Information Flows and Memory in Games*

Pierpaolo Battigalli
Bocconi University and IGIER

Nicolò Generoso
Yale University

February 15, 2022

Abstract

The standard extensive-form partitional representation of information in sequential games fails to distinguish the description of the rules of interaction from the description of players' personal traits. This is because such framework is silent about how the information that is available to players as per the rules of the game blends with players' ability to retain such information. We propose a representation of sequential games that explicitly describes the *flow* of information accruing to players rather than the *stock* of information retained by players, as encoded in information partitions. Then, we add a formal description of players' mnemonic abilities by means of possibility correspondences. Assuming that players have perfect memory, our flow representation gives rise to information partitions satisfying perfect recall, but different combinations of rules about information flows and of players mnemonic abilities may give rise to the same information partitions. All extensive-form representations with information partitions, including those failing perfect recall or featuring absentmindedness, can be generated by some such combinations. Our approach also allows to model situations where players' knowledge cannot be described by information partitions.

1 Introduction

All games whose play can be implemented on IT platforms—including market games, auctions, games played in the lab, and games like poker and chess—are defined by formal rules clearly specifying (*i*) the feasible alternatives of active players according to previous play, (*ii*) what information accrues to players, and (*iii*) the material consequences of each complete play (terminal sequence of actions and, possibly, realizations of chance moves). Such rules should be amenable to a description using a formal mathematical language and this description should be independent of the personal features of the agents playing the game in each role in any particular instance of play. Indeed, we propose the following **separation principle** as a general methodological tenet: *the formal description of the rules of the game should be independent of the personal features of those who happen to play the game.*

This is easily done as far as the concerned personal features are related to tastes and preferences. For example, in simultaneous-move games with monetary consequences, one can describe preferences over monetary lotteries as von Neumann-Morgenstern utility functions, under a set of well understood behavioral assumptions. Appending such utility functions to the formal description of the rules of the game and adding assumptions about players' interactive knowledge (or interactive beliefs) about the relevant aspects of the situation of strategic interaction, one obtains a mathematical structure amenable to game-theoretic analysis.[1]

*We thank Carlos Alós-Ferrer, Manuel Arnese, Edoardo Bonaglia, Tommaso Crosta, and Julien Manili for useful comments.

[1]See, e.g., Osborne and Rubinstein (1994), or Battigalli, Catonini, and De Vito (2021).

Yet, as we consider games with sequential moves, it is not obvious that the extant game-theoretic formalism complies with the aforementioned separation principle. Indeed, with very few notable exceptions concerning multistage games (such as Myerson, 1986, and Myerson & Reny, 2020, whose purposes are different from ours), the information that accrues to players as the play unfolds is described by information partitions of the set of partial plays (or nodes): two partial plays x and y are indistinguishable by player i if and only if they belong to the same cell (equivalence class) of her information partition.[2] Now, one may legitimately wonder what makes two partial plays indistinguishable. In principle, it could be due to a genuine lack of information about how the game is unfolding, or to mnemonic limitations. In order to rule out the latter, it is usually assumed that information partitions satisfy a property called "perfect recall". In the words of Kuhn (1953, p. 213), perfect recall is "equivalent to the assertion that each player is allowed by the rules of the game to remember everything he knew at previous moves and all of his choices at those moves". However, it is unclear how the rules of the game could *prevent* players from remembering things they did or observed, as memory is a *subjective* attribute of players. In general the formalism is silent on where a player's *stock* of information comes from: should it be interpreted as the information that this player is able to retain and use, or as an objective representation of the cumulated information that accrued to her? The former interpretation implies that information partitions are hybrid representations mixing objective features of the rules of the game with a player's personal cognitive abilities, thus violating the separation principle.

In this paper, we put forward and analyze a general mathematical description of the rules of (finite) games, whereby we represent the *flows* of information accruing to players. Specifically, we assume that, throughout the game, players observe some signals informing them of how the play is unfolding. For instance, in a game played in a lab or on an online platform, a mediator may provide players with details about how the game has been played and with instructions about how to proceed (the feasible actions). The sequence of actions taken and signals observed form the stream of information potentially available to players. Importantly, this description is independent of players' personal features, as it only depends on the rules of interaction.

We then put forward a *game-independent*, rudimentary analysis of *memory*. In compliance with the separation principle, this allows us to combine a flow-based description of a game with a formal specification of players' personal mnemonic abilities. An analogy may help clarify our position: just like risk attitudes are framed within the theory of choice and only subsequently embedded in game-theoretic analysis, we believe that a formalization of agents' mnemonic abilities should come from a suitable "theory of memory" conceptually unrelated (although complementary) to game theory. We follow this route by appending to a given game described with information flows a specification of players' mnemonic features and thus obtain a game with possibility correspondences[3] (which may or may not be information partitions).

As a result, what a player deems possible about the play at a given point of the game is given by a combination of (i) the objective information she received during the game (e.g., she may rule out partial plays that are inconsistent with some action she took), and (ii) how such information is processed by her memory (e.g., she may forget which action she played in a given instance). This twofold process yields an accessibility relation among (partial and complete) plays describing what is deemed possible when a given play occurs. Our approach therefore cleanly decouples the objective information streams players have access to as the game

[2]Of course, we are considering work on general games. The literature on repeated games with imperfect monitoring (e.g., Mailath & Samuelson, 2006) provides a whole class of "exceptions", that is, works that describe information in compliance with the separation principle.

[3]This means that the formal description of a player's memory allows us to identify the partial or complete plays (nodes) that are deemed *possible* by such player at some point of the game, based on what she remembers.

unfolds and the subjective retention of information. On the one hand, the rules of interaction include relevant informational aspects – roughly speaking, who knows what about how the play is progressing. On the other hand, potentially imperfect mnemonic abilities may lead to a loss, or distortion of such game-specific information.

An explicit distinction of these two aspects is beneficial in several respects. First, it allows to seamlessly introduce mnemonic failures and heuristics in game-theoretic analysis. For instance, we are able to elucidate "absentmindedness", i.e., the possibility that a player forgets not just what actions she took, but also whether she took some action at all (see, e.g., Chapter 11 in Osborne & Rubinstein, 1994): according to our approach, absentmindedness is the consequence of quite natural personal cognitive limitations that our analysis of memory can easily capture. Second, from a game designer's perspective, one could also ask whether there is an "optimal" way to provide information to players who retain information in an imperfect or biased way: our framework provides the right tools to tackle the issue. Third, our flow approach naturally lends itself to the analysis of issues concerning the information of inactive players, which is key, for example, in the theories of self-confirming equilibrium and of psychological games,[4] or when one wants to model anticipatory feelings (Caplin & Leahy, 2001) and preferences for the temporal resolution of uncertainty (Kreps & Porteus, 1978).

We relate our approach to the traditional one by providing two possible interpretations of the perfect recall property of information sets. The first interpretation is that information partitions satisfying perfect recall are those that are generated by suitable flows of information in a setting where players' memory is perfect (Proposition 1). Alternatively, information partitions satisfying perfect recall can be thought of as those arising in situations where the rules of the game themselves constantly remind players of all the information that they had access to during the game (Proposition 3). The mere assumption of "perfect recall" does not clarify which interpretation one should take.

Moreover, we show that *every* standard information partition (even those that fail perfect recall) can be retrieved in our setting by considering a suitable combination of information flows and descriptions of memory (Proposition 2): this makes our approach at least as expressive as the traditional one. Furthermore, we are able to capture situations where a player's mnemonic failures imply a non-partitional information structure. This occurs, for instance, when the agent rules out the play that actually occurred, or when she is unable to correctly distinguish "similar" plays. Such situations are impossible to model using the standard formalism based on information sets.

Roadmap The remainder of this paper is organized as follows. Section 2 introduces some notation. Section 3 outlines our flow-based description of the rules of interactions. Section 4 proposes a flexible way to model memory. Section 5 relates our approach to the conventional one. Section 6 concludes and discusses the related literature. The Appendix collects proofs.

2 Preliminaries

Maintained assumptions In the following, we assume that the set of players I and the sets of actions A_i ($i \in I$) potentially available to each player i are *finite*. We also assume that the game always terminates after a finite sequence of actions, or profiles of simultaneously chosen actions. We denote by $L \in \mathbb{N}$ the maximal duration of the game. The finiteness assumption is

[4] On psychological games, see the survey by Battigalli and Dufwenberg (forthcoming). The self-confirming equilibrium idea was independently put forward by several authors, including Fudenberg and Levine (1993), who coined the term. See the literature review in the Discussion section of Battigalli, Cerreia-Vioglio, Maccheroni, and Marinacci (2015).

only a simplification; our approach extends seamlessly to infinite games (cf. Myerson & Reny, 2020 for the special case of multistage games).

Sequences and trees For a generic set X and $n \in \mathbb{N}$, we let $X^n := \bigtimes_{k=1}^{n} X$ denote the n-fold Cartesian product of X, with generic element $x^n = (x_k)_{k=1}^{n}$. By convention, we let $X^0 := \{\varnothing_X\}$ denote the singleton with the **empty sequence** of elements of X as its unique element; we often drop the subscript and just write $\varnothing$ when this causes no confusion. The empty set is denoted by the (different) symbol $\emptyset$. For $N \in \mathbb{N}$, $X^{\leq N} := \bigcup_{n=0}^{N} X^n$ denotes the set of sequences of elements of X of length N or less. Superscripts are used to denote the length of a given sequence (i.e., we write x^n to denote a generic element of X^n), but sometimes the length of a given sequence may be left unspecified (i.e., we may write ξ for a generic element of $X^{\leq N}$). For each $\xi \in X^{\leq N}$, we let $\ell(\xi)$ denote the **length** of ξ. Finally, we let $\preceq$ denote the reflexive "prefix of" relation defined as follows: for each $(a_k)_{k=1}^{m}$ and $(b_k)_{k=1}^{n}$ in $X^{\leq N}$, $(a_k)_{k=1}^{m} \preceq (b_k)_{k=1}^{n}$ if $m \leq n$ and $a_k = b_k$ for each $k \in \{1, \dots, m\}$. We say that $V \subseteq X^{\leq N}$ is a **tree** if $v' \in V$ and $v \preceq v'$ imply that $v \in V$ – i.e., V is a tree if it is closed under the "prefix of" relation $\preceq$ (cf. Kechris, 1995, Definition 2.1).[5]

Possibility correspondences For an abstract set X, a **possibility correspondence** on X is a map $\mathcal{P} : X \rightrightarrows X$ associating each element x of X with a corresponding subset $\mathcal{P}(x) \subseteq X$. The conventional interpretation, which justifies the terminology used, is that X is a set of states, and $\mathcal{P}(x)$ $(x \in X)$ is the set of states that are deemed possible by an agent when the true state is x (cf. Chapter 5 of Osborne & Rubinstein, 1994). Note that possibility correspondences on X are equivalent to binary relations over X. In particular, $\mathcal{P} : X \rightrightarrows X$ is represented by the unique binary relation $R_{\mathcal{P}} \subseteq X \times X$ such that $xR_{\mathcal{P}}y$ $(x, y \in X)$ if and only if $y \in \mathcal{P}(x)$. Vice versa, for each binary relation $R \subseteq X \times X$, there is a unique $\mathcal{P}_R : X \rightrightarrows X$ such that, for each $x, y \in X$, $y \in \mathcal{P}_R(x)$ if and only if xRy.

3 Flows of information

We call "**game structure**" the mathematical description of the rules of interaction, *without* the map from complete paths to outcomes, which is irrelevant for our analysis (see Battigalli, Leonetti, & Maccheroni, 2020). We relate descriptions featuring "flows" of information to descriptions featuring "stocks" of information, represented by information sets.

We assume that, throughout the game, each player receives some messages about the previous play, in addition to observing the action she just played. Such messages play a double role: on one hand, they (perhaps imperfectly) inform players of how the game has unfolded; on the other hand, they inform players of the actions they can take. For instance, in an ascending auction played on an online platform, a player may be notified right after the beginning that "the first bid was \$100: bid at least \$101 to continue". Or, more generally, messages could look like: "your opponent moved, now you can choose a or b", or "your opponent chose c, now you can choose a or b". The bottom line is that such messages can be more or less informative about the behavior of others, but they provide players with all the instructions needed to play the game (which requires, of course, knowing one's own feasible actions).

3.1 Key ingredients

We begin by illustrating the mathematical objects forming a game structure.

[5] A tree V defined in this way can be naturally mapped into an equivalent tree defined in a graph-theoretic fashion. In particular, the set of vertices of such graph is isomorphic to V, and every two distinct vertices u, v are connected by a path if and only if $u \preceq v$ or $v \preceq u$.

Players, actions, and messages Recall that we have a finite set of **players** I, and, for each player $i \in I$, a finite set A_i of potentially available **actions**. As already mentioned, we assume that players receive some game-specific messages as the play unfolds: for each $i \in I$, M_i is the finite set of **messages** that player i can receive throughout the play.

Given that not all players necessarily move or receive messages at the same time, it is convenient to define, for each nonempty $J \subseteq I$, the sets $A_J := \bigtimes_{i\in J} A_i$ and $M_J := \bigtimes_{i\in J} M_i$ of profiles of actions and messages of players in subset J. We let $A := \bigcup_{J\in 2^I\setminus\{\emptyset\}} A_J$ and $M := \bigcup_{J\in 2^I\setminus\{\emptyset\}} M_J$. In words, $a \in A$ is a profile of actions $(a_i)_{i\in J}$ for an unspecified and nonempty $J \subseteq I$, and an analogous interpretation applies to elements of M.

Formally, profiles are functions that associate each player in a subset $J \subseteq I$ with a corresponding object, such as an action or a message. We are mostly interested in describing which players receive a message when a generic message profile is generated. To this end, define $\mathcal{D} : M \rightrightarrows I$ to be such that, for each $m \in M$, $\mathcal{D}(m) \subseteq I$ is the *domain* of m. As a matter of terminology, we say that player $i \in I$ is **alert** given message profile $m \in M$ if $i \in \mathcal{D}(m)$. Equivalently, m alerts player i. The interpretation is that alert players are the ones who receive and process information when a message profile is generated.

"Being (un)alert" is a non-standard concept in game theory. To fix ideas, one may consider a game where players interact on some IT platform and they move sequentially, one at a time. The last player to move can be informed in different ways about the moves of others. For instance, she could receive a message after each of her co-players' turn (this message could simply state "player j moved, wait for your turn"): in this case he would be alert after each stage of the game, despite not being able to move when it is not his turn. Or, she could receive no message at all until her turn (and when her turn comes, she could receive a message saying "it is your turn, you can choose your action"): in this case she would be unalert during whole duration of the game until her turn. This is just an heuristic illustration, and we will shortly comment on the difference between being alert and being active, and on why it matters.

A rule to determine feasible actions We said that messages inform players of their feasible actions. In particular, players *do not have to remember* the actions they previously took or the messages they previously received to be able to understand what they can do: the message just received also encodes the set of feasible actions. To formalize this idea, we introduce, for each player $i \in I$, an **action feasibility correspondence** $\mathcal{A}_i : M_i \rightrightarrows A_i$. Thus, $\mathcal{A}_i(m_i) \subseteq A_i$ ($i \in I$, $m_i \in M_i$) is the set of actions that player i can take after receiving message m_i.

Our definition of $\mathcal{A}_i$ ($i \in I$) implies that being alert is a prerequisite for acting – that is, a player can move only if she receives some message, and hence only if she has been alerted by (her coordinate of) the profile $m \in M$ just realized. Thus, messages have the role of "guiding" players as the game unfolds, providing them with some information about the play and about their feasible actions. A message could even simply inform a player that her turn came and that she can take a move – moving only at one's own turn is a prerequisite for playing the game in compliance with the rules. Hence, we say that player $i \in I$ is **active** after $m \in M$ if $i \in \mathcal{D}(m)$ and $|\mathcal{A}_i(\mathrm{proj}_{M_i} m)| > 1$.[6] As mentioned, moves are taken after receiving information, so a player is active after some message profile if (*i*) she received some information, and if (*ii*) she can choose from a set of at least two feasible actions based on the piece of information she received. Note that point (*ii*) allows an alert player to have only one feasible action (i.e., a dummy action "wait", which is neglected in our notation) – in such case, a player is alert but inactive, meaning that she receives and processes information without acting. An empty set of feasible actions means "game over".

