$, respectively.

Given that Mom is inactive at the second stage, she does not need to update her beliefs about her personal external state with (CR). However, (BR-a_i) applies. Focus for the sake of the example on Mom's beliefs about $F := \{s_C : s_C(\varnothing) = V\} \times \Theta \times \mathcal{T}^1_C$ (i.e., "Child played video-games"), and say that she observes $(Y, \neg b)$. Message $\neg b$ is generated after Child's second stage action, that is, after Mom's personal history. It can be checked that

$$g^*_{M,h^1_M,\tau^1_M}(s, \theta, \tau^1_C)[(Y, \neg b)] = \begin{cases} 1 & \text{if } s_C = H.Y; \\ 1 - q & \text{if } s_C = V.Y; \\ 0 & \text{if } s_C \in \{V.N, H.N\}; \end{cases}$$

where $q = \tau_{C,1}\big(\{s_M : s_M\big((Y, b)\big) = B\}|V\big)$. As a result, the probability with which epistemic type τ^∞_M expects to observe $(Y, \neg b)$ is:

$$\alpha(\tau^\infty_M) + \beta(\tau^\infty_M) := \tau_{M,2}(\{H.Y\}|h^1_M) + \int_{\{V.Y\}\times\Theta\times\mathcal{T}^1_C} (1 - q) \cdot \big(\mathrm{marg}_{S_C\times\Theta\times\mathcal{T}^1_C}\, \tau_{M,2}\big)\big(\mathrm{d}(s_C, \theta, \tau^1_C)|h^1_M\big).$$

[43] Recall that $S_i(h_i, a_i)$ is the set of personal external states of i allowing for h_i and prescribing action a_i at h_i.

Then, if $\alpha(\tau_M^\infty) + \beta(\tau_M^\infty) > 0$, (BR-$a_i$) implies that:

$$\tau_{M,2}\big(F|(Y, \neg b)\big) = \frac{\beta(\tau_M^\infty)}{\alpha(\tau_M^\infty) + \beta(\tau_M^\infty)}.$$

Lastly, a point is worth clarifying. 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, Mom's final blame $\tau_{M,1}(L|z_M)$ (i.e., the probability with which she believes Child lied), which concerns Child's actions, actually arises at the second stage, since her belief about L does not change after her action.Similarly, Child's initial confidence does not change after his first-stage action. ▲

4.4 Optimal planning

Before introducing a notion of optimal planning in the present framework, we discuss a prerequisite condition. Specifically, 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 8 *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 $h_i \in \bar{H}_i$,*

$$\tau_{i,K+1}\left(S \times \{\bar{\theta}_i\} \times \Theta_{-i}|h_i\right) = 1.$$

Let KT_i be the set of personal states where player i knows her personal trait.[44]

Next, we retrieve a *plan* of player i (technically a behavior strategy) from her epistemic type τ_i^∞, denoted as $\sigma(\tau_i^\infty) \in \bigtimes_{h_i \in H_i} \Delta(\hat{\mathcal{A}}_i(h_i))$. It is defined, for each $h_i \in H_i$ and $a_i \in \hat{\mathcal{A}}_i(h_i)$, as[45]

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

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. Put differently, a plan specifies how a player expects herself to behave at each contingency she could observe. Moreover, note that a player's plan coincides with her behavioral predisposition s_i if and only if, for each $h_i \in H_i$, $\sigma(\tau_i^\infty)(s_i(h_i)|h_i) = 1$.

Before defining our notion of optimal planning, we need additional pieces of notation. First, we extend the reasoning carried out to retrieve ζ to allow players to assess the distribution over terminal histories induced by a profile (s, θ, τ^K) conditional on having observed a given

[44] 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.

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

personal history. To do so, first define for each $h_i \in H_i$ the set of terminal histories that may realize after h_i as $Z(h_i) := \{z \in Z : \exists\, h_{-i} \in H_{-i}, (h_i, h_{-i}) \preceq z\}$. Then, define functions $(\zeta_{h_i} : S \times \Theta \times \mathcal{T}^K \to \Delta(Z))_{h_i \in H_i}$, to be, for each $h_i \in H_i$, $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$, and $z \in Z$,

$$\zeta_{h_i}(z|s, \theta, \tau^K) := \prod_{t=L(h_i)}^{L(z)} g_{h^t(z)}(s, \theta, \tau^K)[(a_{t+1}(z), m_{p,t+1}(z), m_{t+1}(z))],$$

if $z \in Z(h_i)$, and $\zeta_{h_i}(z|s, \theta, \tau^K) := 0$ otherwise. In words, $\zeta_{h_i}(z|s, \theta, \tau^K)$ is the probability player i assigns to z after observing h_i, conditional on the utility-relevant state being (s, θ, τ^K).

Second, we define for each $i \in I$ the sequence $(u_{i,h_i} : S \times \Theta \times \mathcal{T}^K \to \mathbb{R})_{h_i \in H_i}$ to be such that, for each $h_i \in H_i$ and $(s, \theta, \tau^K) \in S \times \Theta \times \mathcal{T}^K$, $u_{i,h_i}(s, \theta, \tau^K) := \sum_{z \in Z} v_i(z, \theta, \tau^K)\zeta_{h_i}(z|s, \theta, \tau^K)$. Basically, $u_{i,h_i}(s, \theta, \tau^K)$ is the (objectively) expected utility of player i after h_i when the utility-relevant state is (s, θ, τ^K).[46] Then, it is convenient to retrieve a profile of "local" *decision utility functions* of player i, $(\bar{u}_{i,h_i} : S_i \times \Theta_i \times \mathcal{T}_i^{K+1} \to \mathbb{R})_{h_i \in H_i}$: for each $h_i \in H_i$ and $(s_i, \theta_i, \bar{\tau}_i^{K+1}) \in S_i \times \Theta_i \times \mathcal{T}_i^{K+1}$, let[47]

$$\begin{aligned}&\bar{u}_{i,h_i}(s_i, \theta_i, \bar{\tau}_i^{K+1}) := \mathbb{E}_{s_i, \theta_i, \bar{\tau}_i^{K+1}}[u_{i,h_i}|h_i] \\ &= \int_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K} u_{i,h_i}(s_i, s_{-i}, \theta_i, \theta_{-i}, \bar{\tau}_i^K, \tau_{-i}^K) \operatorname{marg}_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K} \bar{\tau}_{i,K+1}(\mathrm{d}s_{-i}, \mathrm{d}\theta_{-i}, \mathrm{d}\tau_{-i}^K|h_i).\end{aligned}$$

For each $i \in I$, $h_i \in H_i$ and $(s_i, \theta_i, \bar{\tau}_i^{K+1}) \in S_i \times \Theta_i \times \mathcal{T}_i^{K+1}$, we interpret $\bar{u}_{i,h_i}(s_i, \theta_i, \bar{\tau}_i^{K+1})$ as player i's expected utility at h_i when her personal external state is s_i, her personal trait is θ_i, and her epistemic type induces the hierarchical system of beliefs $\bar{\tau}_i^{K+1}$.

Remark 8 For each $i \in I$ and $h_i \in H_i$, $\bar{u}_{i,h_i}$ is continuous.[48]

Third, we introduce the profile $\left(r_{i,h_i} : \Theta_i \times \mathcal{T}_i^\infty \rightrightarrows S_i\right)_{h_i \in H_i}$, where, for each $i \in I$ and $h_i \in H_i$, r_{i,h_i} is the correspondence

$$(\theta_i, \tau_i^\infty) \mapsto \arg\max_{s_i \in S_i} \bar{u}_{i,h_i}(s_i, \theta_i, \tau_i^{K+1}),$$

with τ_i^{K+1} being the system of beliefs of order $K + 1$ induced by τ_i^∞. Personal external states in $r_{i,h_i}(\theta_i, \tau_i^\infty)$ are *optimal at personal history* h_i for a player with personal trait θ_i and epistemic type τ_i^∞. We may write $r_{i,h_i}(\theta_i, \tau_i^{K+1})$ instead of $r_{i,h_i}(\theta_i, \tau_i^\infty)$ for convenience, as beliefs of order higher than $K + 1$ are irrelevant for the maximization of local decision utilities.

With this, we say that an *action* $a_i^* \in \hat{\mathcal{A}}_i(h_i)$ is *optimal at personal history* h_i for a player with trait θ_i and epistemic type τ_i^∞ if it is prescribed by some personal external state that is optimal at h_i given θ_i and τ_i^∞ – that is, if there exists $s_i^* \in r_{i,h_i}(\theta_i, \tau_i^\infty)$ such that $s_i^*(h_i) = a_i^*$. At this point, we can give our definition of optimal planning: a plan is optimal if it assigns positive probability, at each non-terminal personal history, only to optimal actions.

[46] Note that the map u_{i,h_i} (for $i \in I$ and $h_i \in H_i$) is well-defined even for utility-relevant states (s, θ, τ^K) where s prevents h_i from happening. This is because what actually matters is the behavior entailed by s *from* h_i *onward* (cf. the definition of the maps $(\zeta_{h_i})_{i \in I, h_i \in H_i}$ given above).

[47] In the following expression, we report s_i and θ_i as subscripts of the expectation operator to indicate that the expectation of function u_i is taken holding s_i and θ_i fixed.

[48] The map u_i is continuous as per Remark 5 and this implies continuity of $\bar{u}_{i,h_i}$.

Definition 9 *Player i* ***plans optimally*** *at personal state $(s_i, \theta_i, \tau_i^\infty)$ if $(s_i, \theta_i, \tau_i^\infty) \in KT_i$ and, for each $h_i \in H_i$:*

$$\operatorname{supp} \sigma(\tau_i^\infty)(\,\cdot\,|h_i) \subseteq \bigcup_{s_i^* \in r_{i,h_i}(\theta_i, \tau_i^\infty)} \{s_i^*(h_i)\}.$$

Let OP_i be the set of personal states where player i plans optimally.

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

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 $h_C \in \{\varnothing, (H), (V)\}$,

$$\operatorname{proj}_{S_M} \tau_{C,2}(\,\cdot\,|h_C) = \delta_{N.B.N}; \tag{5}$$
$$\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L^C|(Y,\neg b)\big)|h_C\big] = \mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,b)\big)|h_C\big] = \mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L^C|(N,\neg b)\big)|h_C\big] = 1. \tag{6}$$

In words, at each history where he is active, Child is sure that Mom would behave according to $N.B.N$ (i.e., that Mom would buy him the ice-cream only if he says "no" without blushing),[49] and that Mom would be sure he is a liar if and only if he blushes. A derivation of Child's local decision utilities is given in Appendix B.1. Here, we note that:

$$\begin{aligned}
\bar{u}_{C,\varnothing}(H.Y, \theta_C, \tau_C^2) =& \tau_{C,2}\big(\{s_M : s_M\big((Y,\neg b)\big) = B\}|\varnothing\big) - \lambda \mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,\neg b)\big)|\varnothing\big] = 1;\\
\bar{u}_{C,\varnothing}(H.N, \theta_C, \tau_C^2) =& \tau_{C,2}\big(\{s_M : s_M\big((N,\neg b)\big) = B\}|\varnothing\big) - \lambda \mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(N,\neg b)\big)|\varnothing\big] = 0;\\
\bar{u}_{C,\varnothing}(V.Y, \theta_C, \tau_C^2) =& \nu + q\left(\tau_{C,2}\big(\{s_M : s_M\big((Y,\neg b)\big) = B\}|\varnothing\big) - \lambda \mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,\neg b)\big)|\varnothing\big]\right) +\\
&+ (1-q)\left(\tau_{C,2}\big(\{s_M : s_M\big((Y,b)\big) = B\}|\varnothing\big) - \lambda \mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,b)\big)|\varnothing\big]\right)\\
=& \nu - \lambda;\\
\bar{u}_{C,\varnothing}(V.N, \theta_C, \tau_C^2) =& \nu + \tau_{C,2}\big(\{s_M : s_M\big((N,\neg b)\big) = B\}|\varnothing\big) - \lambda \mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(N,\neg b)\big)|\varnothing\big]\big) = \nu.
\end{aligned}$$

Suppose that $\nu > 1$ and $\lambda = 1$. His optimal action at the root is V. Moreover, after (V), action N leads to an expected utility of ν, and action Y to an expected utility of $\nu - 1$: the optimal action after (V) is N. Similarly, we can conclude that the optimal action after (H) is Y. Hence, Child plans optimally if he plans to choose with certainty V at the beginning of the game, N afterwards, and Y in case he has to act after (H) (however, this will not happen if he sticks to his plan). If instead we had $\nu = \lambda = 1$, Child's optimal actions after (H) and (V) would not change, but both H and V would be optimal at the beginning of the game. Then, any plan $\big(\sigma(\,\cdot\,|\varnothing), \sigma(\,\cdot\,|H), \sigma(\,\cdot\,|V)\big) \in \Delta(\{H,V\}) \times \{\delta_Y\} \times \{\delta_N\}$ would be optimal. ▲

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.

