The Computational Complexity of Single-Player Imperfect-Recall Games

Emanuel Tewolde[1] , **Caspar Oesterheld**[2] , **Vincent Conitzer**[2,3] and **Paul Goldberg**[4]
[1] First Affiliation
[2] Second Affiliation
[3] Third Affiliation
[4] Fourth Affiliation
{first, second}@example.com, third@other.example.com, fourth@example.com

Abstract

We study single-player extensive-form games with imperfect recall, such as the Sleeping Beauty problem or the Absentminded Driver game. For such games, two natural equilibrium concepts have been proposed as a solution concept alternative to ex-ante optimality. One equilibrium concept uses generalized double halving (GDH) as a belief system and evidential decision theory (EDT), and another one uses generalized thirding (GT) as a belief system and causal decision theory (CDT). Our findings associate those three solution concepts of a game to solution concepts of a polynomial maximization problem - namely - global optima, optimal points with respect to subsets of variables and Karush–Kuhn–Tucker (KKT) points. Based on these correspondences, we are able to settle various complexity-theoretic questions on the computation of ex-ante optimal or equilibrium strategies.

1 Introduction

Most formalisms that are used for reasoning under uncertainty and decision making in AI – for example, HMMs, (dynamic) Bayesian networks, influence diagrams, MDPs, POMDPs, multiagent versions of these – assume what is known as *perfect recall*: the agent does not forget anything it knew before. This may seem to be a very natural assumption: in the design of AI agents, generally we have plenty of reliable memory available. Moreover, the property of perfect recall ensures many nice properties in the context of extensive-form games, including polynomial-time solvability of two-player zero-sum games [Koller and Megiddo, 1992] (and hence, *a fortiori*, single-player games). Finally, even when modeling humans (as in, for example, behavioral game theory [Camerer, 2003]), in spite of our clearly imperfect memory, usually perfect-recall models are used. So why should one use in models with imperfect recall in AI?

It turns out there are a number of reasons why imperfect recall is relevant for AI agents; moreover, in cases where it is relevant, it is clear what the agent will and will not remember – unlike in the case of human memory, which is harder to predict and consequently to model in standard representations of imperfect recall. Imperfect-recall games already appear in the AI literature in the context of solving very large games such as poker: one technique for solving such games is *abstraction* – i.e., reducing the game to a smaller, simplified one to solve instead – and this process can give rise to imperfect recall in the abstracted game [Waugh *et al.*, 2009; Lanctot *et al.*, 2012; Kroer and Sandholm, 2016]. But imperfect recall is also of interest for other reasons. First, we may deliberately choose to have our agents forget: for example, the agent may temporarily need access to data that is sensitive from a privacy perspective, and therefore best forgotten afterwards. An AI agent could also take the form of a highly distributed system operating across many nodes, where not all the nodes have access to the same information; hence, it may act at one node without having access to information that it did have available when acting at another node. Relatedly, the same agent (in the sense of being based on the same source code) may be instantiated multiple times, for example by human users deploying it in multiple contexts. In such cases it can still be useful to consider this family of instantiations as a single agent, but again these instantiations will not all have access to the same information. Finally, again building on the previous case, an agent may be acting not only in the real world, but also in simulations; for example, it may be simulated by another agent that wants to ensure that another instantiation of the same agent will act in a trustworthy fashion in the real world later. In this case, the real-world instantiation of the agent will generally not have access to the information that the simulation had access to earlier.

Being able to make decisions with imperfect recall also represents a technical frontier. Many existing techniques inherently rely on perfect recall. Solving two-player zero-sum games become NP-hard as soon as one player has imperfect recall [Koller and Megiddo, 1992]. Moreover, in these contexts, there remains controversy at the very foundations of how to do probabilistic reasoning and decision making. For example, the Sleeping Beauty problem [Elga, 2000] asks one to give the probability of a state of the world in an imperfect-recall setting; some (Thirders) believe that the correct answer is $1/2$, and others (Halvers) believe it is $1/3$, see Section 3.2. Only recently has a clear picture started to emerge regarding how each of these positions can be combined with a corresponding form of decision theory to make good decisions [Hitchcock, 2004; Draper and Pust, 2008; Briggs, 2010; Conitzer, 2015; Oesterheld and Conitzer, 2022]; we build on that recent conceptual work here to define and study several foundational computational problems.

In this paper, we use extensive-form games to represent settings with imperfect recall. Even though we are considering a single-agent setting, the extensive form is still especially natural to use to model imperfect-recall settings, specifically with the use of information sets. Indeed, as we will discuss, with imperfect recall, game-theoretic phenomena such as notions of equilibrium naturally come up even with just a single agent, intuitively because it is more challenging for that agent to coordinate its actions with those it takes at other times; also, randomization is in general necessary. We consider behavior strategies, which map each information set to a probability distribution over actions. Based on recent literature, we study three distinct solution concepts: *ex ante* optimality, where the behavior strategy is one that maximizes expected utility at the outset; equilibria based on causal decision theory and generalized thirding, in which an agent would not want to change its action at any information set under the assumption that at all other game tree nodes (including ones in the same information set) the agent would follow the original strategy; and equilibria based on evidential decision theory and generalized double halving, in which an agent would not want to switch to a different distribution over actions at a given information set, assuming that the agent would also use the *new* distribution at other nodes in that information set (but would use the original strategy at all other information sets).

Section 2 and 3 define those solution concepts and cover previously known characterizations and hardness results. Section 4 presents our novel results: First we show that the equilibria based on causal decision theory and generalized thirding are exactly the Karush–Kuhn–Tucker points of a corresponding utility maximization problem. Based on that, we derive that the problem of finding such an equilibrium - up to an inverse exponential precision - is complete for the class Continuous Local Search. Finally, we derive various NP-hardness results for maximizing over the set of equilibria and for finding an equilibrium based on evidential decision theory and generalized double halving.

2 Background for Imperfect-Recall Games

2.1 Single-Player Extensive-Form Games with Imperfect Recall

We first define single-player extensive-form games, allowing for imperfect recall. The concepts we use in doing so are standard; for more detail and background, see, e.g., Fudenberg and Tirole [1991], Nisan *et al.* [2007] and Piccione and Rubinstein [1997].

Definition 1. *A single-player extensive-form game with imperfect recall, sometimes also called an extensive decision problem with imperfect recall, consists of:*

1. *A rooted tree of nodes $\mathcal{H}$, where the edges are labeled with actions. The game starts at the root node h_0 and finishes at a leaf node, also called terminal node. The terminal nodes in $\mathcal{H}$ will be denoted as $\mathcal{Z}$. The set A_h refers to the set of actions available at a nonterminal node $h \in \mathcal{H}$.*
2. *A utility function $u : \mathcal{Z} \to \mathbb{R}$, where $u(z)$ represents the payoff that the player receives from finishing the game at terminal node z.*
3. *A partition $\mathcal{H} \setminus \mathcal{Z} = \mathcal{H}_* \sqcup \mathcal{H}_c$ of nonterminal nodes into a set of decision nodes $\mathcal{H}_*$ of the player and a set of chance nodes $\mathcal{H}_c$. This partition indicates whether the single player or exogenous stochasticity determines the action at this node.*
4. *For each chance node $h \in \mathcal{H}_c$, a fixed distribution $\mathbb{P}_c(\cdot \mid h)$ over A_h according to which chance determines an action at h.*
5. *A partition $H_* = \sqcup_{I \in \mathcal{I}_*} I$ of the player's decision nodes into information sets I. We require $A_h = A_{h'}$ for any nodes h, h' of a given info set I.*

Throughout this paper, we use Γ to denote a single-player extensive-form game with imperfect recall. For computational purposes, we assume our games Γ to be represented as a binary encoding of their game tree structure (including the partitions), their utility payoffs and their chance node probabilities. The last two are required to take on rational values only.