[6] We denote as proj_X the canonical projection onto set X. Taking projections of profiles amounts to extracting some coordinates from the profile. For instance, if $m = (m_j)_{j\in J}$ for some $J \subseteq I$ and $i \in J$, then $\mathrm{proj}_{M_i} m = m_i$.

Note that decoupling the concepts of "being alert" and "being active" allows to describe the information that accrues to players when they are inactive, which includes the end-game information. As hinted in the Introduction, the end-game information is key in the theory of self-confirming equilibrium, as well as in settings where utilities depend on terminal beliefs. The information available to inactive players during the game is instead important if one is to model, for example, anticipatory feelings (Caplin & Leahy, 2001) or preferences for the temporal resolution of uncertainty (Kreps & Porteus, 1978).

Lastly, it is convenient to define the joint feasibility correspondence $\mathcal{A} : M \rightrightarrows A$, as $m \mapsto \bigtimes_{i\in\mathcal{D}(m)} \mathcal{A}_i(\mathrm{proj}_{M_i} m)$. Thus, $\mathcal{A}(m)$ ($m \in M$) is the set of action profiles that may be taken (by alert players) after m.

A rule to generate messages We represent how messages are generated by means of a **collective feedback function** $\tilde{f} : A^{\leq L} \to M$. Such function associates each conceivable sequence of action profiles (of length equal to L at most) with the corresponding profile of messages to alert players. For instance, an element of $A^{\leq L}$ may represent a sequences of bids in an ascending auction, and for each of them a message profile is generated according to the map $\tilde{f}$. Such message could inform players, e.g., of the highest bid.

Note that $A^{\leq L}$ may contain sequences of action profiles that are unfeasible given the rules of the game. For instance, in an ascending auction, feasible sequences of bids have to be increasing, but the set $A^{\leq L}$ also contains sequences of bids that do not have such feature. Obviously, we are ultimately interested in considering the restriction of $\tilde{f}$ to the set of *feasible* sequences of action profiles. Yet, such set is a derived object, and it needs to be retrieved in a recursive way by exploiting the action feasibility correspondences $(\mathcal{A}_i)_{i\in I}$ and the feedback function $\tilde{f}$.

Some restrictions We can combine feasibility correspondences and message-generating functions to see how they shape the game unfolding. Before doing so, however, we have to impose some natural joint restrictions on $(\mathcal{A}_i)_{i\in I}$ and $\tilde{f}$.

- *Knowledge that the game started.* It is reasonable to assume that the first message profile to ever be generated alerts everyone. Obviously, not all players need to be active at the beginning of the game, but all players should be informed that the game started. In this regard, the first message player $i \in I$ receives may be thought of as stating "the game starts" and specifying i's initially feasible actions. Formally, we impose:

$$\mathcal{D}(\tilde{f}(\varnothing_A)) = I. \tag{KGS}$$

- *Knowledge that the game ended.* The game ends after $m \in M$ if $\mathcal{A}(m) = \emptyset$. In such case, we say that m **terminates the game**. Specifically, given that $\mathcal{A}(m) = \bigtimes_{i\in\mathcal{D}(m)} \mathcal{A}_i(\mathrm{proj}_{M_i} m)$ for each $m \in M$, we require that whenever $\mathcal{A}_j(\mathrm{proj}_{M_i} m) = \emptyset$ for some $j \in \mathcal{D}(m)$, then $\mathcal{A}_i(\mathrm{proj}_{M_j} m) = \emptyset$ for all $i \in \mathcal{D}(m)$. Therefore, as soon as the game is over for some player (i.e., such player does not have any feasible actions) it is over for everyone. Just like it was reasonable to require that all players be informed of the game start, it is equally compelling to require that all players be informed of the game end. That is, whenever a message profile terminates the game, then it must alert everyone. Furthermore, since the maximal duration of the game is L, the game must end after each sequence of action profiles of length L. Formally:

$$\begin{aligned} &\forall\, m \in M, \quad \mathcal{A}(m) = \emptyset \Longrightarrow \mathcal{D}(m) = I, \\ &\forall\, a^L \in A^L, \quad \mathcal{A}(\tilde{f}(a^L)) = \emptyset. \end{aligned} \tag{KGE}$$

Game structure We can now give the definition of game structure.

Definition 1 *A* ***flow-based game structure*** *is a tuple*

$$\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i \in I} \rangle,$$

where the elements are as above, and satisfy (KGS) *and* (KGE).

Importantly, all the mathematical objects forming a flow-based game structure represent properties of the game that do not hinge in any way on the personal features of those who happen to play the game. To see this, note that it would be possible to give an explicit and objective description of each of the objects in Definition 1 if one were to design, for example, a lab experiment.

3.2 The possible plays

Our objective is now to derive all the possible ways in which the game may unfold given the rules of interaction $\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i \in I} \rangle$. In the following, we distinguish between (partial or complete) **plays**[7] and **extended histories**:[8] the former are taken to be feasible sequences of action profiles, and the latter are instead feasible sequences of *action and message* profiles. We will also define a feedback function f, obtained by restricting the collective feedback function $\tilde{f} : A^{\leq T} \to M$ to the set of feasible sequences of action profiles (i.e., to the set of plays).

The starting point is obvious: define the sets of length-0 plays and extended histories as $P^0 := \{\varnothing\}$ and $H^0 := \{\varnothing\}$.[9] Then, let $f^0 : P^0 \to M$ be such that $f^0(\varnothing) := \tilde{f}^0(\varnothing)$.

Given P^0, H^0, and f^0, we can obtain the set of length-1 plays and extended histories as

$$P^1 := \mathcal{A}(f^0(\varnothing)), \quad H^1 := \{f^0(\varnothing)\} \times \left(\bigcup_{a^1 \in P^1} \{a^1\} \times \{f^1(a^1)\} \right),$$

where, $f^1 : P^1 \to M$ is obtained by restricting $\tilde{f}$ to P^1. Lastly, the set of length-1 terminal extended histories is[10]

$$Z^1 := \{(m_0, a_1, m_1) \in H^1 : \mathcal{A}(m_1) = \emptyset\}.$$

Note that terminal histories feature as last element a message profile with domain I that terminates the game, and this verifies condition (KGE).

Assume now that P^ℓ and H^ℓ have been defined for $0 < \ell < L$. Then,

$$\begin{aligned}
P^{\ell+1} &:= \bigcup_{a^\ell \in P^\ell} \{a^\ell\} \times \mathcal{A}(f^\ell(a^\ell)), \\
H^{\ell+1} &:= \bigcup_{h \in H^\ell} \{h\} \times \left(\bigcup_{a \in \mathcal{A}(f^\ell(\mathrm{proj}_{P^\ell} h))} \{a\} \times \left\{ f^{\ell+1}\big((\mathrm{proj}_{P^\ell} h, a)\big) \right\} \right), \\
Z^{\ell+1} &:= \{(m_0, \ldots, a_{\ell+1}, m_{\ell+1}) \in H^{\ell+1} : \mathcal{A}(m_{\ell+1}) = \emptyset\},
\end{aligned}$$

[7]The term "play" has been used in the literature to refer to complete plays, which correspond to terminal nodes (cf. Kuhn, 1953). However, we find it convenient to use the same term to include also partial plays, which correspond to non-terminal nodes.

[8]Our representation in terms of sequences of (messages and) actions is similar to that of Osborne and Rubinstein (1994), who use the term "history" to denote feasible sequences of (profiles of) actions. Here the adjective "extended" refers to the fact that our formalism includes profiles of messages in addition to profiles of actions. In this way, our terminology clearly distinguishes between sequences of action profiles (plays) and sequences of action and message profiles (extended histories).

[9]Recall that we drop the range subscript from the empty sequence symbol when no confusion may arise; e.g., we write just $\varnothing$ instead of $\varnothing_A$ as the unique element of A^0.

[10]The length of extended histories is understood as the length of the induced play.

where f^ℓ and $f^{\ell+1}$ are the restrictions of $\tilde{f}$ to P^ℓ and $P^{\ell+1}$, respectively.

Wrapping up, we let $P := \bigcup_{\ell=0}^{L} P^\ell$ be the set of feasible **plays**, $H := \bigcup_{\ell=0}^{L} H^\ell$ the set of feasible **extended histories**, and $Z := \bigcup_{\ell=1}^{L} Z^\ell$ the set of feasible **terminal extended histories**. Let $f : P \to M$ denote the map $a^\ell \mapsto f^\ell(a^\ell)$. A **play** $p \in P$ is **terminal** if $f(p)$ terminates the game (or, equivalently, if it is induced by a terminal extended history). Lastly, player $i \in I$ is **alert after play** $p \in P$ if $i \in \mathcal{D}(f(p))$. We let $P_i := \{p \in P : i \in \mathcal{D}(f(p))\}$ denote the set of plays after which player i is alert.

With the recursive construction of set P, it is easy to appreciate that the rules of the game allow play $(a_k)_{k=1}^{m}$, only if they allow every prefix $(a_k)_{k=1}^{\ell}$ ($\ell \leq m$). Therefore:

Remark 1 The set of possible plays $P \subseteq A^{\leq L}$ is a tree.

A generic extended history has the form $h = (m_0, (a_1, m_1), \ldots, (a_\ell, m_\ell))$. Hence, each extended history features an initial profile of messages followed by a sequence of pairs of action and message profiles. This formalism highlights the double role of messages. On one hand, they inform players of their feasible actions, and hence an initial message profile m_0 is needed to inform players of what they can do at the beginning. On the other hand, messages give players information about previous play. In particular, a terminal message profile m_ℓ lets players know the game is over and may give them further information about the just completed play.

We conclude this subsection with an illustrative example.

Example 1 (Committee) We show how to derive the tree of feasible plays from a set of rules of interaction. To appreciate the flexibility of the proposed approach, one may read the present example with the eyes of a game designer (e.g., a company's board constituting a committee) deciding how to shape the interactions: she has to decide who can do what and who knows what about previous play. To further fix ideas, one could even imagine that the players are not physically in the same place, but they interact by means of some IT platform.

At the root, all players get the message "The game is starting, Ann can choose A, D, or O". Then, the set of feasible actions for Ann is $\{A, D, O\}$, while the other players are inactive. If Ann chooses to opt out (i.e., if she plays O), she terminates the game and all players observe a "game over" message. If Ann chooses D, Ann and Dave get to observe the message "It is Dave's turn: he can choose L or R". Hence, Dave is the only active player (with feasible actions set $\{L, R\}$, Ann is alert (because she received some information) but inactive, and Bob and Chloe are not alert. After Dave's choice, the game ends and everyone receives a "Game over" message. If instead Ann chooses A, Bob and Chloe privately receive the (same) message "It is your turn: you can choose between U and D". They simultaneously choose their action from $\{U, D\}$, while Ann and Dave are not alert. If they both choose U, the game ends and all players receive a "Game over" message. If they both choose D, Ann and Dave receive the message "It is Dave's turn: he can choose L or R", and the game terminates after Dave's choice – just like in the eventuality of Ann choosing D. If instead Bob and Chloe choose different actions, Ann and Dave receive the message "Bob and Chloe could not reach an agreement: Dave has to break the tie by choosing U or D". After Dave's decision, players are notified that the game is over.

The tree of feasible (partial and complete) plays is depicted in Figure 3.

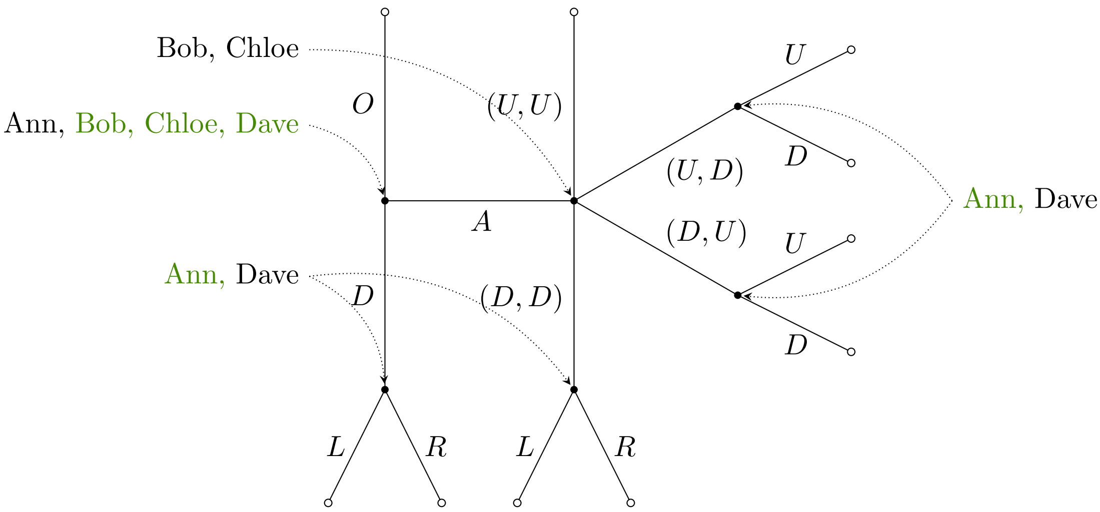

Figure 1 Active players in black, alert but inactive players in green.

Note that when Dave observes the message "Bob and Chloe could not reach an agreement: Dave has to break the tie by choosing U or D", he can infer Ann's action (A) at the beginning of the game, because he knows that Bob and Chloe get to move if and only if Ann chooses A. On the other hand, the message Dave receives before choosing between L and R is silent about whether Bob and Chloe acted at all. Upon seeing such message, the rules of the game do not allow Dave to know if Ann gave him the move at the beginning of the game or if Bob and Chloe played in the meantime. For this to occur, Dave must be unalert after play (A).[11]

This game structure can be interpreted as a way in which a committee of four people can decide between two alternatives (U and D, while L and R correspond to Dave authorizing D). In light of such interpretation, we should think that the player labels we used represent *roles* rather than actual people (e.g., Ann could stand for the chair, who is always informed about the play, and who gets to decide whether to consult Dave or Bob and Chloe, or even if a decision is taken at all). Then, the rules of interaction we described are the procedure the committee follows to make decisions. For instance, such procedure might have been approved by a company's board and spelled out in the company's charter. In compliance with our separation principle, such rules of interactions do not hinge on the personal features of those who will end up serving in the committee, whose identity might even be unknown at the moment in which the procedure is approved. ▲

3.3 Making inferences about the play

Players are clearly interested in inferring how the game is unfolding. However, as the play proceeds, they are informed only of the actions they take and of the messages they receive. That is, an extended history $h = (m_0, (a_1, m_1), \ldots, (a_\ell, m_\ell))$ allowed by the rules of the game induces a play $p = (a_1, \ldots, a_\ell)$, but player i only observes the sequence of actions she played and messages she received (which may be less than the total number of moves made or messages received, as player i may be unalert in some instances). We now formalize the reasoning players can carry out based on the streams of information they observe during the game.

