On Imperfect Recall in Multi-Agent Influence Diagrams

James Fox
University of Oxford
james.fox@cs.ox.ac.uk

Matt MacDermott
Imperial College London
m.macdermott21@imperial.ac.uk

Lewis Hammond
University of Oxford
lewis.hammond@cs.ox.ac.uk

Paul Harrenstein
University of Oxford
paul.harrenstein@cs.ox.ac.uk

Alessandro Abate
University of Oxford
aabate@cs.ox.ac.uk

Michael Wooldridge
University of Oxford
mjw@cs.ox.ac.uk

Multi-agent influence diagrams (MAIDs) are a popular game-theoretic model based on Bayesian networks. In some settings, MAIDs offer significant advantages over extensive-form game representations. Previous work on MAIDs has assumed that agents employ behavioural policies, which set independent conditional probability distributions over actions for each of their decisions. In settings with imperfect recall, however, a Nash equilibrium in behavioural policies may not exist. We overcome this by showing how to solve MAIDs with forgetful and absent-minded agents using mixed policies and two types of correlated equilibrium. We also analyse the computational complexity of key decision problems in MAIDs, and explore tractable cases. Finally, we describe applications of MAIDs to Markov games and team situations, where imperfect recall is often unavoidable.

1 Introduction

Multi-agent influence diagrams (MAIDs) are a graphical representation for dynamic non-cooperative games, which can be more compact and expressive than extensive-form games (EFGs) [25]. Like Bayesian networks (BNs), MAIDs use a directed acyclic graph (DAG) to represent conditional probabilistic dependencies between random variables, but they also specify decision and utility variables for each agent. Each agent selects a behavioural policy – independent conditional probability distributions (CPDs) over actions for each of their decision variables – to maximise their expected utility. A MAID's mechanised graph extends this DAG by explicitly representing each variable's distribution and showing which other variables' distributions matter to an agent optimising a particular decision rule [18, 25, 10].

MAIDs, and their causal variants [18], have been used in the design of safe and fair AI systems [14, 1, 15, 16, 7], to explore reasoning patterns and deception [40, 48], and to identify agents from data [22]. However, to date, agents in MAIDs are usually assumed to have perfect (or, at least, 'sufficient') recall [25]. This assumption is often unreasonable. For example, MAIDs must allow imperfect recall to handle bounded rationality, teams with imperfect communication [13], or memoryless policies in Markov games. However, forgetfulness (of previous observations) or absent-mindedness (about whether previous decisions have even been made) can prevent the existence of a Nash Equilibrium (NE) in behavioural policies. To overcome this, one can consider other solution concepts, such as mixed or correlated equilibria.

In this work, we focus on imperfect recall in MAIDs. Imperfect recall has already been extensively studied in EFGs [41, 26, 49], but a MAID's mechanised graph makes graphically explicit the semantic difference between behavioural and mixed policies (hidden in EFGs) and readily identifies forgetful or absent-minded agents (or teams). Our insights inspire two definitions of *correlated equilibrium* in MAIDs. The first follows from the normal-form game definition [2]. The second, based on von Stengel

R. Verbrugge (Ed.): Theoretical Aspects of
Rationality and Knowledge 2023 (TARK 2023)
EPTCS 379, 2023, pp. 201–220, doi:10.4204/EPTCS.379.17

and Forges' extensive-form correlated equilibrium [47], is more natural for dynamic settings, can yield greater social welfare, and is easier to compute. Again, mechanised graphs clearly depict the assumptions made in both. Next, we examine MAIDs from a computational complexity perspective by studying the decision problems of finding a best response, checking whether a policy profile is an NE, and checking whether each type of NE exists. These provide an insight into what makes particular instances hard, when computations can be made tractable, and rigorously identify which problems are suitable for analysis as MAIDs. Our results also apply to refinements of MAIDs, such as *causal games* [18]. We assume familiarity with EFGs [31], BNs [24], and the complexity classes P, NP, and PP [38]. Proof sketches are provided, but details are deferred to the appendices.

Related Work. There is a rich literature on influence diagrams [23] and imperfect recall has been studied in single-agent influence diagrams [33, 34, 29, 6, 35] as well as in EFGs [3, 21, 26, 41, 49]. However, to our knowledge, we are the first to focus on imperfect recall in influence diagrams with multiple agents.

A full policy profile in a MAID induces a BN, so many of our results inherit from that setting, where the decision problem variant of marginal inference is, in general, PP-complete [30]. However, we care about the cases we encounter in practice, not just the worst case. Marginal inference in a BN can be performed in time exponential in the treewidth of the underlying graph [24], which entails a poly-time algorithm when the treewidth is small. Similarly, we will see that tractable results for computations in MAIDs can be found when problems are restricted to certain settings. We also sometimes reduce from partial order games [50], which can be interpreted as MAIDs without chance nodes, with deterministic decision rules, and where each agent has a single utility node as a child of all the decision nodes.

2 The Model

We use capital letters V for random variables, lowercase letters v for their instantiations, and bold letters $\boldsymbol{V}$ and $\boldsymbol{v}$, respectively, for sets of variables and their instantiations. We let $dom(V)$ denote the (finite, non-singleton) domain of V (for ease, we take this to be binary unless stated otherwise) and $dom(\boldsymbol{V}) := \times_{V \in \boldsymbol{V}} dom(V)$. Parents and children of V in a graph are denoted by $\mathbf{Pa}_V$ and $\mathbf{Ch}_V$, respectively (with $\mathbf{pa}_V$ and $\mathbf{ch}_V$ their instantiations) and $\Delta(X)$ denotes the set of all probability distributions over a set X.

Example 1. *An autonomous taxi decides whether to offer Alice a discount (T) depending on whether its journey count exceeds a quota (Q). Alice decides whether to accept a journey (A) depending on the price. The taxi wants to maximise profit, but if its journey count is less than the quota and Alice rejects it, the taxi pays a penalty (the municipality uses this mechanism to prevent a proliferation of unnecessary taxis). Alice's utility is a function of her decision and the price offered by the taxi.*

Figure 1a shows a MAID for this example. Chance variables (moves by nature), decision variables, and utility variables are represented by white circles, squares, and diamonds, respectively. Full edges leading into chance and utility nodes represent probabilistic dependence, as in a BN. Dotted edges leading into decision nodes identify information available to the agent when a decision D is made, so $\mathbf{pa}_D$, the values of $\mathbf{Pa}_D$, represents the decision context for D. In EFGs, imperfect information is represented using explicitly labelled information sets. In MAIDs, we can infer that Alice is unaware of the value of Q when making her decision by the lack of edge $Q \to A$. A parameterisation defines the CPDs for the chance and utility variables, whereas CPDs of decision nodes are chosen by the agents playing the game.

Definition 1 ([25])**.** *A **multi-agent influence diagram (MAID)** is a structure $\mathscr{M} = (\mathscr{G}, \boldsymbol{\theta})$. $\mathscr{G} = (N, \boldsymbol{V}, E)$ specifies a set of agents $N = \{1, \dots, n\}$ and a DAG $(\boldsymbol{V}, E)$, where $\boldsymbol{V}$ is partitioned into chance variables*

(a) (b)

Figure 1: A MAID (a) and its mechanised graph (b) for Example 1, which is a perfect recall and imperfect, but sufficient, information game.

$\boldsymbol{X}$, *decision variables* $\boldsymbol{D} = \bigcup_{i \in N} \boldsymbol{D}^i$, *and utility variables* $\boldsymbol{U} = \bigcup_{i \in N} \boldsymbol{U}^i$. *The parameters* $\boldsymbol{\theta} = \{\theta_V\}_{V \in \boldsymbol{V} \setminus \boldsymbol{D}}$ *define the CPDs* $\Pr(V \mid \boldsymbol{Pa}_V)$ *for each non-decision variable such that for any setting of the decision variables' CPDs, the resulting joint distribution over* $\boldsymbol{V}$ *is Markov compatible with the DAG, i.e.,* $\Pr(\boldsymbol{v}) = \prod_{V \in \boldsymbol{V}} \Pr(v \mid \boldsymbol{pa}_V)$.

Given a MAID, a **decision rule** π_D for $D \in \boldsymbol{D}$ is a CPD $\pi_D(D \mid \mathbf{Pa}_D)$. A **partial (behavioural) policy profile** $\pi_{\boldsymbol{D}'}$ is a set of decision rules for each $D \in \boldsymbol{D}' \subseteq \boldsymbol{D}$, whereas $\pi_{-\boldsymbol{D}'}$ is the set of decision rules for each $D \in \boldsymbol{D} \setminus \boldsymbol{D}'$. A **(behavioural) policy** $\boldsymbol{\pi}^i$ refers to $\boldsymbol{\pi}_{\boldsymbol{D}^i}$, and a **(full) policy profile** $\boldsymbol{\pi} = (\boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^n)$ is a tuple of policies, where $\boldsymbol{\pi}^{-i} := (\boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^{i-1}, \boldsymbol{\pi}^{i+1}, \ldots, \boldsymbol{\pi}^n)$. A decision rule is **pure** if $\pi_D(d \mid \mathbf{pa}_D) \in \{0, 1\}$, which holds for a policy (profile) if it holds for all decision rules in the policy (profile). For clarity, we use an overhead dot to mark this determinism, e.g., $\dot{\pi}_D, \dot{\boldsymbol{\pi}}^i$, or $\dot{\boldsymbol{\pi}}$.

By combining $\boldsymbol{\pi}$ with the partial distribution Pr over the chance and utility variables, we obtain a joint distribution:

$$\Pr^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{u}) := \prod_{V \in \boldsymbol{V} \setminus \boldsymbol{D}} \Pr(v \mid \mathbf{pa}_V) \cdot \prod_{D \in \boldsymbol{D}} \pi_D(d \mid \mathbf{pa}_D)$$

A full policy profile $\boldsymbol{\pi}$ therefore induces a BN with DAG given by the MAID's graph. Agent i's **expected utility** $EU^i(\boldsymbol{\pi})$ for a given policy profile $\boldsymbol{\pi}$ is defined as the expected sum of their utility variables:

$$EU^i(\boldsymbol{\pi}) := \sum_{U \in \boldsymbol{U}^i} \sum_{u \in dom(U)} \Pr^{\boldsymbol{\pi}}(U = u) \cdot u$$

Utility variables have deterministic CPDs, so can be interpreted as functions $U : dom(\mathbf{Pa}_U) \to \mathbb{R}$ to show their functional dependence on their parents (e.g., Figure 1a). An NE is defined in the usual way.