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

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 $(s_i, \theta_i, \tau_i^\infty) \in C_i \cap KB_i \cap BR_i \cap OP_i \cap CON_i$. 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 $i \in I$.*

Our notion of rationality deserves some comment. First of all, note that it is richer than the one usually adopted in the literature because we distinguish plans from objective behavior (cf. also Battigalli & 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 requirement is motivated by the observation that, in our setting, players do not commit to personal external states (in fact, they do need not even know their true ones) – rather, in planning how to act, they have to figure out which action to choose after each non-terminal personal history.

Note that in light of Lemma 3, rational (hence, coherent) players are able to formulate beliefs over the set of personal states of opponents. Then, as already stressed, 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 meaningful way – this is a prerequisite to define a theory of strategic thinking (cf. Section 6).

5 Strong Δ-rationalizability

The aim of this section is to 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 restrictions to players' beliefs (see Battigalli & Tebaldi, 2019, Battigalli, Corrao, & Dufwenberg, 2019 and relevant references therein). This in turn builds on earlier concepts of rationalizability for extensive-form games (Pearce, 1984). The epistemic foundations of our solution concept will be thoroughly discussed in Section 6 – for the moment, it is enough to note that it captures the behavioral implications of rationality and forward-induction reasoning. In a nutshell, this way of reasoning posits that players interpret unexpected moves as purposeful choices of their opponents: 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 may be 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:[50]

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

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 $h_i \in H_i$, $F \cap \Omega_{-i,\tau_i^K}^K(h_i) \neq \emptyset$ 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 $(\Omega^0, (\Omega_{-i}^n)_{n=1}^{K-1})$.

Consider the following procedure.[51]

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. $(\tau_i^K, \bar{\tau}_{i,K+1}) \in \operatorname{proj}_{\mathcal{T}_i^{K+1}} \left(\mathcal{T}_{i,C}^{\infty} \cap \Delta_{\theta_i}^{\infty} \right)$;
2. *for each* $h_i \in H_i$, $s_i \in r_{i,h_i}(\theta_i, (\tau_i^K, \bar{\tau}_{i,K+1}))$;
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. However, this procedure is carried out on utility-relevant states, rather than on states of the world.

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) $\mathbf{P}_i^{\Delta}(n) \subseteq \mathbf{P}_i^{\Delta}(n-1)$.

Given the emphasis put on measurability in earlier sections, point (i) of Lemma 9 may seem surprising. However, it is important to clarify that our solution procedure is a means to obtain the behavioral predictions implied by relevant epistemic assumptions about, e.g., players'

[50] 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.

[51] 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.

rationality and (strong) belief in the rationality of others. What is crucial is that the epistemic assumptions that justify our solution concept are events, because such events ultimately pin down players' beliefs and behavior. Section 6 shows that strong Δ-rationalizability captures the utility-relevant implications of a set of meaningful epistemic events.

Thanks to Lemma 9, the limit of the sequence $(\mathbf{P}^{\Delta}(n))_{n\in\mathbb{N}\cup\{0\}}$ is well-defined: we say that a utility-relevant state (s,θ,τ^K) is *strongly Δ-rationalizable* if $(s,\theta,\tau^K)\in\mathbf{P}^{\Delta}(\infty):=\bigcap_{n\in\mathbb{N}\cup\{0\}}\mathbf{P}^{\Delta}(n)$. It is important to note that, without additional assumptions, the set of strongly Δ-rationalizable states may be empty.

A slightly different and more convenient procedure has been proposed in the literature.

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*

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

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

3M. for each $h_i\in H_i$, $\mu_i(S_i(h_i,s_i(h_i))|h_i)=1$;

4M. μ_i *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 defined as "naive" strong Δ-rationalizability (Battigalli & 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). It is straightforward to adapt the proof of Lemma 9 to show the following result.

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) $\mathbf{Q}_i^{\Delta}(n)\subseteq\mathbf{Q}_i^{\Delta}(n-1)$.

In light of Remark 9, $\mathbf{Q}^{\Delta}(\infty):=\bigcap_{n\in\mathbb{N}\cup\{0\}}\mathbf{Q}^{\Delta}(n)$ is meaningfully defined. Thus, it is interesting to verify whether equivalence between the two procedures can be established. Similar results have already been proved in previous works (cf. Battigalli & Prestipino, 2013), and here we prove the equivalence for a special case of belief restrictions. 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. Specifically, 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 $\Delta_{\theta_i}=\bigtimes_{n=1}^{K+1}\bigtimes_{h_i\in\bar{H}_i}B_{\theta_i,n,h_i}$.[52] Conceptually, for each $\theta_i\in\Theta_i$, $n\in\{1,\dots,K+1\}$, and $h_i\in\bar{H}_i$, 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.

[52] Note that, with some abuse, we write Ω_{-i}^0 instead of Ω^0 to ease notation.

Proposition 2 *Consider a rectangular profile of belief restrictions Δ. For each $i \in I$ 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 any belief restrictions – that is, for each $i \in \{C, M\}$ and $\theta_i \in \Theta_i$, $\Delta_{\theta_i} = \mathcal{T}_i^2$ –, and we make the simplifying assumption that $\Lambda = \{1\}$ and that $N = \{\nu', \nu''\}$, with $0 < \nu' < 1 < \nu''$. To keep the exposition simple, we describe the procedure only informally.[53] 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 conditions 1 and 3 of Definition 12, our analysis goes through unchanged if we assume that Child directly chooses among $H.Y$, $H.N$, $V.Y$, and $V.N$ at the root of the game (cf. the discussion in Section B.2).

Step 1 It is possible to check that $V.N$ grants Child a strictly higher expected utility than $H.N$ at the root of the game.[54] Thus, $\text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(1) = \{H.Y, V.Y, V.N\} \times \{\nu', \nu''\}$. As for Mom, by condition 1 of Definition 12 (specifically, by knowledge-implies-belief), she needs to be sure that Child played video-games in the first stage whenever she observes (Y, b). Hence, she is better off not buying him the ice-cream in such case. We conclude that $\text{proj}_{S_M} \mathbf{P}_M(1) = \big\{s_M \in S_M : s_M\big((Y, b)\big) = N\big\}$.

Step 2 Child now realizes that Mom will be sure he lied if she sees him blushing, and that he will not get the ice-cream in such case. Therefore, his image concern makes $V.Y$ strictly worse than $V.N$ – at least, in the former case, he will not be seen as a liar. Moreover, $V.N$ ensures a utility of ν coming from video-games: for trait-type ν'', this is higher than the maximum utility that $H.Y$ may lead to (i.e., the utility of 1 coming from the ice-cream). Hence, $\text{proj}_{S_C \times \Theta_C} \mathbf{P}_C(2) = \big(\{H.Y, V.N\} \times \{\nu'\}\big) \cup \big(\{V.N\} \times \{\nu''\}\big)$. On the other hand, Mom's strong belief in $\mathbf{P}_C(1)$ leads her to conclude that, if she observes $(N, \neg b)$, it must be the case that Child played video-games in the first stage – indeed, the only personal external state of Child that survived the first step and that prescribes playing N at the second stage is $V.N$. Thus, upon observing $(N, \neg b)$, she is sure that Child did not do his homework: she will not buy him the ice-cream in such case. We obtain $\text{proj}_{S_M} \mathbf{P}_M(2) = \big\{N.N.N, N.B.N\big\}$ – that is, Mom knows for sure that she will not buy Child an ice-cream if she observes $(N, \neg b)$ or (Y, b).

Step 3 This step has no behavioral implications for Child, because trait-type ν' is not sure of Mom's behavior after $(Y, \neg b)$, so both $H.Y$ and $V.N$ can be optimal for some belief (e.g., the latter is optimal if he is sure that Mom would not buy him the ice-cream also if she observes $(Y, \neg b)$). Mom instead concludes, by strong belief in $\mathbf{P}_C(2)$, that

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

[54] Intuitively, if he correctly expects to play N in the second stage, he would be sure not to blush after his report. Then, his expectation about Mom's behavior (which he knows to depend on the fact that she observes personal history $(N, \neg b)$) will be exactly the same regardless of whether he plays $H.N$ or $V.N$, as they both give rise to Mom's personal history $(N, \neg b)$ – thus, playing video-games allows him to unambiguously increase his utility.

personal history $(Y, \neg b)$ realizes if and only if Child did his homework. Upon observing such personal history, she would be sure that he behaved well, and she would be glad to buy him an ice-cream. Thus, $\mathrm{proj}_{S_M} \mathbf{P}_M(3) = \big\{N.B.N\big\}$.

Step 4 At this point, by strong belief in $\mathbf{P}_M(3)$, Child is sure that Mom will buy him an ice-cream if she observes $(Y, \neg b)$ – in other words, $H.Y$ allows to secure the ice-cream without being blamed. Thus, trait-type ν' finds it optimal to play according to $H.Y$, as the valuation he attaches to playing video-games (i.e., ν') is lower than that of the ice-cream (i.e., 1). Hence, $\mathrm{proj}_{S_C \times \Theta_C} \mathbf{P}_C(4) = \{(H.Y, \nu'), (V.N, \nu'')\}$.

Subsequent steps of the procedure do not yield further implications, and we conclude that:

$$\mathrm{proj}_{S_C \times \Theta_C} \mathbf{P}_C(\infty) = \big\{(H.Y, \nu'), (V.N, \nu'')\big\}, \quad \mathrm{proj}_{S_M} \mathbf{P}_M(\infty) = \{N.B.N\}.$$

This result shows that the possibility of betraying a lie through an emotional signal provides Child with a strong enough incentive to tell the truth. This is a "full disclosure" result: Child privately chooses an action according to his appreciation for video-games, and reveals it to Mom, who can in turn fully believe him. Such result seems interesting, as we believe that this basic structure of interaction can be applied also to other situations – specifically, whenever (*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. However, even with standard preferences, our insights would still apply if player 2 could enforce a punishment. In such case, emotional feedback uncovers an emotion that does not matter directly for players' utilities, but that allows to make crucial inferences about the truthfulness of some statement: this could apply, e.g., to Example 4. In other words, player 2's report would be enriched by an emotional component in such setting. This makes our framework well-suited for the analysis of information transmission in situations where factors like facial mimicry are crucial: we can think for example at politicians delivering speeches, salesmen advertising their products, or individuals engaging in face-to-face bargaining (cf. Section 7). ▲

6 Epistemic justification of strong Δ-rationalizability

The aim of this section is to discuss the epistemic foundations of the solution procedure defined in Section 5. That is, we show that the proposed procedure captures the utility-relevant implications of some meaningful epistemic assumptions (namely, players' rationality and strong belief in rationality, as well as common strong (correct) belief in the restrictions described by Δ).[55] 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

[55] Battigalli and Siniscalchi (2002) provide an epistemic justification of strong rationalizability, therefore neglecting restrictions on players' beliefs. Battigalli and Tebaldi (2019) and Battigalli et al. (2020) extend the analysis to a class of infinite games, and to psychological games, respectively. For an epistemic foundation of strong Δ-rationalizability, see Battigalli and Prestipino (2013).

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[56] and the *strong belief operator*, as:

$$\begin{aligned}
&\mathrm{B}_{i,h_i}(F_{-i}) := \big\{(s_i, \theta_i, \tau_i^{\infty}) \in C_i : \varphi_i(\tau_i^{\infty})(F|h_i) = 1\big\};\\
&\mathrm{SB}_i(F_{-i}) := \big\{(s_i, \theta_i, \tau_i^{\infty}) \in C_i : \forall\, h_i \in H_i,\ \Omega_{-i,\tau_i^K}^{\infty}(h_i) \cap F_{-i} \neq \emptyset \Longrightarrow \varphi_i(\tau_i^{\infty})(F_{-i}|h_i) = 1\big\}.
\end{aligned}$$

The following result establishes that, 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 each* $i \in I$, $h_i \in \bar{H}_i$, *and* $F_{-i} \in \mathcal{B}(S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^{\infty})$.

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

Remark 10 For each $i \in I$, D_i is measurable.[57]

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

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

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 instead maintains whenever possible that her opponents are first-order strategically sophisticated, on top of being rational herself. A third-order strategically sophisticated player instead 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

[56] We use the term "belief operator", but to be precise we should talk about "probability-one belief", or "conditional (on observing h_i) certainty".