To begin with, it is crucial to "extract" from each extended history in H the pieces of information each player is exposed to during such history. To this end, for each player $i \in I$, we

[11] Even if after action A Dave is only told "Ann just moved," after $(A, (D, D))$ he could infer what happened from the fact that he received two messages (with the second one telling him that he can choose L or R), not just one.

let $c_i : M \cup A \to M_i \cup A_i$ be defined, for each $b \in M \cup A$, as

$$c_i(b) := \begin{cases} \operatorname{proj}_{M_i \cup A_i} b & \text{if } i \in \mathcal{D}(b); \\ \varnothing_{M_i \cup A_i} & \text{otherwise.} \end{cases}$$

In words, $c_i(\cdot)$ $(i \in I)$ isolates i's component (if any) from a profile of messages or actions. Then, we can obtain the information available to player i along an extended history with the map $C_i : H \to \{m_0\} \times (A_i \times M_i)^{\leq L}$, which is defined as $(m_0, (a_k, m_k)_{k=1}^{\ell}) \mapsto (m_{i,0}, (c_i(a_k), c_i(m_k))_{k=1}^{\ell})$.[12] Thus, $C_i(h)$ $(i \in I,\ h \in H)$ is the personal history experienced by i given extended history h, and it can be interpreted as the *cumulated information* (i.e., actions played and messages received) player i has access to given h. We may also refer to $C_i(h)$ $(h \in H)$ as the *stream of information, or personal history* experienced by player i within extended history h. The set of **personal histories** player i can experience as the game unfolds is defined as $H_i := C_i(H)$.

How personal histories allow players to make inferences about the realized plays is intuitive: as a player receives messages about how the game is unfolding, she is able to combine such stream of pieces of information with the actions she took to identify the set of plays that are consistent with this evidence.[13] Note that, for play $(a_1, \ldots, a_t) \in P$, the *unique* extended history in H consistent with it is $(m_0, a_1, f^1(a_1), a_2, f^2((a_1, a_2)), \ldots, a_\ell, f^\ell((a_1, \ldots, a_\ell)))$, where uniqueness essentially follows from the fact that feedback is deterministic. It is indeed possible to check that the projection $\operatorname{proj}_P : H \to P$ (that maps each extended history to the corresponding induced play) is a bijection because messages are determined by plays by means of the feedback function.[14] Therefore:

Remark 2 The set of feasible plays P and the set of feasible extended histories H are isomorphic, so that the isomorphism preserves the partial order induced on the two sets by the "prefix of" relation $\preceq$.[15]

For each play $p \in P$, we let $E(p)$ denote the unique extended history inducing it (uniqueness follows from Remark 2). Formally, $E : P \to H$ is the inverse of the projection $\operatorname{proj}_P : H \to P$ mentioned earlier. Note that $E(\cdot)$ is well-defined because proj_P is a bijection.

Note that two extended histories $g, h \in H$ cannot be distinguished by player i $(i \in I)$ if they correspond to the same personal history (or cumulated information) of i, that is, if $C_i(g) = C_i(h)$. Thus, we say that two plays after which player $i \in I$ is alert $p, q \in P_i$ are **indistinguishable** for i if the extended histories in H inducing them result in the same stream of information for i – that is, if $C_i(E(p)) = C_i(E(q))$. To ease notation, for each $i \in I$, we let $F_i := C_i \circ E : P_i \to H_i$ denote the map from plays where i is alert to personal histories for i. Finally, let $\mathcal{F}_i : P_i \rightrightarrows P_i$ $(i \in I)$ be the possibility correspondence defined, for each $p \in P_i$, as

$$\mathcal{F}_i(p) := \{q \in P_i : F_i(p) = F_i(q)\}.$$

[12]Note that for a generic set X and for each sequence ξ of elements of X, the following holds: $(\xi, \varnothing_X) = (\varnothing_X, \xi) = \xi$. This is why we can neglect every $\varnothing_{M_i \cup A_i}$ we may have in the sequence $(C_i(b_k)_{k=1}^{t})$.

[13]A necessary prerequisite for doing that is that pieces of evidence have to be retained/memorized. Given that we have not yet introduced our account of players' mnemonic abilities, the inferential reasoning described here should be understood as the reasoning that is in principle possible given the rules of interaction.

[14]To check that proj_P is injective, consider $h, h' \in H$ such that $h \neq h'$. If the two extended histories have different lengths, then so will their projections, proving $\operatorname{proj}_P h \neq \operatorname{proj}_P h'$. Assume then $\ell(h) = \ell(h')$ and proceed by induction. For the basis step, let $\ell(h) = \ell(h') = 1$, so that $h = (a, f(a))$ and $h' = (b, f(b))$ for some $a, b \in A$. Obviously, $h \neq h'$ if and only if $a \neq b$, in which case we also have $\operatorname{proj}_P h = a \neq b = \operatorname{proj}_P h'$. The proof of the inductive step is analogous. Surjectivity of proj_P is obvious.

[15]That is, if a play p is a predecessor of play q, then the extended history induced by p is a predecessor of the extended history induced by q.

It is immediate to notice that, for each $i \in I$, $\mathcal{F}_i$ represents an equivalence relation, so that its range is a partition of P_i. In light of this, we say that $\mathcal{F}_i$ is *partitional*.[16]

From a conceptual point of view, $\mathcal{F}_i$ represents the inferences player i can carry out about the true play based on the rules of the game. Hence, the notion of indistinguishability of plays is in some sense "objective", because it is derived from the rules of interaction without hinging on the personal features of those who will happen to play the game.

Example 2 (Committee, continued) By inspection of the rules of interaction of our running example, it is easy to see that Dave only gets to observe two personal histories (obviously in addition to the empty history), which correspond to the two messages m' = "It is Dave's turn: he can choose L or R" and m'' = "Bob and Chloe could not reach an agreement: Dave has to break the tie by choosing U or D". After the personal histories (m') and (m''), he is also active. The plays consistent with (m') are (D) and $(A,(D,D))$ – that is, $F_{Dave}(D) = F_{Dave}(A,(D,D)) = (m')$. After message m' he cannot know if Ann gave him the move straight away or if Bob and Chloe also acted in the meantime because he receives no message if Ann chooses A. Similarly, it is possible to see that $F_{Dave}(A,(D,U)) = F_{Dave}(A,(U,D)) = (m'')$. Therefore, the set of plays where Dave is alert $P_{Dave} = \{\varnothing,(D),(A,(D,D)),(A,(D,U)),(A,(U,D))\}$ is partitioned through the correspondence $\mathcal{F}_{Dave}$ as $\big\{\{\varnothing\},\{(D),(A,(D,D))\},\{(A,(D,U)),(A,(U,D))\}\big\}$. ▲

3.4 More on feedback functions

To give a better sense of the feedback functions, we introduce some terminology and we discuss some examples. To this end, it is convenient to refer to a player's **individual feedback function**, which directly specifies the message a given player $i \in I$ would observe after each feasible play (if any).[17] Recall that P_i $(i \in I)$ is the set of plays alerting player i. Then, for each player $i \in I$, let $f_i : P_i \to M_i$ be defined, for each $p \in P_i$ as $f_i(p) := \mathrm{proj}_{M_i} \tilde{f}(p)$.

We say that feedback is **perfect** if for each player the feedback received allows that player to exactly infer the actions chosen by others, regardless of the actions she played. Formally, for each player $i \in I$ and for each pair of plays $p, p' \in P_i$, $\mathrm{proj}_{A_{-i}^{\leq L}}\, p \neq \mathrm{proj}_{A_{-i}^{\leq L}}\, p'$ implies $f_i(p) \neq f_i(p')$. That is, whenever two sequences of action profiles differ in the actions chosen by players other than i, they result in different messages for i.

We say that feedback is **cumulative** for a given player if new messages remind that player of previously available pieces of information (i.e., actions chosen and messages received). Formally, for each pair of plays $p, p' \in P_i$ $(i \in I)$ with $\ell(F_i(p)) = \ell(F_i(q))$, $\mathrm{proj}_{A_i^{\leq L}}\, p \neq \mathrm{proj}_{A_i^{\leq L}}\, p'$ or $f_i(p) \neq f_i(p')$ imply that, for each pair of successors $q \succeq p$ and $q' \succeq p'$ with $\ell(F_i(q)) = \ell(F_i(q'))$, $f_i(q) \neq f_i(q')$. The condition says that, for each player, whenever two sequences of action profiles differ in the information they convey to that player (via either the actions chosen or the feedback observed), then also subsequent action profile sequences will result in different messages. To put it differently, future feedback incorporates past information, which includes information about own actions in previous stages.

Note that to assess cumulativeness of feedback it makes sense to compare only plays inducing personal histories of the same length – if p and p' in the definition above were such that $p \prec p'$, $\mathrm{proj}_{A_i^{\leq L}}\, p \neq \mathrm{proj}_{A_i^{\leq L}}\, p'$ would trivially hold, but the requirement $f_i(q) \neq f_i(q')$ would fail for $q = q' \succ p' \succ p$ (if such q exists). Moreover, note that with cumulative feedback $\mathrm{proj}_{A_i^{\leq L}}\, p \neq \mathrm{proj}_{A_i^{\leq L}}\, p'$ (for $p, p' \in P_i$ with $\ell(F_i(p)) = \ell(F_i(p'))$) by itself implies $f_i(p) \neq f_i(p')$.

[16] A possibility correspondence $\mathcal{P} : X \rightrightarrows X$ is partitional if it represents an equivalence relation. Equivalently, it has to satisfy: (*i*) reflexivity (for each $x \in X$, $x \in \mathcal{P}(x)$), (*ii*) symmetry (for each $x, y \in X$, $x \in \mathcal{P}(y)$ implies $y \in \mathcal{P}(x)$), and (*iii*) transitivity (for each $x, y, z \in X$, $y \in \mathcal{P}(x)$ and $z \in \mathcal{P}(y)$ imply $z \in \mathcal{P}(x)$).

[17] The individual feedback map is a derived object, so we directly define its domain to be a subset of the set of plays (another derived object).

This is because we can take $p \succeq p$ and $p' \succeq p'$, and cumulativeness requires $f_i(p) \neq f_i(p')$. This is consistent with the idea that cumulative feedback reminds players also of the actions they took. All in all, when feedback is cumulative, two plays that induce different personal histories for a player must also result in different messages for that player:

Remark 3 Fix $i \in I$ and $p, q \in P_i$. If feedback is cumulative, $f_i(p) = f_i(q)$ if and only if $F_i(p) = F_i(q)$.

The proof of this remark is in the appendix (see Lemma A1). A couple of examples may shed light on the terminology, as well as on the nature of feedback.

Example 3 (Chess) In chess, players observe all the moves performed at each stage, as well as the resulting positions of pieces on the chessboard. However, information about the previous play differs based on how and where the game is played.

Consider first a *friendly, in-person match* played by amateurs, where we assume that players do not write down the moves they take. In such setting, in each stage they observe the move of the active player and the end-of-stage positions of pieces on the board, but they are not reminded of past play. If players recall some past moves, then it is just because they memorized them. Hence, feedback is perfect but not cumulative.

If instead the game is played in a *competitive tournament*, or on an *online platform*, the rules of the game themselves provide players with a complete account of the game unfolding. Indeed, in both such settings the log of moves taken throughout the game is publicly available. This makes feedback both perfect and cumulative.

Lastly, *blind chess* provides yet another feedback structure: in such game, the only available feedback pertains to the last move taken, and players have to remember past moves and figure out the positions of pieces on the board.[18] As in the first case, feedback is perfect but not cumulative. ▲

Example 4 (Auctions) Suppose that a number of agents repeatedly engage in first-price sealed-bid auctions. This may be due, for example, to the fact that multiple items are being sold. After each round of bids, players may have access to different forms of feedback. For instance, players may be informed only of the current-round winning bid: in such case feedback is neither perfect nor cumulative. If instead players are told the sequence of winning bids of all previous rounds, feedback is cumulative but not perfect. Specifically, "cumulativeness" comes from the fact that the feedback of a given stage incorporates previous feedback, as the communication of the previous-rounds winning bids is repeated as the interaction progresses. For feedback to be both perfect and cumulative, players must be informed after each round of *all* the bids made in previous rounds. ▲

4 Memory

In this section, we present a very basic and unstructured way to model individuals' ability to memorize information, which is *independent* of any specific context or game. As already stressed, our flow-based approach allows to cleanly and explicitly isolate players' game-specific information as implied by the rules of interaction, in compliance with the separation principle. Therefore, a rigorous and expressive language can be derived by combining these two accounts of objectively

[18] Of course, the rules have to account for the possibility that a player with imperfect memory, or an imperfect ability to figure out the positions of pieces on the board, might attempt an illegal move. For example, they could stipulate that moves consist of instructions, every instruction can be given, and instructions to execute illegal moves terminate the game with the loss of the moving player.

provided and subjectively retained information. In particular, appending a description of players' cognitive features to a flow-based game structure allows to embed a wide array of cognitive limitations into specific interactive situations, and it allows to analyze them in a rigorous way. We believe this to be a key step towards the meaningful integration of cognitive failures into game-theoretic analyses.

To keep our analysis as general as possible, we are going to model memory using possibility correspondences. To be precise, memory can be thought of as a two-step process involving both the *storage* and the subsequent *retrieval* of information (see Kahana, 2012 for an overview). The latter process may well depend on the environmental cues an individual may observe: for instance, a cue may facilitate the retrieval of similar past experiences, while inhibiting that of less similar ones (see, e.g., Bordalo, Gennaioli, & Shleifer, 2020). In this setting, we are focusing on the first channel – that is, on how information is stored by agents. However, the analysis of the current section is abstract enough that it can easily be enriched in order to model phenomena involving cued recall of stored information. We first present a game-independent account of memory, and subsequently we embed such analysis in a game-theoretic framework.

4.1 A game-independent analysis of memory

Consider a generic set X, and the set $X^{<\mathbb{N}} := \bigcup_{n\in\mathbb{N}} X^n$ of sequences of elements of X of finite length. To ease the comparison with previous sections, X should be interpreted as a universal set of actions and messages – that is, $X = A^\star \cup M^\star$, where $A^\star$ and $M^\star$ are universal sets of actions and of messages, respectively. With the term "universal" we want to stress the idea that these are sets of conceivable actions the agent can take or of messages the agent can imagine to receive, and such actions and messages are not tied to any specific context or situation. Obviously, when a description of an agent's memory is embedded in a game tree, the set of actions she can take and of messages she can receive will be determined by the rules of interaction. Elements of $X^{<\mathbb{N}}$ instead represent sequences of pieces of information an agent can observe. In general, we may refer to such sequences of pieces of information as "information streams". A **memory correspondence**[19] for agent i is a nonempty-valued possibility correspondence $\mathcal{M}_i : X^{<\mathbb{N}} \rightrightarrows X^{<\mathbb{N}}$. The interpretation is the following: after observing an information stream $\xi \in X^{<\mathbb{N}}$, $\mathcal{M}_i(\xi) \subseteq X^{<\mathbb{N}}$ is the set of sequences that are consistent with what agent i retained of the actual sequence ξ. In other words, after observing an information stream ξ, agent i cannot rule out the information streams in the set $\mathcal{M}_i(\xi)$. Depending on her ability to memorize information, the set $\mathcal{M}_i(\xi)$ will vary in size – the larger such set the higher the number of information streams the agent is unable to rule out. In this sense, $\mathcal{M}_i$ is ultimately a description of how precisely agent i stores and retains the information she receives.[20]

The most demanding specification is the following.

Example 5 (Perfect memory) Agent i exhibits *perfect memory* if $\mathcal{M}_i$ is the identity correspondence $\xi \mapsto \{\xi\}$. ▲

This way of formalizing an agent's cognitive abilities allows to flexibly relax the assumption of perfect retention of information and retrieval in interesting ways. For instance, allowing for bounded memory may be plausible in many economic applications involving interactions that are repeated many times. In such cases, it may be reasonable to assume that players fail to keep track of all the outcomes of previous interactions.

[19] Note that we use the term "memory" when we consider an agent's personal cognitive feature, and we keep "(perfect) recall" to mean a property of information partitions in standard game structures.

[20] As mentioned, we are focusing on the "information storage" step, rather than on the "information retrieval" one. Our formalism could be augmented to model also the latter process, but we leave this to future research.