Definition 2 ([25])**.** *A (behavioural) policy profile* $\boldsymbol{\pi}$ *is a* ***Nash equilibrium (NE)*** *(in behavioural policies) if for every agent* $i \in N$ *and every alternative (behavioural) policy* $\boldsymbol{\varpi}^i$*:* $EU^i(\boldsymbol{\pi}^{-i}, \boldsymbol{\pi}^i) \geq EU^i(\boldsymbol{\pi}^{-i}, \boldsymbol{\varpi}^i)$

Collectively, the decision rules of decision variables and the CPDs of chance or utility nodes are known as mechanisms. A mechanism M_V for V is **strategically relevant** to a decision rule for D if the choice of the CPD at M_V can affect the optimal choice of this decision rule. Koller and Milch [25] define an associated sound and complete graphical criterion for strategic relevance, ***s*-reachability**, based on d-separation which can be checked in $\mathcal{O}(|\boldsymbol{V}| + |E|)$ time [43] (see Appendix A for formal definitions).

A MAID's regular graph $\mathcal{G}$ captures the probabilistic dependencies between **object-level** variables in the game's environment, but its **mechanised graph** $\mathsf{m}\mathcal{G}$ is an enhanced representation which adds an explicit representation of the strategically relevant dependencies between agents' decision rules and the game's parameterisation (see [18] for details). Each object-level variable $V \in \boldsymbol{V}$ has a mechanism parent M_V representing the distribution governing V: each decision D has a new *decision rule* parent $\Pi_D = \mathsf{M}_D$ and each non-decision V has a new *parameter* parent $\Theta_V = \mathsf{M}_V$, whose values parameterise the CPDs.

Agents select a decision rule π_D (i.e., the value of a decision rule variable Π_D) based on both the parameterisation of the game (i.e., the values of the parameter variables) and the selection of the other

decision rules $\boldsymbol{\pi}_{-D}$ – these dependencies are captured by the edges from other mechanisms into decision rule nodes. s-reachability determines which of these edges are necessary, so $\mathsf{M}_V \rightarrow \Pi_D$ exists if and only if Π_D strategically relies on M_V. The mechanised graph for Example 1 (in Figure 1b) shows that Π_T strategically relies on Θ_{U^T} and Π_A, whereas Π_A only strategically relies on Θ_{U^A}. In contrast to a MAID's regular graph $\mathcal{G}$, which is a DAG, there may exist cycles between mechanisms (e.g., Figure 3a).

For convenience, we denote the set of agent i's behavioural policies as $\boldsymbol{P}^i := dom(\boldsymbol{\Pi}^i)$, with sets of pure policies denoted as $\dot{\boldsymbol{P}}^i$ and (pure) policy profiles denoted by $\boldsymbol{P}$ ($\dot{\boldsymbol{P}}$).

2.1 Concise Representations

A concise representation of MAIDs is needed for three reasons. First, real numbers may obscure the true complexity of the problems [5], so we assume that all probability parameters are given by a fraction of two integers, both expressed in finite binary notation. This is realistic since the probabilities are normally either assessed by domain experts or estimated by a learning algorithm and means that all CPDs can be read in poly-time. Second, even with binary variables, a joint distribution across $\boldsymbol{V}$ requires $2^{|\boldsymbol{V}|} - 1$ parameters. A MAID or BN's graphical Markov factorisation reduces this to $\sum_{V \in \boldsymbol{V}} 2^{|\mathbf{Pa}_V|}$, but this can still be exponential in $|\boldsymbol{V}|$. Therefore, it is standard [45, 42, 28, 24] to assume that the maximum in-degree in the graph is much less than $|\boldsymbol{V}|$ (or constant), so that the size of the CPDs are polynomial in $|\boldsymbol{V}|$. This means that the total representation of our MAID (including all CPDs) is polynomial in our chosen complexity parameter $|\boldsymbol{V}|$. Finally, as in BNs, our complexity results are strongly affected by the DAG's **treewidth**. The **treewidth** of a DAG measures its resemblance to a tree and is given by the number of vertices in the largest clique of the corresponding triangulated moral graph minus one [4].

3 Imperfect Recall in MAIDs

Agents may possess different degrees of information about the state of a game. A game has **perfect recall** if each agent remembers all their past decisions and observations, and it has **perfect information** if each agent is aware of *every* agent's past decisions and observations.

Definition 3 ([25])**.** *Agent i in a MAID $\mathcal{M}$ is said to have **perfect recall** if there exists a total ordering $D_1 \prec \cdots \prec D_m$ over $\boldsymbol{D}^i$ such that $(\boldsymbol{Pa}_{D_j} \cup D_j) \subseteq \boldsymbol{Pa}_{D_k}$ for any $1 \leq j < k \leq m$. $\mathcal{M}$ is a perfect recall game if all agents in $\mathcal{M}$ have perfect recall. $\mathcal{M}$ is a **perfect information** game if there exists such an ordering over $\boldsymbol{D}$.*

A MAID with perfect information (recall) can be transformed into an EFG with perfect information (recall), and vice versa [17]. Hence, these information conditions also guarantee the existence of an NE in pure (behavioural) policies in the MAID ([26] gives the equivalent results in EFGs). However, the mechanised representation of a MAID enables weaker criteria to be defined – **sufficient information** and **sufficient recall**. Later, in Proposition 3, we will see that these criteria preserve the NE existence results of perfect information and perfect recall games, respectively.

Definition 4. *Agent i in a MAID $\mathcal{M}$ has **sufficient recall** [36] if the subgraph of the mechanised graph $\mathsf{m}\mathcal{G}$ restricted to just agent i's decision rule nodes $\boldsymbol{\Pi}_{\boldsymbol{D}^i}$ is acyclic. $\mathcal{M}$ is a sufficient recall game if all agents in $\mathcal{M}$ have sufficient recall. $\mathcal{M}$ is a **sufficient information** game if the subgraph of $\mathsf{m}\mathcal{G}$ restricted to contain only and all decision rule nodes $\boldsymbol{\Pi}_{\boldsymbol{D}}$ is acyclic.*[1]

[1]Note that since previous work on influence diagrams has not modelled absent-mindedness (see our Definition 5 in Section 3.1), this definition implicitly assumes each mechanism variable has a single child.

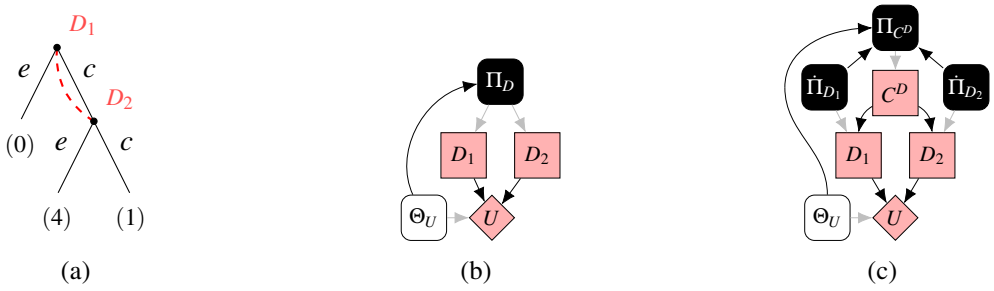


Figure 2: The EFG (a) and the mechanised graphs for an absent-minded driver choosing behavioural (b) or mixed (c) policies.

3.1 Forgetfulness and Absent-Mindedness

Previous work on MAIDs has assumed perfect or sufficient recall. We now begin the contributions of this paper by distinguishing between two types of imperfect recall in MAIDs. **Forgetfulness** applies when an agent forgets an observation or the *outcome* of one of their previous decisions. **Absent-mindedness** applies when an agent cannot even remember whether they have previously made a decision. To make this distinction, we leverage the following insight: *mechanism nodes represent the CPDs governing object-level variables. Every edge between a mechanism and object-level node represents an independent draw from the mechanism's distribution.* We now provide formal definitions.

Definition 5. *Agent i has **imperfect recall** in a MAID $\mathcal{M}$ if for every total ordering $D_1 \prec \cdots \prec D_m$ over $\boldsymbol{D}^i$ there exists some $j < k$ such that $(\boldsymbol{Pa}_{D_j} \cup D_j) \not\subseteq \boldsymbol{Pa}_{D_k}$ (i.e., if agent i does not have perfect recall). Agent i is **forgetful** if such a D_j and D_k have distinct decision rules and is **absent-minded** if in $\mathcal{M}$'s mechanised graph, a decision rule node has more than one outgoing edge to a decision node.*

To motivate our definition of absent-mindedness in MAIDs, we revisit Piccione and Rubinstein's absent-minded driver game [41] (its EFG is in Figure 2a). A driver on a highway may take one of two exits. Taking the first, second, or no exit yields a payoff of 0, 4, or 1, respectively. Adopting Aumann [3]'s *modified multi-selves approach* (i.e., that the driver should only be able to control her current action, not her future actions), the driver does not know which junction she is facing, so she must have the same decision rule at both junctions. We make absent-mindedness explicit with a shared decision rule node Π_D for D_1 and D_2 in the mechanised graph (Figure 2b) (note this is consistent with our mechanised graph definition). Π_D's *two outgoing edges now represent two independent draws from the same distribution*. For D_i and D_j to share a decision rule, it is necessary that $dom(D_i) = dom(D_j)$ and $dom(\boldsymbol{Pa}_{D_i}) = dom(\boldsymbol{Pa}_{D_j})$. Note that perfect recall implies that for any two decisions belonging to the same agent, one's set of parents is a strict superset of the other's, so their decision rules have a different type signature, which rules out absent-mindedness.

In the following examples, used just to explain this paper's concepts, Alice and Bob play variations of matching pennies with the usual payoffs given according to the *final* state of their two coins (where a/b and $\bar{a}/\bar{b}$ represent heads and tails, respectively). Example 2 illustrates a consequence of Bob being forgetful – meaning he cannot remember the *outcome* of his previous decision. In Example 3, Bob is absent-minded – he cannot remember whether he has made a decision at all.