Any node $h \in \mathcal{H}$ uniquely corresponds to a (node, action)-history hist(h) from root h_0 to h in the game tree. Define functions d, ν and a such that the node history and action history from h_0 to h are exactly the sequences $\big(\nu(h,0)\,,\,\nu(h,1)\ldots,\,\nu(h,d(h))\big)$ and $\big(a(h,0)\,,\,a(h,1)\,,\,\ldots,\,a(h,d(h)-1)\big)$. In other words, function $d : \mathcal{H} \to \mathbb{N}_0$ determines the tree depth of a node, $\nu : \mathcal{H} \times \mathbb{N}_0 \to \mathcal{H}$ the node of a history at a specified depth, and, by abuse of notation, $a : \mathcal{H} \times \mathbb{N}_0 \to \sqcup_{h \in \mathcal{H}} A_h$ the action of a history at a specified depth. In particular, for all $h \in \mathcal{H}$, we have $\nu(h,0) = h_0$ and $\nu(h,d(h)) = h$. Moreover, restrict the domain of functions ν and a to inputs (h,k) with $k \leq d(h)$ and $k \leq d(h)-1$ respectively, and let function a always send (h,k) into $A_{\nu(h,k)}$.

The depth of Γ is defined to be the maximal depth of one of the leaf nodes. For notational convenience, we will add an info set I to Γ for each chance node in Γ. The collection of these info sets, each consisting of a single element in $\mathcal{H}_c$, shall be denoted by $\mathcal{I}_c$. For each nonterminal node $h \in \mathcal{H} \setminus \mathcal{Z}$, let $I_h \in \mathcal{I}_* \sqcup \mathcal{I}_c$ denote its info set. For each info set $I \in \mathcal{I}_* \sqcup \mathcal{I}_c$, let A_I denote its action set.

Nodes of the same info set are assumed to be indistinguishable to the player during the game (even though the player is always aware of the full game structure). There may be information about the history of play that the player holds at some node, and that the player forgets somewhere further down its subtree. For instance, consider the game in Figure 1. Once the player arrives at node h_3, she cannot distinguish it from possibly being at node h_2. Thus she has already forgotten that she has only taken one action (action C) throughout the game so far. In contrast to that, games with perfect recall have every info set reflect that the player remembers all her earlier actions. In particular, the player does not forget which info sets she entered in which order in the history of play.

With imperfect recall, it could furthermore be the case that multiple nodes of the same history (of some terminal node) belong to the same info set, as it is the case with info set I_1

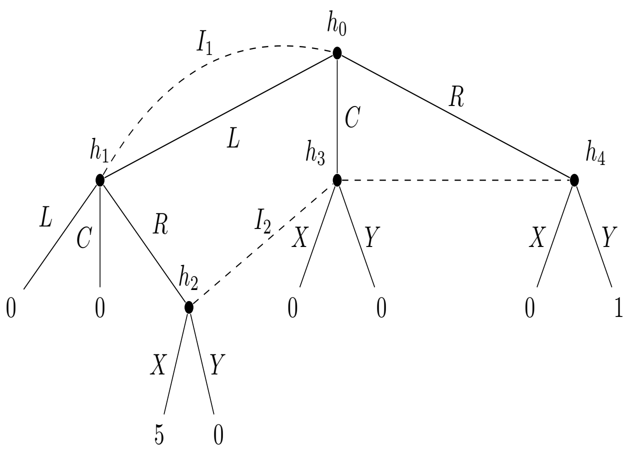

Figure 1: Main example of a single-player extensive-form game with imperfect recall. It has two info sets and five nonterminal nodes.

in the game of Figure 1. The inability of a player to distinguish between two nodes on the same history is a property that we will refer to as *absentmindedness* [cf. the absent-minded driver from Piccione and Rubinstein [1997]].

Let $\Delta(A_I)$ denote the set of all probability distributions over actions a in A_I. A (behavioural) *strategy* $\mu : \mathcal{I}_* \to \sqcup_{I \in \mathcal{I}_*} \Delta(A_I)$ of the player assigns to each info set I a probability distribution $\mu(\cdot \mid I) \in \Delta(A_I)$. At info set I, the player will then randomly draw an action according to $\mu(\cdot \mid I)$. By abuse of notation, we extend any strategy μ of the player to info sets $I_h \in \mathcal{I}_c$ of chance nodes $h \in \mathcal{H}_c$ by setting $\mu(\cdot \mid I_h) := \mathbb{P}_c(\cdot \mid h)$.

Given that the player is currently at node $\bar{h} \in \mathcal{H} \setminus \mathcal{Z}$ and that she plays according to strategy μ, we can calculate the probability of reaching node $h \in \mathcal{H}$ as the product of the probabilities of the action on the path from $\bar{h}$ to h:

$$\mathbb{P}(h \mid \mu, \bar{h}) = \prod_{k=d(\bar{h})}^{d(h)-1} \mu\big(a(h,k) \mid I_{\nu(h,k)}\big) \quad \text{if } \bar{h} \in \text{hist}(h)$$

and $\mathbb{P}(h \mid \mu, \bar{h}) = 0$ otherwise.

As a special case, we define the reach probability $\mathbb{P}(h \mid \mu) := \mathbb{P}(h \mid \mu, h_0)$ of a node $h \in \mathcal{H}$ to be its reach probability from the root h^0 of Γ. Naturally, the reach probability of the root is 1.

The expected utility payoff of being at node $h \in \mathcal{H} \setminus \mathcal{Z}$ and using strategy μ from then on can be determined by

$$Q(\mu \mid h) := \sum_{z \in \mathcal{Z}} \mathbb{P}(z \mid \mu, h) \cdot u(z)\,. \qquad (1)$$

Let $Q : \mu \mapsto Q(\mu) := Q(\mu \mid h_0)$ be the function that takes a strategy μ of Γ and returns the expected utility payoff of the player from following μ from the game start to game termination. $Q(\mu)$ is also called *ex-ante*[1] (expected) utility of μ.

[1]Ex ante is Latin and translates to "from before the event".

2.2 Correspondence with Polynomial Functions

Let us describe the strategy space of Γ and the strategy utility function Q in more detail. Fix an ordering $I_1, \dots, I_l$ of the info sets in $\mathcal{I}_*$ and denote $m_i := |A_{I_i}| - 1$ for all $i \in [l]$. Moreover, fix an ordering $a_1, \dots, a_{m_i+1}$ of the actions in A_{I_i} for all $i \in [l]$.

We can uniquely describe a strategy μ of Γ by its probability values that it assigns to action a_j at info set I_i. In other words, strategies μ are vectors $\mu = (\mu_{ij})_{i,j=1}^{l,m_i+1} \in \times_{i=1}^{l} \mathbb{R}^{m_i+1}$ such that each subvector $\mu_{i \cdot} = (\mu_{ij})_{j=1}^{m_i+1}$ for $i \in [l]$ lies in $\Delta^{m_i} := \{y \in \mathbb{R}^{m_i+1} : y_j \geq 0 \,\forall j\,, \sum_{j=1}^{m_i+1} y_j = 1\}$. The strategy space of Γ becomes exactly $\times_{i=1}^{l} \Delta^{m_i}$.

With (1), the strategy utility function Q of a strategy μ in Γ can be written out to

$$Q(\mu) = \sum_{z \in \mathcal{Z}} \Big(u(z) \cdot \prod_{k=0}^{d(z)-1} \mu\big(a(z,k) \mid I_{\nu(z,k)}\big) \Big)\,.$$

As noted by Piccione and Rubinstein [1997], this is a polynomial function in the variables $(\mu_{ij})_{i,j=1}^{l,m_i+1} \in \times_{i=1}^{l} \mathbb{R}^{m_i+1}$. Note that whenever $\nu(z,i)$ is a chance node ($\in \mathcal{I}_c$), an occurrence $\mu\big(a(z,k) \mid I_{\nu(z,k)}\big)$ in Q does not correspond to one of these variables μ_{ij}. Instead $\mu\big(a(z,k) \mid I_{\nu(z,k)}\big)$ will then be fixed scalar factors $\mathbb{P}_c\big(a(z,k) \mid I_{\nu(z,k)}\big)$ determined by the chance node. Thus, the degree of the polynomial function Q is less than or equal to the maximum number of times the player of Γ might have to take a decision in order to reach a terminal node.