Example 6 (Bounded memory) Agent i exhibits *k-bounded memory* if $\mathcal{M}_i$ is the correspondence

$$(x_\ell)_{\ell=1}^n \mapsto \big\{(y_\ell)_{\ell=1}^n \in X^{\leq N} : \forall\, j \in \{0,\dots,\min\{k,n\}\},\ y_{n-j} = x_{n-j}\big\}.$$

For an illustration, consider $\{0,1\}^{\leq 3}$, and assume agent i exhibits 1-bounded memory. Then, $\mathcal{M}_i\big((0,0,1)\big) = \big\{(1,1,1),(1,0,1),(0,1,1),(0,0,1)\big\}$.

Note that this definition assumes that agent i can recall the number of pieces of information she observed. That is, all the sequences in $\mathcal{M}_i(\xi)$ ($\xi \in X^{<\mathbb{N}}$) must have the same length of ξ. This assumption can easily be dispensed with. In such case, we would have, for example, $\mathcal{M}_i\big((0,0,1)\big) = \big\{(1),(0,1),(1,1),(1,1,1),(1,0,1),(0,1,1),(0,0,1)\big\}$. ▲

Another interesting case is when an agent is able to remember all the pieces of information she received, and how many times she received them, but fails to memorize their order. In other words, the length of the sequence of pieces of information observed is retained, and so is the frequency of each observed piece of information.[21]

Example 7 (Statistical memory) An agent with statistical memory memorizes a statistical distribution over the set of conceivable pieces of information X, and this justifies our terminology.

For each $n \in \mathbb{N}$, denote as S_n the set of all permutations on $\{1,\dots,n\}$. Then, agent i exhibits *statistical memory* if $\mathcal{M}_i$ is the correspondence

$$\xi \mapsto \big\{\rho \in X^{<\mathbb{N}} : \exists\, \pi \in S_{\ell(\xi)},\ \rho = \xi \circ \pi\big\}.$$

The mathematical intuition is as follows. First note that ξ is a map from $\{1,\dots,n\}$ to X, where we let n denote the length of ξ. Then, each element $\rho \in \mathcal{M}_i(\xi)$ is obtained by first permuting $\{1,\dots,n\}$ through some $\pi \in S_n$, and then by mapping each permuted element to X through ξ, obtaining $(\xi_{\pi(1)},\dots,\xi_{\pi(n)})$. Thus, the elements of $\mathcal{M}_i\,(\xi)$ are the sequences ρ with the same number of occurrences of each element of X as ξ.

For an illustration, consider $\{0,1\}^{\leq 3}$, and $\xi = (1,0,1)$. If agent i exhibits statistical memory, $\mathcal{M}_i(\xi) = \{(1,1,0),(1,0,1),(0,1,1)\}$. ▲

If we allow an agent to retain the pieces of information she observed but not their frequencies we obtain a more imprecise information storage than that of statistical memory.

Example 8 (Range memory) For a generic function $g : X \to Y$, its *range* is the image of its domain. That is, $\operatorname{range} g := g(X)$. Then, we say that agent i satisfies *range memory* if $\mathcal{M}_i$ is the correspondence

$$\xi \mapsto \big\{\rho \in X^{<\mathbb{N}} : \operatorname{range}\rho = \operatorname{range}\xi\big\}.$$

In words, agent i retains the range of the observed information stream. Hence, the agent is able to memorize which pieces of information she observed, but not their frequencies nor anything about their order. Thus, it is not required that the agent retains the length of the observed sequence, which can be imposed as an additional assumption. For an example, consider $\{0,1\}^{\leq 3}$, and $\xi = (1,0,1)$: $\mathcal{M}_i(\xi) = \{(0,1),(1,0),(1,1,0),(1,0,1),(0,1,1),(0,0,1),(0,1,0),(1,0,0)\}$. ▲

Note that the correspondences proposed so far are partitional. In particular, they are such that the agent never rules out the observed sequence. However, we may also think of more severe failures in the process of information storage.

[21] Obviously, retaining the number of instances a given piece of information is received implies being able to remember how many pieces of information were received, and this explains why we assume that the length of the sequence is retained.

Example 9 (Information distortion) Agent i *distorts information* if $\mathcal{M}_i$ is not reflexive. That is, if there exist $\xi \in X^{<\mathbb{N}}$ such that $\xi \notin \mathcal{M}_i(\xi)$. This amounts to assuming that agent i's stored information does not necessarily include the actual experienced stream of pieces of information. ▲

A non-partitional correspondence can be obtained also by allowing an agent to retain sequences of pieces of information that are somewhat "similar" to what she actually observed. In general, similarity is *not* a transitive notion, and therefore the correspondence formalizing this failure of perfect memory would not be partitional.

Example 10 (Approximate memory) Suppose that the set $X^{<\mathbb{N}}$ is equipped with some distance d. For each $\varepsilon \in \mathbb{R}_+$, we say that agent i exhibits *ε-approximate memory* if $\mathcal{M}_i$ is the correspondence

$$\xi \mapsto \{\rho \in X^{<\mathbb{N}} : d(\xi, \rho) \leq \varepsilon\}.$$

In words, the sequence ξ of pieces of information that is actually observed leads the agent to recall sequences that are "close enough" to ξ. To put it differently, ξ provides the agent with some evidence, but agent i's ability to store information to be "fuzzy": she retains information that is close to (but not necessarily equal to) what she observed. Note that perfect memory is obtained by setting $\varepsilon = 0$. ▲

The ability to retain information may also be influenced by the emotional valence of information, or on an individual's ability to establish a link between the information received and some emotionally-relevant aspect or event. If we allow for set X to be richer, this mechanisms can be captured by memory correspondences.

Example 11 (Emotional memory) Assume that $X = (A^\star \cup M^\star) \times E$, where $E = \{0, 1\}$. For each observed $(x, e) \in X$, we say that x has (positive or negative) emotional valence if $e = 1$. That is, the pieces of information observed by an agent are now enriched by an admittedly rough description of whether they hold some sort of emotional valence for the agent.

We say that agent i exhibits emotional memory if $\mathcal{M}_i$ is the correspondence

$$(x_k, e_k)_{k=1}^n \mapsto \{(y_k, e_k)_{k=1}^n \in X^{<\mathbb{N}} : \forall\, k \in \{1, \ldots, n\},\ e_k = 1 \Longrightarrow x_k = y_k\}.$$

In words, agent i retains only the observed pieces of information that bear some emotional valence for her or whose observation was associated with a relevant emotional state, as well as the length of the observed sequence. The last assumption can easily be relaxed.

Obviously, a more detailed account of emotions may be conceived. For instance, one could posit a richer spectrum of emotional states, ranging from negative to positive ones, and then allow for an individual to remember only the pieces of information that were associated with an emotional state that is similar to the one being currently experienced. With such formalism, for instance, happy individuals would be able to retrieve from their memory those pieces of information they observed when they were happy. Proceeding in this way would allow to shift the emphasis from the storage to the retrieval of information. ▲

We conclude by mentioning a slight modification of the approach presented so far that is achieved by endowing the set of conceivable pieces of information with a different structure. Specifically, we can partition the set of conceivable pieces of information in several categories, representing different kinds of information (e.g., actions taken versus messages received, or even messages pertaining to different topics). Our analysis in the next section will be based on such approach. That is, the sequence of pieces of information available to a given agent will be a sequence of pairs of messages and actions, which obviously represent two different sources of information.

Example 12 (Memory and categories) Suppose that the set X is endowed with a finite partition $\mathscr{C} = \{C_1, C_2, \ldots, C_K\}$, whose elements are called *categories*. Building on the formalism used so far, the set X may be partitioned into two categories, one being that of "actions" and the other one being that of "messages", but we can also imagine finer partitions (e.g., one could envisage categories such as "political news" or "football results" in which messages may be further partitioned).

Then, consider the set $\tilde{X} := \bigtimes_{k=1}^{K} C_k$. An element of such set may be thought of as the information received in some situation, divided into the relevant categories. Obviously, an agent need not be exposed to information from all categories at the same time, so we might as well assume that each category includes a dummy element that specifies that no information from such category is observed at some point. For example, if $\mathscr{C} = \{A^\star, M^\star\}$, it may well be that in some instances a given agent is exposed to some information (i.e., receives a message) without necessarily taking an action.

As before, it is natural to consider the set $\tilde{X}^{<\mathbb{N}}$, which represents a sequence of (categorized) information a given agent may receive. The product structure given to set $\tilde{X}$ is then convenient to allow an agent to treat information of different categories in different ways. For an illustration, assume that $\mathscr{C} = \{A^\star, M^\star\}$. If we want agent i to correctly memorize the actions she took but not the messages she received, we can let $\mathcal{M}_i$ be the correspondence $\xi \mapsto \{\rho \in \tilde{X}^{<\mathbb{N}} : \mathrm{proj}_{(A^\star)^{<\mathbb{N}}}\, \rho = \mathrm{proj}_{(A^\star)^{<\mathbb{N}}}\, \xi\}$.

Moreover, virtually all the examples mentioned so far can be analyzed within this category-enriched approach. Statistical memory could then be used to formalize the idea that an agent memorizes the frequencies with which she receives information belonging to each category. Or, fuzzy memory could be leveraged to allow for more or less precise storage of different kinds of information. Lastly, note that the emotional message described in Example 11 is formally a special case of this approach, where pieces of information are partitioned based on whether they are emotionally relevant or not. ▲

4.2 Memory in games

The memory correspondences just introduced fit naturally in our framework. If we fix a flow-based game structure $\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i\in I}\rangle$, we can consider a profile of memory correspondences $\mathcal{M} = (\mathcal{M}_i)_{i\in I}$ such that, for each $i \in I$, $\mathcal{M}_i$ is a nonempty-valued possibility correspondence defined over the set $(A_i \cup M_i)^{\leq 2L+1}$. Obviously, we are ultimately interested in the restriction of $\mathcal{M}_i$ to the set of personal histories of i, H_i. The obtained tuple $\langle \Gamma_F, \mathcal{M}\rangle$ is a **flow-based game structure with memory**. A minimal restriction that we impose to a game structure with memory $\langle \Gamma_F, \mathcal{M}\rangle$ is that players cannot forget the last message they receive. In this way, they are always able to play the game, given that each message provides them with the information needed to figure out their feasible actions. The assumption is reasonable if we think that players can directly observe the last message received when they have to make a choice (e.g., a player can look at the disposition of pieces on the chessboard when she has to move). However, it does impose some restrictions on players' ability to process information: one may imagine situations where the "last available message" is a complex object, and perfect processing of all the information encoded in some message requires significant effort. Formally, fix $i \in I$ and $h_i, h'_i \in H_i$ with $h_i = (m_{i,0}, \ldots, a_{i,n}, m_{i,n})$ and $h'_i = (m_{i,0}, \ldots, a'_{i,k}, m'_{i,k})$. We impose the following "knowledge of the last message" condition:

$$h_i \in \mathcal{M}_i(h'_i) \Longrightarrow m_{i,n} = m'_{i,k} \tag{KLM}$$

This assumption will be maintained henceforth.

Consider a flow-based game structure with memory $\langle \Gamma_F, \mathcal{M}\rangle$, as well as the set of plays P derived from Γ_F. Recall that two plays $p, q \in P_i$ are **indistinguishable** for player $i \in I$ if and only if $F_i(p) = F_i(q)$. That is, two plays (after which a player is alert) are indistinguishable for such player if they induce the same realized stream of information for such player. Importantly, such stream is entirely determined by the rules of the game. The objective indistinguishability relation is described, for each player $i \in I$, by the partitional possibility correspondence $\mathcal{F}_i$:

$$\mathcal{F}_i(p) = \{q \in P_i : F_i(p) = F_i(q)\}.$$

Now, for each $i \in I$, $\mathcal{M}_i$ represents a binary relation over the set H_i: we say that $h_i \in H_i$ is **mistakable for** (or, **confused with**) $g_i \in H_i$ if $h_i \in \mathcal{M}_i(g_i)$. That is, player i's mnemonic abilities prevent her from distinguishing the sequences of pieces of information h_i and g_i. The indistinguishability relation was however defined on the set of plays after which a given player is alert. It is therefore convenient to retrieve a relation (or, equivalently a possibility correspondence) on the set of play that embodies the possibility of player i misremembering the personal histories she observes.

To this end, we say that $q \in P_i$ is **possible** (for $i \in I$) given $p \in P_i$, if $F_i(q) \in \mathcal{M}_i(F_i(p))$. In words, play q can be deemed possible when play p realizes if the stream of information it induces (i.e., $F_i(q)$) is consistent with what player i recalls of the flow induced by p (i.e., if it belongs to $\mathcal{M}_i(F_i(p))$). We can define a possibility correspondence on the set of plays where player $i \in I$ is alert formalizing this idea. Specifically, let $\mathcal{P}_i : P_i \rightrightarrows P_i$ be defined, for each $p \in P_i$, as

$$\mathcal{P}_i(p) := \{q \in P_i : F_i(q) \in \mathcal{M}_i(F_i(p))\}.$$

It is worth comparing the correspondences $\mathcal{F}_i$ and $\mathcal{P}_i$ ($i \in I$). The former depends only on the objective informational aspects entailed by the rules of the game – indeed, the indistinguishability equivalence relation corresponds to the partition of P_i induced by F_i^{-1} (where F_i is the function mapping each play where i is alert to the personal history of i it induces). The latter instead depends also on $\mathcal{M}_i$, which encodes a description of how player i subjectively perceives and retains the information she receives. Hence, even if the two plays can be perfectly distinguished based on the information available to player i (e.g., because $\mathcal{F}_i(p) = \{p\}$ and $\mathcal{F}_i(q) = \{q\}$), i's cognitive limitations may prevent her from recognizing this. In other words, correspondence $\mathcal{F}_i$ ($i \in I$) describes the inferences that player i is in principle allowed to make by the rules of the game. Correspondence $\mathcal{P}_i$ instead describes how such inferences are mediated by i's mnemonic abilities. Clearly, if player i exhibits perfect memory (i.e., she perfectly retains the observed pieces of information, cf. Example 5), the two correspondences coincide. This is because, under perfect memory, $\mathcal{M}_i(F_i(p)) = \{F_i(p)\}$ for each $i \in I$ and $p \in P_i$, and $\mathcal{P}_i(p) = \{q \in P_i : F_i(q) \in \mathcal{M}_i(F_i(p))\} = \{q \in P_i : F_i(q) = F_i(p)\} = \mathcal{F}_i(p)$. Hence:

Remark 4 Fix $i \in I$. If $\mathcal{M}_i$ satisfies perfect memory, $\mathcal{P}_i = \mathcal{F}_i$.

The converse is not true in general. Intuitively, there may be situations where a player's mnemonic failures make her forget only some completely uninformative messages: this may not prevent her from drawing the same inferences she could make if she had perfect memory.

Importantly, the relation represented by $\mathcal{P}_i$ may even fail to be partitional (this may happen, e.g., if player i distorts information, cf. Example 9). However, the following holds:

Lemma 1 *Fix $i \in I$. The correspondence $\mathcal{P}_i$ is partitional if and only if $\mathcal{M}_i$ is partitional.*

Lemma 1 ensures that whenever the memory correspondence of a given player is "well-behaved", then so is that player's induced possibility correspondence. Most of the examples presented in Section 4.1 involved partitional memory correspondences: therefore, interesting phenomena (e.g., bounded memory, or statistical memory) can be modeled in a relatively convenient way.

It is natural to conjecture that a player's subjective treatment of game-specific flows of information – as described by her memory correspondence – can only make the available information in some sense "less precise". The reason for that is that a player may forget some of the pieces of information she observed. Lemma 2 formalizes this intuition. In the following, for $i \in I$, we let $\mathscr{P}_i$ and $\mathscr{F}_i$ be the partitions induced by $\mathcal{P}_i$ and $\mathcal{F}_i$, respectively (clearly under the assumption that the former is partitional).[22]

Lemma 2 *Fix $i \in I$. If $\mathcal{M}_i$ is partitional, $\mathscr{P}_i$ is a coarsening of $\mathscr{F}_i$.*

Up until now, we showed that partitional but imperfect mnemonic ability can only make the partition of possible plays coarser. This clearly does not occur if players have perfect memory (Remark 4). However, there is another way to prevent failures in information retention, and this intuitively consists in constantly reminding players of all the pieces of information available to them at some point of the play – in our terminology, this occurs whenever feedback is cumulative (cf. Section 3.4). Therefore, if a player correctly remembers the last message observed and if such message encodes all the available information, then it must be the case that what players deem possible coincides with what is consistent with the available information, as determined by the rules of the game. In some sense, cumulative feedback prevents players from forgetting.