[57] The remark follows from the fact that D_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.

her opponents are only first-order strategically sophisticated. The reasoning can be generalized, but the bottom line is that, under our epistemic assumptions, players ascribe to opponents the highest level of strategic sophistication consistent with evidence.

The following is straightforward.

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 $n \in \mathbb{N}$.[58]

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 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 $n \in \mathbb{N}$.*

Therefore, the steps of the procedure described by Definition 12 capture the utility-relevant implications of rationality, of the belief restrictions, and of mutual strong belief of a finite order in rationality and in the belief restrictions. The limit of the strong Δ-rationalizability procedure instead identifies the utility-relevant implications of rationality, belief restrictions Δ, and common strong belief in both these features. In light of Theorem 1, the proposed procedure is convenient because it allows to focus on beliefs of order up to $K+1$, whereas the epistemic assumptions we formalized in the present sections involve infinite hierarchies of systems of beliefs.

7 Conclusion

In this paper, we introduced a novel framework to incorporate noisy emotional feedback into games, that may effectively be adapted to relevant applications, as we sketched out in the introduction. On the one hand, it is possible to formalize testable theoretical predictions about the extent to which the appraisal of others' emotion affects choices, to be validated experimentally (cf. Examples 1, 2, and 3). On the other hand, our framework is well suited to analyze an important economic problem such as information transmission. When communication occurs during face-to-face interactions, misreports may be betrayed by emotional signals, and this could shape incentives in interesting ways, that would not be captured by standard models (cf. Examples 4 and 5). In this respect, our framework can reasonably be applied to relevant settings such as court hearings, presidential debates, political speeches, bargaining, product advertisement by salesmen, and physician-patient interactions (cf. p. 34).

[58] 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 induction. As for the basis step, note that $\mathbf{R}_i^{\Delta}(1) = R_i \cap D_i$, and both R_i and D_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 $k \in \{1,\dots,n\}$, 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.

Our framework naturally calls for applied models, but 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. Perhaps even more interestingly, our language is rich enough to model situations in which players may reason about cognitive failures of opponents. We believe such expressiveness to be a key step toward a rigorous analysis of the implications of failures of rationality in strategic settings. In this regard, future research may consist in capturing the utility-relevant implications of different sets of assumptions about players' cognitive and behavioral features.

Second, our previous-play-message approach in the definition of the game tree gives a transparent description of the flow of information that accrues to players according to the rules of the game. Modeling game-specific information as a flow rather than as a stock allows to explicitly remove any dependence of the game form on the cognitive ability of players, as discussed in detail by Battigalli and Generoso (2021). The traditional choice of describing players' information by means of information sets on the other hand needs the implicit assumption that players recall *all* the information they received for them to be able to play the game (hence, for the rules of the game to be meaningfully defined).

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. 20)

Fix $h = (h_i)_{i\in I} \in H$, $(s,\theta) \in S \times \Theta$ and $\bar{m} \in M$. Recall that we can write $\bar{m} = ((\bar{m}_{i,j})_{i\in I})_{j\in I\setminus\{i\}}$ (cf. footnote 5), where $m_{i,j}$ is a message i observes about j. Then, to ease notation, let $\bar{\ell}_i = (\bar{m}_{j,i})_{j\in I\setminus\{i\}}$ – in words, $\bar{\ell}_i$ is i's emotional leakage (i.e., the profile of messages about i received by her opponents) implied by $\bar{m}$. Note that $\bar{\ell}_i$ belongs to the set $L_i := \bigtimes_{j\in I\setminus\{i\}} M_{j,i}$, and that $\bar{m} = (\bar{\ell}_i)_{i\in I}$.

Consider now $\{(s,\theta)\} \times \{\tau^1 \in \mathcal{T}^1 : \bar{m} \in \operatorname{supp} f_h(s,\theta,\tau^1)\}$, where we let $K = 1$ because feedback is simple (cf. point (i) of Definition 3). It is possible to check that:

$$\{\tau^1 \in \mathcal{T}^1 : \bar{m} \in \operatorname{supp} f_h(s,\theta,\tau^1)\} = \bigcap_{i\in I} \big\{\tau^1 \in \mathcal{T}^1 : \bar{\ell}_i \in \operatorname{supp}(\operatorname{marg}_{L_i} f_h(s,\theta,\tau^1))\big\}. \qquad (7)$$

Simplicity of feedback implies that, for each $i \in I$, $\big\{\tau^1 \in \mathcal{T}^1 : \bar{\ell}_i \in \operatorname{supp}(\operatorname{marg}_{L_i} f_h(s,\theta,\tau^1))\big\}$

depends exclusively on $\tau_i^1(\,\cdot\,|h_i)$ (cf. point (ii) of Definition 3). 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 $i \in I$,

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

Then, expression (7) can be rewritten as

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

which is a rectangle. However, $\{\tau^1 \in \mathcal{T}^1 : \bar{m} \in \operatorname{supp} f_h(s,\theta,\tau^1)\}$ 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, $\{\tau^1 \in \mathcal{T}^1 : \bar{m} \in \operatorname{supp} f_h(s,\theta,\tau^1)\}$ is a measurable rectangle, proving regularity. ■

Proof of Lemma 1 (p. 21)

We focus on correspondences $(\mathbf{A}^t)^\ell : 2^{\operatorname{proj}_{A^t} \bar{H}} \to 2^{S\times\Theta\times\mathcal{T}^K}$ and $(\mathbf{H}_i^t)^\ell : 2^{\bar{H}_i^t} \to 2^{S\times\Theta\times\mathcal{T}^K}$, and we show that, for each $t \in \{1,\dots,T\}$, $a^t \in \operatorname{proj}_{A^t}\bar{H}$, and $h_i^t \in \bar{H}_i^t$, $(\mathbf{A}^t)^\ell(a^t) \in \mathcal{B}(S\times\Theta\times\mathcal{T}^K)$ and $(\mathbf{H}_i^t)^\ell(h_i^t) \in \mathcal{B}(S\times\Theta\times\mathcal{T}^K)$. We proceed by induction.

Consider first $t = 1$, fix $a^1 \in \operatorname{proj}_{A^1}\bar{H}$ and . We have:

$$(\mathbf{A}^1)^\ell(a^1) = \{(s,\theta,\tau^K) : a^1 \in \mathbf{A}^1(s,\theta,\tau^K)\} = \{s \in S : (s_i(\varnothing))_{i\in I} = a^1\} \times \Theta \times \mathcal{T}^K,$$

which is clearly measurable as $\{s \in S : (s_i(\varnothing)) = a^1\}$ is measurable in the discrete σ-algebra of S (indeed, any subset of S is measurable), and $\Theta\times\mathcal{T}^K$ is trivially measurable. The same holds for each $a^1 \in \operatorname{proj}_{A^1}\bar{H}$.

Next, fix $i \in I$ and $h_i^1 \in \bar{H}_i^1$. We have:

$$\begin{aligned}&(\mathbf{H}_i^t)^\ell(h_i^1) = \{(s,\theta,\tau^K) : h_i^1 = (a_i, m_{i,p}, m_i) \in \mathbf{H}_i^1(s,\theta,\tau^K)\}\\ &=\{(s,\theta,\tau^K) :\ (1)\ a_i^1 = s_i(\varnothing),\ (2)\ m_{i,p} = p_i(\mathbf{A}^1(s,\theta,\tau^K)),\ (3)\ m_i \in \operatorname{supp}(f_{i,\varnothing}(\mathbf{A}^1(s,\theta,\tau^K),\theta,\tau^K))\},\end{aligned}$$

where recall that for simplicity we write p_i to denote player i's previous play messages, as generated by function p. We can write such set as the intersection of the family of sets $\{G_i \subseteq S\times\Theta\times\mathcal{T}^K : (i) \text{ holds}\}_{i=1}^3$. Clearly, G_1 is measurable because it coincides with $(\mathbf{A}^1)^\ell(a)$, which we already showed to be measurable. G_2 is measurable as well, as it can be rewritten as $T\times\Theta\times\mathcal{T}^K$ for some $T \subseteq S$ (i.e., it is a product of measurable sets, hence it is measurable). Lastly, G_3 is measurable by semi-regularity (cf. Definition 4). Thus, $(\mathbf{H}_i^1)^\ell(h_i^1)$ is measurable, and the same holds for each $h_i^1 \in \bar{H}_i^1$

Suppose now that, for each $i \in I$, $n \in \{1,\dots,t-1\}$ (with $t \leq T$), $a^n \in \operatorname{proj}_{A^n}\bar{H}$, and $h_i^n \in \bar{H}_i^n$, $(\mathbf{A}^n)^\ell(a^n)$ and $(\mathbf{H}_i^n)^\ell(h_i^n)$ are measurable. Fix $a^t \in \operatorname{proj}_{A^t}\bar{H}$ and consider:

$$\begin{aligned}(\mathbf{A}^t)^\ell(a^t) =& \Big\{(s,\theta,\tau^K) : a^t = (a^{t-1},(a_i)_{i\in I}) \in \mathbf{A}^t(s,\theta,\tau^K)\Big\}\\ =& \Big\{(s,\theta,\tau^K) : (1)\ a^{t-1} \in \mathbf{A}^{t-1}(s,\theta,\tau^K),\end{aligned}$$

$$(2)\ \forall\, i \in I, a_i \in \bigcup_{\substack{(m_{i,p}^{t-1}, m_i^{t-1}):\\ (a_i^{t-1}, m_{i,p}^{t-1}, m_i^{t-1}) \in \mathbf{H}_i^{t-1}(s,\theta,\tau^K)}} \{s_i(a_i^{t-1}, m_{i,p}^{t-1}, m_i^{t-1})\}\Big\}.$$

Again, define the family of sets $\{G_i \subseteq S \times \Theta \times \mathcal{T}^K : (i) \text{ holds}\}_{i=1,2}$. G_1 is measurable by our inductive hypothesis. As for G_2, we rewrite it as:

$$\begin{aligned}G_2 =& \bigcap_{i\in I}\Big\{(s,\theta,\tau^K) : \exists\, (m_{i,p}^{t-1}, m_i^{t-1}) \in M_{i,p}^{t-1} \times M_i^{t-1},\\ & (a_i^{t-1}, m_{i,p}^{t-1}, m_i^{t-1}) \in \mathbf{H}_i^{t-1}(s,\theta,\tau^K),\ a_i = s_i(a_i^{t-1}, m_{i,p}^{t-1}, m^{t-1})\}\\ =& \bigcap_{i\in I} \bigcup_{(m_{i,p}^{t-1}, m_i^{t-1}) \in M_{i,p}^{t-1} \times M_i^{t-1}} \Big(\big\{(s,\theta,\tau^K) : (a_i^{t-1}, m_{i,p}^{t-1}, m_i^{t-1}) \in \mathbf{H}_i^{t-1}(s,\theta,\tau^K)\big\}\\ & \cap \big\{(s,\theta,\tau^K) : a_i = s_i(a_i^{t-1}, m_{i,p}^{t-1}, m_i^{t-1})\big\}\Big).\end{aligned}$$

At this point, note that the first set intersected within parentheses is measurable by our inductive hypothesis. The second one can be written as $T \times \Theta \times \mathcal{T}^K$ for some $T \subseteq S$, and hence it is measurable. We conclude that the intersection between parentheses in the last formula is measurable. The intersection over I and the union over $M_{i,p}^{t-1} \times M_i^{t-1}$ are finite. Hence, G_2 is measurable, and so is $(\mathbf{A}^t)^\ell(a^t)$. The same holds for each $a^t \in \operatorname{proj}_{A^t} H$.