Example 2 (Figures 3a-3c)**.** *Bob is told he must submit a move in advance (B_1) and then confirm it on game day (B_2). If his moves agree, payoffs correspond with normal matching pennies, but if his moves disagree, he must forfeit and always loses (these payoffs are shown in Figure 3c). Bob is forgetful, so on game day he cannot remember his advance choice (i.e., the edge $B_1 \to B_2$ is missing in Figure 3a).*

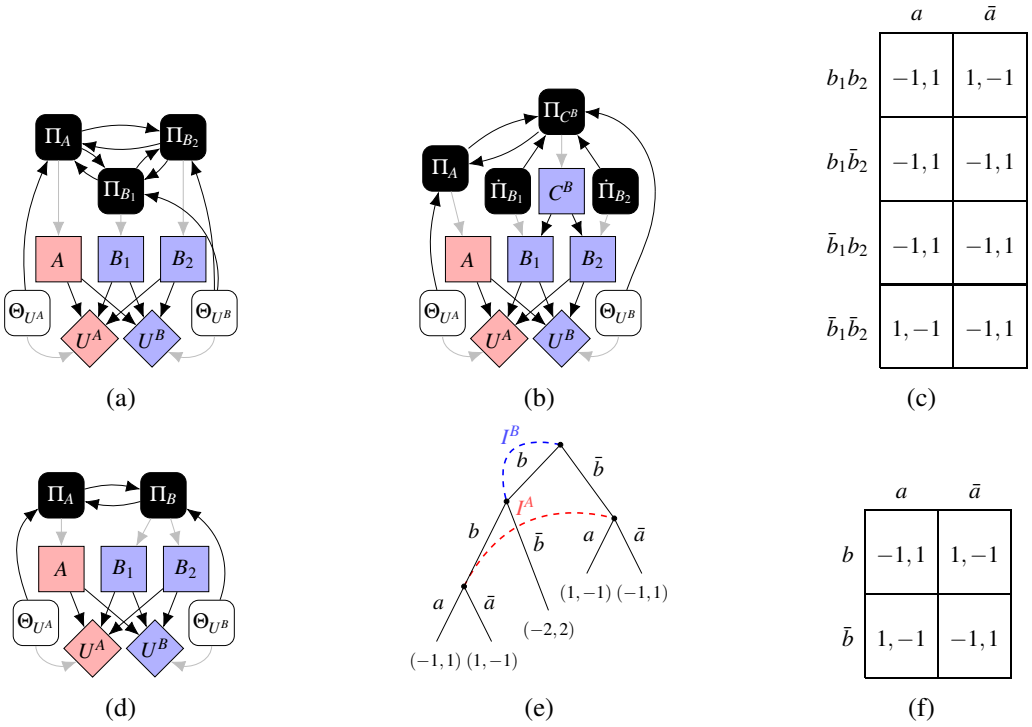


Figure 3: The mechanised graphs for forgetful Bob (Example 2) using (a) behavioural or (b) mixed policies, with normal-form in (c). (d) The mechanised graph for absent-minded Bob (Example 3) using a behavioural policy, with EFG and normal-form representations in (e) and (f).

Example 3 (Figures 3d-3f)**.** *In a new game, the pennies start heads up, and Bob decides whether or not to turn the coin over (B_1). He is absent-minded, so when he sees heads he cannot remember whether he has already made his move, and he decides again (B_2). If he turns the coin having previously chosen to keep heads, Bob gets a -2 penalty and Alice a $+2$ bonus. In all other cases, the payoffs correspond with normal matching pennies (payoffs are shown at the leaves of the EFG in Figure 3e).*

Observe that the MAID's regular graph (just the object-level variables) is identical for both Figures 3a and 3d with the missing $B_1 \to B_2$ edge implying imperfect recall. The difference between forgetfulness and absent-mindedness is only revealed by the mechanised graph. Forgetful Bob has two independent decision rules Π_{B_1} and Π_{B_2} for B_1 and B_2. Absent-minded Bob only has one shared decision rule Π_B.

Examples 2 and 3 demonstrate that both types of imperfect recall can mean an NE in behavioural policies may not exist, even in zero-sum two agent MAIDs with binary decisions. The normal-form games (in Figures 3c and 3f) show that neither contains an NE in pure policies. It is also easy to prove non-existence in behavioural policies (see Appendix B). This arises due to the grand best response function being non-convex valued, which violates a condition of Kakutani's fixed point theorem.

Proposition 1. *Both forgetfulness and absent-mindedness can prevent the existence of an NE in behavioural policies.*

4 Solution Concepts for MAIDs under Imperfect Recall

To overcome the fact that a behavioural policy NE may not exist in imperfect recall MAIDs, one can use mixed or correlated policies. These ensure that the grand best response function always satisfies the

conditions of Kakutani's fixed point theorem, so an equilibrium always exists. We show how the assumptions behind mixed policies, behavioural mixtures, and correlated equilibria (well-studied in EFGs [21, 47], but unexplored in MAIDs) are made graphically explicit in mechanised graphs.

4.1 Mixed Policies and Behavioural Mixtures

Behavioural policies allow agents to randomise independently at every decision node. By contrast, a **mixed policy** $\mu^i \in \Delta(\dot{\boldsymbol{P}}^i)$ is a distribution over pure policies. It allows an agent to coordinate their choice of decision rules at different decisions by randomising once at the game's outset and then committing to the assigned pure policy. More generally, **behavioural mixtures** in $\Delta(\boldsymbol{P}^i)$ are distributions over all behavioural policies. They allow agents to randomise *both* at the outset of the game and before each decision. The outcome of the first randomisation determines the distributions for the others.

A behavioural mixture changes the specification of the game because it can require correlation between different decision rules. At the object-level, a behavioural mixture for agent i requires a new (correlation) decision variable C^i with $\mathbf{Pa}_{C^i} = \varnothing$, $\mathbf{Ch}_{C^i} = \boldsymbol{D}^i$, and $dom(C^i) = \boldsymbol{P}^i$ (the set of all behavioural policies). The decision rules for each D^i become conditional on C^i, so each value of C^i determines a behavioural policy. This explains why C^i and still every $D \in \boldsymbol{D}^i$ are decision nodes – the agent chooses the CPDs for both. Even in the mixed policy case, where each D^i depends deterministically on C^i, the agent chooses the dependence independently from choosing the distribution over C^i. In the mechanised graph (see Figure 2c), C^i gets an associated mechanism variable Π_{C^i} for the distribution C^i is drawing from (its mechanism parents are again determined by s-reachability).

In EFGs, the mechanism by which agents decide on their decision rules is not explicitly shown. Mechanised graphs, however, show clearly when an agent chooses to randomise. Behavioural and mixed policies are the limiting cases of behavioural mixtures: the former where the distribution over $\boldsymbol{P}^i$ is deterministic; the latter where the decision rules $\boldsymbol{\Pi}_{\boldsymbol{D}^i}$ are deterministic. The difference between forgetful Bob in Example 2 using a behavioural or mixed policy is shown in Figures 3a and 3b. For Bob's behavioural policy, C^B and Π_{C^B} are omitted as the decision rules Π_{B_1} and Π_{B_2} are independent. This leaves a normal mechanised graph. Whereas, if Bob uses a mixed policy, he only randomises once from Π_{C^B} at the start of the game to select a pure policy at C^B. This fixes deterministic decision rules at $\dot{\Pi}_{B_1}$ and $\dot{\Pi}_{B_2}$.

Proposition 2. *Given a MAID $\mathscr{M}$ with any partial profile $\boldsymbol{\pi}^{-i}$ for agents $-i$, then if agent i is not absent-minded, for any behavioural policy $\boldsymbol{\pi}^i$ there exists a pure policy $\dot{\boldsymbol{\pi}}^i$ which yields a payoff at least as high against $\boldsymbol{\pi}^{-i}$. On the other hand, if agent i is absent-minded in $\mathscr{M}$ across a pair of decisions with descendants in $\boldsymbol{U}^i$, then there exists a parameterisation of $\mathscr{M}$ and a behavioural policy $\boldsymbol{\pi}^i$ which yields a payoff strictly higher than any payoff achievable by a pure policy.*

Proposition 2 says that a non-absent-minded agent cannot achieve more expected utility by using a behavioural rather than a pure (or mixed) policy, but an absent-minded agent often can. Consider Figure 2c, where $dom(C^D) = \dot{\boldsymbol{P}}^D$, the set of all the driver's pure policies. Π_{C^D} represents the distribution over $dom(C^D)$, so D_1 and D_2 must both be e or both be c. Therefore, $EU^D \leq 1$ under any mixed policy. Whereas, under the behavioural policy $\pi^1_D(e) = \frac{1}{3}$, $EU^D = \frac{4}{3}$. This highlights an important difference between absent-mindedness and forgetfulness. Under perfect recall, every mixed policy has an equivalent behavioural policy, in the sense of inducing the same distribution over outcomes against every opposing policy profile [18]. Under forgetfulness, whilst a mixed policy might not have an equivalent behavioural policy, a behavioural policy always has an equivalent mixed policy [26], so there must exist a pure policy which performs just as well. On the other hand, under absent-mindedness, neither mixed nor behavioural policies are guaranteed to have an equivalent of the other type, so there can be a behavioural policy which outperforms every mixed policy against a given policy profile.

We introduce mixed policies (and behavioural mixtures) to MAIDs to allow more generality in modelling when agents randomise and to guarantee an NE. However, a mixed policy can require exponentially more parameters $\mathcal{O}(2^{2^{|\boldsymbol{V}|}})$ than a behavioural policy $\mathcal{O}(2^{|\boldsymbol{V}|})$ to define. Moreover, single agents are often more naturally modelled as randomising once they meet decision points [26] (this changes for team situations described in Section 6). It is therefore important to know when existence of each type of NE is guaranteed. The sufficient recall result was proved by [18], which we adapt to get the sufficient information result (in Appendix B). The mixed policies result follows directly from Nash's theorem [37].

Proposition 3. *A MAID with sufficient information always has an NE in pure policies, a MAID with sufficient recall always has an NE in behavioural policies, and every MAID has an NE in mixed policies.*

Since both sufficient recall and sufficient information (Definition 4) can be checked in poly-time[2], they expand the class of games that have simple NEs beyond those identifiable using an EFG. For example, we can check in poly-time that the MAID in Figure 1a is an imperfect, but sufficient, information game, and hence know that there must exist an NE in pure policies.