Lemma 3 *Fix $i \in I$. If f_i is cumulative, $\mathcal{P}_i = \mathcal{F}_i$.*

Note that Lemma 3 is independent of player i's mnemonic abilities. The only assumption needed is that, after any play, player i remembers the last message received – and this assumption is maintained whenever we talk about flow-based game structures with memory. As mentioned in the discussion of the "knowledge of last message" assumption, a message can encode complex information. This may be especially true under cumulative feedback, where players are constantly reminded of what happened as the game unfolded. Therefore, Lemma 3 dispenses with assumptions on players' mnemonic abilities at the cost of an implicit assumption about the ability of players to perfectly process the (potentially complex) information received.

5 Information flows, memory, and perfect recall

In this section, we relate the dual account of objective information flows and subjective memory we described so far with the conventional formalism. We first describe the conventional approach used to model information. In such formalism, information is described by means of information sets, as is standard for the description of abstract sequential games since the seminal work of von Neumann and Morgenstern (1944). The collection of information sets of a player is essentially a partition of the nodes of the game tree where such player is active. As already stressed, however, where such partition comes from is unclear. In light of the foregoing discussion, one can imagine that these partitions arise from some flow of information players are exposed to (cf. the possibility correspondences $(\mathcal{F}_i)_{i \in I}$ introduced earlier on). However, absent additional requirements, this is not true in general. Moreover, as clarified by our analysis of memory, there are cases where players' "knowledge" (as derived from both objective information flows and personal cognitive features) is not represented by a partitional possibility correspondence.

Before giving the definition of standard game structure, we introduce some notation. For a tree $V \subseteq A^{\leq L}$ and all vertices/sequences $v \in V$ and players $i \in I$, let $A(v) := \{a \in A : (v, a) \in$

[22] We will stick to this notation also in subsequent sections: calligraphic letters will denote correspondences and script letters will denote collections of sets (typically partitions induced by partitional possibility correspondences). Note that the partition induced by a partitional possibility correspondence is simply the range of such correspondence.

$V\}$ and $A_i(v) := \{a_i \in A_i : \exists\, a_{-i} \in A_{-i}, (v, (a_i, a_{-i})) \in V\}$ (i.e., the projection of $A(v)$ onto A_i) respectively denote the sets of action profiles and actions of i consistent with V given v.

Definition 2 *A* ***standard game structure*** *is a tuple*

$$\Gamma_S = \langle I, V, \mathcal{I}, (A_i, \mathscr{Q}_i)_{i\in I}\rangle,$$

where

- *I is the set of* ***players*** *and, for each $i \in I$, A_i is the set of i's potentially feasible actions;*
- *$V \subseteq A^{\leq T}$ is a tree (call it "the* ***tree****", and its elements "****plays****");*
- *$\mathcal{I} : V \rightrightarrows I$ is the* ***alert-player correspondence****, and it satisfies the following properties:*

$$\mathcal{I}(\varnothing) = I; \tag{KGS-S}$$

$$\forall\,(v,a) \in V, \quad \mathcal{I}(v) \neq \emptyset,\ a \in A_{\mathcal{I}(v)}; \tag{APM}$$

$$\forall\, v \in V, \quad A(v) = \bigtimes_{i\in\mathcal{I}(v)} A_i(v); \tag{APF}$$

- *for each $i \in I$, $\mathscr{Q}_i$ is a partition of the set $V_i := \{v \in V : i \in \mathcal{I}(v)\}$ (call it "the* ***information structure****" and its element "the* ***information sets****" of player i), satisfying*

$$\forall\, v, w \in V_i, \quad Q_i(v) = Q_i(w) \Longrightarrow A_i(v) = A_i(w), \tag{KfA}$$

  where $Q_i(v)$ and $Q_i(w)$ are the cells of $\mathscr{Q}_i$ containing v and w, respectively.[23]

A few comments are in order. First, the tree V specifies what sequences of action profiles can be played throughout the game. Moreover, the correspondence $\mathcal{I}$ specifies which players are alert after some play (recall that "being alert" means that some information is received after some play). Condition (KGS-S) then imposes that all players be alert at the beginning of the game – this is the natural counterpart of the condition we imposed for flow-based game structures in Section 3. Consistently with the approach adopted so far, being alert is a prerequisite for being active, and, for each $v \in V$ and $i \in \mathcal{I}(v)$, $A_i(v) \subseteq A_i$ is the set of player i's available actions at v.[24] Condition (APM) (alert players move) then requires that only alert player move after some play. Finally, condition (APF) (action profile feasibility) imposes that what is feasible for i given v is independent of what is feasible for j given v. A play (or node, or vertex) v is terminal if and only if $A(v) = \emptyset$. As before, being inactive at some play v amounts to having only one available action.

Second, an information set $Q_i \in \mathscr{Q}_i$ of player i is interpreted as a set of plays where i is alert that she cannot distinguish. That is, $Q_i(v)$ ($i \in I$, $v \in V_i$) is the set of plays that player i deems possible given that v occurred. Condition (KfA) (knowledge of feasible actions) requires that two plays be indistinguishable for a player only if the set of actions available to such player after the two plays is the same – otherwise, players would not even able to play the game, as they might be unsure of the actions they can take at some point of the play.

[23]We use script letters to denote collections of sets and calligraphic letters to denote correspondences. Clearly, we could have equivalently given the definition of information structure $\mathscr{Q}_i$ using the partitional possibility correspondence $\mathcal{Q}_i$ that induces (i.e., whose range is) $\mathscr{Q}_i$.

[24]Note, in this *axiomatic* approach, a (game) tree V is posited as primitive element of the analysis, and the feasible action profiles are derived from V. This should be contrasted with the more *constructive* approach of Definition 1, whereby action feasibility correspondences are taken as primitives.

It is worth noting that the traditional formalism dispenses with the alert-player correspondence $\mathcal{I}$ because it only models the information available to players when they are active – being alert conventionally coincides with being active. That is, $\mathscr{Q}_i$ is defined to be a partition of the plays after which player i is active. Given that we distinguish between active and alert players, we introduce the correspondence $\mathcal{I}$ to make the comparison between the different approaches more straightforward.

As already hinted in the introduction, information sets are hybrid concepts that may fail to adhere to the separation principle, as they may represent situations of genuine ignorance about the game unfolding, or cognitive failures in information retention, or both. The following one-player game illustrates a failure of memory.

Example 13 (Did I lock the door?) After leaving her home, Alice no longer remembers whether she locked the door or not. When she realizes this, she can either go back and check or not. The standard multistage game structure portraying this situation is as follows, where shaded areas are used to represent information sets.

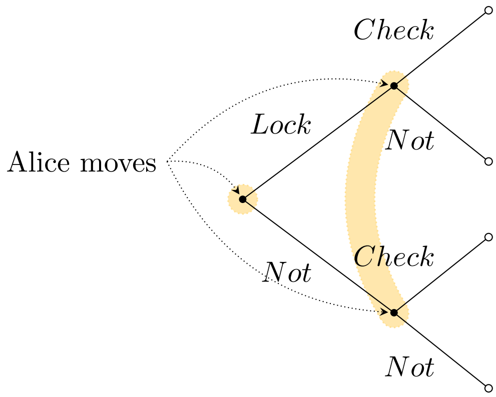

Figure 2 Alice does not remember if she locked the door.

The game tree V is depicted in the figure, and the alert-player correspondence can be trivially taken to be such that Alice is alert if and only if she is active. Then, it is easy to check that all the conditions of Definition 2 are met. Specifically, (APF) is trivially satisfied because the set of players is a singleton. Similarly, (APM) holds because Alice is the only player and she is always alert throughout the game. Moreover, (KfA) holds because $Q_{Alice}((Lock)) = Q_{Alice}((Not))$, and $A_{Alice}(Lock) = \{Check, Not\} = A_{Alice}(Not)$.

However, Alice's failure in distinguishing plays $(Lock)$ and (Not) is a personal cognitive shortcoming made relevant by the fact that the rules of the game do not provide Alice with an automatic reminder of what she did. This clarifies the hybrid nature of information sets: in general, we have no guarantee at all that they represent a situation where players' ignorance about the realized play is induced by the rules of the game alone. ▲

To rule out situations where information sets encode some sorts of cognitive failures such as the one represented in Example 13, the notion of **perfect recall** has been proposed (cf., e.g., Kuhn, 1953, or Selten, 1975). As the name suggests, perfect recall aims to rule out all the situations where information sets incorporate failures in players' ability to retain information.

The notion of perfect recall we are going to employ is formalized by means of the concept of "experience", as introduced by Osborne and Rubinstein (1994). As the game unfolds, a player goes through a sequence of information sets. Such sequence of information sets, coupled with the actions played at each such information set, forms the "experience" of a player within a play. To provide a formal definition of experience, we introduce the following notation. For each pair of plays $(v', v'') \in V \times V$ such that $v' \prec v''$ and for each player $i \in I$, we denote by $a_i(v', v'')$ the unique action such that $(v', (a_i(v', v''), a_{-i})) \preceq v''$ for some $a_{-i} \in A_{-i}$. In words, $a_i(v', v'')$ is the unique action of player i that does not prevent v'' from realizing when taken at v'. With this, the **experience function** X_i of player $i \in I$ can be defined recursively as follows. Fix a generic play $v \in V_i \cup \{\varnothing\}$.

- If $\ell(v) = 0$ (i.e., $v = \varnothing$), we define $X_i(\varnothing) := (Q_i(\varnothing))$.
- Assume that $X_i(u)$ has been defined for each $u \in V_i$ with $0 \leq \ell(u) \leq k$. Fix $v \in V_i$ with $\ell(v) = k+1$ (if any), and let $\mathrm{last}_i\, v$ be the longest predecessor of v where i is alert (if any), or the empty play (otherwise). Note that $\mathrm{last}_i\, v \in V_i$ and $\ell(\mathrm{last}_i\, v) \leq k$. With this, we can define $X_i(v) := (X_i(\mathrm{last}_i\, v), a_i(\mathrm{last}_i\, v, v), Q_i(v))$.

The definition of perfect recall of Osborne and Rubinstein (1994) requires that, whenever two plays belong to the same information set of a player, then they induce the same experience for such player.

Definition 3 *Fix a standard game structure* $\Gamma_S = \langle I, V, \mathcal{I}, (A_i, \mathscr{Q}_i)_{i \in I} \rangle$*;* ***perfect recall*** *holds if:*

$$\forall\, i \in I, \forall\, v, w \in V_i, \quad Q_i(v) = Q_i(w) \Longrightarrow X_i(v) = X_i(w). \tag{PR}$$

Note that perfect recall is taken to be a property of the *game structure* (i.e., of the rules of the game), even though it arguably pertains to some individual features of the players who happen to play the game. This is why we adopt a different terminology and we use the term "memory" to refer to the subjective mnemonic abilities of those who play a given game.

5.1 Flow-based and standard game structures

As one may expect, there should be some relation between the flow-based account we provided and the standard one just described. In fact, our approach is more general as it allows for situations where players' game-specific "knowledge" is non-partitional – therefore it allows to model situations that do not fall under the conditions of Definition 2. In this subsection, we show how standard game structures are retrieved from our memory-augmented flow-based approach.

Fix a flow-based game structure $\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i \in I} \rangle$, a profile of partitional memory correspondences $(\mathcal{M}_i)_{i \in I}$,[25] and consider the derived objects P, f, $(\mathcal{P}_i)_{i \in I}$ – respectively, the sets of feasible plays, the restriction of the feedback function to such set, and the possibility correspondences derived from information flows and players' mnemonic abilities (cf. Section 4.2). Then, a standard game structure is retrieved as follows.

First, recall that P is a tree by construction (Remark 1). Second, we retrieve an alert-player correspondence $\mathcal{I}$ on P by letting, for each $p \in P$, $\mathcal{I}(p) = \mathcal{D}(f(p))$, where the correspondence $\mathcal{D}$ extracts the domain of a given function (cf. Section 3.1). The interpretation is straightforward: after $p \in P$, the message profile $f(p)$ is generated, and players who are alerted by $f(p)$ are those in $\mathcal{D}(f(p))$. Hence, the alert players at p are the ones that are alerted by the message profile generated immediately after p. With this, it is possible to verify that the alert-player correspondence $\mathcal{I}$ satisfies conditions (KGS-S), (APM), and (APF): this follow by inspection of the derivation of P (Section 3.2). Third, let $\mathscr{P}_i := \mathrm{range}\, \mathcal{P}_i$ (i.e., the partition induced by $\mathcal{P}_i$) be the information structure of player $i \in I$. Note that (KfA) holds by construction. This is because the memory correspondence $\mathcal{M}_i$ satisfies (KLM) and by definition the feasibility correspondence $\mathcal{A}_i$ the actions available to a given player only depend on the last message received. Therefore,

[25]The fact that $\mathcal{M}_i$ ($i \in I$) is partitional implies that $\mathcal{P}_i$ is partitional (Lemma 1). Moreover, recall that memory correspondences are assumed to satisfy the "know the last message" condition (KLM).

two play belong to the same cell of $\mathscr{P}_i$ only if they result in the same last message for player i – that is, only if i's feasible actions after the two plays are the same. Therefore, the standard game structure implied by the flow-based one with memory is $\langle I, P, \mathcal{D} \circ f, (A_i, \mathscr{P}_i)_{i \in I} \rangle$.

Example 14 (Committee, continued) The tree of possible has already been derived. Our focus here is to clarify how information sets are derived from information streams. Given that we did not discuss the feedback induced by terminal plays, we restrict attention to partial plays. For the sake of the illustration, focus on Dave and assume that he has perfect memory,[26] and recall that the partition induced by the flows of information he receives is $\mathscr{F}_{Dave} = \mathscr{P}_{Dave} = \{\{\varnothing\}, \{(D), (A, (D, D))\}, \{(A, (D, U)), (A, (U, D))\}\}$.

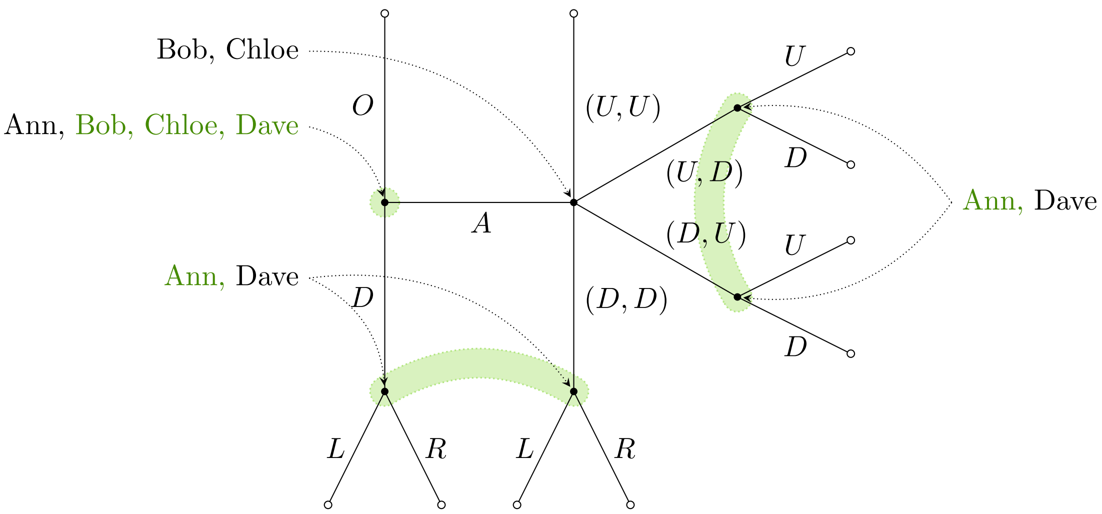

Figure 3 Active players in black, alert but inactive players in green, induced Dave's information sets depicted.