We now turn to $(\mathbf{H}_i^t)^\ell$. Fix $h_i^t = (\bar{a}_i^{t-1}, \bar{m}_{i,p}^{t-1}, \bar{m}_i^{t-1}, a_{i,t}, m_{i,p,t}, m_{i,t}) \in H_i^t$ and consider:

$$\begin{aligned}(\mathbf{H}_i^t)^\ell(h_i^t) =& \big\{(s,\theta,\tau^K) : h_i^t \in \mathbf{H}_i^t(s,\theta,\tau^K)\big\}\\ =& \Big\{(s,\theta,\tau^K) : (1)\ (\bar{a}_i^{t-1}, \bar{m}_{i,p}^{t-1}, \bar{m}_i^{t-1}) \in \mathbf{H}_i^{t-1}(s,\theta,\tau^K),\ (2)\ a_{i,t} = s_i(\bar{a}_i^{t-1}, \bar{m}_{i,p}^{t-1}, \bar{m}_i^{t-1}),\\ & (3)\ m_{i,p,t} \in \bigcup_{a_{-i}^t : (a_i^t, a_{-i}^t) \in \mathbf{A}^t(s,\theta,\tau^K)} \{p_i(a_i^t, a_{-i}^t)\}\\ & (4)\ m_{i,t} \in \bigcup_{\substack{h_{-i}^{t-1} : h^{t-1} =\\ = (h_i^{t-1}, h_{-i}^{t-1}) \in (\mathbf{H}_i^{t-1}(s,\theta,\tau^K))_{i\in I}}} \operatorname{supp} f_{i,h^{t-1}}(s,\theta,\tau^K)\Big\}.\end{aligned}$$

Let $\{G_i\}_{i=1}^4$ be defined as usual. G_1 is measurable by our inductive hypothesis, and G_2 is measurable because it can be written as $T \times \Theta \times \mathcal{T}^K$ for some $T \subseteq S$. As for G_3, we write it as:

$$\begin{aligned}G_3 =& \big\{(s,\theta,\tau^K) : \exists\, a_{-i}^t \in A_{-i}^t,\ (a_i^t, a_{-i}^t) \in \mathbf{A}^t(s,\theta,\tau^K),\ m_{i,p,t} = p_i(a_i^t, a_{-i}^t)\big\}\\ =& \bigcup_{a_{-i}^t \in A_{-i}^t} \Big(\big\{(s,\theta,\tau^K) : (a_i^t, a_{-i}^t) \in \mathbf{A}^t(s,\theta,\tau^K)\big\} \cap\\ & \cap \big\{(s,\theta,\tau^K) : m_{i,p,t} = p_i(a_i^t, a_{-i}^t)\big\}\Big),\end{aligned}$$

and we note that the first intersected set within parentheses is $(\mathbf{A}^t)^\ell(a_i^t, a_{-i}^t)$, which we showed to be measurable. The second set is trivially measurable, because it is either $S \times \Theta \times \mathcal{T}^K$ or the empty set. The union over A_{-i}^t is finite, and thus G_3 is measurable. As for G_4, we rewrite it as:

$$G_4 = \big\{(s,\theta,\tau^K) : \exists\, h_{-i}^{t-1} \in \bar{H}_i^{t-1},\ (h_i^{t-1}, h_{-i}^{t-1}) \in (\mathbf{H}_j^{t-1}(s,\theta,\tau^K))_{j\in I},\ m_{i,t} \in \operatorname{supp} f_{i,h^{t-1}}(s,\theta,\tau^K)\big\}$$

$$= \bigcup_{h_{-i}^{t-1} \in \bar{H}_{-i}^{t-1}} \Bigg(\big\{(s,\theta,\tau^K) : (h_i^{t-1}, h_{-i}^{t-1}) \in (\mathbf{H}_j^{t-1}(s,\theta,\tau^K))_{j\in I}\big\}$$

$$\cap \big\{(s,\theta,\tau^K) : m_{i,t} \in \operatorname{supp} f_{i,h^{t-1}}(s,\theta,\tau^K)\big\}\Bigg).$$

Of the sets intersected within parentheses, the first one can be checked to be measurable by our inductive hypothesis, and the second one is measurable by semi-regularity of feedback. The union over H_{-i}^{t-1} is finite, and G_4 is therefore measurable. Wrapping up, we conclude that $(\mathbf{H}_i^t)^\ell(h_i^t)$ is measurable, and the same holds for each $h_i^t \in \bar{H}_i^t$.

All in all, we proved that $(\mathbf{A}^t)^\ell(a^t)$ and $(\mathbf{H}_i^t)^\ell(h_i^t)$ are measurable for each $t \in \{1,\dots,T\}$, $a^t \in \operatorname{proj}_{A^t} \bar{H}$, and $h_i^t \in \bar{H}_i^t$. Then, measurability of the correspondences $\mathbf{A}^t$ and $\mathbf{H}_i^t$ (see footnote 33 for a definition) follows straightforwardly for each $t \in \{1,\dots,T\}$, as any closed subset of the finite sets $\operatorname{proj}_{A^t} \bar{H}$ and $\bar{H}_i^t$ (with $t \in \{1,\dots,T\}$) can be written as a finite union of singletons – which are closed in the discrete topology –, so that $(\mathbf{A}^t)^\ell(P) = \bigcup_{a^t\in P}(\mathbf{A}^t)^\ell(a^t) \in \mathcal{B}(S\times\Theta\times\mathcal{T}^K)$ and $(\mathbf{H}_i^t)^\ell(Q) = \bigcup_{h_i^t\in Q}(\mathbf{H}_i^t)^\ell(h_i^t) \in \mathcal{B}(S\times\Theta\times\mathcal{T}^K)$ for each closed $P \subseteq \operatorname{proj}_{A^t} \bar{H}$ and $Q \subseteq \bar{H}_i^t$. ■

Proof of Lemma 2 (p. 23)

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

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

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. 23)

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 $\varphi_i := (\varphi_{h_i})_{h_i \in H_i} : \mathcal{T}_{i,C}^{\infty} \to \big[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\big]^{H_i}$. We want to show that it is indeed an homeomorphism.[59]

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 $\varphi_i(\mathcal{T}_{i,C}^{\infty}) = \big[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\big]^{H_i}$. Indeed, $\varphi_i(\mathcal{T}_{i,C}^{\infty}) \subseteq \big[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\big]^{H_i}$ holds by definition. To see that $\big[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\big]^{H_i} \subseteq \varphi_i(\mathcal{T}_{i,C}^{\infty})$, take $t_i \in \big[\Delta(\Omega^0 \times \mathcal{T}_{-i}^{\infty})\big]^{H_i}$ 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. 23)

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

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

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

$$\Omega_{-i,\tau_i^K}^K(h_i^t) = \bigcup_{(s,\theta)\in S\times\Theta} \left(\Omega_{-i,\tau_i^K,s,\theta}^K(h_i^t)\right) = \bigcup_{(s,\theta)\in S\times\Theta} \left(\{s\} \times \{\theta\} \times (\mathbf{H}_{i,\tau_i^K,s,\theta}^t)^{\ell}(h_i^t)\right). \tag{8}$$

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

$$\begin{aligned}(\mathbf{H}_{i,\tau_i^K,s,\theta}^t)^{\ell}(h_i^t) := \Bigg\{ \tau_{-i}^K \in \mathcal{T}_{-i}^K : \forall\, k \in \{1,\dots,t\},\ &(1,k)\ a_{i,k} = s_i(h_i^{k-1}),\\ &(2,k)\ m_{i,p,k} \in \bigcup_{a_{-i}^{k-1}:(a_i^{k-1},a_{-i}^{k-1})\in\mathbf{A}^{k-1}(s,\theta,\tau^K)} \{p_i(a_i^{k-1}, a_{-i}^{k-1})\},\\ &(3,k)\ m_{i,k} \in \bigcup_{\substack{h_{-i}^{k-1}:(h_i^{k-1},h_{-i}^{k-1})\\ \in\mathbf{H}^{k-1}(s,\theta,\tau^K)}} \operatorname{supp} f_{i,(h_i^{k-1},h_{-i}^{k-1})}(s,\theta,\tau^K)\Bigg\}.\end{aligned}$$

Sticking to the notation introduced in the proof of Lemma 1, define the collection of sets $\{\{G_{i,k} \subseteq \mathcal{T}_{-i}^K : (i,k) \text{ holds}\}_{i=1}^3\}_{k=1}^t$.

Note that $G_{1,k}$ is independent from players' hierarchical systems of beliefs for each $k \in \{1,\dots,t\}$. On the other hand, it is easy to check that, for each $k \in \{1,\dots,t\}$, $G_{2,k} \cap G_{3,k}$ belongs to following collection:

$$\begin{aligned}&\big\{\tau_{-i} \in \mathcal{T}_{-i}^K : m_{i,p,k} = p_i(a_i^{k-1}, a_{-i}^{k-1}),\\ &\quad m_{i,k} \in \operatorname{supp} f_{i,(h_i^{k-1},h_{-i}^{k-1})}\big((a_{i,k}, a_{-i}), \theta, (\tau_i^K, \tau_{-i}^K)\big)\big\}_{a_{-i}^{k-1}\in A_{-i}^{k-1},\, h_{-i}^{k-1}\in \bar{H}_{-i}^{k-1}},\end{aligned}$$

which is easily seen to be 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).

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

Thus, since $(\mathbf{H}^t_{i,\tau_i^K,s,\theta})^\ell(h_i^t) = \bigcap_{k=1}^t \bigcap_{i=1}^3 G_{i,k}$, the foregoing argument allows us to conclude that the collection $\{(\mathbf{H}^t_{i,\tau_i^K,s,\theta})^\ell(h_i^t)\}_{\tau_i^K \in \mathcal{T}_i^K}$ is finite. With equation (8) 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 $t \in \{1, \ldots, T\}$, $i \in I$, $(s, \theta) \in S \times \Theta$, $h_i^t \in \bar{H}_i$, and $a^t \in \mathrm{proj}_{A^t} \bar{H}$, $(\mathbf{A}^t_{i,s,\theta})^\ell(h_i^t)$ and $(\mathbf{H}^t_{i,s,\theta})^\ell(h_i^t)$ are unions 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 $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$.*[60] *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, 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

[60] 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.

$i^* \in D(\alpha)$ 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, $[\tau^K_i]_{h_i}$ is measurable for each $i \in I$ and $h_i \in \bar{H}_i$.*

Proof of Lemma A5. Fix generic $i \in I$, $h^t_i \in \bar{H}^t_i$, and $\bar{\tau}^K_i \in \mathcal{T}^K_i$, and note that, for each $\tau^K_i \in \mathcal{T}^K_i$,

$$\Omega^K_{-i,\tau^K_i}(h^t_i) = \bigcup_{(s,\theta)\in S\times\Theta} \left(\Omega^K_{-i,\tau^K_i,s,\theta}(h^t_i)\right),$$

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

$$\Omega^K_{-i,\tau^K_i,s,\theta}(h^t_i) = \{s\} \times \{\theta\} \times (\mathbf{H}^t_{i,\tau^K_i,s,\theta})^\ell(h^t_i),$$

where we use subscripts to denote sections of the correspondence $\mathbf{H}^t_i$. Then, we can say that, for each $\tau^K_i \in \mathcal{T}^K_i$, $\Omega^K_{-i,\tau^K_i,s,\theta}(h^t_i) = \Omega^K_{-i,\bar{\tau}^K_i,s,\theta}(h^t_i)$ if and only if $(\mathbf{H}^t_{i,\tau^K_i,s,\theta})^\ell(h^t_i) = (\mathbf{H}^t_{i,\bar{\tau}^K_i,s,\theta})^\ell(h^t_i)$.

Note that for each $i \in I$, $\tau_i \in \mathcal{T}^K_i$ and $h^t_i \in \bar{H}_i$ we can write $(\mathbf{H}^t_{i,s,\theta})^\ell(h^t_i)$ as:

$$\begin{aligned}(\mathbf{H}^t_{i,s,\theta})^\ell(h^t_i) = \Big\{ &\tau^K \in \mathcal{T}^K : \forall\, k \in \{1,\dots,t\},\ (1,k)\ a_{i,k} = s_i(h^{k-1}_i),\\ &(2,k)\ \exists\, a^k_{-i} \in \mathbf{A}^k_{-i,s,\theta}(\tau^K_{-i}), m_{i,p,t} = p_i(a^t_i, a^t_{-i}),\\ &(3,k)\ \exists\, h^{k-1}_{-i} \in \mathbf{H}^{k-1}_{-i,s,\theta}(\tau^K_{-i}), m_{i,k} \in \operatorname{supp} f_{i,(h^{k-1}_i,h^{k-1}_{-i})}(s,\theta,\tau^K)\Big\}.\end{aligned}$$