Polynomial Q and parameters $l, m_1, \dots, m_l$ can be constructed in polynomial time in the encoding size of Γ.

Example 2. *In the game of Figure 1, we get* $l = 2$, $m_1 = 2$, $m_2 = 1$. *Let the actions be ordered as* (L, C, R) *and* (X, Y). *Then, for any point* $\mu \in \mathbb{R}^3 \times \mathbb{R}^2$, *we have* $Q(\mu) = 5\mu_{11}\mu_{13}\mu_{21} + \mu_{13}\mu_{22}$.

We show in the appendix that one can also reduce a polynomial $p : \times_{i=1}^{l} \mathbb{R}^{m_i+1} \to \mathbb{R}$ to a single-player extensive-form game Γ with imperfect recall such that $p = Q^{\Gamma}$ on $\times_{i=1}^{l} \mathbb{R}^{m_i+1}$.

2.3 (Computing) Ex-ante Optimal Strategies

Say, we are given a game Γ and we would like to *solve* it. From a planning perspective, one would naturally search for a strategy that promises the highest payoff at a time before the player enters the game.

Definition 3. *We say a strategy* μ^* *is ex-ante optimal for* Γ *if it maximizes the ex-ante utility in* Γ, *that is, if it solves*

$$\max_{\mu \in \times_{i=1}^{l} \mathbb{R}^{m_i+1}} Q(\mu) \quad \text{s.t.} \quad \mu_{i \cdot} \in \Delta^{m_i} \quad \forall i \in [l]\,. \qquad (2)$$

Due to Koller and Megiddo [1992], we can find an ex-ante optimal strategy for a single-player game with perfect recall in polynomial time. This will not be the case anymore in the presence of imperfect recall, as we will show next.

The class ZPP is defined as the set of all those decision problems that can be solved in expected polynomial time by a randomized (Las Vegas) algorithm. Say, OPT is an optimization problem with max and min values $\bar{q}$ and $\underline{q}$. Then, a fully polynomial-time approximation scheme (FPTAS) for OPT computes a solution to an instance of OPT with an objective value that is at most $\epsilon \cdot (\bar{q} - \underline{q})$ away from the optimal value. This computation must take polynomial time in $1/\epsilon$ and the encoding size of the instance. For a more precise definition, see de Klerk [2008].

Proposition 4. *Consider the problem that takes a game Γ and target value $t \in \mathbb{Q}$ (encoded in binary) as inputs and asks whether there is a strategy μ for Γ with ex-ante expected payoff $Q(\mu) \geq t$. This problem is NP-hard. Moreover:*

(1.) Unless NP = ZPP, there is no FPTAS for this problem. NP-hardness and conditional inapproximability hold even if the game instances Γ have a tree depth of 3 and only one info set.

(2.) NP-hardness holds even if Γ has no absentmindedness, a tree depth of 4 and the player has 2 actions per info set.

(3.) NP-hardness holds even if Γ has no absentmindedness, a tree depth of 3 and the player has 3 actions per info set.

An early NP-hardness result of that kind was given by Koller and Megiddo [1992]. Note that finding the ex-ante optimal strategy of Γ is at least as hard as this NP-hard decision problem of whether a target value can be achieved in Γ. With an efficient solver of the decision version, one can at least recover the optimal ex-ante utility $Q^* := \max_\mu Q(\mu)$ through binary search.

A proof of Proposition 4 can be found in the appendix. Result (1.) is based on our reduction from polynomials to games and with the known hardness of maximizing a polynomial function over a simplex [de Klerk, 2008]. Result (3.) reduces from the hard problem of finding an optimal joint strategy in multiplayer common payoff games [Chu and Halpern, 2001].

3 Equilibria in Imperfect-Recall Games

Proposition 4 shows that we cannot hope to find or approximate ex-ante optimal strategies in efficient time for single-player extensive-form games with imperfect recall. In light of these limitations, we will relax the space of solutions to *equilibrium* strategies. This solution concept argues that, whenever the player finds herself in an info set, she will have no influence over which actions she chooses at other info sets. Therefore, at an equilibrium strategy μ, the player shall play the best action at each info set, assuming that she has been playing according to μ up to the current decision point and that she will continue to do so at future decision points. Prior work has given a detailed description of viable equilibrium concepts in single-player games with imperfect recall [Piccione and Rubinstein, 1997; Briggs, 2010; Oesterheld and Conitzer, 2022]. We will consider two well-motivated equilibrium concepts that have been proposed and that include the ex-ante optimal strategy as an equilibrium. The two concepts only differ in games with absentmindedness. More specifically, they differ in how expected utilities are computed for an action a at a current info set I, given that the player plays according to a strategy μ anywhere "else". Computing such expected utilities require

1. a method to form beliefs (that is, a probability distribution) over being at a specific node / history of Γ given that the player is at info set I (Belief System); and
2. an understanding of how an action choice at the current node affects the freedom to choose at other nodes of *the same* info set (Decision Theory).

Throughout this section, let Γ be a single-player extensive-form game with imperfect recall (and with absentmindedness), let the player enter the game with strategy μ, and find herself at info set I.

3.1 Decision Theories

Causal Decision Theory (CDT) postulates that the player can take an action $\alpha \in \Delta(A_I)$ at the current node without violating that the player has been playing according to $\mu(\cdot \mid I)$ at past arrivals at I, nor that she will be playing according to $\mu(\cdot \mid I)$ at future arrivals at I. The intuition behind CDT is that your choice to deviate away from μ at the current node does not *cause* any change in behaviour at any other node of the same info set I.

In contrast to that, Evidential Decision Theory (EDT) postulates that if the player takes an action $\alpha \in \Delta(A_I)$ at the current node, then she will have also deviated to α whenever she arrived in I in past play, and she will deviate to α whenever she arrives in I again in future play. That is because according to EDT, the choice for α now is evidence for the player taking the same choice in the past and future.

Denote with $\mu_{I \mapsto \alpha}$ an EDT deviation, i.e., the strategy of Γ that plays according to μ at every info set except at the info set $I \in \mathcal{I}_*$ where it plays according to $\alpha \in A_I$. Observe that a CDT deviation may result in different actions taken at the same info set. This would not make a valid strategy that the player could have picked before the game started.

Example 5. *Consider the game in Figure 1 and suppose the player enters the game with the strategy $\mu = (R, X)$. Say, upon visiting info set I_1, the player plans to deviate from the μ-prescribed action R to the action L this one time only. Then, CDT argues that the player will stick to her μ-prescribed action R at the other node of I_1, leading to the action histories (L, R, X) or (R, X). EDT, on the other hand, argues that such a deviation will then happen at both nodes of I, leading to the action histories (L,L).*

3.2 Self-locating Belief Systems

Let $I^{1\text{st}} \subseteq I$ refer to those nodes $h \in I$ that are the first node of their respective history to enter info set I. Define the reach probability and (expected) visit frequency of I under μ as $\mathbb{P}(I \mid \mu) := \sum_{h \in I^{1\text{st}}} \mathbb{P}(h \mid \mu)$ and $\text{Fr}(I \mid \mu) := \sum_{h \in I} \mathbb{P}(h \mid \mu)$ respectively. Note that the reach probability and the visit frequency can only differ in games with absentmindedness, and that the visit frequency can be greater than 1. Nonetheless, we have $\mathbb{P}(I \mid \mu) > 0$ if and only if $\text{Fr}(I \mid \mu) > 0$.

Let function $\chi : P \mapsto \{0, 1\}$ take a Boolean property P as input and evaluate 1 if and only if P is true, and let Boolean property "I in $\text{hist}(z)$" evaluate as true if and only if I occurs in history $\text{hist}(z)$ of terminal node $z \in \mathcal{Z}$ at least once.