4.2 Correlated Equilibria

We have just shown how mechanised graphs can explicitly represent the assumption behind mixed policies: a *single* agent uses a source of randomness to correlate their decision rules. We now do the same for when *multiple* agents can use the same source of randomness, so the choice of pure policy made by each agent may be correlated. An equilibrium in such a game is called a *correlated equilibrium (CE)* [2], which is a distribution κ over the set of all pure policy profiles, i.e., $\kappa \in \Delta(\dot{\boldsymbol{P}})$. A mediator samples $\dot{\boldsymbol{\pi}}$ according to κ, then recommends to each agent i the pure policy $\dot{\boldsymbol{\pi}}^i$. The distribution κ is a CE if no agent, given their information, has an incentive to unilaterally deviate from their recommended policy $\dot{\boldsymbol{\pi}}^i$.

Definition 6. *In a MAID, $\kappa \in \Delta(\dot{\boldsymbol{P}})$ is a **correlated equilibrium (CE)** if and only if $\forall i$, $\forall \dot{\boldsymbol{\pi}}^i, \dot{\boldsymbol{\omega}}^i \in \dot{\boldsymbol{P}}^i$:*

$$\sum_{\dot{\boldsymbol{\pi}}^{-i} \in \dot{\boldsymbol{P}}^{-i}} \kappa(\dot{\boldsymbol{\pi}}^i, \dot{\boldsymbol{\pi}}^{-i}) EU^i(\dot{\boldsymbol{\pi}}^i, \dot{\boldsymbol{\pi}}^{-i}) \geq \sum_{\dot{\boldsymbol{\pi}}^{-i} \in \dot{\boldsymbol{P}}^{-i}} \kappa(\dot{\boldsymbol{\pi}}^i, \dot{\boldsymbol{\pi}}^{-i}) EU^i(\dot{\boldsymbol{\pi}}^{-i}, \dot{\boldsymbol{\omega}}^i)$$

We illustrate how MAIDs and their mechanised graphs make explicit the assumptions used for a CE using a costless-signal variation of Spence's job market game [46].

Example 4. *Alice is hardworking or lazy (X) with equal probability. She applies for a job with Bob by deciding which costless signal (A) to send. Bob can distinguish between the signals, but does not know Alice's true temperament. He decides whether to offer the job (B) to Alice. The utility functions for Alice and Bob are $U^A = (6-2X)\cdot B$ and $U^B = 6+(10X-6)\cdot B$, respectively.*

The mechanised graph for the original game's MAID is shown in Figure 4c. The cycle between Π_A and Π_B reveals that each agent's decision rule strategically relies on the other agent's decision rule.[3] Therefore, the MAID has insufficient information and no proper subgames, making it difficult to solve.

To find the CE of this game, a trusted mediator is added using a *correlation variable* C with $\mathbf{Pa}_C = \varnothing$, $\mathbf{Ch}_C = \boldsymbol{D}$, and $dom(C) = \dot{\boldsymbol{P}}$. In the mechanised graph, C's associated mechanism variable K_C represents the distribution $\kappa \in \Delta(\dot{\boldsymbol{P}})$ that the mediator draws a pure policy profile according to. This time, since

[2]The mechanised graph is constructed using s-reachability, which uses the poly-time graphical criterion d-separation [43].

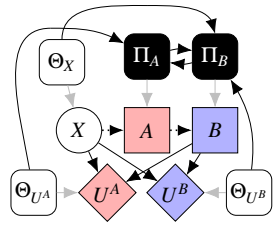
[3]That Bob strategically relies on Alice's decision rule might be less obvious than the fact that Alice strategically relies on Bob's decision rule. The dependency occurs because since Bob can observe A, this unblocks an active path $\Pi_A \to A \leftarrow X \to U^B$ in the independent mechanised graph, so Π_A is s-reachable from Π_B.

	$b_a b_{\bar{a}}$	$b_a \bar{b}_{\bar{a}}$	$\bar{b}_a b_{\bar{a}}$	$\bar{b}_a \bar{b}_{\bar{a}}$
$a_x a_{\bar{x}}$	5,5	5,5	0,6	0,6
$a_x \bar{a}_{\bar{x}}$	5,5	2,8	3,3	0,6
$\bar{a}_x a_{\bar{x}}$	5,5	3,3	2,8	0,6
$\bar{a}_x \bar{a}_{\bar{x}}$	5,5	0,6	5,5	0,6

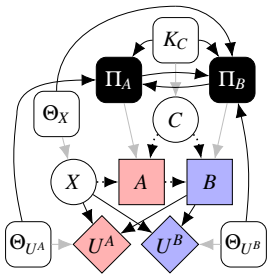
(a)

	$b_a b_{\bar{a}}$	$b_a \bar{b}_{\bar{a}}$	$\bar{b}_a b_{\bar{a}}$	$\bar{b}_a \bar{b}_{\bar{a}}$
$a_x a_{\bar{x}}$	α_1	α_2	α_3	α_4
$a_x \bar{a}_{\bar{x}}$	β_1	β_2	β_3	β_4
$\bar{a}_x a_{\bar{x}}$	γ_1	γ_2	γ_3	γ_4
$\bar{a}_x \bar{a}_{\bar{x}}$	δ_1	δ_2	δ_3	δ_4

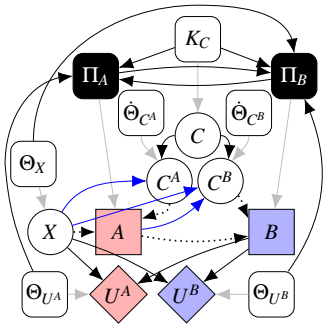
(b)



(c)



(d)



(e)

Figure 4: The sub-figures (a) and (b) give the expected payoff for each agent under each pure policy profile and the parameterisation of the distribution κ, respectively. The mechanised graph for Example 4's original MAID is shown in (c), and the mechanised graphs for when a trusted mediator gives public or private recommendations to find a CE are shown in (d) and (e), respectively. The blue edges are added to the graph in (e) for a MAID-CE's staggered recommendations.

K_C is fixed as κ at the game outset instead of being chosen by any agent, C acts as a chance variable (in contrast to the correlation decision variable introduced for mixed policies and behavioural mixtures).

There is a well-known difference between public and private recommendations. If public, every payoff in the convex hull of the set of NE payoffs can be attained by a CE; however, if the recommendations are private, then the payoffs to each agent in a CE can lie outside this convex hull (e.g., Aumann's game of chicken [2]). This distinction is made explicit in the MAID's graph. If the recommendations are public, then the full outcome of C (the pure policy profile chosen by the mediator) is known by every agent (shown by the dotted edges between C and both A and B in Figure 4d). If the recommendations are private, then each agent only observes their decision rules (action recommendations) in C's outcome, i.e., all recommendations given to other players are hidden (at C^A and C^B in Figure 4e). In this latter case, the agent infers, using Bayes' rule, a posterior over the pure policy profile that was chosen (and also which action was recommended to the other agent(s)). If κ is a CE, then each agent picks for their decision D's decision rule the mediator's recommendation, i.e., $\dot{\pi}_D$ where $c = \dot{\boldsymbol{\pi}}$. The set of variables $\boldsymbol{D}$ remain as decisions because agents are free to deviate from their recommendation and pick any CPDs as decision rules for their decisions.

This mediator's distribution $\kappa \in \Delta(\dot{\boldsymbol{P}})$ can be parameterised according to that in Figure 4b. Note that $b_a \bar{b}_{\bar{a}}$ denotes the pure policy profile where Bob offers the job (b) to Alice if she selects a and Bob does not offer the job ($\bar{b}$) if Alice selects $\bar{a}$. Using the expected payoff for Alice and Bob under each pure

policy profile (Figure 4a), Definition 6's incentive constraints define 24 inequalities that must be satisfied by the CE distribution. After some algebra, we find that $\alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = \gamma_1 = \gamma_2 = \gamma_3 = 0$; $\alpha_4, \beta_4, \gamma_4, \delta_4 \geq 0$; $\alpha_4 - 2\beta_4 + 3\gamma_4 \geq 0$, and $3\beta_4 - 2\gamma_4 + \delta_4 \geq 0$. Any CE, therefore, has Bob never offering a job to Alice because they play the pure policy $\bar{b}_a \bar{b}_{\bar{a}}$ with probability 1, i.e., Bob's decision rule has $\pi^B(B = \bar{b} \mid A = a) = \pi^B(B = \bar{b} \mid A = \bar{a}) = 1$. The remaining constraints require Alice not to give any incentive for Bob to offer her a job by making the conditional probability of Alice being hardworking too high relative to the conditional probability of her being lazy when he receives the signal a or $\bar{a}$. These constraints find that every CE will result in $EU^A = 0$ and $EU^B = 6$. This is unsurprising because, in a signaling game with costless signals, every CE will be a 'pooling equilibrium' [8] (an equilibrium in which Alice chooses the same action regardless of their temperament).

Whilst the CE is among the best-known solution concepts for normal-form games, and is efficiently computable in that setting (e.g., via linear programming [19]), there can be an exponential number of pure policies (so an exponential number of incentive constraints) in EFGs and even in bounded treewidth MAIDs. It is therefore currently unknown if a CE can be found in an EFG or MAID in poly-time. Motivated by these tractability concerns, Von Stengel and Forges proposed an *extensive-form correlated equilibrium (EFCE)* [47]. Along similar lines, we define a *MAID correlated equilibrium*.

Instead of revealing the entire recommendation $\boldsymbol{\pi}^i$ to each agent i immediately, we let the mediator *stagger* their recommendations. This is made visible in the mechanised graph by adding the blue edges in Figure 4e. Importantly, if an agent deviates from any recommendation, then the mediator will *cease giving further recommendations to that agent* (but will still give recommendations to all other agents). Thus, the incentive constraints are now tied to the threat of the mediator withholding future information.