This example may further clarify the importance of "alertness". If Dave received a message after (A) saying that Ann moved, he could understand that the chosen action was A because otherwise he would be active. The information set $\{(A)\}$ may not be relevant because Dave is not active after such history. However, Dave's alertness after (A) would "split" the information set $\{(D), (A, (D, D))\}$ into two singletons. ▲

In light of Remark 4, we can say that:

Remark 5 If $\mathcal{M}_i$ satisfies perfect memory for each $i \in I$, the perfect recall property (PR) holds in $\langle I, P, \mathcal{D} \circ f, (A_i, \mathscr{P}_i)_{i \in I} \rangle$.

More generally, given a flow-based game structure $\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i \in I} \rangle$ and a profile of memory correspondences $\mathcal{M} = (\mathcal{M}_i)_{i \in I}$, we define the **generalized standard game form induced by** $\langle \Gamma_F, \mathcal{M} \rangle$ as $G(\Gamma_F, \mathcal{M}) := \langle I, P, \mathcal{D} \circ f, (A_i, \mathcal{P}_i)_{i \in I} \rangle$. The adjective "generalized" reminds that we allow for the (derived) possibility correspondences $(\mathcal{P}_i)_{i \in I}$ to be non-partitional. If such correspondences are partitional, $G(\Gamma_F, \mathcal{M})$ is a standard game form, as clarified by the foregoing discussion. In that case, we equivalently denote player i's information structure by either the possibility correspondence $\mathcal{P}_i$ or by the induced partition $\mathscr{P}_i$.

[26] In fact, no assumption on Dave's memory is needed to partition the non-terminal nodes where he is alert, because he is always alert and active only after receiving a single message, and we assume that players always correctly remember the last message they received.

The following result proves the converse of Remark 5 and gives a characterization of the perfect recall property of standard information structures in terms of flows of information and players' mnemonic abilities.

Proposition 1 *Fix a standard game structure* Γ_S*. The perfect recall property* (PR) *holds in* Γ_S *if and only if there exist a flow-based game structure* Γ_F *and a profile of memory correspondences* $\mathcal{M}$ *satisfying perfect memory such that* $\Gamma_S = G(\Gamma_F, \mathcal{M})$*.*

We can say more: *any* standard game structure can be retrieved from a suitable flow-based game structure with memory where the memory correspondences are partitional.

Proposition 2 *Fix a standard game structure* Γ_S*. There exist a flow-based game structure* Γ_F *and a profile of partitional memory correspondences* $\mathcal{M}$ *such that* $\Gamma_S = G(\Gamma_F, \mathcal{M})$*.*

Note that Proposition 2 ensures that any situation that can be described by a standard formalism can be framed within our memory-enriched approach. On the other hand, as already mentioned, there are situations that are easily described with our language while at the same time being inexpressible in a standard setting – that is, all the situations where a player's game-specific "knowledge" cannot be expressed in terms of information partitions. A notable case is information distortion (cf. Example 9): in that case, when some play realizes, player i deems such play impossible. This is inexpressible in a standard setting, where each player's information structure is a partition of the set of plays where such player is alert.

Proposition 2 also ensures that game structures that satisfy the perfect recall property can be retrieved from some flow-based game structure with memory. Notably, the memory correspondences of such structure *may fail to satisfy perfect memory.* Combining this observation with Proposition 1, one can conclude that standard game structures with perfect recall are in general consistent with multiple combinations of information flows and players' mnemonic abilities. Proposition 1 however ensures that such differences are ultimately immaterial, as game structures with the perfect recall property can always be interpreted as arising from situations where players actually exhibit perfect memory.[27]

It is worth stressing that Proposition 2 applies to every game structure, and this includes of course those where perfect recall fails, as well as those encoding "problematic" cognitive failures such as absentmindedness.

Example 15 (Absentminded driver) The following game structure is taken from Piccione and Rubinstein (1997). A driver is heading back home and he needs to get off the highway at the second exit. However, when he reaches an exit, he cannot understand if it is the first or the second one – that is, he does not remember if such exit is the first one he crosses or if he already passed past one. The graphical representation is as follows.

[27] This discussion nonetheless further clarifies how the formalism of information sets may fail to comply with the separation principle: while it describes how informed players are during the game unfolding, it does not fully specify where their "knowledge" comes from. Hence, even when the information structure of a standard game structure is well-behaved enough to satisfy perfect recall, it may arise from different combinations of objectively-determined rules of interaction and subjective cognitive features.

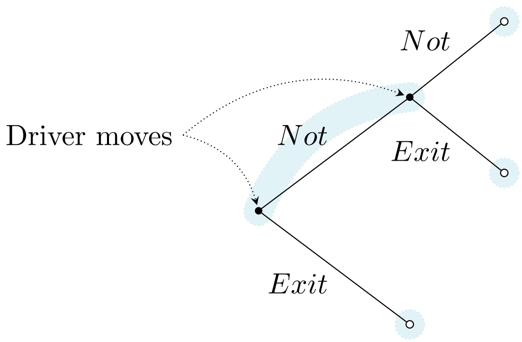

Figure 4 Driver's information sets.

To retrieve a flow-based game structure with memory inducing such situation, focus on the following flow of information.

Play	Message	Available actions
$\varnothing$	$m^* =$ *You are at a crossing.*	$\{Exit, Not\}$
$(Exit)$	$m_{\mathrm{w}} =$ *Wrong exit!*	$\emptyset$
(Not)	$m^* =$ *You are at a crossing.*	$\{Exit, Not\}$
$(Not, Exit)$	$m_{\mathrm{h}} =$ *You arrived home!*	$\emptyset$
(Not, Not)	$m_{\mathrm{m}} =$ *You missed your exit!*	$\emptyset$

Table 1 A flow of information for Driver.

The feasible extended histories are straightforwardly derived. Moreover it is easy to verify that distinct plays are always distinguishable. For instance, consider $\varnothing$ and (Not): their induced streams of information are $F_D(\varnothing) = (m^*)$ and $F_D(Not) = (m^*, Not, m^*)$, which are obviously different. Hence, confusion between such two plays *cannot* arise as a byproduct of the rules of the game. Rather, we have to introduce a memory correspondence $\mathcal{M}_D$ such that Driver only retains the last message induced by a play, and nothing more. Specifically, this implies that $\mathcal{M}_D((m^*)) = \mathcal{M}_D((m^*, Not, m^*)) = \{(m^*), (m^*, Not, m^*)\}$. Therefore, despite being sometimes regarded as a pathological feature that can arise from a flawed specification of the information structure of a standard game structure,[28] absentmindedness can be modeled smoothly in our memory-enriched framework. ▲

We conclude by stating a result that provides yet another interpretation of perfect recall as a natural consequence of Lemma 3. Recall that Lemma 3 ensures that, under cumulative feedback, the possibility partition obtained from a player's memory has to coincide with the indistinguishability partition induced by the rules of the game. Therefore, standard game structures with the perfect recall property can also be thought of as arising from some flow-based game structure with cumulative feedback. This is formalized as follows.

Proposition 3 *Fix a standard game structure Γ_S. The perfect recall property* (PR) *holds in Γ_S if there exist a flow-based game structure Γ_F with cumulative feedback and a profile of memory correspondences $\mathcal{M}$ such that $\Gamma_S = G(\Gamma_F, \mathcal{M})$.*

Interestingly, Proposition 3 does not require players' mnemonic abilities to be perfect. All is needed is that players correctly remember the last message they receive, which encodes all

[28] For instance, absentmindedness is ruled out by the defining features of the modeling framework introduced by Kuhn (1953). See also the discussion in Alós-Ferrer and Ritzberger (2016), pp 75-78.

the information that has been available to them during the play by cumulativeness of feedback. Therefore, we can interpret perfect recall as the property of (partitional) information structures that arise in such settings.

6 Conclusion

Related literature The present paper relates to several strands of the literature. First of all, it is closely tied to the theory of the detailed representation of sequential games (or, extensive-form games, as they are commonly called).[29] The first definition of "extensive-form games" is due to the seminal work of von Neumann and Morgenstern (1944), who start with a set of "outcomes" that are progressively refined by players' choices. This yields as a derived object a graph-theoretic representation with trees satisfying a multistage structure, i.e., all nodes in the same information set have the same number of predecessors. Information is assumed to have a partitional structure and perfect recall is not a maintained assumption. To remove the built-in multistage assumption of von Neumann and Morgenstern (1944), Kuhn (1953) posited the tree representation as primitive and defined the perfect recall property of information partitions. Alós-Ferrer and Ritzberger (2008, 2013) generalized the representations of von Neumann and Morgenstern (1944) and Kuhn (1953). Like the former, they start with a set of outcomes and let choices select subsets of outcomes, like the latter they do not assume a multistage structure; furthermore, they allow for all kinds of infinite games studied in applications and not impose a partitional information structure. The book by Alós-Ferrer and Ritzberger (2016) provides a broad overview of the field. Finally, the seminal work of Harris (1985) introduced the sequence representation later used by Osborne and Rubinstein (1994) in their texbook. Like them, we use the sequence representation, but—as we explained at length—, we represent information in a crucially different way.

As for perfect recall, several definitions of the same concept have been proposed in addition to the recent one by Osborne and Rubinstein (1994, Definition 203.3) used here. The first one is due to Kuhn (1953, Definition 17), and it leverages the derived concept of strategy, rather than primitive elements of the analysis. Other notions are proposed by Selten (1975) and Perea (2001, Definition 2.1.2). Alós-Ferrer and Ritzberger (2016, Proposition 6.6) and Alós-Ferrer and Ritzberger (2017, Theorem 1 and Corollary 1) prove the equivalence of all the aforementioned notions. Remarkably, the equivalence continues to hold in games with infinite horizon, and in games where agents can choose their actions from an infinite set.

Other works focused on the interpretation and on the characterizations of perfect recall (Alós-Ferrer & Ritzberger, 2017; Bonanno, 2003, 2004; Okada, 1987; Ritzberger, 1999). The most important insight that comes from this branch of the literature is that perfect recall indeed captures a situation where (*i*) players never forget what they did, (*ii*) players never forget what they knew, and (*iii*), past, present, and future have an unambiguous meaning (Ritzberger, 1999). A similar decomposition of perfect recall in "memory of past knowledge" and "memory of past actions" is obtained by Bonanno (2003, 2004) following a syntactic approach that relies on tools from temporal logic. Our framework provides another perspective to look at the same issues, and our results are complementary to the aforementioned ones.

A burgeoning body of literature analyzed the role of memory in decision or learning problems (Bordalo, Coffman, Gennaioli, Schwerter, & Shleifer, 2020; Bordalo, Gennaioli, & Shleifer, 2019, 2020; Fudenberg, Lanzani, & Strack, 2022). Such works underline the role of environmental cues in facilitating the retrieval of similar past experiences from memory, and they show how this

[29] We are critical of the "normal/extensive form game" terminology: as von Neumann and Morgenstern (1944) make clear, the normal and extensive form are different kinds of *representations* of games, not different kinds of games.

process influences an agent's assessment of given information. Our general analysis of memory is somewhat complementary: while this literature focused on the *retrieval* of stored information, we aimed at modeling the process of information *storage*. Given the generality of the approach proposed in the present paper, we believe that blending the two perspectives may be both feasible and insightful.

As already hinted, our approach allows to describe the information that may accrue to alert but inactive players, which is instead usually neglected. This relates our analysis to two strands of the literature. On the one hand, the information an inactive player receives throughout the play may well be relevant for psychological reasons (cf. Battigalli & Dufwenberg, forthcoming for a survey of the literature on psychological games). On the other hand, the end-game information available to player is key when self-confirming equilibrium is studied. In the setting presented so far, such information is the cumulated information available to a given player after some terminal play (i.e., the personal history induced by such terminal play). The literature on self-confirming equilibrium indeed usually posits an end-game feedback function about the play, but, when sequential games are studied, the information available to players during the game is described by means of information sets (cf. Battigalli, Catonini, Lanzani, & Marinacci, 2019). A flow-based description of information may help harmonize such hybrid representation.

Lastly, we mentioned that a flow-based approach has already been used to describe the information accrual to players throughout a game. In particular, in the vast majority of models of repeated games information is modeled as a flow. Specifically, after each round of interaction players obtain a novel piece of information, which may be somewhat revealing of the actions chosen by co-players. For instance, in the oligopoly model of Green and Porter (1984) firms observe the price realization at each period, which is a noisy signal of the competitors' production choices. A sequence of such signals and of chosen actions is the information available to a given firm at a given point in time: the similarity with our approach is obvious. Outside of the literature on repeated games, the same approach was used in Myerson (1986) and Myerson and Reny (2020).

Concluding remarks This paper proposed a framework to explicitly describe players' information in sequential games as provided by the rules of interaction. In doing so, we focused on *flows* of information, as opposed to the standard information-set-based representation which treats information as a *stock*. While this approach is admittedly not new, we tried to offer a systematic exposition, and we enriched it with a formal description of players' ability to retain information. We argued that flows of information provide an explicit and complete description of the information that *objectively* accrues to players in a sequential way as the game unfolds, while the memory correspondences we introduced give a formal definition of players' *subjective* ability to retain the observed information. Decoupling objective and subjective informational aspects allows us to comply with the separation principle, it makes our language more expressive, and it is a key step to formally introduce cognitive limitations in game theoretic analyses.

In this regard, we believe that one of the most promising avenues for future research would consist in trying to enrich our understanding of and our ability to formalize the cognitive aspects linked to memory. This may hopefully help to combine our model of information storage with the already mentioned works on information retrieval given some environmental cues, or to shed light on the emotional aspects influencing mnemonic abilities.

Moreover, it would be interesting to embed such memory-related elements of bounded rationality in interactive situations. In particular, it could be possible to model situations where (not necessarily rational) players wonder about others' cognitive abilities when reasoning strategically. Expressing this kind of assumptions in a formal way would require an expressive language – in particular, one should work with a rich space of states of the world, where irrationality and cognitive failures are allowed to persist at some states.[30] We believe this to be crucial for a better understanding of the implications of bounded rationality and cognitive limitations in strategic interactions.

Lastly, it is worth noting that several of our assumptions may be relaxed in a straightforward way. In particular, we can seamlessly allow for infinite sets of actions and messages, as well as for an infinite length of the game of interest: this would imply virtually no changes as far as definitions are concerned, while some additional requirements may be needed to generalize our results.[31] We can also allow for stochastic elements in a relatively simple way: it is enough to allow for some chance moves by including chance in the set of players. Given that we are interested in the representation of a game structure with possibly some chance moves, we are not even required to specify which are the probabilities of such moves. Note that this device could allow us to make players' mnemonic abilities depend on stochastic elements.

A Proofs

A.1 Proof of Lemma 1 (p. 17)

"If" direction. Assume $\mathcal{M}_i$ is partitional. We check that $\mathcal{P}_i$ satisfies reflexivity, symmetry, and transitivity.

1. *Reflexivity.* Fix $p \in P_i$. Then, note that $\mathcal{P}_i(p) = \{q \in P_i : F_i(q) \in \mathcal{M}_i(F_i(p))\}$. Given that $\mathcal{M}_i$ is partitional, it satisfies reflexivity, meaning that $F_i(p) \in \mathcal{M}_i(F_i(p))$. This implies that $p \in \mathcal{P}_i(p)$.

2. *Symmetry.* Fix $p, q \in P_i$ with $q \in \mathcal{P}_i(p)$. This implies that $F_i(q) \in \mathcal{M}_i(F_i(p))$. By symmetry of $\mathcal{M}_i$, $F_i(p) \in \mathcal{M}_i(F_i(q))$. Hence, $p \in \mathcal{P}_i(q)$, showing symmetry of $\mathcal{P}_i$.