The first belief system argues that one should consider the visit frequencies:

Definition 6. *Let I be an info set with $\mathrm{Fr}(I \mid \mu) > 0$ under μ, and $h \in \mathcal{H}_*$ a player node. Then, Generalized Thirding (GT) determines the probability of the player to be at h, given that she uses μ and is currently in I, through $\mathbb{P}_{\mathrm{GT}}(h \mid \mu, I) := \chi(h \in I) \cdot \mathbb{P}(h \mid \mu)/\mathrm{Fr}(I \mid \mu)$.*

The second belief system argues that one should only take reach probability into account:

Definition 7. *Let I be an info set with $\mathbb{P}(I \mid \mu) > 0$ under μ, and $z \in \mathcal{Z}$ a terminal node. Then, Generalized Double Halving (GDH) determines the probability of the player to be in the history $\mathrm{hist}(z)$ of z, given that she uses μ and is currently in I, through $\mathbb{P}_{\mathrm{GDH}}(\,\mathrm{hist}(z) \mid \mu, I\,) := \chi(I \text{ in } \mathrm{hist}(z)) \cdot \mathbb{P}(z \mid \mu)/\mathbb{P}(I \mid \mu)$.*

GT and GDH were introduced as "consistency" and "z-consistency" by [Piccione and Rubinstein, 1997].

With the current definitions, GT and GDH assign probabilities to different events (to be at player node versus to be at a history of a terminal node). In the appendix, we phrase GT and GDH in each others language. In the language of GDH, GT assigns a history $\mathrm{hist}(z)$ of a terminal node z a higher probability if the reach probability of z under μ is higher (same as GDH) and if $\mathrm{hist}(z)$ visits info set I very often (GDH only cares about I being visited at least once).

Example 8. *Consider the game in Figure 1 again and suppose the player enters the game with the strategy $\mu = (\frac{1}{2}L + \frac{1}{2}R, X)$. Say, the player observes to be in info set I_1. Then a GT player believes to be at node h_0 in the history (R, X) with probability $\frac{1}{3}$ whereas a GDH player believes to be at h_0 in (R, X) with probability $\frac{1}{2}$. The names "Halving" and "Thirding" originate from this phenomenon but for a different example called Sleeping Beauty [Elga, 2000].*

3.3 Two equilibrium concepts

Any claims made in this section are proven in the appendix.

We start with the equilibrium concept that uses Casual Decision Theory and Generalized Thirding. Denote with $h \circ a$ the child node reached in Γ by following action $a \in A_h$ from player node $h \in \mathcal{H}_*$. Then $Q(\mu \mid h \circ a)$ is the expected utility the player receives from being at h, playing a now, and playing according to μ afterwards.

Definition 9. *Let the player currently be at an info set I with $\mathrm{Fr}(I \mid \mu) > 0$, and let $\alpha \in \Delta(A_I)$ be a mixed action. Then, the (CDT,GT)-expected utility of playing α now and according to μ otherwise is*

$$\mathrm{EU}_{\mathrm{CDT,GT}}(\alpha \mid \mu, I) := \sum_{h \in I} \mathbb{P}_{\mathrm{GT}}(h \mid \mu, I) \cdot \Big(\sum_{a \in A_I} \alpha(a) \cdot Q(\mu \mid h \circ a) \Big).$$

Definition 10. *We say a strategy μ^* of Γ is a (CDT,GT)-Equilibrium if for all info sets $I \in \mathcal{I}_*$ with $\mathrm{Fr}(I \mid \mu^*) > 0$ under μ, we have*

$$\mu^*(\cdot \mid I) \in \underset{\alpha \in \Delta(A_I)}{\mathrm{argmax}}\, \mathrm{EU}_{\mathrm{CDT,GT}}(\alpha \mid \mu^*, I)\,. \tag{3}$$

Instead of (3), we can check the equivalent condition that for all pure actions $a \in A_I$ with $\mu^*(a \mid I) > 0$, we have

$$a \in \underset{a' \in A_I}{\mathrm{argmax}}\, \mathrm{EU}_{\mathrm{CDT,GT}}(a' \mid \mu^*, I)\,.$$

Next, we introduce the equilibrium concept that uses Evidential Decision Theory and Generalized Double Halving.

Definition 11. *Let the player currently be at an info set I with $\mathbb{P}(I \mid \mu) > 0$, and let $\alpha \in \Delta(A_I)$ be a mixed action. Then, the (EDT,GDH)-expected utility of playing α now and according to μ otherwise is*

$$\mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha \mid \mu, I) := \sum_{z \in \mathcal{Z}} \mathbb{P}_{\mathrm{GDH}}(\,\mathrm{hist}(z) \mid \mu_{I \mapsto \alpha}, I\,) \cdot u(z)\,.$$

The probabilities in Definition 11 are well-defined due to $\mathbb{P}(I \mid \mu) = \mathbb{P}(I \mid \mu_{I \mapsto \alpha})$.

Definition 12. *We say a strategy μ^* of Γ is a (EDT,GDH)-Equilibrium if for all info sets $I \in \mathcal{I}_*$ with $\mathbb{P}(I \mid \mu^*) > 0$ under μ, we have*

$$\mu^*(\cdot \mid I) \in \underset{\alpha \in \Delta(A_I)}{\mathrm{argmax}}\, \mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha \mid \mu^*, I)\,. \tag{4}$$

For (EDT,GDH), it is not sufficient to only check for optimality of pure actions that are in the support of μ^*. For instance, take the game in Figure 1 and suppose the player enters the game with the strategy $\mu = (C, X)$. Say, the player observes to be in info set I_1. Then action C is optimal among {L, C, R}. But the player would strictly benefit from deviating to $\frac{1}{2}$L + $\frac{1}{2}$C.

Finally, observe that in games without absentmindedness, the following notions coincide: CDT and EDT, GT and GDH, Definitions 9 and 11, and therefore also (CDT,GT)-Equilibria and (EDT,GDH)-Equilibria.

3.4 Equilibria from the Ex-Ante Perspective

Recall from Section 2.2 that the (ex-ante) strategy utility function Q of Γ is a polynomial function from $\times_{i=1}^{l} \mathbb{R}^{m_i+1}$ to $\mathbb{R}$. In this section, we give characterizations for (CDT,GT)- and (EDT,GDH)-Equilibria in terms of Q, as presented by Oesterheld and Conitzer [2022] and Piccione and Rubinstein [1997]. We reprove these results in the appendix since our setup and end goal differs slightly.

Polynomial Q is continuously differentiable[2]. For $i \in [l]$ and $j \in [m_i + 1]$, let $\nabla_{ij}\, Q$ stand for the partial derivative in direction (i, j), that is, the linear change of Q at a point μ if you infinitesimally increase probability $\mu(a_j \mid I_i)$.

Lemma 13. *Let I_i be an info set, $a_j \in A_{I_i}$ an action, and $\mu \in \times_{i=1}^{l} \Delta^{m_i}$ a strategy. Then:*

1. *$\nabla_{ij}\, Q(\mu) = 0$ if $\mathrm{Fr}(I_i \mid \mu) = 0$, and*
2. *$\nabla_{ij}\, Q(\mu) = \mathrm{Fr}(I_i \mid \mu) \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i)$ otherwise.*

Note that an infinitesimal increase of $\mu(a_j \mid I_i)$ means that the decision at every node of info set I_i is altered. This resembles an EDT type of deviation power but restricted to mixed actions that are very close to the current action $\mu(a_j \mid I_i)$. Then, Lemma 13 says that a CDT deviation - rescaled by $\mathrm{Fr}(I_i \mid \mu)$ - accurately captures the linear (=dominant) effect of such a "local EDT deviation".

[2]Here, we identify $\times_{i=1}^{l} \mathbb{R}^{m_i+1}$ with $\mathbb{R}^{\sum_{i=1}^{l}(m_i+1)}$.

Lemma 14. *Strategy $\mu \in \bigtimes_{i=1}^{l} \Delta^{m_i}$ of Γ is an (EDT,GDH)-Equilibrium if and only if for all $i \in [l]$:*

$$\mu_{i\cdot}^* \in \underset{y \in \Delta^{m_i}}{\operatorname{argmax}} \; Q(\mu_{1\cdot}^*, \dots, \mu_{i-1\cdot}^*, y, \mu_{i+1\cdot}^*, \dots, \mu_{l\cdot}^*) \,.$$

One possible interpretation of Lemma 14 is that (EDT,GDH)-Equilibria of Γ are exactly the Nash Equilibria of an l-player simultaneous and identical-interest game G: Each player i shall have the continuous action space Δ^{m_i} and the (single) utility payoff - as a function of the chosen action profile $(\mu_{i\cdot})_{i=1}^{l} \in \bigtimes_{i=1}^{l} \Delta^{m_i}$ - shall be polynomial Q.