Definition 7. *Given a distribution $\kappa \in \Delta(\dot{\boldsymbol{P}})$, consider the MAID with an additional correlation variable C with $\boldsymbol{Pa}_C = \varnothing$, $\boldsymbol{Ch}_C = \{C_D\}_{D \in \boldsymbol{D}}$, and $\boldsymbol{Ch}_{C_D} = \{D\}$ for each D. Let a pure policy profile $\dot{\boldsymbol{\pi}}$ be selected at C according to κ. Then, when each decision context $\boldsymbol{pa}_D$ is reached, agent i receives a recommended move $d \in dom(D)$ specified by $\dot{\boldsymbol{\pi}}_D \in \dot{\boldsymbol{\pi}}$ (C_D hides all other recommendations $\dot{\boldsymbol{\pi}}_{-D} \in \dot{\boldsymbol{\pi}}$). A **MAID correlated equilibrium (MAID-CE)** is an NE of this game in which no agent has an incentive to deviate from their recommendations.*

The localised recommendations in a MAID-CE pose weaker incentive constraints compared to a CE, so the set of MAID-CE outcomes is larger. As such, MAID-CEs can lead to Pareto-improvements over the CEs (and NEs) in a game. We now give one such MAID-CE. The mediator chooses a signal s with equal probability for type $X = x$, i.e., $\Pr(c_A = a \mid X = x) = \Pr(c_A = \bar{a} \mid X = x) = 0.5$. Bob is recommended to offer Alice a job (b) when Alice's action matches s and to reject otherwise ($\bar{b}$). If $X = \bar{x}$, then the recommendation to Alice is arbitrary and is independent of the signal s, which is only shown to hardworking Alice. Because the mediator only gives Alice her recommendation once her decision context $\mathbf{Pa}_A$ is set, lazy Alice cannot know s. Therefore, in any situation, lazy Alice's action will match s with probability $\frac{1}{2}$. Consequently, when Bob is called to play (i.e., the decision context $\mathbf{Pa}_B$ is set), and Alice's action matches s, Alice is twice as likely to be hardworking than lazy (so $EU^B = \frac{20}{3}$ for offering Alice a job rather than $EU^B = 6$ for rejecting her). If instead, Alice's action does not match s, then he knows with certainty that Alice is lazy, so his best response is to reject. Overall, Alice's expected payoff in this MAID-CE is 3.5, and Bob's is 6.5 (higher than 0 and 6, respectively, for all CEs).

A MAID-CE can be computed in poly-time if the treewidth is bounded, via a reduction to a linear program. We follow Huang et al [20]'s method because the information sets in an EFG are in bijection with the decision contexts in a MAID, but relax beyond their conditions as MAIDs only require sufficient (rather than perfect) recall [20]. Any distribution over pure policies induced by an NE can be represented using a distribution κ, and hence any mixed NE (or equivalent behavioural NE) is also a CE and MAID-CE. As every MAID has an NE in (mixed) policies, every MAID must also have a CE and a MAID-CE.

Proposition 4. *A MAID-CE in bounded treewidth MAIDs with sufficient recall can be found in poly-time.*

5 Complexity Results in MAIDs

We now give some complexity results in MAIDs. Our first follows from the known result in normal-form games [9]. Any normal-form game $\mathscr{N}$ can be reduced to a MAID where each agent has one utility node (which copies the payoffs in $\mathscr{N}$) and one decision node. The domains of the decision variables are the set of each agent's pure strategies in $\mathscr{N}$. Edges are added from every $D \in \boldsymbol{D}$ to every $U \in \boldsymbol{U}$.

Proposition 5. *In a MAID, finding an NE in mixed policies is* PPAD*-hard.*

Problem	Input	Question
IS-BEST-RESPONSE	$\mathscr{M}, i, \boldsymbol{\pi}^{-i}, q \in \mathbb{Q}$	Is there some $\hat{\boldsymbol{\pi}}^i$ such that $EU^i(\hat{\boldsymbol{\pi}}^i, \boldsymbol{\pi}^{-i}) > q$?
IS-NASH	$\mathscr{M}, \boldsymbol{\pi}$	Is $\boldsymbol{\pi}$ a (behavioural) NE of $\mathscr{M}$?
NON-EMPTINESS:	$\mathscr{M}$	Does $\mathscr{M}$ have a (behavioural) NE?

Table 1: Three decision problems in MAIDs with behavioural policies.

In the following results, we focus on the complexity of the decision problems in Table 1.

Proposition 6. IS-BEST-RESPONSE *is* NP^{PP}*-complete,* NP*-complete when restricted to MAIDs with graphs of bounded treewidth, and* PP*-complete if both* $|\boldsymbol{D}^i|$ *and the in-degrees of* $\boldsymbol{D}^i$ *are bounded.*

Proof sketch. IS-BEST-RESPONSE is in NP^{PP} because given $\hat{\boldsymbol{\pi}}^i$, we can verify that $EU^i(\hat{\boldsymbol{\pi}}^i, \boldsymbol{\pi}^{-i}) > q$ in poly-time using a PP oracle for inference in a BN [30]. With bounded treewidth, verification can be done in poly-time. The final setting is in PP by analogy with Kwisthout's PARAMETER TUNING [27]. For the general case's hardness, we can reduce from E-MAJSAT as in [39], where MAP-nodes are replaced by agent i's decision nodes; for bounded treewidth, we can reduce from MAXSAT as in [12]; and for the final case, IS-BEST-RESPONSE with $|\boldsymbol{D}^i| = 0$ is the same as inference in a BN. □

Proposition 6 suggests IS-BEST-RESPONSE is, in general, only tractable if inference is easy *and* $|\boldsymbol{D}^i|$ is bounded by a constant. Proposition 7 then explains the decision problem's name.

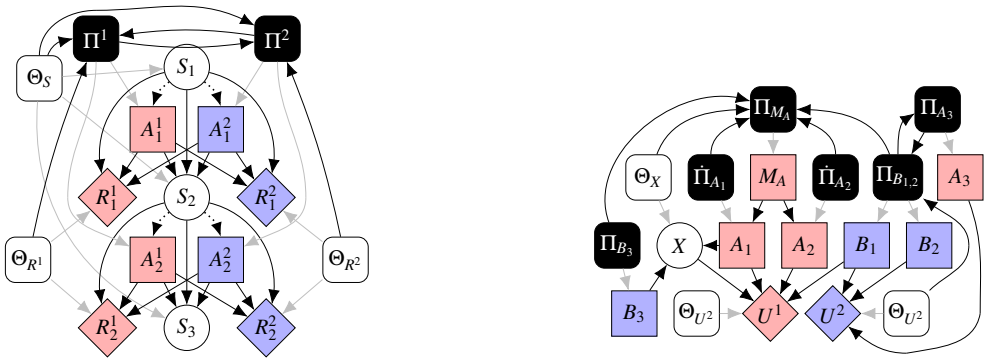
Proposition 7. *If the in-degrees of* $\boldsymbol{D}^i$ *are bounded and* IS-BEST-RESPONSE *can be solved in poly-time, then a best response policy for agent i to a partial profile* $\boldsymbol{\pi}^{-i}$ *can be found in polynomial time.*

Proposition 8. IS-NASH *is* coNP^{PP}*-complete, and* coNP*-complete when restricted to MAIDs with graphs of bounded treewidth. The general problem remains* coNP^{PP}*-hard in sufficient information MAIDs. In MAIDs without chance variables, the problem remains* coNP*-hard.*

Proof sketch. For membership, we can check that $\boldsymbol{\pi}$ is *not* an NE by guessing an agent i and checking if $\boldsymbol{\pi}^{\boldsymbol{i}} \in \boldsymbol{\pi}$ is a best response in poly-time using a PP-oracle (this is unnecessary if the graph has bounded treewidth). Hardness comes from the single-agent setting where it is the complement of IS-BEST-RESPONSE. In MAIDs without chance variables, we reduce from partial order games [50]. □

Proposition 3 shows when NON-EMPTINESS is vacuous. However, in an insufficient recall MAID, NON-EMPTINESS is, in general, intractable even without chance variables.

Proposition 9. NON-EMPTINESS *is* NEXPTIME*-hard and becomes* NEXPTIME*-complete if we restrict to MAIDs without chance variables.*



(a) (b)

Figure 5: Mechanised graphs for a CE with (a) public and (b) private recommendations, where the blue edges are added for a MAID-CE; (c) a Markov game;(d) a team setting with imperfect communication.

Proof sketch. For hardness, we can reduce from partial order games. Without chance variables, we can determine NON-EMPTINESS using a similar algorithm to that in [50]. It exploits the setting's determinism: payoffs are poly-time computable and the number of policy profiles is reduced to $\mathcal{O}(2^{|\boldsymbol{V}|})$. □

Proposition 10. *In a MAID with sufficient information, if the in-degrees of* $\boldsymbol{D}$ *are bounded and* IS-BEST-RESPONSE *can be solved in poly-time, then a pure NE can be found in poly-time.*

This result suggests an NE can be found efficiently in certain MAIDs, but even in games without sufficient information, NEs can be found more efficiently in a MAID than in an EFG. The mechanised graph dependencies reveal more 'subgames' – parts of the MAID that can be solved independently from the rest – to which dynamic programming can be applied [25, 17]. As finding an NE in both EFGs and MAIDs depends significantly on the game's size, this can empirically lead to large compute savings [25].

6 Applications and Conclusion

We introduced forgetfulness and absent-mindedness as properties of individual agents (due to imperfect memory). However, imperfect recall also commonly arises in *team situations*; each team consists of several agents targeting a common goal with imperfect communication. Forgetfulness or absent-mindedness occurs when an agent does not know their teammates' actions (or observations) or whether they have acted at all. Mechanised graphs represent these situations where teams often employ a mix of randomisation strategies (e.g., Figure 5b). For mixed policies, the random seed is chosen at the start, before the agents set out following their distinct policies. For behavioural policies, agents pick a new random seed at every decision point. Behavioural mixtures correspond to randomising at both stages.

Another application of imperfect recall in MAIDs is to *Markov (or 'stochastic') games* [44], in which the agents move between different states over time (e.g., Figure 5a). At each time step t, each agent i selects an action A_t^i, and the game probabilistically transitions to a new state S_{t+1}, depending on the previous state S_t and the actions selected, and each agent receives a payoff R_t^i. Each S_{t+1} and R_t^i has parents $\{S_t, A_t^1, \ldots, A_t^n\}$ and must be identically distributed for all t, again represented using shared mechanism variables. Often, the agent must learn a memoryless, stationary policy $\pi^i : S \to \Delta(A^i)$, where S is the set of states and $\Delta(A^i)$ the set of probability distributions over agent i's actions. Hence, the agents are absent-minded (every decision A_{t+1}^i of agent i shares the same decision rule) and use *behavioural* policies (since the action selected in each state is independently stochastic). In light of Proposition 1,

it is therefore natural to ask whether a Markov game may not have an NE in memoryless stationary policies. It is known that infinite-horizon Markov games might not (for a counterexample see [11]). Although infinite games lie outside of the scope of this paper, it is nonetheless insightful to note that this possible non-existence is due to absent-mindedness: if agents can choose a different decision rule at each time step, a behavioural NE is guaranteed [32].