3. *Transitivity.* Fix $p, q, r \in P_i$ with $q \in \mathcal{P}_i(p)$ and $r \in \mathcal{P}_i(q)$. This means that $F_i(q) \in \mathcal{M}_i(F_i(p))$ and $F_i(r) \in \mathcal{M}_i(F_i(q))$. By transitivity of $\mathcal{M}_i$, $F_i(r) \in \mathcal{M}_i(F_i(p))$, meaning that $r \in \mathcal{P}_i(p)$. This proves transitivity of $\mathcal{P}_i$.

"Only if" direction. In light of the proof of the previous direction, the contrapositive statement is easily verified. ∎

A.2 Proof of Lemma 2 (p. 18)

Recall that given two partitions $\mathscr{A}$ and $\mathscr{B}$ of a set X, $\mathscr{B}$ is a **coarsening** of $\mathscr{A}$ if for each $A \in \mathscr{A}$ there is $B \in \mathscr{B}$ such that $A \subseteq B$.

Now, consider a generic $A \in \mathscr{F}_i$. We check that there is $B \in \mathscr{P}_i$ such that $A \subseteq B$. By definition of $\mathscr{F}_i$, this means that $A = F_i^{-1}(h_i)$ for some $h_i \in H_i$. Given that $\mathcal{M}_i$ is partitional on H_i (denote as $\mathscr{M}_i$ the induced partition), there is a unique cell $C \in \mathscr{M}_i$ such that $h_i \in C$. Now consider $B := F_i^{-1}(C) = \bigcup_{h'_i \in C} F_i^{-1}(h'_i)$. Clearly, $A \subseteq B$. It remains to show that $B \in \mathscr{P}_i$. That is, that $B = \mathcal{P}_i(p) = \{q \in P_i : F_i(q) \in \mathcal{M}_i(F_i(p))\}$ for some $p \in P_i$. Given that $\mathscr{P}_i$ is a partition (Lemma 1), we can take a generic $p \in B$ to show that $B = \mathcal{P}_i(p)$. Choose $p \in F_i^{-1}(h_i) \subseteq B$, so that $F_i(p) = h_i$. We prove two opposite inclusions. First, pick $p' \in B$. This implies that there is $h'_i \in C$ such that $F_i(p') = h'_i$. But C is the equivalence class of h_i in the partition induced by $\mathcal{M}_i$. Therefore, $h'_i \in \mathcal{M}_i(h_i)$. Since $h'_i = F_i(p')$ and $h_i = F_i(p)$, we have

[30] This formalism is employed by Battigalli, Corrao, and Sanna (2020), even though attention is restricted to rational players there.

[31] For instance, one should impose appropriate measurability conditions to the feedback functions.

$F_i(p') \in \mathcal{M}_i(F_i(p))$. Hence, $p' \in \mathcal{P}_i(p)$. Second, pick $p' \in \mathcal{P}_i(p)$. We have $F_i(p') \in \mathcal{M}_i(F_i(p))$, so that $F_i(p') \in C$. By definition of C and F_i^{-1}, $p' \in F_i^{-1}(F_i(p')) \subseteq F_i^{-1}(C)$, where the inclusion follows from the fact that $F_i(p') \in C$. ■

A.3 Proof of Lemma 3 (p. 18)

Recall that feedback is cumulative for player $i \in I$ if, for each pair of plays $p, p' \in P_i$ with $\ell(F_i(p)) = \ell(F_i(q))$, $\operatorname{proj}_{A_i^{\leq T}} p \neq \operatorname{proj}_{A_i^{\leq T}} p'$ or $f_i(p) \neq f_i(p')$ imply that, for each pair of successors $q \succeq p$ and $q' \succeq p'$ with $\ell(F_i(q)) = \ell(F_i(q'))$, $f_i(q) \neq f_i(q')$.

We first state the following auxiliary result.

Lemma A1 *Consider a flow-based game structure with cumulative feedback, and fix $i \in I$ and $q, q' \in P_i$. Then, $f_i(q) = f_i(q')$ if and only if $F_i(q) = F_i(q')$.*

Proof of Lemma A1. The "if" direction is obvious. For the converse, we show the contrapositive. Fix $i \in I$ and $q, q' \in P_i$ with $\ell(F_i(q)) = \ell(F_i(q'))$ and $F_i(q) \neq F_i(q')$. We show by induction on $\ell(F_i(q)) = \ell(F_i(q'))$ that $f_i(q) \neq f_i(q')$. For the basis step, consider $\ell(F_i(q)) = \ell(F_i(q')) = 3$.[32] Therefore, $F_i(q) = (m_{i,0}, a_i, f_i(q))$ and $F_i(q') = (m_{i,0}, a_i', f_i(q'))$. Given that $F_i(q) \neq F_i(q')$, we have that $\operatorname{proj}_{A_i^{\leq T}} q = a_i \neq a_i' = \operatorname{proj}_{A_i^{\leq T}} q'$ or $f_i(q) \neq f_i(q')$. By cumulativeness of feedback, $f_i(q) \neq f_i(q')$. Assume now that the statement holds for $\ell(F_i(q)) = \ell(F_i(q')) = t < T$ and consider $\ell(F_i(q)) = \ell(F_i(q')) = t+1$. Write $F_i(q) = (F_i(p), a_i, f_i(q))$ and $F_i(q') = (F_i(p'), a_i', f_i(q'))$, where p and p' are the longest predecessors of, respectively, q and q' where player i is alert. Assume $F_i(q) \neq F_i(q')$. Then, (i) $F_i(p) \neq F_i(p')$, (ii) $a_i \neq a_i'$, or (iii) $f_i(q) \neq f_i(q')$. If (i) holds, $f_i(p) \neq f_i(p')$ by the inductive hypothesis. Then, since $\succeq p$ and $q' \succeq p'$, cumulativeness of feedback implies that $f_i(q) \neq f_i(q')$. If (ii) holds, it means that $\operatorname{proj}_{A_i^{\leq T}} q \neq \operatorname{proj}_{A_i^{\leq T}} q'$, and cumulativeness of feedback implies $f_i(q) \neq f_i(q')$. If (iii) holds, the result is obvious. This concludes the induction. ■

Now fix a flow-based game structure with memory $\langle \Gamma_F, \mathcal{M} \rangle$ where feedback is cumulative. Fix $i \in I$ and $p \in P_i$. We want to show that $\mathcal{F}_i(p) = \mathcal{P}_i(p)$. Take $q \in \mathcal{P}_i(p)$. Given that, for each $i \in I$, $\mathcal{M}_i$ satisfies the "knowledge of last message" condition, it must be the case that $f_i(p) = f_i(q)$. Thanks to Lemma A1, this implies $F_i(p) = F_i(q)$. Therefore, $q \in \mathcal{F}_i(p)$. For the converse inclusion, take $q \in \mathcal{F}_i(p)$. Note that it has to be the case that $\mathcal{M}_i(F_i(p))$ is nonempty (by definition, memory correspondences are taken to be nonempty-valued). Then, there must exist $r \in P_i$ such that $F_i(r) \in \mathcal{M}_i(F_i(p))$. By the "knowledge of last message" condition, $F_i(r) \in \mathcal{M}_i(F_i(p))$ implies $f_i(r) = f_i(p) = f_i(q)$, or, equivalently, $F_i(r) = F_i(p) = F_i(q)$ (Lemma A1). It follows that $F_i(q) \in \mathcal{M}_i(F_i(p))$, and therefore $q \in \mathcal{P}_i(p)$. ■

A.4 Proof of Proposition 1 (p. 23)

"Only if" direction Consider a standard game structure $\Gamma_S = \langle I, V, \mathcal{I}, (A_i, \mathscr{Q}_i)_{i \in I} \rangle$ satisfying the perfect recall property. By inspection of the defintion of experience functions and of the perfect recall property (PR),

$$\forall\, i \in I, \forall\, v, w \in V_i, \quad Q_i(v) = Q_i(w) \Longleftrightarrow X_i(v) = X_i(w). \qquad \text{(Q}\leftrightarrow\text{X)}$$

Now, we want to construct a flow-based game structure $\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i \in I} \rangle$ that gives $G(\Gamma_F, \mathcal{M}) = \Gamma_S$ when coupled with a profile of memory correspondences $\mathcal{M} = (\mathcal{M}_i)_{i \in I}$ satisfying perfect memory. We begin with the construction.

[32] The length of personal histories is always an odd number. We start from 3 because there is only one personal history of length 1 – the empty history.

First of all, for each $i \in I$, let $M_i := \mathscr{Q}_i$.

Then, define $\tilde{f}_i : A^{\leq T} \to M_i$ to be, for each $a^t \in A^{\leq T}$,

$$\tilde{f}_i(a^t) := \begin{cases} Q_i(a^t) & \text{if } a^t \in V_i; \\ Q_i & \text{otherwise;} \end{cases}$$

where Q_i is a generic element of $\mathscr{Q}_i$. Now let $\tilde{f} : A^{\leq T} \to M$ be, for each $a^t \in A^{\leq T}$,

$$\tilde{f}(a^t) := \begin{cases} (\tilde{f}_i(a^t))_{i \in \mathcal{I}(a^t)} & \text{if } a^t \in V; \\ (\tilde{f}_i(a^t))_{i \in I} & \text{otherwise.} \end{cases}$$

Moreover, for each $i \in I$ define $\mathcal{A}_i : M_i \rightrightarrows A_i$ to be such that, for each $m_i \in M_i$,

$$\mathcal{A}_i(m_i) := A_i(w),$$

where w is a generic element of $\{v \in V_i : Q_i(v) = m_i\}$ and $A_i(w)$ is the set of i's available actions at w. Note that player i's feasible actions ($i \in I$) are the same for each play in such set thanks to property (KfA), so that $\mathcal{A}_i$ is always well-defined.

Consider now $\Gamma_F := \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i \in I} \rangle$, where M_i, $\tilde{f}$ and $(\mathcal{A}_i)_{i \in I}$ are defined in the way just described. Consider also the derived objects P (set of feasible plays), f (collective feedback function restricted to P), and $(F_i : P_i \to H_i)_{i \in I}$ (functions mapping plays to the induced personal histories). Also, let $\mathcal{M} = (\mathcal{M}_i)_{i \in I}$ satisfy perfect memory. Then, the derived possibility correspondences $(\mathcal{P}_i)_{i \in I}$ coincide with $(\mathcal{F}_i)_{i \in I}$ (Remark 4), where $\mathcal{F}_i$ is the partition induced by (i.e., the range of) F_i^{-1}. By definition (cf. Section 5.1), $G(\Gamma_F, \mathcal{M}) = \langle I, P, \mathcal{D} \circ f, (A_i, \mathscr{P}_i)_{i \in I} \rangle$, where $\mathscr{P}_i$ is the partition induced by $\mathcal{P}_i = \mathcal{F}_i$.

Notice the following. By construction, we have $P = V$. Moreover, $\mathcal{D}(f(v)) = \mathcal{I}(v)$ for each $v \in V = P$, so that $\mathcal{D} \circ f = \mathcal{I}$. To check that $\mathscr{P}_i = \mathscr{Q}_i$ ($i \in I$), we proceed in three steps.

To check condition (*iii*), we proceed in two steps. The way in which we defined the feedback function $\tilde{f}$ implies the following:

$$\forall\, i \in I, \forall\, v, w \in V_i, \quad X_i(v) = X_i(w) \iff F_i(v) = F_i(w). \tag{X$\leftrightarrow$F}$$

To see why this holds, fix $i \in I$ and $v, w \in V_i$. Then:

- Assume $X_i(v) = X_i(w)$. This implies $\ell(X_i(v)) = \ell(X_i(w))$, so we can proceed by induction on such length.
  Basis step. Assume $\ell(X_i(v)) = \ell(X_i(w)) = 1$. This can only happen if $v = w = \varnothing$, so that $X_i(v) = X_i(w) = (Q_i(\varnothing))$. Then, $F_i(v) = F_i(w)$ trivially holds as $v = w$.
  Inductive step. Now assume that the claim holds for $\ell(X_i(v)) = \ell(X_i(w)) \in \{1, \dots, n\}$ for some $n \in \mathbb{N}$, and suppose $\ell(X_i(v)) = \ell(X_i(w)) = n + 1$. By perfect recall, it can be checked that $X_i(t) = X_i(u)$ for each $t \preceq v$ and $u \preceq w$ where player i is alert. Hence, $X_i(\mathrm{last}_i\, v) = X_i(\mathrm{last}_i\, w)$, which implies $F_i(\mathrm{last}_i\, v) = F_i(\mathrm{last}_i\, w)$ thanks to the inductive hypothesis (as $\ell(X_i(\mathrm{last}_i\, v)) = \ell(X_i(\mathrm{last}_i\, w)) < n$). Therefore,

  $$(X_i(\mathrm{last}_i\, v), a_i(\mathrm{last}_i\, v, v), Q_i(v)) = (X_i(\mathrm{last}_i\, w), a_i(\mathrm{last}_i\, w, w), Q_i(w)),$$

  which implies $a_i(\mathrm{last}_i\, v, v) = a_i(\mathrm{last}_i\, w, w)$ and $Q_i(v) = Q_i(w)$. By construction of $\tilde{f}$, $Q_i(x) = \tilde{f}_i(x) = f_i(x)$ for $x \in \{v, w\}$, and this allows to say that $f_i(v) = f_i(w)$. Wrapping up, we write

  $$F_i(v) = (F_i(\mathrm{last}_i\, v), a_i(\mathrm{last}_i\, v, v), f_i(v)),$$
  $$F_i(w) = (F_i(\mathrm{last}_i\, w), a_i(\mathrm{last}_i\, w, w), f_i(w)),$$

  and we conclude that $F_i(v) = F_i(w)$ in light of the foregoing observations.

- Assume $F_i(v) = F_i(w)$. Again, we proceed by induction on $\ell(F_i(v)) = \ell(F_i(w)) \in \mathbb{N}_0$.
  Basis step. The result trivially holds for $\ell(F_i(v)) = \ell(F_i(w)) = 0$.
  Inductive step. By arguments analogous to the proof above.

At this point, (Q↔X) and (X↔F) yield

$$\forall\, i \in I, \forall\, v, w \in V_i, \quad Q_i(v) = Q_i(w) \Longleftrightarrow F_i(v) = F_i(w).$$

This means that the partitions $\mathscr{Q}_i$ and $\mathscr{F}_i$ coincide, and this concludes the proof that $\Gamma_S = G(\Gamma_F, \mathcal{M})$.

"If" direction Consider a flow-based game structure Γ_F, a profile of memory correspondences $\mathcal{M} = (\mathcal{M}_i)_{i\in I}$ satisfying perfect memory, and the derived (standard) game structure $G(\Gamma_F, \mathcal{M}) = \langle I, P, \mathcal{D} \circ f, (A_i, \mathscr{F}_i)_{i\in I}$. Note that we directly write $\mathscr{F}_i$ instead of $\mathscr{P}_i$ because the two coincide when $\mathcal{M}_i$ satisfies perfect memory (Remark 4). We show that the perfect recall property (PR) holds in $G(\Gamma_F, \mathcal{M})$.

Fix a player $i \in I$, and $p, q \in P = V$ such that $\mathcal{F}_i(p) = \mathcal{F}_i(q)$ (i.e., p and q belong to the same cell of $\mathscr{F}_i$). Now note that $F_i(p) = F_i(q)$ implies that $\ell(F_i(p)) = \ell(F_i(q))$. We can then proceed by induction on such length. Recall that the lengths of (personal) histories are always odd numbers (cf. Section 3.2).

Basis step. Assume that $\ell(F_i(p)) = \ell(F_i(q)) = 1$. This implies that $p = q = \varnothing$. Hence, $X_i(p) = X_i(q)$ trivially holds.