3.5 Computational Considerations

One of our main results addresses the complexity of finding a (CDT,GT)-Equilibrium. There are problem instances where all (CDT,GT)-Equilibria take on irrational numbers even though the game is easy to encode (see the appendix). Therefore, we relax our search to ϵ-approximate (CDT,GT)-Equilibria for a numerical precision parameter ϵ.

Definition 15. *An instance of the problem* (CDT,GT)-EQUILIBRIUM *consists of a single-player extensive-form game Γ with imperfect recall and a precision parameter $\epsilon > 0$ encoded in binary. A solution consists of a strategy μ for Γ that satisfies for all $I \in \mathcal{I}_*$ with $\mathrm{Fr}(I \mid \mu) > 0$:*

$$\mathrm{EU}_{\mathrm{CDT,GT}}\big(\mu(\cdot \mid I) \mid \mu, I\big) \geq \max_{a' \in A_I} \mathrm{EU}_{\mathrm{CDT,GT}}(a' \mid \mu, I) - \epsilon \,.$$

We will show that (CDT,GT)-EQUILIBRIUM is complete for the class *Continuous Local Search* (CLS). CLS was introduced by Daskalakis and Papadimitriou [2011] and sits somewhere in the function problem counterparts of the classes P and NP. We can characterize CLS through one of its complete problems: CLS contains all those search problems that can be solved by performing a Gradient Descent algorithm on a bounded domain. Fearnley *et al.* [2021] showed that it is CLS-complete to find Karush-Kuhn-Tucker (KKT) points, that is, the solution concept that Gradient Descent searches for.

In order to define KKT points, consider a general non-linear maximization problem

$$\max_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Bx + b \leq 0 \,, \, Cx + c = 0 \qquad (5)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, $B \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{l \times n}$, and $c \in \mathbb{R}^l$, and the domain is bounded. A point $x \in \mathbb{R}^n$ is then said to satisfy the KKT-conditions (and, be a KKT point) for (5) if there exist KKT multipliers $\tau_1, \dots, \tau_m, \kappa_1, \dots, \kappa_l \in \mathbb{R}$ such that

$$Bx + b \leq 0 \text{ and } Cx + c = 0$$
$$\tau_j \geq 0 \quad \forall j \in [m]$$
$$\tau_j = 0 \quad \text{or} \quad B_{j\cdot}x + bj = 0 \quad \forall j \in [m]$$
$$\nabla f(x) = \sum_{j=1}^{m} \tau_j \cdot (B_{j\cdot})^T + \sum_{i=1}^{l} \kappa_i \cdot (C_{i\cdot})^T = 0 \,.$$

The KKT conditions are necessary first-order conditions for a point to be a local optimum of (5). Stationary points in the domain satisfy the KKT conditions. Fearnley *et al.* [2021] show that finding an ϵ-KKT point for inverse exponential precision parameter ϵ is CLS-complete. Even though Gradient Descent is used extensively in practice, such as in many machine learning tasks, it can still require exponential time to find such a solution point. CLS-hard problems are generally believed to be hard to solve (by any algorithm) as their hardness can be based on the cryptographic assumption of indistinguishability obfuscation [Hubácek and Yogev, 2017].

4 Main Results

To our knowledge, the results of this section are all novel unless explicitly stated otherwise. All proofs can be found in the appendix.

4.1 Complexity of the Search Problems

First, we use can use Lemma 13 to give a characterization of (CDT,GT)-equilibria in terms of ex-ante utility. For that, recall the ex-ante maximization problem (2).

Theorem 1. *Strategy $\mu \in \bigtimes_{i=1}^{l} \Delta^{m_i}$ of Γ is a (CDT,GT)-Equilibrium if and only μ is a KKT point of (2).*

In the appendix, we give a visualization of finding a (CDT,GT)-Equilibria through KKT conditions. Note that in continuous optimization, there can be KKT-points that are not locally optimal. An analogous effect can also happen in games with imperfect recall: In the game of Figure 1, strategy $\mu = (C, X)$ is a (CDT,GT)-equilibrium. But it is not a local optimum from an ex-ante perspective because for any $\epsilon > 0$ a shift from $\mu(\cdot \mid I_1) = C$ to $\epsilon \cdot (\frac{1}{2}\mathrm{L} + \frac{1}{2}\mathrm{R}) + (1 - \epsilon) \cdot \mathrm{C}$ would yield the player ex-ante utility $5 \cdot \frac{\epsilon}{2} \cdot \frac{\epsilon}{2} > 0$. However, a (CDT,GT) player is satisfied with her choice at I_1. In terms of the original definition of CDT, this is because deviating exactly *once* in I_1 never suffices to attain a utility of 5. In terms of Lemma 13 and Theorem 1, the issue is that the first order effect of increasing the probabilities of R and L is 0.

We obtain a hierarchy among our three solution concepts, which was already shown by Oesterheld and Conitzer [2022] [cf. Piccione and Rubinstein, 1997]:

Lemma 16. *Ex-ante optimal strategies μ of a game Γ are also (EDT,GDH)-Equilibria of Γ. (EDT,GDH)-Equilibria μ of a game Γ are also (CDT,GT)-Equilibria of Γ. It follows that a game Γ always admits an (EDT,GDH)-Equilibrium and a (CDT,GT)-Equilibrium.*

The implication chain of Lemma 16 does not hold in the reverse direction: Consider the game in Figure 1. Then strategy $\mu = (C, X)$ is a (CDT,GT)-Equilibrium, but not an (EDT,GDH)-equilibrium. Moreover, strategy $\mu' = (R, Y)$ is an (EDT,GDH)-Equilibrium with ex-ante utility 1. This is not ex-ante optimal because strategy $\mu'' = (\frac{1}{2}L + \frac{1}{2}C, X)$ achieves the (optimal) ex-ante utility $5/4$.

Lemma 16 also implies that finding an ex-ante optimal strategy must be at least as hard as finding an (EDT,GDH)-Equilibrium which must be at least as hard as finding a (CDT,GT)-Equilibria. For the latter, we get the following classification:

Theorem 2. (CDT,GT)-EQUILIBRIUM *is CLS-complete. CLS-hardness holds even for game instances that*

(1.) have a tree depth of 5 *and the player has* 2 *actions per info set*
(2.) have no absentmindedness and a tree depth of 6
(3.) have no chance nodes, a tree depth of 5*, and only one info set*

In (2.), one can replace the restriction on tree depth by the restriction that the player has 2 actions per info set. We prove (1.) by a reduction from finding a KKT point of a polynomial function over the hypercube. For (2.) and (3.), we reduce from finding a Nash equilibrium of a polytensor identical interest game. The problems we reduce from were both shown to be CLS-complete by Babichenko and Rubinstein [2021].

Theorem 2 shows in particular that absentmindedness is not the reason for why it is hard to find (CDT,GT)-equilibria in games with imperfect recall. Moreover, the CLS membership of (CDT,GT)-EQUILIBRIUM implies that unless NP = co-NP, the problem cannot be hard for the class NP [Megiddo and Papadimitriou, 1991].

Corollary 17. *Finding an approximate (EDT,GDH)-Equilibrium for games without absentmindedness is CLS-complete.*

Finding (EDT,GDH)-equilibria in games with absentmindedness. Lemma 14 gives a characterization of (EDT,GDH)-equilibria phrased in terms of polynomials over the product of simplices. The authors are not aware of any results that put the problem of finding an approximate (EDT,GDH)-equilibrium in the class FNP (which would immediately yield TFNP membership since an (EDT,GDH)-equilibrium is guaranteed to exist). The difficulty lies in the fact that given a candidate (EDT,GDH)-equilibrium μ and an information set I_i with $\mathbb{P}(I_i \mid \mu) > 0$, it is unclear how to check in polynomial time whether $\mu(\cdot \mid I_i)$ indeed achieves ϵ-optimal (EDT,GDH)-expected utility in the global optimization (sub-)task (4) for I_i. The same conceptual difficulties arise with finding an ex-ante optimal strategy.