We have shown how to handle imperfect recall in MAIDs by overcoming the potential lack of NEs in behavioural policies using mixed and correlated equilibria. EFGs leave many assumptions about how agents play games hidden, but mechanised graphs make explicit the assumptions behind imperfect recall (both forgetfulness and absent-mindedness), mixed policies, and two types of correlated equilibria. Our complexity results highlight the importance of restricting the use of MAIDs to those with a limited number of decision variables and bounded treewidth. Finally, our applications to Markov games and team situations show that imperfect recall broadens the scope of what can be modelled using MAIDs.

Acknowledgements The authors wish to thank Ryan Carey, Tom Everitt, and Francis Rhys Ward for invaluable feedback, as well as three anonymous reviewers for their helpful comments. Fox was supported by the EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines and Systems (Reference: EP/S024050/1), MacDermott was supported by the UKRI Centre for Doctoral Training in Safe and Trusted Artificial Intelligence (Reference: EP/S023356/1), Hammond was supported by an EPSRC Doctoral Training Partnership studentship (Reference: 2218880), and Wooldridge was supported by a UKRI Turing AI World Leading Researcher Fellowship (Reference: EP/W002949/1).

References

[1] Carolyn Ashurst, Ryan Carey, Silvia Chiappa & Tom Everitt (2022): *Why fair labels can yield unfair predictions: Graphical conditions for introduced unfairness*. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 36, pp. 9494–9503, doi:10.1609/aaai.v36i9.21182.

[2] Robert J Aumann (1974): *Subjectivity and correlation in randomized strategies*. *Journal of mathematical Economics* 1(1), pp. 67–96, doi:10.1016/0304-4068(74)90037-8.

[3] Robert J Aumann, Sergiu Hart & Motty Perry (1997): *The absent-minded driver*. *Games and Economic Behavior* 20(1), pp. 102–116, doi:10.1006/game.1997.0577.

[4] Hans L Bodlaender (1993): *A linear time algorithm for finding tree-decompositions of small treewidth*. In: *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pp. 226–234, doi:10.1145/167088.167161.

[5] Hans L Bodlaender, Frank van den Eijkhof & Linda C van der Gaag (2002): *On the complexity of the MPA problem in probabilistic networks*. In: *ECAI*, pp. 675–679.

[6] Cassio P de Campos & Qiang Ji (2008): *Strategy selection in influence diagrams using imprecise probabilities*. In: *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, pp. 121–128.

[7] Micah Carroll, Alan Chan, Henry Ashton & David Krueger (2023): *Characterizing Manipulation from AI Systems*. *arXiv preprint arXiv:2303.09387*.

[8] In-Koo Cho & David M Kreps (1987): *Signaling games and stable equilibria*. *The Quarterly Journal of Economics* 102(2), pp. 179–221, doi:10.2307/1885060.

[9] Constantinos Daskalakis, Paul W Goldberg & Christos H Papadimitriou (2009): *The complexity of computing a Nash equilibrium*. *SIAM Journal on Computing* 39(1), pp. 195–259, doi:10.1145/1132516.1132527.

[10] A. P. Dawid (2002): *Influence Diagrams for Causal Modelling and Inference*. *International Statistical Review* 70(2), pp. 161–189, doi:10.1111/j.1751-5823.2002.tb00354.x.

[11] Luca De Alfaro & Rupak Majumdar (2001): *Quantitative Solution of Omega-Regular Games*. In: *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pp. 675–683, doi:10.1016/j.jcss.2003.07.009.

[12] Cassio Polpo De Campos & Fabio Gagliardi Cozman (2005): *The inferential complexity of Bayesian and credal networks*. In: *IJCAI*, 5, Citeseer, pp. 1313–1318.

[13] Apiruk Detwarasiti & Ross D Shachter (2005): *Influence diagrams for team decision analysis*. *Decision Analysis* 2(4), pp. 207–228, doi:10.1287/deca.1050.0047.

[14] Tom Everitt, Ryan Carey, Eric D Langlois, Pedro A Ortega & Shane Legg (2021): *Agent incentives: A causal perspective*. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 35, pp. 11487–11495, doi:10.1609/aaai.v35i13.17368.

[15] Tom Everitt, Marcus Hutter, Ramana Kumar & Victoria Krakovna (2021): *Reward tampering problems and solutions in reinforcement learning: A causal influence diagram perspective*. *Synthese* 198(Suppl 27), pp. 6435–6467, doi:10.1007/s11229-021-03141-4.

[16] Sebastian Farquhar, Ryan Carey & Tom Everitt (2022): *Path-specific objectives for safer agent incentives*. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 36, pp. 9529–9538, doi:10.1609/aaai.v36i9.21186.

[17] Lewis Hammond, James Fox, Tom Everitt, Alessandro Abate & Michael Wooldridge (2021): *Equilibrium Refinements for Multi-agent Influence Diagrams: Theory and Practice*. In: *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems*, pp. 574–582.

[18] Lewis Hammond, James Fox, Tom Everitt, Ryan Carey, Alessandro Abate & Michael Wooldridge (2023): *Reasoning about causality in games*. *Artificial Intelligence* 320, p. 103919, doi:10.1016/j.artint.2023.103919.

[19] Sergiu Hart & David Schmeidler (1989): *Existence of correlated equilibria*. *Mathematics of Operations Research* 14(1), pp. 18–25, doi:10.1287/moor.14.1.18.

[20] Wan Huang & Bernhard von Stengel (2008): *Computing an extensive-form correlated equilibrium in polynomial time*. In: *International Workshop on Internet and Network Economics*, Springer, pp. 506–513, doi:10.1007/978-3-540-92185-1_56.

[21] Mamoru Kaneko & J Jude Kline (1995): *Behavior strategies, mixed strategies and perfect recall*. *International Journal of Game Theory* 24(2), pp. 127–145, doi:10.1007/bf01240038.

[22] Zachary Kenton, Ramana Kumar, Sebastian Farquhar, Jonathan Richens, Matt MacDermott & Tom Everitt (2022): *Discovering Agents*. *arXiv preprint arXiv:2208.08345*.

[23] Uffe B Kjaerulff & Anders L Madsen (2008): *Bayesian networks and influence diagrams*. *Springer Science+ Business Media* 200, p. 114.

[24] Daphne Koller & Nir Friedman (2009): *Probabilistic graphical models: principles and techniques*. MIT press.

[25] Daphne Koller & Brian Milch (2003): *Multi-agent influence diagrams for representing and solving games*. Games and economic behavior 45(1), pp. 181–221, doi:10.1016/s0899-8256(02)00544-4.

[26] Harold W. Kuhn (1953): *Extensive Games and the Problem of Information*. In: Contributions to the Theory of Games (AM-28), 2, Princeton University Press, pp. 193–216, doi:10.1515/9781400881970-012.

[27] Johan Kwisthout & Linda C van der Gaag (2008): *The computational complexity of sensitivity analysis and parameter tuning*. In: Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence, pp. 349–356.

[28] Johan Henri Petrus Kwisthout et al. (2009): *The computational complexity of probabilistic networks*. Utrecht University.

[29] Steffen L Lauritzen & Dennis Nilsson (2001): *Representing and solving decision problems with limited information*. Management Science 47(9), pp. 1235–1251, doi:10.1287/mnsc.47.9.1235.9779.

[30] Michael L Littman, Stephen M Majercik & Toniann Pitassi (2001): *Stochastic boolean satisfiability*. Journal of Automated Reasoning 27(3), pp. 251–296.

[31] Michael Maschler, Shmuel Zamir & Eilon Solan (2020): *Game theory*. Cambridge University Press.

[32] Eric Maskin & Jean Tirole (2001): *Markov perfect equilibrium: I. Observable actions*. Journal of Economic Theory 100(2), pp. 191–219, doi:10.1006/jeth.2000.2785.

[33] Denis Deratani Mauá, Cassio P de Campos & Marco Zaffalon (2012): *Solving limited memory influence diagrams*. Journal of Artificial Intelligence Research 44, pp. 97–140, doi:10.1613/jair.3625.

[34] Denis Deratani Mauá & Fabio Gagliardi Cozman (2016): *Fast local search methods for solving limited memory influence diagrams*. International Journal of Approximate Reasoning 68, pp. 230–245, doi:10.1016/j.ijar.2015.05.003.

[35] Chris van Merwijk, Ryan Carey & Tom Everitt (2022): *A Complete Criterion for Value of Information in Soluble Influence Diagrams*. Proceedings of the AAAI Conference on Artificial Intelligence 36(9), pp. 10034–10041, doi:10.1609/aaai.v36i9.21242.

[36] Brian Milch & Daphne Koller (2008): *Ignorable Information in Multi-agent Scenarios*. Technical Report MIT-CSAIL-TR-2008-029, Computer Science and Artificial Intelligence Laboratory, MIT.

[37] J. F. Nash (1950): *Equilibrium Points in N-person Games*. Proceedings of the National Academy of Sciences 36(1), pp. 48–49.

[38] Christos Papadimitriou (1994): *Computational Complexity*. Addison Wesley.

[39] James D Park & Adnan Darwiche (2004): *Complexity results and approximation strategies for MAP explanations*. Journal of Artificial Intelligence Research 21, pp. 101–133, doi:10.1613/jair.1236.

[40] Avi Pfeffer & Ya'akov Gal (2007): *On the reasoning patterns of agents in games*. In: AAAI, pp. 102–109.

[41] Michele Piccione & Ariel Rubinstein (1997): *On the interpretation of decision problems with imperfect recall*. Games and Economic Behavior 20(1), pp. 3–24, doi:10.1016/0165-4896(96)81573-3.

[42] Dan Roth (1996): *On the hardness of approximate reasoning*. Artificial Intelligence 82(1-2), pp. 273–302, doi:10.1016/0004-3702(94)00092-1.

[43] Ross D Shachter (1998): *Bayes-ball: Rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams)*. In: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, pp. 480–487.

[44] Lloyd S Shapley (1953): *Stochastic games*. Proceedings of the national academy of sciences 39(10), pp. 1095–1100.

[45] Solomon Eyal Shimony (1994): *Finding MAPs for belief networks is NP-hard*. Artificial intelligence 68(2), pp. 399–410, doi:10.1016/0004-3702(94)90072-8.