As we did in the proof of Lemma 1, define $\{\{G_{j,k} \subseteq \mathcal{T}^K : (j,k) \text{ holds}\}^3_{j=1}\}^t_{k=1}$ and $G^* := \bigcap^t_{k=1} \bigcap^3_{j=1} G_{j,k}$, and let $G^*_{\tau^K_i}$ denote the section of G^* at a generic $\tau^K_i \in \mathcal{T}^K_i$. With this, we observe that, for each $\tau^K_i \in \mathcal{T}^K_i$, $(\mathbf{H}^t_{i,\tau^K_i,s,\theta})^\ell(h^t_i) = (\mathbf{H}^t_{i,\bar{\tau}^K_i,s,\theta})^\ell(h^t_i)$ if and only if $G^*_{\tau^K_i} = G^*_{\bar{\tau}^K_i}$. Next, note that $G_{1,k}$ is either empty or equal to $\mathcal{T}^K$ for each $k \in \{1,\dots,t\}$. On the other hand, $G_{2,k}$ and $G_{3,k}$ are (finite) unions of measurable rectangles as per Lemma A3. Hence, both $\bigcap_{k=1}^{t}\bigcap_{j=1}^{3} G_{j,k}$ and $\bigcap_{k=1}^{t}\bigcap_{j=1}^{3}(G_{j,k}(\bar{\tau}_i^K))$ are (finite) intersections 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:

$$\begin{aligned}
\mathcal{T}_{i,KB}^{\infty} &= \left\{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \forall\, h_i \in \bar{H}_i, \tau_{i,K+1}\left(\Omega^K_{-i,\tau_i^K}(h_i)\middle|h_i\right) = 1\right\}\\
&= \bigcap_{h_i\in\bar{H}_i}\left\{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \exists\, [\bar{\tau}_i^K]_{h_i} \subseteq \mathcal{T}_i^K, \tau_i^K \in [\bar{\tau}_i^K]_{h_i}, \tau_{i,K+1}\left(\Omega^K_{-i,[\bar{\tau}_i^K]_{h_i}}(h_i)\middle|h_i\right) = 1\right\}\\
&= \bigcap_{h_i\in\bar{H}_i}\ \bigcup_{[\bar{\tau}_i^K]_{h_i}} \left(\left([\bar{\tau}_i^K]_{h_i} \times \bigtimes_{k\geq K+1} \mathcal{T}_{i,k}\right) \cap \left\{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \tau_{i,K+1}\left(\Omega^K_{-i,[\bar{\tau}_i^K]_{h_i}}(h_i)\middle|h_i\right) = 1\right\}\right). \qquad (9)
\end{aligned}$$

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 $\{\tau_{i,K+1}(\,\cdot\,|h_i) \in \Delta(\Omega^K_{-i}) : \tau_{i,K+1}(\Omega^K_{-i,\tau_i^K}(h_i)|h_i) = 1\}$ is measurable as per Remark A1. Then, the intersection and the union of equation (9) are countable (in particular, Corollary A1 ensures that the union over equivalence classes is finite). All in all, we conclude that $\mathcal{T}_{i,KB}^{\infty}$ can be written as the countable intersection and union of measurable sets, hence it is measurable. $KB_i = S_i \times \Theta_i \times \mathcal{T}_{i,KB}^{\infty}$ is measurable as well, and the same is true for each $i \in I$. ■

Proof of Lemma 5 (p. 26)

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

$$\mathcal{T}_{i,CR}^{\infty} := \bigcap_{h_i \in H_i} \bigcap_{a_i \in \hat{\mathcal{A}}_i(h_i)} \bigcap_{h_i' \in \bar{H}_i(h_i)} \bigcap_{s_i \in S_i(h_i,a_i)} \{\tau_i^{\infty} \in \mathcal{T}_i^{\infty} : \text{(CR) holds}\}; \tag{10}$$

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

Note that $\mathcal{T}_{i,CBU}^{\infty} = \mathcal{T}_{i,CR}^{\infty} \cap \mathcal{T}_{i,BR}^{\infty}$. 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 (10):

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

Note that the intersections in (10) 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 (10) 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. 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 & 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:

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

$$= \int_B g^*_{i,h_i,s_i^*,\tau_i^K}(\,\cdot\,)[m_i^*] \cdot \big(\operatorname{marg}\tau_{i,K+1}\big)\big(\mathrm{d}(s_{-i},\theta,\tau_{-i}^K)|h_i\big)\Big\}, \tag{11}$$

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

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

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 $g^*_{i,h_i,s_i^*,\tau_i^K}(\,\cdot\,)[m_i^*] : S_{-i}\times\Theta\times\mathcal{T}_{-i}^K \to [0,1]$ 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_2 be the map

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

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 (11) 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 $\mathcal{T}_{i,BR}^\infty$.

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

Proof of Lemma 6 (p. 29)

We fix $i \in I$ and we start by rewriting:

$$\begin{aligned} OP_i =& \left\{ (s_i, \theta_i, \tau_i^\infty) \in KT_i : \forall\, h_i \in H_i, \operatorname{supp} \sigma(\tau_i^\infty)(\,\cdot\,|h_i) \subseteq \bigcup_{s_i^* \in r_{i,h_i}(\theta_i, \tau_i^\infty)} \{s_i^*(h_i)\} \right\} \\ =& S_i \times \left\{ (\theta_i, \tau_i^\infty) \in \operatorname{proj}_{\Theta_i \times \mathcal{T}_{i,C}^\infty} KT_i : \forall\, h_i \in H_i, \operatorname{supp} \sigma(\tau_i^\infty)(\,\cdot\,|h_i) \subseteq \bigcup_{s_i^* \in r_{i,h_i}(\theta_i, \tau_i^\infty)} \{s_i^*(h_i)\} \right\} \\ =&:\, S_i \times \widetilde{OP}_i. \end{aligned}$$

Consider a sequence $(\theta_{i,n}, \tau_{i,n}^\infty)_{n\in\mathbb{N}}$ of elements of $\widetilde{OP}_i$ converging to $(\bar{\theta}_i, \bar{\tau}_i^\infty)$. First, we have that (i) there exists $n_1 \in \mathbb{N}$ such that, for each $n \geq n_1$, $\theta_{i,n} = \bar{\theta}_i$, and (ii) $\tau_{i,n}^\infty$ converges to $\bar{\tau}_i^\infty$. It is easy to see that $(\bar{\theta}_i, \bar{\tau}_i^\infty) \in \operatorname{proj}_{\Theta_i \times \mathcal{T}_i^\infty} KT_i$.

For each $h_i \in H_i$, we can then define $\sigma(\bar{\tau}_i^\infty)(\,\cdot\,|h_i)$ and $\sigma(\tau_{i,n}^\infty)(\,\cdot\,|h_i)$ as usual. For each $h_i \in H_i$, $\sigma(\tau_{i,n}^\infty)(\,\cdot\,|h_i)$ converges to $\sigma(\bar{\tau}_i^\infty)(\,\cdot\,|h_i)$ because $\tau_{i,K+1,n}(\,\cdot\,|h_i)$ converges to $\bar{\tau}_{i,K+1}(\,\cdot\,|h_i)$.

At this point, fix a generic $h_i \in H_i$ and note 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 $\hat{\mathcal{A}}_i(h_i) \subseteq A_i$. Theorem 15.3 of Aliprantis and Border (2006) ensures that $\sigma(\tau_{i,n}^\infty)(\,\cdot\,|h_i)$ converges to $\sigma(\bar{\tau}_i^\infty)(\,\cdot\,|h_i)$ if and only if $\sigma(\tau_{i,n}^\infty)(B|h_i) \to \sigma(\bar{\tau}_i^\infty)(B|h_i)$ for each Borel set B whose boundary has measure zero. However, in the relative topology of $\hat{\mathcal{A}}_i(h_i)$, each subset of $\hat{\mathcal{A}}_i(h_i)$ is clopen, and thus has empty boundary (hence, its boundary has measure zero). Moreover, each subset B of $\hat{\mathcal{A}}_i(h_i)$ can be written as a finite union of singletons. Therefore, we can prove that $\sigma(\tau_{i,n}^\infty)(\,\cdot\,|h_i)$ converges to $\sigma(\bar{\tau}_i^\infty)(\,\cdot\,|h_i)$ by showing 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 $a_i \in \hat{\mathcal{A}}_i(h_i)$ such that $\sigma(\bar{\tau}_i^\infty)(a_i|h_i) > 0$. By the observation we just made, we have that $\lim_{n\to\infty} \sigma(\tau_{i,n}^\infty)(a_i|h_i) > 0$. This means that there exists $n_2 \in \mathbb{N}$ such that, for each $n \geq n_2$, $\sigma(\tau_{i,n}^\infty)(a_i|h_i) > 0$. Consider then $\bar{n} := \max\{n_1, n_2\}$: for each $n \geq \bar{n}$, we have that $\theta_{i,n} = \bar{\theta}_i$ and that $\sigma(\tau_{i,n}^\infty)(a_i|h_i) > 0$. Since the sequence $(\theta_{i,n}, \tau_{i,n}^\infty)_{n\in\mathbb{N}}$ is made of elements of $\widetilde{OP}_i$, we conclude that, for each $n \geq \bar{n}$, there exists $s_{i,n}^* \in r_{i,h_i}(\bar{\theta}_i, \tau_{i,n}^\infty)$ such that $a_i = s_{i,n}^*(h_i)$ (we use the subscript n to remind that such personal external state may vary). Thus:

$$\forall\, s_i \in S_i(h_i), \forall\, n \geq \bar{n},\ \bar{u}_{i,h_i}(s_{i,n}^*, \bar{\theta}_i, \tau_{i,n}^{K+1}) \geq \bar{u}_{i,h_i}(s_i, \bar{\theta}_i, \tau_{i,n}^{K+1}),$$

where $\tau_{i,n}^{K+1}$ is the $(K+1)$-th-order hierarchical system of beliefs induced by $\tau_{i,n}^\infty$.

Note that $(s_{i,n}^*)_{n\in\mathbb{N}}$ is a sequence in the finite (hence, compact) set $S_i(h_i)$. Thus, it admits a subsequence $(s_{i,n_k}^*)_{k\in\mathbb{N}}$, that converges to $s_i^* \in S_i(h_i)$. We can write:

$$\forall\, s_i \in S_i(h_i), \forall\, n_k \geq \bar{n},\ \bar{u}_{i,h_i}(s_{i,n_k}^*, \bar{\theta}_i, \tau_{i,n_k}^{K+1}) \geq \bar{u}_{i,h_i}(s_i, \bar{\theta}_i, \tau_{i,n_k}^{K+1}).$$

Since function $\bar{u}_{i,h_i}$ is continuous by Remark 8, we can take limits for $k \to \infty$. Noting that subsequence $(\tau_{i,n_k}^{K+1})_{k\in\mathbb{N}}$ obviously converges to $\bar{\tau}_i^{K+1}$ by our starting assumption about sequence $(\theta_{i,n}, \tau_{i,n}^\infty)_{n\in\mathbb{N}}$, we obtain:

$$\forall\, s_i \in S_i(h_i),\ \bar{u}_{i,h_i}(s_i^*, \bar{\theta}_i, \bar{\tau}_i^{K+1}) \geq \bar{u}_{i,h_i}(s_i, \bar{\theta}_i, \bar{\tau}_i^{K+1}).$$

Lastly, we conclude that $s_i^*(h_i) = a_i$, as we argued that $s_{i,n}^*(h_i) = a_i$ for each $n \geq \bar{n}$.

All in all, we showed that $\sigma(\bar{\tau}_i^\infty)(a_i|h_i) > 0$ ultimately implies that there exists $s_i^* \in S_i(h_i)$ such that $s_i^* \in \arg\max_{s_i \in S_i(h_i)} \bar{u}_{i,h_i}(s_i, \bar{\theta}_i, \bar{\tau}_i^{K+1})$ and such that $s_i^*(h_i) = a_i$. Clearly, the same holds for each non-terminal personal history, as the chosen $h_i \in H_i$ was generic. This in turn proves that $(\bar{\theta}_i, \bar{\tau}_i^\infty) \in \widetilde{OP}_i$, showing that $\widetilde{OP}_i$ is closed. OP_i is then easily seen to be closed as well, and the same holds for each player $i \in I$. ■

Proof of Lemma 7 (p. 30)

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

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

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 finititeness 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. 30)

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. 31)

We first state and prove an auxiliary result.

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

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

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

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.

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

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 as per Remark 8 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 $\mathbf{P}_i^{\Delta}(n+1) = \operatorname{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K} P^*$: 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 $P_4(q) = S_i \times \Theta_i \times \big\{\tau_i^{K+1} \in \mathcal{T}_{i,C}^{K+1} : \forall k \in \{1, \ldots, q-1\}, \tau_{i,K+1} \text{ strongly believes } \mathbf{P}_{-i}^{\Delta}(k)\big\}$. Note that, for each $k \in \mathbb{N}$, we can write $\mathbf{P}_i^{\Delta}(k) = \operatorname{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K}(P_1 \cap P_2 \cap P_3 \cap P_4(k-1))$, and that, for each $k \in \mathbb{N}$, $P_4(k) \subseteq P_4(k-1)$. With this, we conclude that $\mathbf{P}_i^{\Delta}(n+1) = \operatorname{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K}(P_1 \cap P_2 \cap P_3 \cap P_4(n)) \subseteq \operatorname{proj}_{S_i \times \Theta_i \times \mathcal{T}_i^K}(P_1 \cap P_2 \cap P_3 \cap P_4(n-1)) = \mathbf{P}_i^{\Delta}(n)$, which yields the desired result. ■

Proof of Proposition 2 (p. 33)

We begin this proof by introducing some terminology and by proving auxiliary results. Moreover, 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 $F^1, F^2 \subseteq \Omega^K_{-i}$. 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 $F \subseteq \Omega^K_{-i}$ and $\mu_i \in \mathcal{T}_{i,K+1}$, we say that F is *compatible* with μ_i and h_i if

$$F \cap \Omega^K_{-i,\operatorname{marg}\mu_i}(h_i) \neq \emptyset,$$

where $\operatorname{marg}\mu_i$ is a shorthand to denote the hierarchical system of beliefs of order K obtained by taking the marginals of μ_i over the sets $(\Omega^0, (\Omega^n_{-i})_{n=1}^{K-1})$. 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 $k \in \{1,\ldots,n\}$, 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 both the chain rule and Bayes rule are satisfied by $\bar{\mu}_i$.