4.2 Complexity of the Decision Problems

Next, we show that maximizing expected utility in an info set or maximizing over the space of equilibria is NP-hard. In the following problem formulations, any target value $t \in \mathbb{Q}$ shall be encoded in binary.

Theorem 3. *The following problems are all NP-hard. Unless NP = ZPP, there is also no FPTAS for these problems.*

(1a.) Given Γ *and* $t \in \mathbb{Q}$*, is there an (CDT,GT)-Equilibrium of* Γ *with ex-ante utility* $\geq t$*?*
(1b.) Given Γ*, an info set* I *of* Γ *and* $t \in \mathbb{Q}$*, is there a (CDT,GT)-Equilibrium* μ *such that* $\mathrm{Fr}(I \mid \mu) > 0$*, and such that the player has a (CDT,GT)-expected utility* $\geq t$ *upon reaching* I*?*
(1c.) Given Γ*, an info set* I *of* Γ *and* $t \in \mathbb{Q}$*, is there a strategy* μ *of* Γ *such that* $\mathrm{Fr}(I \mid \mu) > 0$*, and such that the player has a (CDT,GT)-expected utility* $\geq t$ *upon reaching* I*?*
(2a.) Given Γ *and* $t \in \mathbb{Q}$*, is there an (EDT,GDH)-Equilibrium of* Γ *with ex-ante utility* $\geq t$*?*
(2b.) Given Γ*, an info set* I *of* Γ *and* $t \in \mathbb{Q}$*, is there a (EDT,GDH)-Equilibrium* μ *such that* $\mathbb{P}(I \mid \mu) > 0$*, and such that the player has a (EDT,GDH)-expected utility* $\geq t$ *upon reaching* I*?*
(2c.) Given Γ*, an info set* I *of* Γ *and* $t \in \mathbb{Q}$*, is there a strategy* μ *of* Γ *such that* $\mathbb{P}(I \mid \mu) > 0$*, and such that the player has a (EDT,GDH)-expected utility* $\geq t$ *upon reaching* I*?*
(3a.) Given Γ *and* $t \in \mathbb{Q}$*, do all (EDT,GDH)-Equilibria of* Γ *have ex-ante utility* $\geq t$*?*
(3b.) Given Γ*, an info set* I *of* Γ *and* $t \in \mathbb{Q}$*, do all (EDT,GDH)-Equilibria* μ *with* $\mathbb{P}(I \mid \mu) > 0$ *promise the player a (EDT,GDH)-expected utility* $\geq t$ *upon reaching* I*?*

All these results follow from Proposition 4. Therefore, NP-hardness and conditional inapproximability remain for problems of the form (-a.) even if we restrict the game instances as described in Proposition 4. The same holds for problems of the form (-b.) and (-c.) except that we have to add one information set and one tree depth level to the game instances.

Hardness of decision problem (3a.) relies on the observation that if Γ has one info set only, because then, the (EDT,GDH)-Equilibria coincide with the ex-ante optimal strategies (see Lemma 14). In particular, unless P = NP, there will not be a polynomial time solver for finding an (EDT,GDH)-Equilibrium of a game Γ.

5 Conclusion

Games of imperfect recall have traditionally often been considered a theoretical curiosity; it is hard to model settings with human actors as imperfect-recall games, because, while most of us frequently forget things, we do not *reliably* forget things according to well-specified rules. For AI agents, however, this is no longer true; moreover, because they can be instantiated many times, sometimes in simulation, one instantiation will generally not know what another knew earlier. All this motivates the *computational* study of games of imperfect recall, which we initiated here for the single-player case. We are aided in this endeavor by recent conceptual work that specifies and motivates several natural solution concepts, and we based our work on these. Standard polynomial-time algorithms such as ones based on the sequence form are known to no longer work with imperfect recall, and indeed in this paper we found various complexity-theoretic evidence that indeed, single-player imperfect-recall games are hard to solve. Some of this evidence is, intriguingly, based on the complexity class CLS whose careful study is only very recent. On the positive side, we also provided insights into solving such games by drawing close connections to several problems about maximizing polynomial functions.

There remain many avenues for future work. What can be said about these computational problems for representation schemes other than the extensive form? Are there special cases of imperfect-recall games that can be solved more efficiently, whether they are single-player or multi-player? One may also ask whether our results give insight into the more conceptual questions. For example, does the fact that (CDT,GT)-equilibria are (under reasonable complexity assumptions) easier to compute than (EDT,GDH)-equilibria provide support for using the former solution concept, at least for certain purposes? We hope that the work we have done in this paper can serve as a springboard for further research into this fascinating and important topic.

A On Section 2.2: Reductions from Polynomial Maximization to Games

At the end of Section 2.2, we mention that we can reduce a polynomial $p : \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1} \to \mathbb{R}$ to a single-player extensive-form game Γ with imperfect recall such that $p = Q^\Gamma$ on $\bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1}$. Let us show that by giving two variants of the same reduction idea.

A general polynomial function $p : \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1} \to \mathbb{R}$ of degree d can be uniquely represented in terms of the standard monomial basis $\left\{ \prod_{i,j=1}^{l,m_i+1} x_{ij}^{D_{ij}} \right\}_{D \in \mathrm{MB}(d,\mathbf{m}+\mathbf{1})}$. Here, we summarized $(m_i+1)_{i=1}^{l}$ to the vector $\mathbf{m}+\mathbf{1}$, and we denote with $\mathrm{MB}(d,\mathbf{m}+\mathbf{1})$ the set

$$\{D = (D_{ij})_{ij} \in \bigtimes_{i=1}^{l} \mathbb{N}_0^{m_i+1} \, : \, \sum_{i,j=1}^{l,m_i+1} D_{ij} \leq d\}$$

of all variations to draw up to $d \in \mathbb{N}$ elements out of the set $\{(i,j)\}_{i,j=1}^{l,m_i+1}$, with replacement and without regard to draw order.

Throughout this paper (or more accurately, throughout the appendix), if the instance to a problem contains a polynomial function $p : \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1} \to \mathbb{R}$, then we assume it to be represented as a binary encoding of $l, (m_i+1)_{i=1}^{l}$, and the polynomials coefficients $(\lambda_D)_{D \in \mathrm{MB}(d,\mathbf{m}+\mathbf{1})}$, which need to be rational. This is also called the Turing (bit) model. Denote $\mathrm{supp}(p) := \{D \in \mathrm{MB}(d,\mathbf{m}+\mathbf{1}) \, : \, \lambda_D \neq 0\})$ and, for each $D \in \mathrm{MB}(d,\mathbf{m}+\mathbf{1})$, $\mathrm{supp}(D) := \{(i,j) \text{ with } i \in [l], j \in [m_i+1] \, : \, D_{ij} > 0\})$ as well as $|D| := \sum_{i,j=1}^{l,m_i+1} D_{ij}$.

Given such a polynomial function, let us construct a corresponding single-player extensive-form game Γ with imperfect recall. It shall have info sets I_i for $i \in [l]$ with action sets $A_{I_i} := \{a_1, \dots, a_{m_i+1}\}$, and a tree depth of up to $d+1$. Let the root h_0 be a chance node that has one outgoing edge to depth 1-node h_D for each monomial index $D \in \mathrm{supp}(p)$. An outgoing edge is drawn uniform randomly. First, handle the special case of D being the zero vector. There, h_D will be a terminal node with a utility payoff of $|\mathrm{supp}(p)|$. So consider $D \neq \mathbb{0}$ from now on, where h_D will be a nonterminal node. Denote the subtree rooted at h_D with T_D. Keep in mind that it is associated with monomial $\prod_{i,j=1}^{l,m_i+1} x_{ij}^{D_{ij}}$. We will now build T_D - a tree of depth $|D|$ - depth layer by depth layer by lexicographically going through the set $\mathrm{supp}(D)$. Otherwise, depth $k-1$ of T_D corresponds to the k-th element of $\mathrm{supp}(D)$, referred to as $\big(i(k), j(k)\big)$. We present two variants with how to continue building Γ:

Variant 1: There will be at most one nonterminal node h of depth $k-1$ of T_D (it would be h_D on depth 0). Assign h to info set $I_{i(k)}$. Create $m_{i(k)}+1$ outgoing edges out of h into a new node of depth k, and the edges shall be labeled with $\{a_1, \dots, a_{m_{i(k)}+1}\}$ of $A_{I_{i(k)}}$. The created node $h \circ a_{j(k)}$ shall be non-terminal and the created nodes $h \circ a_j$ for $j \neq j(k)$ shall be terminal with utility payoff 0. With this procedure, subtree T_D will have depth $\sum_{(i,j) \in \mathrm{supp}(D)} D_{ij}$. The nonterminal node of the last depth layer has action history $\big(a_{j(k)} \in A_{I_{i(k)}}\big)_{k \in [|\mathrm{supp}(D)|]}$ in T_D. Now reverse the fact that it is nonterminal, and make it terminal instead. Denote it as z_D and assign it a utility of $\lambda_D \cdot |\operatorname{supp}(p)|$.