[46] Michael Spence (1978): *Job market signaling*. In: Uncertainty in economics, Elsevier, pp. 281–306.

[47] Bernhard Von Stengel & Françoise Forges (2008): *Extensive-form correlated equilibrium: Definition and computational complexity*. Mathematics of Operations Research 33(4), pp. 1002–1022, doi:10.1287/moor.1080.0340.

[48] Francis Rhys Ward, Francesca Toni & Francesco Belardinelli (2022): *On Agent Incentives to Manipulate Human Feedback in Multi-Agent Reward Learning Scenarios.* In: *AAMAS*, pp. 1759–1761.

[49] Kevin Waugh, Martin Zinkevich, Michael Johanson, Morgan Kan, David Schnizlein & Michael H Bowling (2009): *A Practical Use of Imperfect Recall.* In: *SARA*.

[50] Valeria Zahoransky, Julian Gutierrez, Paul Harrenstein & Michael Wooldridge (2021): *Partial order games*. Games 13(1), p. 2, doi:10.3390/g13010002.

A Strategic Relevance and Subgames

Koller and Milch define **strategic relevance** to infer whether the choice of a decision rule can affect the optimality of another decision rule [25]. Hammond et al. extend strategic relevance to also consider whether the parameterisation of non-decision nodes can affect the decision rule's optimality [18]. Intuitively, a mechanism M_V is strategically relevant to the decision rule Π_D of $D \in \boldsymbol{D}^i$ if the choice of CPD at M_V can affect agent i's utility nodes that are downstream of D (i.e., those in $\boldsymbol{U}^i \cap \mathbf{Desc}_D$). Formally:

Definition 8 ([25, 18])**.** *Recall that $dom(\Pi_D)$ gives the set of possible decision rules at Π_D for decision node D. Given a MAID with $D \in \boldsymbol{D}^i$ and $V \neq D \in \boldsymbol{D}$, the mechanism M_V for V is **strategically relevant** to Π_D if there exist two joint distributions over $\boldsymbol{V}$ parameterised by mechanisms m and m' respectively such that:*

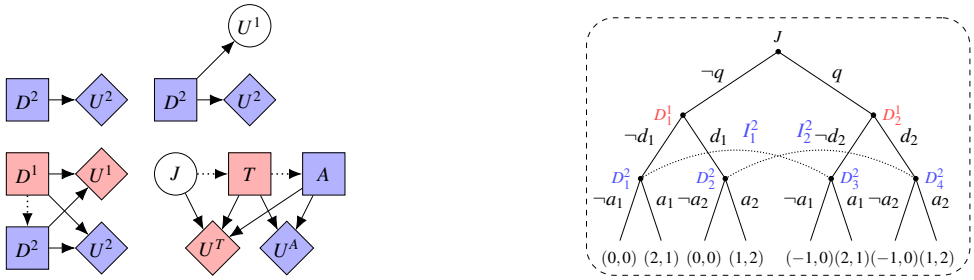
- $\pi_D \in \arg\max_{\varpi_D \in dom(\Pi_D)} EU^i((\varpi_D, \boldsymbol{\pi}_{-D}) \mid \mathsf{m})$
- m *differs from* m' *only at* M_V*,*
- $\pi_D \notin \arg\max_{\varpi_D \in dom(\Pi_D)} EU^i((\varpi_D, \boldsymbol{\pi}_{-D}) \mid \mathsf{m}')$*, and neither does any decision rule ϖ_D that agrees with π_D on all $\boldsymbol{pa}_D$ such that* $\Pr(\boldsymbol{pa}_D \mid \mathsf{m}') > 0$.

The first two conditions say: if the decision rule π_D is optimal for the MAID parameterisation (i.e., the setting of all mechanism variables) m, and Π_D does not strategically rely on M_V, then π_D must also be optimal for any other parameterisation m' that differs from m only at M_V. The third condition deals with sub-optimal decision rules in response to zero-probability decision contexts (i.e., non-credible threats).

Koller and Milch [25] also derive a graphical criterion for strategic relevance, called *s-reachability*, which is sound (if M_V is strategically-relevant to Π_D, then M_V is *s*-reachable from Π_D) and complete (if M_V is *s*-reachable from Π_D, then there is some parameterisation m of the MAID and some policy profile π such that M_V is strategically-relevant to Π_D). This uses the *independent mechanised graph* $\mathsf{m}_{\perp}\mathscr{G}$, which contains a separate mechanism parent for each variable in the original MAID graph, but no edges between the mechanism variables.

Definition 9 ([25])**.** M_V *is s-reachable from* Π_D *if* $\mathsf{M}_V \not\perp_{\mathsf{m}_{\perp}\mathscr{G}} \boldsymbol{U}^i \cap \boldsymbol{Desc}_D \mid D, \boldsymbol{Pa}_D$.

s-reachability determines which inter-mechanism edges are present in the MAID's mechanised graph; $\mathsf{M}_V \to \Pi_D$ exists in the mechanised graph if and only if Π_D strategically relies on M_V.



(a) (b)

Figure 6: (a) shows the four subdiagrams (three of which are 'proper') of the MAID in Figure 1a and (b) shows the corresponding EFG in which none of the MAID's proper subgames can be recognised.

We now briefly introduce subgames (see [18]) for more details) because they simplify the presentation of some of our proofs in Appendix B. Subgames in EFGs represent parts of the game that can be solved independently from the rest. In MAIDs, they fulfil the same purpose: they identify parts of the game that can be solved independently (and allow a subgame-perfect equilibrium refinement to be defined). Subgames in MAIDs are found by exploiting *s*-reachability to find the graphs underlying the subgames, called sub-diagrams. To then find the subgames for each subdiagram, the parameterisation of the remaining variables is updated to be consistent with the original game and graph structure.

Importantly, because MAIDs explicitly represent conditional independencies between variables, we can often find more subgames in a MAID than in a corresponding EFG. This is the case for Example 1's MAID (shown in Figure 1a) with the four subdiagrams (three proper) in Figure 6a. Each subdiagram has a set of associated subgames, one for each instantiation of the variables outside of the subdiagram. None of the proper MAID subgames can be recognised as subgames in the corresponding EFG (in Figure 6b).

Definition 10. *Given a MAID* $\mathscr{M} = (\mathscr{G}, \boldsymbol{\theta})$*, with* $\mathscr{G} = (N, \boldsymbol{V}, E)$*, the subgraph* $(\boldsymbol{V}', E')$ *of* $\mathscr{G}$*, along with the set of agents* $N' \subseteq N$ *possessing decision variables in that subgraph, is known as a* ***subdiagram*** $\mathscr{G}' = (N', \boldsymbol{V}', E')$ *if:*

- $\boldsymbol{V}'$ *contains every variable* Z *such that* M_Z *is s-reachable from some* Π_D *with* $D \in \boldsymbol{V}'$*,*
- $\boldsymbol{V}'$ *contains, for all* $X, Y \in \boldsymbol{V}'$*, every variable that lies on a directed path* $X \dashrightarrow Y$ *in* $\mathscr{G}$*.*

*A **subgame** of $\mathscr{M}$ is a new MAID $\mathscr{M}' = (\mathscr{G}', \boldsymbol{\theta}')$ where $\mathscr{G}'$ is a subdiagram of $\mathscr{G}$ and $\boldsymbol{\theta}'$ is defined by $\Pr'(\boldsymbol{v}'; \boldsymbol{\theta}') := \Pr(\boldsymbol{v}' \mid \boldsymbol{z}; \boldsymbol{\theta})$, where $\boldsymbol{z}$ is some instantiation of the variables $\boldsymbol{Z} = \boldsymbol{V} \setminus \boldsymbol{V}'$. A subgame is **feasible** if there exists a policy profile $\boldsymbol{\pi}$ where $\Pr^{\boldsymbol{\pi}}(\boldsymbol{z}) > 0$.*

The first condition on $\boldsymbol{V}'$ ensures that for any decision variable D in the subdiagram, any variable whose mechanism may impact the optimal decision rule for D is also included in the graph. The second condition says that additional variables may also be included in the subdiagram as long as mediators are included too. This ensures that the CPDs for all the variables in the subgame remain consistent.

B Proofs

Proposition 1. *Both forgetfulness and absent-mindedness can prevent the existence of an NE in behavioural policies.*

Proof. Example 2 (Figures 3a-3c) and Example 3 (Figures 3d-3f) are counterexamples for each case.

Proof for Example 2 (forgetfulness): The normal-form game showing the payoffs for each agent is shown in Figure 3c. First, observe that there are no NE in pure policies. Now, suppose that there does exist an NE in behavioural policies. If Alice always plays a or always $\bar{a}$ – i.e., $\pi^A(a) = 1$ or $\pi^A(a) = 0$ – then Bob's best response is always $\bar{b}_1\bar{b}_2$ or always b_1b_2, respectively. However, this does not form an NE. So, Alice must select a stochastic decision rule π_A and be indifferent (by the principle of indifference) between a and $\bar{a}$.

Letting Π_{B_1} and Π_{B_2} be parameterised by $p, q \in [0,1]$ where $\pi_{B_1}(b_1) = p$ and $\pi_{B_2}(b_2) = q$, we obtain two constraints on p and q. On the one hand, by virtue of Alice's indifference, Bob's behavioural policy $\boldsymbol{\pi}^B$ must result in $\pi^B(\neg b_1, \neg b_2) = \pi^B(b_1, b_2)$, and so: $(1-p)(1-q) = pq \implies p+q = 1$. On the other hand, Bob receives utility -1 if his policy $\boldsymbol{\pi}^B$ results in any outcome with $B_1 = \neg b_1$ and $B_2 = b_2$, or $B_1 = b_1$ and $B_2 = \neg b_2$, whatever the choice of $\boldsymbol{\pi}^A$. Therefore, we must have that $\pi^B(\neg b_1, b_2) + \pi^B(b_1, \neg b_2) < \pi^B(b_1, b_2) + \pi^2(\neg b_1, \neg b_2)$ and thus, by substituting in the result that $p+q = 1$: $(1-p)q + p(1-q) < pq + (1-p)(1-q) \implies (2p-1)^2 < 0$.. This contradiction implies that the MAID for Example 2 has no NE in behavioural policies.