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 $F^2 \cap \Omega^K_{-i,\mu_i^2}(h_i') = \emptyset$ and $F^2 \cap \Omega^K_{-i,\mu_i^2}(h_i) \neq \emptyset$, $(\mu_i^2)^*(F^2|h_i) = 1$ and $(\mu_i^2)^*(\Omega^K_{-i,\mu_i^2}(h_i')|h_i) = 0$.[61] Second, since $h_i' \in \bar{H}_i(h_i,a_i)$, each $s_i' \in \Omega^K_{-i,\mu_i^2}(h_i')$ must also belong to $S_i(h_i,a_i)$. Taken together, these observations yield $\mu_i^2(s_i|h_i) = \mu_i^2(S_i(h_i,a_i)|h_i) = 0$.[62] Therefore

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

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

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

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: the 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 $g_{i,h_i,s_i^*,\operatorname{marg}\mu_i} : S_{-i}\times\Theta\times\mathcal{T}_{-i}^K \to \Delta(M_i^*)$ as g^* for simplicity (s_i^* is a generic element of $S_i(h_i, a_i)$). We want to show that

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

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

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

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 $g^*(s_{-i},\theta,\tau_{-i})[m_i^*] > 0$ only if $(s_{-i},\theta,\tau_{-i}) \in \operatorname{proj}_{S_{-i}\times\Theta\times\mathcal{T}_{-i}^K}\Omega_{-i,\operatorname{marg}\mu_i^2}(h_i') =: X$.[63] 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, $g^*(s_{-i},\theta,\tau_{-i})[m_i^*] > 0$ only if $(s_{-i},\theta,\tau_{-i}) \in Y$.

At this point, it is easy to check that

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

where the first equality follows from the consideration that $g^*(\,\cdot\,)[m_i^*]$ 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 that $G \subseteq S_{-i}\times\Theta\times\mathcal{T}_{-i}^K$ and $g^*(\,\cdot\,)[m_i^*] : S_{-i}\times\Theta\times\mathcal{T}_{-i}^K \to [0,1]$ 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.

[63] Note that X is analytic.

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 $k \in \{1, \ldots, n\}$, 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 $k \in \{1, \ldots, n\}$. Therefore, for each $k \in \{1, \ldots, n\}$, 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 $k \in \{1, \ldots, n\}$, $(\tau_i, \mu_i^k) \in \operatorname{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$, then $(\tau_i, \bar{\mu}_i) \in \operatorname{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$. To see why this holds, consider that, for each $h_i \in \bar{H}_i$, there is $\bar{k} \in \{1, \ldots, n\}$ such that $\bar{\mu}_i(\,\cdot\,|h_i) = \mu_i^{\bar{k}}(\,\cdot\,|h_i)$. Given that $(\tau_i, \mu_i^{\bar{k}}) \in \operatorname{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$, then $\operatorname{marg}_{\Omega_{-i}^{K-1}} \mu_i^{\bar{k}}(\,\cdot\,|h_i) = \tau_{i,K}(\,\cdot\,|h_i)$ for each $h_i \in \bar{H}_i$, and the same holds for each $k \in \{1, \ldots, n\}$ (cf. condition 1M of Definition 13). Therefore, we can conclude that, for each $h_i \in \bar{H}_i$, $\operatorname{marg}_{\Omega_{-i}^{K-1}} \bar{\mu}_i(\,\cdot\,|h_i) = \tau_{i,K}(\,\cdot\,|h_i)$. 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 $(\tau_i, \bar{\mu}_i^k) \in \operatorname{proj}_{\mathcal{T}_i^{K+1}} \mathcal{T}_{i,C}^\infty$.

   Similarly, we have that, for each $k \in \{1, \ldots, n\}$, $(\tau_i, \mu_i^k) \in \Delta_{\theta_i}$. Recall that Δ is rectangular, so that we can write $\Delta_{\theta_i} = \times_{n=1}^{K+1} \times_{h_i \in \bar{H}_i} B_{\theta_i,n,h_i}$ for a suitable profile of measurable sets. Thanks to this, we can conclude that $\tau_i \in \operatorname{proj}_{\mathcal{T}_i^K} \Delta_{\theta_i} = \times_{n=1}^{K} \times_{h_i \in \bar{H}_i} B_{\theta_i,n,h_i}$. Moreover, note that, for each $h_i \in \bar{H}_i$, there is $\bar{k} \in \{1, \ldots, n\}$ such that $\bar{\mu}_i(\,\cdot\,|h_i) = \mu_i^{\bar{k}}(\,\cdot\,|h_i)$, and that, for each $k \in \{1, \ldots, n\}$, $\mu_i^k \in \operatorname{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i} = \times_{h_i \in \bar{H}_i} B_{\theta_i,K+1,h_i}$. 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 $\bar{\mu}_i \in \operatorname{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i}$. Wrapping up, $\tau_i \in \operatorname{proj}_{\mathcal{T}_i^K} \Delta_{\theta_i}$ and $\bar{\mu}_i \in \operatorname{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i}$ imply $(\tau_i, \bar{\mu}_i) \in \operatorname{proj}_{\mathcal{T}_i^K} \Delta_{\theta_i} \times \operatorname{proj}_{\mathcal{T}_{i,K+1}} \Delta_{\theta_i} = \Delta_{\theta_i}$, 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 $(s_i, \theta_i, (\tau_i, \mu_i)) \in S_i \times \Theta_i \times \mathcal{T}_i^{K+1}$, $\bar{u}_{i,h_i}(s_i, \theta_i, (\tau_i, \mu_i)) = \int_{S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K} u_{i,h_i,s_i,\theta_i} \mathrm{d}\nu_i(\,\cdot\,|h_i)$, 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 $h_i \in H_i$, $r_{i,h_i} : \Theta_i \times \mathcal{T}_i^{K+1} \rightrightarrows S_i$ is the correspondence $(\theta_i, (\tau_i, \mu_i)) \mapsto \arg\max_{s_i'} \bar{u}_{i,h_i}(s_i', \theta_i, (\tau_i, \mu_i))$.

   Since $(s_i, \theta_i, \tau_i) \in \bigcap_{k=1}^n \mathbf{Q}_i^\Delta(k)$, we have $s_i \in \bigcap_{h_i \in H_i} r_{i,h_i}(\theta_i, (\tau_i, \mu_i^k))$ for each $k \in \{1, \ldots, n\}$. Fix $h_i \in H_i$: we know that there is $k \in \{1, \ldots, n\}$ such that $\bar{\mu}_i(\,\cdot\,|h_i) = \mu_i^k(\,\cdot\,|h_i)$. Moreover, $s_i \in r_{i,h_i}(\theta_i, (\tau_i, \mu_i^k))$, meaning that s_i maximizes $\int u_{i,h_i,s_i,\theta_i} \mathrm{d}\nu_i^k(\,\cdot\,|h_i)$ (ν_i^k is the marginal of μ_i^k over $S_{-i} \times \Theta_{-i} \times \mathcal{T}_{-i}^K$). By the foregoing observation, $\int u_{i,h_i,s_i,\theta_i} \mathrm{d}\nu_i^k(\,\cdot\,|h_i) = \int u_{i,h_i,s_i,\theta_i} \mathrm{d}\bar{\nu}_i(\,\cdot\,|h_i)$, as $\mu_i^k(\,\cdot\,|h_i) = \bar{\mu}_i(\,\cdot\,|h_i)$ implies $\nu_i^k(\,\cdot\,|h_i) = \bar{\nu}_i(\,\cdot\,|h_i)$. Therefore, $s_i \in r_{i,h_i}(\theta_i, (\tau_i, \bar{\mu}_i))$, and the same goes for each $h_i \in H_i$. This implies that $\bar{\mu}_i$ satisfies condition 3 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. 35)

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:

$$\begin{aligned}
\mathrm{SB}_i(F_{-i}) =& \Big\{(s_i, \theta_i, \tau_i^{\infty}) \in C_i : (\exists\, \bar{\tau}_i^K \in \mathcal{T}_i^K, \tau_i^K \in [\bar{\tau}_i^K]), \\
& (\forall\, h_i \in H_i, \Omega_{-i,[\bar{\tau}_i^K]}^{\infty}(h_i) \cap F_{-i} \neq \emptyset \Longrightarrow \varphi_i(\tau_i^{\infty})(F_{-i}|h_i) = 1)\Big\} \\
=& \bigcup_{[\bar{\tau}_i^K]} \Bigg(\bigg(S_i \times \Theta_i \times [\bar{\tau}_i^K]_{h_i} \times \mathop{\times}_{k \geq K+1} \mathcal{T}_{i,k}\bigg) \\
& \cap \bigg(\bigcap_{h_i : \Omega_{-i,[\bar{\tau}_i^K]}^{\infty}(h_i) \cap F_{-i} \neq \emptyset} \left\{(s_i, \theta_i, \tau_i^{\infty}) \in C_i : \varphi_i(\tau_i^{\infty})(F_{-i}|h_i) = 1)\right\}\bigg)\Bigg) \\
=& \bigcup_{[\bar{\tau}_i^K]} \Bigg(\bigg(S_i \times \Theta_i \times [\bar{\tau}_i^K]_{h_i} \times \mathop{\times}_{k \geq K+1} \mathcal{T}_{i,k}\bigg) \cap \bigg(\bigcap_{h_i : \Omega_{-i,[\bar{\tau}_i^K]}^{\infty}(h_i) \cap F_{-i} \neq \emptyset} \mathrm{B}_{i,h_i}(F_{-i})\bigg)\Bigg). \qquad (12)
\end{aligned}$$

In (12), 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 (12) 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. 36)

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 $\operatorname{marg}_X \nu$ as a shorthand for $(\operatorname{marg}_X \nu(\,\cdot\,|C))_{C\in\mathcal{C}}$. 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 $(\operatorname{proj}_C D_k)_{k=1}^n$, then there exists a CPS ν on $(C \times X, \mathcal{C})$ that strongly believes $(D_k)_{k=1}^n$ and such that $\operatorname{marg}_C \nu = \mu$.*

We now proceed with the proof of Theorem 1.

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

First, we show $\mathbf{P}_i^{\Delta}(n) \subseteq \operatorname{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n)$. 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 $(\Omega^0, \Omega^1_{-i}, \dots, \Omega^{K-1}_{-i})$. Moreover, $\tau_{i,K+1}$ strongly believes $(\mathbf{P}^{\Delta}_{-i}(1), \dots, \mathbf{P}^{\Delta}_{-i}(n-1)) = (\operatorname{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(1), \dots, \operatorname{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n-1))$, 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 $(\mathbf{R}^{\Delta}_{-i}(1), \dots, \mathbf{R}^{\Delta}_i(n-1))$ such that $\operatorname{marg}_{\Omega^K_{-i}} \mu = \tau_{i,K+1}$. 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 $\operatorname{marg}_{\Omega^K_{-i}} \mu = \tau_{i,K+1}$. 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 $(\mathbf{R}^{\Delta}_{-i}(1), \dots, \mathbf{R}^{\Delta}_i(n-1))$ as already mentioned. Hence, $(s_i, \theta_i, \bar{\tau}_i^\infty) \in \mathbf{R}_i^{\Delta}(n)$, so that $(s_i, \theta_i, \tau_i^K) \in \operatorname{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n)$.