Variant 2: This variant does not have terminal nodes in depth layers $0, \dots, |D| - 1$. Each node h of depth $k - 1$ of T_D shall belong to info set $I_{i(k)}$ (Depth 0 only has one node, namely h_D). Create $m_{i(k)} + 1$ outgoing edges out of each h into a new node of depth k, and the. The edges shall be labeled with $\{a_1, \dots, a_{m_{i(k)}+1}\}$ of $\in A_{I_{i(k)}}$. The nodes of the last depth layer (depth $|D|$) shall be terminal nodes. One of those nodes has the action history $\big(a_{j(k)} \in A_{I_{i(k)}}\big)_{k \in [|\operatorname{supp}(D)|]}$ in T_D, which we will refer to as z_D. Assign it a utility of $\lambda_D \cdot |\operatorname{supp}(p)|$. Assign all other terminal nodes of T_D a utility of 0.

In both variants, we have that any point $x \in \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1}$ of p that is also in $\bigtimes_{i=1}^{l} \Delta^{m_i}$ naturally makes a strategy μ in Γ with probabilities $\mu(a_j \mid I_i) = x_{ij}$. Moreover, the strategy utility function Q of both variants of Γ satisfies

$$
\begin{aligned}
Q(\mu) &= \sum_{z \in \mathcal{Z}} \mathbb{P}(z \mid \mu) \cdot u(z) = \\
&= \sum_{D \in \operatorname{supp}(p)} \mathbb{P}(z_D \mid \mu) \cdot \lambda_D \cdot |\operatorname{supp}(p)| \\
&= \sum_{D \in \operatorname{supp}(p)} \left[\Big(\frac{1}{|\operatorname{supp}(p)|} \cdot \prod_{D_{ij} > 0} x_{ij}^{D_{ij}} \Big) \cdot \lambda_D \cdot |\operatorname{supp}(p)| \right] \\
&= \sum_{D \in \mathrm{MB}(d, \mathbf{m}+\mathbf{1})} \lambda_D \cdot \prod_{i,j} x_{ij}^{D_{ij}} \\
&= p(x)
\end{aligned}
$$

for corresponding x and μ. This extends to $Q = p$ on all $\bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1}$.

The construction of the first variant of Γ takes polynomial time in the encoding size of $p : \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1} \to \mathbb{R}$. That is, because the game tree has size $\leq |\operatorname{supp}(p)| \cdot \deg(p) \cdot (\max_i m_i + 1)$. The second variant of Γ can have exponential game tree size in general. But, if we know that polynomial instances p have fixed degree d for example, then the tree size is $\leq |\operatorname{supp}(p)| \cdot (\max_i m_i + 1)^d$ which makes the whole construction of Γ polynomial time in the size of the input instances again. We will make use of this observation later.

B On Section 2.3: Proof of Proposition 4

We recall and prove Proposition 4:

Proposition. *Consider the problem that takes a game Γ and target value $t \in \mathbb{Q}$ (encoded in binary) as inputs and asks whether there is a strategy μ for Γ with ex-ante expected payoff $Q(\mu) \geq t$. This problem is NP-hard. Moreover:*

(1.) Unless NP = ZPP, there is no FPTAS for this problem. NP-hardness and conditional inapproximability hold even if the game instances Γ have a tree depth of 3 and only one info set.

(2.) NP-hardness holds even if Γ has no absentmindedness, a tree depth of 4 and the player has 2 actions per info set.

(3.) NP-hardness holds even if Γ has no absentmindedness, a tree depth of 3 and the player has 3 actions per info set.

Proof. (1.)
Appendix A gives a reduction from maximizing polynomial $p : \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1} \to \mathbb{R}$ over the product of simplices to maximizing ex-ante utility in a single-player extensive-form game Γ with imperfect recall (take the first variant of the reduction). de Klerk [2008] gives a survey on maximizing polynomials over popular compact domains. Consider the decision problem that takes as an instance a polynomial function $p : \Delta^m \to \mathbb{R}$ and a target value $t \in \mathbb{Q}$, and that has to decide whether there exists a point $x \in \Delta^m$ with $p(x) \geq t$. de Klerk [2008] note that this problem is NP-hard and has no FPTAS unless NP=ZPP, even if p is known to be quadratic functions only. They derive the conditional inapproximability from Håstad [1997]. In terms of our reduction, such polynomials reduce to a game of depth 3 and one info set. Moreover, a instance of the polynomial problem is a yes instance if and only if the reduced instance of ex-ante utility game problem of this proposition is a yes instance.

(2.)
Reduce from 3SAT. Let $x_1, \dots, x_l$ be the variables of a 3CNF formula ϕ having n clauses. We construct the corresponding game instance Γ as follows. Each variable x_i has corresponding info set I_i whose nodes have 2 outgoing edges with action labels T, F. The root of the tree is a chance node that selects amongst n subtrees, corresponding to the clauses, with equal probability $1/n$. Each subtree is a binary tree of depth 3. If the clause C associated with the subtree contains variables x_i, x_j, x_k, then the root node of the subtree shall belong to info set I_i, its two children belong to info set I_j, and their four children belong to info set I_k. Each leaf has a path that is associated by truth assignments of x_i, x_j, x_j. A leaf shall yield a utility of 1 if the associated truth assignment satisfies the clause C, else value 0. The utility target value for the game instance shall be 1.

Then, ϕ is satisfiable if and only if there is a strategy in the corresponding game with ex-ante utility 1. Note that every clause subtree is reached with positive probability and that an ex-ante utility of 1 means that any realization of the candidate strategy must only reach terminal node with utility payoff 1. For the backward implication, say μ is a behavioural (=randomizing) strategy with an ex-ante utility of 1. Take any pure-action strategy μ' that only uses actions that are in the support of μ. Then μ' makes a satisfying truth assignment for ϕ.

(3.)
Consider the following type of 2-player games G: A game G consists of a family $(G_s)_{s \in S}$ of 2-player simultaneous games where $|S| \in \mathbb{N}$ is finite. There is a probability distribution over S determining which game the two players will play. There is a partition $S = \sqcup_{I \in \mathcal{I}_*} I$ for player 1 and $S = \sqcup_{J \in \mathcal{J}_*} J$ for player 2 representing the info sets of each player within which the respective player cannot differentiate. For any $I \in \mathcal{I}_*$, games G_s with $s \in I$ have the same action set A_I for player 1. Analogously, for any $J \in \mathcal{J}_*$, games G_s with $s \in J$ have the same action set A_J for player 2. First,

nature draws $s \in S$ according to the probability distribution, then each player simultaneously takes an action in G_s. Such an outcome $(s, a \in A_{I_s}, a' \in A_{J_s})$ yields each player the same payoff $u(s, a, a')$. A strategy for player 1 (resp. player 2) returns a mixed action $\alpha \in \Delta(A_I)$ to each set $I \in \mathcal{I}_*$ (resp. an $\alpha \in \Delta(A_J)$ to each set $J \in \mathcal{J}_*$.