To further understand this example, let us again write Bob's policy as a tuple (p, q), and suppose $\pi_A(a) = 0.5$. Then, either pure policy $(1,1)$ and $(0,0)$ is a best response for Bob with $EU^B = 0$. But, consider the convex combination of these best responses $0.5 \cdot (1,1) + 0.5 \cdot (0,0) = (0.5, 0.5)$. Under this policy, each of the eight outcomes in the payoff matrix is equally likely and so Bob's expected payoff drops to $(-1-1-1+1+1-1-1-1)/8 = -0.5$. Since a convex combination of best responses is no longer a best response, Bob's best response function is not convex-valued, and so nor is the grand best response function. The conditions of Kakutani's fixed point theorem are not satisfied, which explains why a Nash equilibrium need not exist.

Proof for Example 3 (absent-mindedness): First, observe from the normal-form game in Figure 3f that there is no NE in pure policies in this game. Next, suppose there exists a NE in behavioural policies and let Π_B be parameterised by $p \in [0,1]$, where $\pi_B(b) = p$ for $p \in [0,1]$. Alice's payoff only depends on her policy π^A when Bob plays bb or $\bar{b}\bar{b}$, for which Alice has pure best responses. This implies that, at an NE, $p^2 = (1-p)^2 \implies p = 0.5$. Therefore, Alice's policy is irrelevant and $EU^B = -1$ ($EU^B = 0$) if he does (doesn't) forfeit, which happens with probability 0.5. Therefore, Bob's policy is dominated by his pure policies, with worst-case payoff $EU^B = -1$. This contradicts the assumption of an NE in behavioural policies.

Explanation: If $\pi_A(a) = 0.5$, then $p = 0$ and $p' = 1$ are both best responses for Bob with $EU^B = 0$. However, the convex combination $0.5p + 0.5p'$ gives expected payoff to Bob $EU^B = 0.25 \cdot 1 + 0.25 \cdot (-1) + 0.5 \cdot (-10) = -5$ and is therefore not a best response. Again this is due to the fact that under behavioural policies, in situations of imperfect recall, a convex combination of pure policies can introduce outcomes that could not occur under either pure policy. Under a mixed combination of pure policies, Alice will always follow one or the other, and so no new outcomes are introduced. However, under a behavioural combination, two independent absent-minded draws from the same distribution over actions can come out differently, introducing new potential outcomes—in this case forfeit. □

Proposition 2. *Given a MAID $\mathscr{M}$ with any partial profile $\boldsymbol{\pi}^{-i}$ for agents $-i$, then if agent i is not absent-minded, for any behavioural policy $\boldsymbol{\pi}^i$ there exists a pure policy $\dot{\boldsymbol{\pi}}^i$ which yields a payoff at least as high against $\boldsymbol{\pi}^{-i}$. On the other hand, if agent i is absent-minded in $\mathscr{M}$ across a pair of decisions with descendants in $\boldsymbol{U}^i$, then there exists a parameterisation of $\mathscr{M}$ and a behavioural policy $\boldsymbol{\pi}^i$ which yields a payoff strictly higher than any payoff achievable by a pure policy.*

Proof. Let $\boldsymbol{\pi}^i$ be a behavioural policy and begin with any decision node $D \in \boldsymbol{D}^i$ with decision rule $\pi_D \in \boldsymbol{\pi}^i$. Now $\pi^i_D(d \mid \mathbf{pa}_D)$ is the probability of choosing $d \in dom(D)$ at D when $\mathbf{Pa}_D = \mathbf{pa}_D$ according to $\boldsymbol{\pi}^i$. Since agent i is not absent-minded, the expected payoff for agent i can be written $EU^i(\boldsymbol{\pi}^i, \boldsymbol{\pi}^{-i}) = \sum_{d \in dom(D)} \pi^i(d \mid \mathbf{pa}_D)\lambda_d + \nu$, where each coefficent λ_d and ν are independent of $\pi^i_D(d \mid \mathbf{pa}_D)$. Consider the action $\hat{d} \in dom(D)$ which achieves the highest λ_d (i.e., contributes most the expected utility) Setting $\pi^i_D(\hat{d} \mid \mathbf{pa}_D) = 1$ therefore yields a payoff at least as high. The first claim therefore follows by repeating this argument for every $D \in \boldsymbol{D}^i$.

For the converse claim, agent i is absent-minded, which means that at least two of agent i's decision nodes must draw from an identical distribution. Without loss of generality, call these D_l and D_m. Recall that for this to be the case, $dom(D_l) = dom(D_m)$ and $dom(\boldsymbol{Pa}_{D_l}) = dom(\boldsymbol{Pa}_{D_m})$. Now consider an outcome of the game $\hat{\boldsymbol{v}} \in dom(\boldsymbol{V})$ where $\mathbf{pa}_{D_l} = \mathbf{pa}_{D_m}$, but $d_l \neq d_m$. Since D_l and D_m have descendants in $\boldsymbol{U}^i$, Parameterise the MAID $\mathscr{M}$ such that $EU^i = 1$ if and only if $\boldsymbol{V} = \hat{\boldsymbol{v}}$. For all other game outcomes $\boldsymbol{v} \neq \hat{\boldsymbol{v}}$, let $EU^i = 0$. The claim follows since the outcome $\hat{\boldsymbol{v}}$ cannot be instantiated by any pure policy for agent i, but can be instantiated by any behavioural policy for agent i that has a (shared) decision rule for D_l and D_m that assigns a positive probability to both actions d_l and d_m. □

Proposition 3. *A MAID with sufficient information always has an NE in pure policies, a MAID with sufficient recall always has an NE in behavioural policies, and every MAID has an NE in mixed policies.*

Proof. The mixed policies case follows from Nash's theorem since all the finite number of random variables in a MAID have finite domains [37]. Hammond et al. proved the case with sufficient recall [18].

We now consider the sufficient information case where we show that a NE in pure policies must exist. Begin with an arbitrary policy profile across all decision nodes in the original MAID, $\mathscr{M}$. Decision rules associated with each $D \in \boldsymbol{D}$ can be optimised by iterating backwards through a subdiagram ordering $\mathscr{G}_1 \prec \cdots \prec \mathscr{G}_m$ of $\mathscr{M}$'s subdiagrams such that $\mathscr{G}_j \prec \mathscr{G}_k$ implies that $\mathscr{G}_j$ is *not* a subdiagram of $\mathscr{G}_k$. When $\mathscr{M}$ is a sufficient information game, this means that $\mathscr{G}_m$ contains just one decision node for some agent $i \in N$, and, for each subdiagram $\mathscr{G}_j$ where $1 \leq j < m$, $\mathscr{G}_{j-1}$ contains *at most* one additional decision variable. Several subdiagrams can have the same set of decisions, $\boldsymbol{D}_k$, so we choose a single subdiagram $\mathscr{G}_k$ (one with the fewest nodes $\boldsymbol{V}'$) for each $\boldsymbol{D}_k$ and discard the others. Each subdiagram in this ordering has an associated subgame for each setting of the nodes which have a child in $\boldsymbol{V}'$.

When considering each subgame $\mathscr{M}_{m-j}$ for $\mathscr{G}_{m-j}$, the decision rules for all decision nodes in proper subgames of $\mathscr{M}_{m-j}$ will have already been optimised and fixed in previous iterations, so these are now

chance nodes in $\mathscr{M}_{m-j}$. In addition, the decision node D_{m-j} in $\mathscr{M}_{m-j}$ does not strategically rely on any of the decision nodes outside of $\mathscr{M}_{m-j}$. Therefore, this step is localised to computing only the optimal decision rule for D_{m-j}. Since this is a single-agent single-decision optimisation, we know that there must exist a pure decision rule best response. In the case of a tie, pick one arbitrarily. After repeating this optimisation process for all subgames in the MAID, we know that every decision node must have a pure decision rule, so we have found a NE in pure policies, as required. □

Proposition 4. *A MAID-CE in bounded treewidth MAIDs with sufficient recall can be found in poly-time.*

Proof sketch. We follow Huang and von Stengel's method for this result [20]. Our result comes from the observation that if there is sufficient recall in a MAID, then: (i) the set of decision contexts of every decision node in the MAID is in bijection with the set of all information sets in a corresponding EFG; and (ii) sufficient recall is sufficient for the ordering of decision contexts analogous to Huang and von Stengel's ordering of information sets. □

Lemma 1. *If* IS-BEST-RESPONSE *can be solved in poly-time, then agent i's expected utility under a best response to a partial policy profile $\boldsymbol{\pi}^{-i}$ in a MAID can be found in poly-time.*

Proof. This follows immediately from using binary search over agent i's policies and uses the fact that we are restricting parameters in the MAID to be rational numbers. □

Proposition 7. *If the in-degrees of $\boldsymbol{D}^i$ are bounded and* IS-BEST-RESPONSE *can be solved in poly-time, then a best response policy for agent i to a partial policy profile $\boldsymbol{\pi}^{-i}$ can be found in poly-time.*

Proof. Begin by constructing the MAID $\mathscr{M}(\boldsymbol{\pi}^{-i})$ by replacing decision nodes $\boldsymbol{D} \setminus \boldsymbol{D}^{-i}$ as chance nodes with CPDs given by $\boldsymbol{\pi}^{-i}$. Next, use Lemma 1 to compute agent i's expected utility under a best response policy in $\mathscr{M}(\boldsymbol{\pi}^{-i})$ and use this value as q. Take each of agent i's decision variables $D \in \boldsymbol{D}^i$ and build a new MAID $\mathscr{M}(\boldsymbol{\pi}^{-i}, \pi_D)$ for every possible decision rule of D (i.e., replace D as a chance node with CPD π_D). The fact that the in-degrees of agent i's decision nodes are bounded, bounds the number of these MAIDs. For each induced MAID, we can then use a poly-time algorithm for IS-BEST-RESPONSE to determine any decision rule π_D that makes up the best response policy for agent i. □

Proposition 10. *In a MAID with sufficient information, if the in-degrees of $\boldsymbol{D}$ are bounded and* IS-BEST-RESPONSE *can be solved in poly-time, then a pure NE can be found in poly-time.*

Proof. First, note that we can check whether a MAID is a sufficient information game in poly-time using s-reachability, a graphical criterion based on d-separation [43]. We can then follow the constructive procedure given for the proof of Proposition 3. Given Proposition 7, each optimisation step must take poly-time and since the in-degrees of all decision nodes are bounded by a constant, the number of subgames is also bounded by a constant. Therefore, the entire procedure takes poly-time. □