Second, we show $\operatorname{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n) \subseteq \mathbf{P}_i^{\Delta}(n)$. Take $(s_i, \theta_i, \tau_i^K) \in \operatorname{proj}_{S_i\times\Theta_i\times\mathcal{T}_i^K} \mathbf{R}_i^{\Delta}(n)$. Then, by definition there exists $\mu = (\mu_k)_{k\geq K+1} \in \bigtimes_{k\geq K+1} \mathcal{T}_{i,k}$ such that $(s_i, \theta_i, \tau_i^K, \mu) \in \mathbf{R}_i^{\Delta}(n) = R_i \cap \left(\bigcap_{k=1}^{n-1} \mathrm{SB}_i(\mathbf{R}^{\Delta}_{-i}(k))\right)$. (τ_i^K, μ_{K+1}) satisfies conditions 1, 2, 3 of Definition 12, 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

$(\mathbf{P}^{\Delta}_{-i}(1), \ldots, \mathbf{P}^{\Delta}_{-i}(n-1))$. 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 $\mathbf{P}^{\Delta}_{i}(k) \cap \Omega^K_{-i,\tau_i}(h_i) \neq \emptyset$. By the inductive hypothesis, the coherence of (τ_i^K, μ), and the definition of inference sets, this is equivalent to writing $\mathbf{R}^{\Delta}_{-i}(k) \cap \Omega^{\infty}_{-i,\tau_i^K}(h_i) \neq \emptyset$. However, if such condition holds, we have that $\varphi_i\big((\tau_i^K, \mu)\big)(\mathbf{R}^{\Delta}_{-i}(k)|h_i) = 1$, because (τ_i^K, μ) strongly believes $(\mathbf{R}^{\Delta}_{i}(1), \ldots, \mathbf{R}^{\Delta}_{-i}(n-1))$. At this point, we can write:

$$\begin{aligned}\mu^*_{K+1}(\mathbf{P}^{\Delta}_{-i}(k)|h_i) &= \mathrm{marg}_{\Omega^K_{-i}} \varphi_i\big((\tau_i^K, \mu)\big)(\mathbf{P}^{\Delta}_{-i}(k)|h_i) = \mathrm{marg}_{\Omega^K_{-i}} \varphi_i\big((\tau_i^K, \mu)\big)(\mathrm{proj}_{\Omega^K_{-i}} \mathbf{R}^{\Delta}_{-i}(k)|h_i) = \\ =&\varphi_i\big((\tau_i^K, \mu)\big)\big(\,\mathrm{proj}^{-1}_{\Omega^K_{-i}}(\mathrm{proj}_{\Omega^K_{-i}} \mathbf{P}^{\Delta}_{-i}(k))\big) \geq \varphi_i\big((\tau_i^K, \mu)\big)(\mathbf{R}^{\Delta}_{-i}(q)) = 1.\end{aligned}$$

The same holds for each $k \in \{1, \ldots, n-1\}$ and $h_i \in H_i$, proving that μ_{K+1} strongly believes $(\mathbf{P}^{\Delta}_{-i}(1), \ldots, \mathbf{P}^{\Delta}_{i}(n-1))$. Hence, $(s_i, \theta_i, \tau_i^K) \in \mathbf{P}^{\Delta}_{i}(n)$, 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^N (resp., z^B) denote a generic terminal history in which Mom plays N (resp., B) – that is, an element of $\{(a_{C,1}, a_{C,2}, m_M, a_M) \in Z : a_M = N\}$ (resp., $\{(a_{C,1}, a_{C,2}, m_M, a_M) \in Z : a_M = B\}$), and let z_M be the length-two personal history of Mom induced by a generic terminal history z. In the following, consider a generic $\theta_C = (\lambda, \nu) \in \Theta_C$, and $\tau^1 \in \mathcal{T}^1$. Then:

1. A terminal history z^N occurs with certainty if: (i) $s_C = H.N$ and $s_M\big((N, \neg b)\big) = N$; (ii) $s_C = H.Y$ and $s_M\big((Y, \neg b)\big) = N$; (iii) $s_C = V.N$ and $s_M\big((N, \neg b)\big) = N$; (iv) $s_C = V.Y$ and $s_M\big((Y, b)\big) = s_M\big((Y, \neg b)\big) = N$. In such case, $u_M(s, \theta, \tau^1) = 0$, and
$$u_C(s, \theta, \tau^1) = \begin{cases} -\lambda \tau_{M,1}(L|z^N_M) & \text{if } s_C(\varnothing) = H; \\ \nu - \lambda \tau_{M,1}(L|z^N_M) & \text{if } s_C(\varnothing) = V. \end{cases}$$

2. A terminal history z^B occurs with certainty if: (i) $s_C = H.N$ and $s_M\big((N, \neg b)\big) = B$; (ii) $s_C = H.Y$ and $s_M\big((Y, \neg b)\big) = B$; (iii) $s_C = V.N$ and $s_M\big((N, \neg b)\big) = B$; (iv) $s_C = V.Y$ and $s_M\big((Y, b)\big) = s_M\big((Y, \neg b)\big) = B$. Then, $u_M(s, \theta, \tau^1) = 2\tau_{M,1}(G|z^B_M) - 1$, and
$$u_C(s, \theta, \tau^1) = \begin{cases} 1 - \lambda \tau_{M,1}(L|z^B_M) & \text{if } s_C(\varnothing) = H; \\ 1 + \nu - \lambda \tau_{M,1}(L|z^B_M) & \text{if } s_C(\varnothing) = V. \end{cases}$$

3. A terminal history z^B (resp., z^N) occurs with probability q (resp., $1 - q$) if $s_C = V.Y$, $s_M\big((Y, b)\big) = B$, and $s_M\big((Y, \neg b)\big) = N$. Hence,
$$\begin{aligned} u_C(s, \theta, \tau^1) =& q\big(1 + \nu - \lambda \tau_{M,1}(L|z^B_M)\big) + (1 - q)\big(\nu - \lambda \tau_{M,1}(L|z^N_M)\big) = \\ &= q + \nu - q\lambda\big(\tau_{M,1}(L|(Y, \neg b))\big) - (1 - q)\lambda\big(\tau_{M,1}(L|(Y, b))\big); \\ u_M(s, \theta, \tau^1) =& q\big(2\tau_{M,1}(G|z^B_M) - 1\big) = q\big(2\tau_{M,1}(G|(Y, \neg b)) - 1\big). \end{aligned}$$

4. A terminal history z^B (resp., z^N) occurs with probability $1-q$ (resp., q) if $s_C = V.Y$, $s_M\big((Y,b)\big) = N$, and $s_M\big((Y,\neg b)\big) = B$. Hence,

$$\begin{aligned} u_C(s,\theta,\tau^1) =&(1-q)\big(1+\nu-\lambda\tau_{M,1}(L|z^B_M)\big) + q\big(\nu-\lambda\tau_{M,1}(L|z^N_M)\big) = \\ =&(1-q)+\nu-(1-q)\lambda\big(\tau_{M,1}(L|(Y,\neg b))\big)-q\lambda\big(\tau_{M,1}(L|(Y,b))\big);\\ u_M(s,\theta,\tau^1) =&(1-q)\big(2\tau_{M,1}(G|z^B_M)-1\big) = (1-q)\big(2\tau_{M,1}(G|(Y,\neg b))-1\big). \end{aligned}$$

In words, Child's personal external state unambiguously defines the first two actions of a terminal history. Then, the only case in which multiple terminal histories may arise is when Child plays according to $V.Y$ and Mom according to a personal external state that prescribes different actions after observing (Y,b) and $(Y,\neg b)$.

Local decision utilities In the present setting, distortions are absent. Then, recall that, for each $i \in I$ and $h_i \in H_i$, $\bar{u}_{i,h_i}$ is the map $(s_i,\theta_i,\tau_i^{K+1}) \mapsto \mathbb{E}_{s_i,\theta_i,\tau_i^{K+1}}[u_i|h_i]$. We start with Child. By building on the previous paragraph, it is straightforward to check that the following are Child's decision utilities at the root of the game (cf. p. 29):

$$\begin{aligned} \bar{u}_{C,\varnothing}(H.Y,\theta_C,\tau_C^2) =&\tau_{C,2}\big(\{s_M : s_M\big((Y,\neg b)\big)=B\}|\varnothing\big)-\lambda\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,\neg b)\big)|\varnothing\big];\\ \bar{u}_{C,\varnothing}(H.N,\theta_C,\tau_C^2) =&\tau_{C,2}\big(\{s_M : s_M\big((N,\neg b)\big)=B\}|\varnothing\big)-\lambda\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(N,\neg b)\big)|\varnothing\big];\\ \bar{u}_{C,\varnothing}(V.Y,\theta_C,\tau_C^2) =&\nu+q\Bigg(\tau_{C,2}\big(\{s_M : s_M\big((Y,\neg b)\big)=B\}|\varnothing\big)-\lambda\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,\neg b)\big)|\varnothing\big]\Bigg)+\\ &+(1-q)\Bigg(\tau_{C,2}\big(\{s_M : s_M\big((Y,b)\big)=B\}|\varnothing\big)-\lambda\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,b)\big)|\varnothing\big]\Bigg);\\ \bar{u}_{C,\varnothing}(V.N,\theta_C,\tau_C^2) =&\nu+\tau_{C,2}\big(\{s_M : s_M\big((N,\neg b)\big)=B\}|\varnothing\big)-\lambda\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(N,\neg b)\big)|\varnothing\big]\big). \end{aligned}$$

Child's decision utilities after personal histories (H) and (V) are easily retrieved by considering his beliefs after (H) and (V), respectively, as well as the action prescribed by each personal external state after such histories.

As for Mom, for each $h_M \in \big\{(Y,\neg b),(Y,b),(N,\neg b)\big\}$, $s_M \in S_M$ and $\tau_M^2 \in \mathcal{T}_M^2$,

$$\bar{u}_{M,h_M}(s_M,\tau_M^2) = \begin{cases} 2\tau_{M,2}\big(G|(h_M,s_M(h_M))\big)-1 & \text{if } s_M\big(h_M\big)=B;\\ 0 & \text{if } s_M\big(h_M\big)=N; \end{cases}$$

where $(h_M,s_M(h_M))$ indicates that the relevant belief is the one held by Mom at the end of the game.[64] However, when the Bayes rule is considered in conjunction with consistency, we have $\tau_{M,2}\big(G|(h_M,s_M(h_M))\big) = \tau_{M,2}\big(G|h_M\big)$ (cf. p. 27).

B.2 Solution procedure

In light of our simplifying assumptions (that is, $\Lambda = \{1\}$ and $N = \{\nu',\nu''\}$), we identify Θ_C with N and we thus write $\theta_C \in \{\nu',\nu''\}$. A few preliminary observations are in order. First, in the following, we consider systems of first-order beliefs such that condition 3 of Definition 12 is

[64] Recall that the set of terminal personal histories of Mom is (isomorphic to) $\big\{(Y,\neg b),(Y,b),(N,\neg b)\big\}\times\big\{B,N\big\}$.

satisfied, without mentioning it at every step. This implies that, by the Bayes rule, he will not change his beliefs after acting. Second, we only check condition 2 for Child with respect to the empty history. In light of the previous observation, a personal external state that is optimal at the root of the game will be optimal also at the personal history it does not prevent. Indeed, considering the previous two observations in conjunction, it is possible to coalesce Child's moves as it is usually done in multistage games whenever players are called to act twice in a row: in this way, it is as if Child were to choose among actions $H.Y$, $H.N$, $V.Y$, and $V.N$ at the root of the game.[65] Third, we implement the procedure of Definition 12 by replacing condition 4 with a requirement that, for each $i \in I$, $\tau_{i,K+1}$ strongly believes *only the last step* of the procedure: this makes no difference when no restrictions on beliefs are imposed (cf. Proposition 2).

First step It is easy to see that $H.N$ is not optimal for any of Child's trait-types, as $V.N$ yields a strictly higher local decision utility at the root of the game – hence, condition 2 of Definition 12 is failed by each profile $(s_C, \theta_C, \tau_C^1)$ with $s_C = H.N$. Moreover, if Child's beliefs are as those described at page 29 but with $\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(N,\neg b)\big)|\varnothing\big]\big) = 1$, $H.Y$ is optimal for both Child's trait-types. Such second-order system of beliefs trivially strongly believes $S_M \times \Theta_M \times \mathcal{T}_M^1$, 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 $V.N$ maximizes $\bar{u}_{C,\varnothing}\big(\,\cdot\,, \theta_C, (\tau_C^1, \tau_{C,2})\big)$ for both trait-types if $\tau_{C,2}$ is such that $\tau_{C,2}\big(\big\{s_M \in S_M : s_M\big((Y,b)\big) = B\big\}|\varnothing\big) = 1$ and $\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}(L|(Y,\neg b))|\varnothing\big] = 0$. Again, such $\tau_{C,2}$ strongly believes $S_M \times \Theta_M \times \mathcal{T}_M^1$, 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 $\mathrm{proj}_{S_C \times \Theta_C} \mathbf{P}_C(1) = \{H.Y, V.Y, V.N\} \times \{\nu', \nu''\}$.