Inductive step. Assume that the claim holds for $\ell(F_i(p)) = \ell(F_i(q)) = k \in \{1, 3, \dots, t\}$, and focus on $p, q \in P$ such that $\ell(F_i(p)) = \ell(F_i(q)) = t + 2 \in \mathbb{N}$ and $Q_i(p) = Q_i(q)$. Note that $Q_i(p) = Q_i(q)$ implies that $Q_i(r) = Q_i(s)$ for each $r \preceq p$ and $s \preceq q$ where player i is alert. Given that $\mathrm{last}_i\, p \preceq p$ and $\mathrm{last}_i\, q \preceq q$, we have $Q_i(\mathrm{last}_i\, p) = Q_i(\mathrm{last}_i\, q)$. Since $\ell(\mathrm{last}_i\, p) = \ell(\mathrm{last}_i\, q) \leq t$, the inductive hypothesis implies that $X_i(\mathrm{last}_i\, p) = X_i(\mathrm{last}_i\, q)$. Now note that

$$X_i(p) = (X_i(\mathrm{last}_i\, p), a_i(\mathrm{last}_i\, p, p), Q_i(p)),$$
$$X_i(q) = (X_i(\mathrm{last}_i\, q), a_i(\mathrm{last}_i\, q, q), Q_i(q)).$$

We already showed that $X_i(\mathrm{last}_i\, p) = X_i(\mathrm{last}_i\, q)$. Moreover, $Q_i(p) = Q_i(q)$ by assumption. Lastly, $a_i(\mathrm{last}_i\, p, p) = a_i(\mathrm{last}_i\, q, q)$ because $F_i(p) = F_i(q)$. Therefore, $X_i(p) = X_i(q)$, and this concludes the proof. ■

A.5 Proof of Proposition 2 (p. 23)

Consider a standard game structure $\Gamma_S = \langle I, V, \mathcal{I}, (A_i, \mathscr{Q}_i)_{i\in I}\rangle$. We want to find a flow-based game structure $\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i\in I}\rangle$ and a profile of memory correspondences $\mathcal{M} = (\mathcal{M}_i)_{i\in I}$ such that $G(\Gamma_F, \mathcal{M}) = \Gamma_S$. Recall that $G(\Gamma_F, \mathcal{M}) = \langle I, P, \mathcal{D} \circ f, (A_i, \mathscr{P}_i)_{i\in I}\rangle$, where P, f, and $(\mathscr{P}_i)_{i\in I}$ are as usual. For the result to hold, the memory correspondences clearly have to be partitional.

We prove the result in a way similar to Proposition 1. In this case however, we let feedback be perfect – with this, $\mathscr{F}_i$ is finest partition of P_i. Then, we conclue that there exists a partitional memory correspondence $\mathcal{M}_i$ such that $\mathscr{P}_i = \mathscr{Q}_i$.

As a matter of notation, denote as $V^t \subseteq A^t$ ($t \in \{0, \dots, T\}$) the set of t-long plays in V: formally $V^t := \{v \in V : \ell(v) = t\}$. Similarly, let $V_i^t \subseteq A^t$ ($i \in I$, $t \in \{0, \dots, T\}$) the set of t-long plays in V after which player i is alert: formally $V_i^t := \{v \in V_i : \ell(v) = t\}$.

Step 1: construction of the flow-based game structure. To construct the desired flow-based game structure, we proceed as follows. First, for each $i \in I$, let $M_i := A^{\leq T}$. Second, let $\tilde{f} : A^{\leq T} \to A^{\leq T}$ be the map defined for each $a^t \in A^{\leq T}$ as

$$\tilde{f}(a^t) := \begin{cases} (a^t)_{i\in\mathcal{I}(a^t)} & \text{if } a^t \in V; \\ (a^t)_{i\in I} & \text{otherwise.} \end{cases}$$

Third, for each $i \in I$, let $\mathcal{A}_i$ be defined for each $a^t \in A^{\leq T}$ as

$$\mathcal{A}_i(a^t) := \begin{cases} A_i(a^t) & \text{if } a^t \in V_i; \\ \emptyset & \text{otherwise;} \end{cases}$$

where $A_i(a^t)$ is the set of player i's available actions at a^t as specified by V.

Step 2: verify that $P = V$. Let P be the set of feasible plays induced by the flow-based game structure Γ_F. We proceed by induction on the length of such plays. Let $P^t := \{p \in P : \ell(p) = t\}$ ($t \in \{0, \ldots, T\}$) be the set of feasible plays with length t.

Then, $P^0 = \{\varnothing\} = V^0$ holds trivially. Assume by way of induction that $P^t = V^t$ for $t \in \{0, \ldots, T-1\}$: we want to show that $P^{t+1} = V^{t+1}$.

Consider $p = (p', a) \in P^{t+1}$. We have that (i) $\mathcal{D}(a) = \mathcal{D}(\tilde{f}(p')) = \mathcal{I}(p')$, and (ii) $a \in \bigtimes_{i \in \mathcal{I}(p')} \mathcal{A}_i(p') = \bigtimes_{i \in \mathcal{I}(p)} A_i(p')$. In particular, this implies that $a \in A(p')$. Hence, we have $p'^t \subseteq V$, and $a \in A(p')$, so that $(p', a) \in V$ – specifically, $(p', a) \in V^{t+1}$.

Conversely, fix $v = (v', a) \in V^{t+1}$. We have that $v'^t = P^t$. Moreover, by condition (APF) of Definition 2, we have $A(v') = \bigtimes_{i \in \mathcal{I}(v')} A_i(v')$. Lastly, note that $\mathcal{I}(v') = \mathcal{D}(a) = \mathcal{D}(\tilde{f}(v'))$, where the first equality follows from condition (APM) of Definition 2 and the second one from the definition of $\tilde{f}$. Therefore, we conclude that $a \in \bigtimes_{i \in \mathcal{D}(\tilde{f}(v'))} A_i(v') = \bigtimes_{i \in \mathcal{D}(\tilde{f}(v'))} \mathcal{A}_i(p')$, with the second equality following from the definition of $(\mathcal{A}_i)_{i \in I}$. As a result, we obtain that $v = (v', a) \in \{v'\} \times \left(\bigtimes_{i \in \mathcal{D}(\tilde{f}(v'))} \mathcal{A}_i(v')\right) \subseteq P$, where the inclusion holds by the recursive definition of set P (cf. Section 3.2). In particular, $p \in P^{t+1}$, and this concludes the induction.

It is also easy to verify that $P_i = V_i$ for each $i \in I$.

Step 3: verify that, for each player, distinct plays are always distinguishable. Note that the feedback function $\tilde{f}$ is injective. Hence, for each $p, q \in P_i = V_i$ with $p \neq q$, $F_i(p) \neq F_i(q)$.

Step 4: check that there is a profile of memory corresponences inducing $(\mathscr{Q}_i)_{i \in I}$. In light of the previous step, we can say that $\mathscr{F}_i$ ($i \in I$) is the finest partition of P_i. Now, $\mathscr{Q}_i$ is also a partition of P_i (one where condition (KfA) holds). As such, it must be a coarsening of $\mathscr{F}_i$. Denote by $\mathcal{Q}_i$ the partitional possibility correspondence generating $\mathscr{Q}_i$. It is easy to see that letting $\mathcal{M}_i = \mathcal{Q}_i$ delivers the desired result $\mathscr{P}_i = \mathscr{Q}_i$. This concludes the proof. ■

Proof of Proposition 3 (p. 24)

"Only if" direction The proof of this direction of the statement follows essentially the same lines as that of Proposition 1. Fix a standard game structure $\Gamma_S = \langle I, V, \mathcal{I}, (A_i, \mathscr{Q}_i)_{i \in I} \rangle$ where the perfect recall property holds. For each $i \in I$, let X_i be the experience function of player i as defined in Section 5.

We construct the desired flow-based game structure as follows. First of all, for each $i \in I$, let $M_i := \text{range} X_i$. Then, define $\tilde{f}_i : A^{\leq T} \to M_i$ to be, for each $a^t \in A^{\leq T}$,

$$\tilde{f}_i(a^t) := \begin{cases} X_i(a^t) & \text{if } a^t \in V_i; \\ x_i & \text{otherwise;} \end{cases}$$

where x_i is a generic element of range X_i. Now let $\tilde{f} : A^{\leq T} \to M$ be, for each $a^t \in A^{\leq T}$,

$$\tilde{f}(a^t) := \begin{cases} (\tilde{f}_i(a^t))_{i \in \mathcal{I}(a^t)} & \text{if } a^t \in V; \\ (\tilde{f}_i(a^t))_{i \in I} & \text{otherwise.} \end{cases}$$

It is possible to check that such feedback is cumulative (see Section 3.4 for a definition – intuitively, cumulative feedback has to always incorporate past feedback as the play progresses, and this is granted by the fact that we are using experience functions to define it).

Moreover, for each $i \in I$ define $\mathcal{A}_i : M_i \rightrightarrows A_i$ to be such that, for each $m_i \in M_i$,

$$\mathcal{A}_i(m_i) := A_i(w),$$

where w is a generic element of $\{v \in V_i : X_i(v) = m_i\}$ and $A_i(w)$ is the set of i's available actions at w. Note that player i's feasible actions ($i \in I$) are the same for each play in such set. This is because, if $X_i(v) = X_i(w) = m_i$ for $v, w \in V$, then $Q_i(v) = Q_i(w)$ – that is, the two nodes belong to the same information set. Therefore, property (KfA) ensures that $A_i(v) = A_i(w)$.

Consider now the flow-based game structure $\Gamma_F = \langle I, \tilde{f}, (A_i, M_i, \mathcal{A}_i)_{i\in I}\rangle$ defined in the way just described. Consider also the derived objects P, f, and $(F_i : P_i \to H_i)_{i\in I}$. By construction (and like in the proof of Proposition 1), we have $P = V$ and $P_i = V_i$ for each $i \in I$. Now let $\mathcal{M} = (\mathcal{M}_i)_{i\in I}$ be a 1-bounded profile of memory correspondences. That is, players only correctly remember the last piece of information they observed – this verifies condition (KLM) as well. Then, the derived possibility correspondences $(\mathcal{P}_i)_{i\in I}$ are such that, for $i \in I$ and $p \in P_i = V_i$,

$$\mathcal{P}_i(p) = \{q \in P_i : X_i(p) = X_i(q)\}.$$

As usual, let $\mathscr{P}_i$ ($i \in I$) denote the partition induced by $\mathcal{P}_i$.

Now consider $G(\Gamma_F, \mathcal{M}) = \langle I, P, \mathcal{D} \circ f, (A_i, \mathscr{P}_i)_{i\in I}\rangle$. We already mentioned that $P = V$ by construction. Moreover, $\mathcal{D}(f(v)) = \mathcal{I}(v)$ for each $v \in V = P$, so that $\mathcal{D} \circ f = \mathcal{I}$. To check that $\mathscr{P}_i = \mathscr{Q}_i$ ($i \in I$), we leverage the perfect recall property that is satisfied by Γ_S. As stated at the beginning of the proof of Proposition 1, perfect recall implies the condition (Q↔X) – that is,

$$\forall\, i \in I, \forall\, v, w \in V_i, \quad Q_i(v) = Q_i(w) \Longleftrightarrow X_i(v) = X_i(w).$$

By construction of $\mathscr{P}_i$ ($i \in I$), we have that

$$\forall\, i \in I, \forall\, v, w \in V_i, \quad X_i(v) = X_i(w) \Longleftrightarrow \mathcal{P}_i(v) = \mathcal{P}_i(w).$$

Putting together these two conditions, one concludes that $\mathscr{P}_i = \mathscr{Q}_i$ for each $i \in I$.

"If" direction Consider a flow-based game structure with memory $\langle \Gamma_F, \mathcal{M}\rangle$, and the induced generalized standard game structure Γ_S. If feedback is cumulative, Lemma 3 implies that $\mathcal{P}_i = \mathcal{F}_i$ for each $i \in I$. Let $\mathscr{F}_i$ ($i \in I$) denote the induced partition – this is the collection of player i's information sets in the standard game structure Γ_S. The proof of the "if" direction of Proposition 1 shows that $\mathscr{F}_i$ satisfies the perfect recall property (PR) for each $i \in I$, and this concludes the proof. ■

References

Alós-Ferrer, C., & Ritzberger, K. (2008). Trees and extensive forms. *Journal of Economic Theory*, *143*(1), 216–250.

Alós-Ferrer, C., & Ritzberger, K. (2013). Large extensive form games. *Economic Theory*, *52*(1), 75–102.

Alós-Ferrer, C., & Ritzberger, K. (2017). Characterizations of perfect recall. *International Journal of Game Theory*, *46*(2), 311–326.

Alós-Ferrer, C., & Ritzberger, K. (2016). *The theory of extensive form games.* Berlin Heidelberg: Springer-Verlag.

Battigalli, P., Catonini, E., & De Vito, N. (2021). *Game Theory: Analysis of Strategic Thinking.* Mimeo.

Battigalli, P., Catonini, E., Lanzani, G., & Marinacci, M. (2019). Ambiguity attitudes and self-confirming equilibrium in sequential games. *Games and Economic Behavior*, *115*, 1–29.

Battigalli, P., Cerreia-Vioglio, S., Maccheroni, F., & Marinacci, M. (2015). Self-confirming equilibrium and model uncertainty. *American Economic Review*, *105*(2), 646–77.

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

Battigalli, P., Leonetti, P., & Maccheroni, F. (2020). Behavioral equivalence of extensive game structures. *Games and Economic Behavior*, *121*, 533–547.

Bonanno, G. (2003). A syntactic characterization of perfect recall in extensive games. *Research in Economics*, *57*(3), 201–217.

Bonanno, G. (2004). Memory and perfect recall in extensive games. *Games and Economic Behavior*, *47*(2), 237–256.

Bordalo, P., Coffman, K., Gennaioli, N., Schwerter, F., & Shleifer, A. (2020). Memory and representativeness. *Psychological Review*.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2019). Memory and reference prices: An application to rental choice. *AEA Papers and Proceedings*, *109*, 572–76.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2020). Memory, attention, and choice. *The Quarterly Journal of Economics*, *135*(3), 1399–1442.

Caplin, A., & Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, *116*(1), 55–79.

Fudenberg, D., Lanzani, G., & Strack, P. (2022). *Selective memory equilibrium* (mimeo).

Fudenberg, D., & Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica*, *61*(3), 523–545.

Green, E. J., & Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica*, *52*(1), 87–100.

Harris, C. (1985). Existence and characterization of perfect equilibrium in games of perfect information. *Econometrica*, *53*(2), 613–628.

Kahana, M. J. (2012). *Foundations of Human Memory.* Oxford: Oxford University Press.

Kechris, A. (1995). *Classical Descriptive Set Theory.* New York, NY: Springer-Verlag.

Kreps, D. M., & Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 185–200.

Kuhn, H. W. (1953). Extensive games and the problem of information. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games* (p. 193-216). Princeton, NJ: Princeton University Press.

Mailath, G. J., & Samuelson, L. (2006). *Repeated Games and Reputations: Long-run Relationships.* Oxford university press.

Myerson, R. B. (1986). Multistage games with communication. *Econometrica*, *54*(2), 323–358.

Myerson, R. B., & Reny, P. J. (2020). Perfect conditional ε-equilibria of multi-stage games with infinite sets of signals and actions. *Econometrica*, *88*(2), 495–531.

Okada, A. (1987). Complete inflation and perfect recall in extensive games. *International Journal of Game Theory*, *16*(2), 85–91.

Osborne, M. J., & Rubinstein, A. (1994). *A Course in Game Theory.* Cambridge, MA: MIT Press.

Perea, A. (2001). Rationality in extensive form games. In *Theory and Decision Library C.* Boston, MA: Kluwer Academic Publisher.

Piccione, M., & Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, *20*(1), 3–24.

Ritzberger, K. (1999). Recall in extensive form games. *International Journal of Game Theory*, *28*(1), 69–87.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, *4*(1), 25–55.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.