As for Mom, it is immediate to notice that condition 1 of Definition 12 implies that, to survive this deletion step, a profile (s_M, τ_M^1) has to be such that $\tau_{M,1}\big(\{s_C \in S_C : s_C(\varnothing) = V\}|(Y,b)\big) = 1$. This in turn implies that $\tau_{M,1}\big(G|(Y,b)\big) = 0$ and $\tau_{M,1}\big(L|(Y,b)\big) = 1$. Then, any $\tau_{M,2}$ that we may look for to carry out the procedure has to conform to such features. Moreover, by the Bayes rule, the same holds for each (personal) terminal history that realizes after Mom's move. Therefore, $\bar{u}_{M,(Y,b)}\big(\,\cdot\,, (\tau_M^1, \tau_{M,2})\big)$ is maximized by any s_M such that $s_M\big((Y,b)\big) = N$. On the other hand, if $\tau_{M,2}\big(G|(Y,\neg b)\big) = \tau_{M,2}\big(G|(N,\neg b)\big) = \frac{1}{2}$, any s_M maximizes both $\bar{u}_{M,(Y,\neg b)}\big(\,\cdot\,, (\tau_M^1, \tau_{M,2})\big)$ and $\bar{u}_{M,(N,\neg b)}\big(\,\cdot\,, (\tau_M^1, \tau_{M,2})\big)$. Thus, $\mathrm{proj}_{S_M} \mathbf{P}_M(1) = \big\{s_M \in S_M : s_M\big((Y,b)\big) = N\big\}$.

Second step We now have to restrict attention to $\tau_{C,2}$ such that, for each non-terminal personal history $h_C \in H_C$, $\tau_{C,2}\big(\big\{s_M : s_M\big((Y,b)\big) = B\big\}|h_C\big\} = 0$ and $\mathbb{E}_{\tau_{C,2}}\big[\tau_{M,1}\big(L|(Y,b)\big)|h_C\big] = 1$. Hence, we can conclude that lying after having played video-games makes Child blush with certainty. It is easy to check that $\bar{u}_{C,\varnothing}\big(V.Y, \theta_C, (\tau_C^1, \tau_{C,2})\big) = \theta_C - 1 < \theta_C = \bar{u}_{C,\varnothing}\big(V.N, \theta_C, (\tau_C^1, \tau_{C,2})\big)$ for each $\tau_{C,2}$ satisfying the aforementioned restrictions and for each $\theta_C \in \Theta_C$. Thus, any $(s_C, \theta_C, \tau_C^1) \in \mathbf{P}_C(1)$ with $s_C = V.Y$ fails condition 2 of Definition 12. Moreover, it can be

[65] Then, we can still easily recover the optimal action for the personal history precluded by a given external state. For instance, the unique strongly rationalizable personal external state for trait-type ν'' prescribes playing V at the first stage and N afterwards. However, by the reasoning we will carry out to apply the procedure, we can conclude that the optimal action to be played after (H) is Y. Thus, it must be the case that the unique strongly rationalizable personal external state for trait-type ν'' prescribes Y after the prevented history (H).

checked that plying according to $H.Y$ yields a utility of at most 1 (cf. previous page): for trait-type ν'', such personal external state is never optimal, as $V.N$ yields a utility of $\nu'' > 1$. It follows that $\mathrm{proj}_{S_C \times \Theta_C} \mathbf{P}_C(2) = \big(\{H.Y, V.N\} \times \{\nu'\}\big) \cup \big(\{V.N\} \times \{\nu''\}\big)$.

As for Mom, any $\tau_{M,2}$ strongly believing $\mathbf{P}_C(1)$ must be such that $\tau_{M,2}\big(\{V.N\}|(N, \neg b)\big) = 1$ – that is, she is now sure that, if Child answers "no", he must have played video-games at the first stage. This also implies that the above-mentioned $\tau_{M,2}$ must be such that $\tau_{M,2}\big(G|(N, \neg b)\big) = 0$. It is now easy to realize that $\bar{u}_{M,(N,\neg b)}\big(\,\cdot\,, (\tau^1_M, \tau_{M,2})\big)$ is maximized by any s_M with $s_M\big((N, \neg b)\big) = N$, for each $\tau_{M,2}$ strongly believing $\mathbf{P}_C(1)$ and $(\tau^1_M, \tau_{M,2})\big)$ satisfying condition 1 of Definition 12. Hence, $\mathrm{proj}_{S_M} \mathbf{P}_M(2) = \big\{s_M \in S_M : s_M\big((Y, b)\big) = s_M\big((N, \neg b)\big) = N\big\} = \big\{N.N.N, N.B.N\big\}$.

Third step We now have to consider only $\tau_{C,2}$ strongly believing $\mathbf{P}_M(2)$, and we shall focus on trait-type ν'. This means that $\tau_{C,2}\big(\big\{s_M : s_M\big((N, \neg b)\big) = B\big\}|h_C\big) = 0$ for each non-terminal $h_C \in H_C$. This implies that $\bar{u}_{C,\varnothing}\big(V.Y, \nu', (\tau^1_C, \tau_{C,2})\big) = \nu'$ for each $\tau_{C,2}$ strongly believing $\mathbf{P}_M(2)$ and for each $(\tau^1_C, \tau_{C,2})$ meeting requirement 1 of Definition 12. Moreover, note that, if $\tau_{C,2}$ has to strongly believe $\mathbf{P}_M(2)$, we obtain $\bar{u}_{C,\varnothing}\big(H.Y, \nu', (\tau^1_C, \tau_{C,2})\big) = \tau_{C,1}(\{N.B.N\}|\varnothing)$. Thus, both $H.Y$ and $V.N$ can be optimal for trait-type ν', and this leads us to conclude that $\mathrm{proj}_{S_C \times \Theta_C} \mathbf{P}_C(3) = \mathrm{proj}_{S_C \times \Theta_C} \mathbf{P}_C(2) = \big(\{H.Y, V.N\} \times \{\nu'\}\big) \cup \big(\{V.N\} \times \{\nu''\}\big)$.

On the other hand, any $\tau_{M,2}$ strongly believing $\mathbf{P}_C(2)$ is such that $\tau_{M,2}\big(\{H.Y\}|(Y, \neg b)\big) = 1$. Therefore, $\tau_{M,2}\big(L|(Y, \neg b)\big) = 0$ and $\tau_{M,2}\big(G|(Y, \neg b)\big) = 1$. With this, $\bar{u}_{M,(N,\neg b)}\big(\,\cdot\,, (\tau^1_M, \tau_{M,2})\big)$ is maximized by any s_M with $s_M\big((N, \neg b)\big) = N$, for each $\tau_{M,2}$ strongly believing $\mathbf{P}_C(1)$ and $(\tau^1_M, \tau_{M,2})\big)$ satisfying condition 1 of Definition 12. Hence, $\mathrm{proj}_{S_M} \mathbf{R}^{\Delta}_M(3) = \{N.B.N\}$.

Fourth step At this point, any $\tau_{C,2}$ strongly believing $\mathbf{P}_M(3)$ must assign probability one to $N.B.N$ at each non-terminal personal history. Hence, $\bar{u}_{C,\varnothing}\big(H.Y, \nu', (\tau^1_C, \tau_{C,2})\big) = 1 > \nu' = \bar{u}_{C,\varnothing}\big(V.N, \nu', (\tau^1_C, \tau_{C,2})\big)$ for each $\tau_{C,2}$ satisfying the above mentioned restrictions and for each $\theta_C \in \Theta_C$. Therefore, we conclude that $\mathrm{proj}_{S_C \times \Theta_C} \mathbf{R}^{\Delta}_C(4) = \big\{(H.Y, \nu'), (V.N, \nu'')\big\}$.

C Notation: a recap

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
$E^{\leq T}$	Sequences of emotion profiles of length up to T
$M_i,\ m_i$	Emotional messages receivable by i

Notation	**Meaning**
$M_{i,p},\ m_{i,p}$	Previous play messages receivable by i
$M_i^* = M_{i,p} \times M_i,\ m_i^*$	Message pairs receivable by i
$\tilde{f} : A \times \Theta \times E^{\leq T} \to \Delta(M)$	Game-independent feedback function
$\tilde{v}_i : Y \times \Theta \times E^{\leq T} \to \mathbb{R}$	Game-independent psychological utility of i
$p : \bigcup_{t=1}^{T} A^t \to M_p$	Previous play messages generating function
$\mathcal{A}_i : M_{i,p} \cup \{\varnothing_{M_{i,p}}\} \rightrightarrows A_i$	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
$S_i = \mathop{\times}_{h_i \in H_i} \hat{\mathcal{A}}_i(h_i), s_i$	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 $K+1$
$\tau_{i,K+1}(\,\cdot \mid h_i)$	Belief of i of order $K+1$ held at h_i
$\mathop{\times}_{h_i \in H_i} \Delta(\hat{\mathcal{A}}_i(h_i)),\ \sigma(\tau_i^{\infty})$	Plans of i
$\varepsilon : H \times \mathcal{T}^{\infty} \to \Delta(E^{\leq T})$	Emotion-generating function
$\Omega^{\infty} = \mathop{\times}_{i \in I}(S_i \times \Theta_i \times \mathcal{T}_i^{\infty})$	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
$(f_h : S \times \Theta \times \mathcal{T}^{K} \to \Delta(M))_{h \in H}$	Game-dependent 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
$(g_h^* : S \times \Theta \times \mathcal{T}^{K} \to \Delta(M^*))_{h \in H}$	State-history-dependent distribution of message-pair profiles
$\zeta : S \times \Theta \times \mathcal{T}^{K} \to \Delta(Z)$	State-dependent distribution of terminal histories
$\zeta_{h_i} : S \times \Theta \times \mathcal{T}^{K} \to \Delta(Z)$	State-dependent distribution of terminal histories conditional on h_i
$\eta_{h_i} : S \times \Theta \times \mathcal{T}^{K} \to \Delta(H)$	State-dependent distribution of histories conditional on h_i
$f_{i,h},\ g_{i,h},\ g_{i,h}^*$	State-dependent distributions over M_i, $A_i \times M_i^*$, and M_i^* derived from f_h, g_h and g_h^*
$f_{i,h_i},\ g_{i,h_i},\ g_{i,h_i}^*$	Expected state-dependent distributions over M_i, $A_i \times M_i^*$, and M_i^* after h_i, derived from $f_{i,h}$, $g_{i,h}$ 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
$\bar{u}_{i,h_i} : S_i \times \Theta_i \times \mathcal{T}_i^{K+1} \to \mathbb{R}$	Decision utility of i at h_i
$r_{i,h_i} : \Theta \times \mathcal{T}_i^{\infty} \rightrightarrows S_i$	Optimality correspondence of i at h_i
$\mathbf{A} : S \times \Theta \times \mathcal{T}^K \rightrightarrows \mathrm{proj}_{A^{\leq T}} \bar{H}$	State-dependent action-profile-sequence correspondence
$\mathbf{H}_i : S \times \Theta \times \mathcal{T}^K \rightrightarrows \bar{H}_i$	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

References

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

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

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

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

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

Battigalli, P., & Dufwenberg, M. (forthcoming). Belief-dependent motivations and psychological game theory. *Journal of Economic Literature*.

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

Battigalli, P., & Generoso, N. (2021). *Information flows and memory in games* (Working Paper No. 678). IGIER.

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

Battigalli, P., & Siniscalchi, M. (1999). Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, *88*(1), 188–230.

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

Battigalli, P., & Tebaldi, P. (2019). Interactive epistemology in simple dynamic games with a continuum of strategies. *Economic Theory*, *68*(3), 737–763.

Behrens, F., & Kret, M. E. (2019). The interplay between face-to-face contact and feedback on cooperation during real-life interactions. *Journal of Nonverbal Behavior*, *43*(4), 513–528.

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

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

Dekel, E., & Siniscalchi, M. (2015). Epistemic game theory. In *Handbook of Game Theory with Economic Applications* (Vol. 4, pp. 619–702). Elsevier.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118.

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

Dubins, L., & Freedman, D. (1964). Measurable sets of measures. *Pacific Journal of Mathematics*, *14*(4), 1211–1222.

Elfenbein, H. A., Der Foo, M., White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior*, *31*(4), 205–223.

Elster, J. (1996). Rationality and the emotions. *The Economic Journal*, *106*(438), 1386–1397.

Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature*, *36*(1), 47–74.

Gadea, M., Aliño, M., Espert, R., & Salvador, A. (2015). Deceit and facial expression in children: the enabling role of the "poker face" child and the dependent personality of the detector. *Frontiers in Psychology*, *6*, 1089.

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

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

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

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

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89*(1), 137.

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

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

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

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

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

Warren, G., Schertler, E., & Bull, P. (2009). Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior*, *33*(1), 59–69.