Chu and Halpern [2001] show that given a game G as described above, it is NP-hard to decide whether there is a strategy profile (μ_1, μ_2) that yields a payoff ≥ 1 in G. After a closer inspection of their reduction from 3SAT, we can see that their NP-hardness result holds even for game instances where each player only has up to three actions in each G_s and where the all payoffs are as simple as 0 or 3. They also show that such a game G can be transformed into an extensive-form game in polynomial time. We will use almost the same transformation to in our upcoming reduction from their problem to our problem of interest.

Take an instance G as described above. Assume further that G has up to three actions in each G_s. Create a single-player extensive-form game Γ with imperfect recall in the following way: Γ has depth 3 and information sets $\mathcal{I}_* \cup \mathcal{J}_*$. The root of Γ is a chance node that chooses a subtree T_s associated with $s \in S$ according to the given probability distribution over S. The root of a subtree T_s shall be assigned to info set $I_s \in \mathcal{I}_*$ and it shall have A_{I_s} outgoing edges. Each node of subtree T_s of subtree depth level 1 shall be assigned to info set $J_s \in \mathcal{J}_*$ and they shall have A_{J_s} outgoing edges. Each node of the whole game Γ of depth level 3 shall be a terminal node, and if its associated action history is $(s, a \in A_{I_s}, a' \in A_{J_s})$, then it shall yield the single player a payoff of $u(s, a, a')$. Also, let the target value t for Γ be 1. A strategy (μ_1, μ_2) of G shall correspond to the strategy μ of Γ that returns $\mu_1(I) \in \Delta(A_I)$ for info sets $I \in \mathcal{I}_* \subset \mathcal{I}_* \cup \mathcal{J}_*$ and returns $\mu_1(J) \in \Delta(A_J)$ for info sets $J \in \mathcal{J}_* \subset \mathcal{I}_* \cup \mathcal{J}_*$

Then, there is a strategy profile (μ_1, μ_2) of G that yields a payoff ≥ 1 if and only if its corresponding strategy μ in Γ yields an ex-ante utility $\geq t$.

Note that in this reduction, Γ has no absentmindedness, a tree of depth 3, and at most three actions for each player info set.

□

C On Section 3.2: Comparing Generalized Thirding and Generalized Double Halving

This section stays close to the exposition of Oesterheld and Conitzer [2022].

Let Γ be a single-player extensive-form game with imperfect recall, μ be the strategy with which the player entered the game, and let the player find herself at an info set $I \in \mathcal{I}_*$ with $\mathbb{P}(I \mid \mu) > 0 \iff \text{Fr}(I \mid \mu) > 0$.

Recall that then, GT determines the probability of being at player node $h \in \mathcal{H}_*$ specifically as

$$\mathbb{P}_{\text{GT}}(h \mid \mu, I) = \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I \mid \mu)}\,.$$

Denote the number of times info set I occurs in the history of a terminal node z with $\#(I \text{ in hist}(z)) \in \mathbb{N} \cup \{0\}$. Furthermore, let "at h in hist(z)" be the event that the player is currently at node h and on the root-to-end path hist(z) for some given terminal node z. Then we get the following GT probabilities for being in the history hist(z) of a specified terminal node $z \in \mathcal{Z}$:

$$\begin{aligned}
\mathbb{P}_{\text{GT}}(\,\text{hist}(z) \mid \mu, I\,) &= \sum_{h \in \mathcal{H}_*} \mathbb{P}_{\text{GT}}(\,\text{at}\, h \,\text{in hist}(z) \mid \mu, I\,) \\
&:= \sum_{h \in \mathcal{H}_* \cap \text{hist}(z)} \mathbb{P}_{\text{GT}}(h \mid \mu, I\,) \cdot \mathbb{P}(z \mid \mu, h) \\
&= \sum_{h \in I \cap \text{hist}(z)} \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I \mid \mu)} \cdot \mathbb{P}(z \mid \mu, h) \\
&= \#(I \text{ in hist}(z)) \cdot \frac{\mathbb{P}(z \mid \mu)}{\text{Fr}(I \mid \mu)}
\end{aligned}$$

Compare this to the definition of GDH:

$$\mathbb{P}_{\text{GDH}}(\,\text{hist}(z) \mid \mu, I\,) = \chi(I \text{ in hist}(z)) \cdot \frac{\mathbb{P}(z \mid \mu)}{\mathbb{P}(I \mid \mu)}$$

So GT assigns a history hist(z) of a terminal node z a higher probability if the reach probability of z under μ is higher (same as GDH) and if hist(z) visits info set I very often (GDH only cares about I being visited at least once).

As we can see, GT assigns being in the history of a terminal node z a higher probability if the reach probability $\mathbb{P}(z \mid \mu)$ under μ is higher and/or if the history hist(z) enters info set I very often.

When it comes to events of the form "at h in hist(z)", then GDH will just uniformly distribute the probability of being in hist(z) among all those nodes in hist(z) that are also in I. With this, GDH can also assign probabilities to the event of being at a specified node $h \in \mathcal{H}_*$, namely, as

$$\begin{aligned}
&\mathbb{P}_{\text{GDH}}(h \mid \mu, I\,) \\
&\quad = \sum_{z \in \mathcal{Z}} \mathbb{P}_{\text{GDH}}(\,\text{at}\, h \,\text{in hist}(z) \mid \mu, I\,) \\
&\quad = \sum_{\substack{z \in \mathcal{Z} \\ h \in I \cap \text{hist}(z)}} \mathbb{P}_{\text{GDH}}(\,\text{at}\, h \,\text{in hist}(z) \mid \mu, I\,) \\
&\quad := \sum_{\substack{z \in \mathcal{Z} \\ h \in I \cap \text{hist}(z)}} \frac{\mathbb{P}_{\text{GDH}}(\,\text{hist}(z) \mid \mu, I\,)}{\#(I \text{ in hist}(z))} \\
&\quad = \chi(h \in I) \cdot \sum_{\substack{z \in \mathcal{Z} \\ h \in \text{hist}(z)}} \frac{\mathbb{P}(z \mid \mu)}{\mathbb{P}(I \mid \mu)} \cdot \frac{1}{\#(I \text{ in hist}(z))} \\
&\quad = \chi(h \in I) \cdot \sum_{\substack{z \in \mathcal{Z} \\ h \in \text{hist}(z)}} \frac{\mathbb{P}(h \mid \mu) \cdot \mathbb{P}(z \mid \mu, h)}{\mathbb{P}(I \mid \mu) \cdot \#(I \text{ in hist}(z))} \\
&\quad = \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\mathbb{P}(I \mid \mu)} \cdot \sum_{\substack{z \in \mathcal{Z} \\ h \in \text{hist}(z)}} \frac{\mathbb{P}(z \mid \mu, h)}{\#(I \text{ in hist}(z))}
\end{aligned}$$

D On Section 3.3: Proofs

This section stays close to the ideas of Oesterheld and Conitzer [2022].

Mixed vs Pure Action Deviations in (CDT,GT). We have

$$\begin{aligned} &\text{EU}_{\text{CDT,GT}}(\alpha \mid \mu, I) \\ &= \sum_{h \in I} \mathbb{P}_{\text{GT}}(h \mid \mu, I) \cdot \Big(\sum_{a \in A_I} \alpha(a) \cdot Q(\mu \mid h \circ a) \Big) \\ &= \sum_{a \in A_I} \alpha(a) \cdot \Big(\sum_{h \in I} \mathbb{P}_{\text{GT}}(h \mid \mu, I) \cdot Q(\mu \mid h \circ a) \Big) \\ &= \sum_{a \in A_I} \alpha(a) \cdot \text{EU}_{\text{CDT,GT}}(a \mid \mu, I) \,. \end{aligned} \tag{6}$$

Therefore, $\alpha^* \in \text{argmax}_{\alpha \in \Delta(A_I)} \text{EU}_{\text{CDT,GT}}(\alpha \mid \mu, I)$ if and only if it mixes only over optimal pure actions, i.e., $\alpha^*(a^*) > 0 \implies a^* \in \text{argmax}_{a \in A_I} \text{EU}_{\text{CDT,GT}}(a \mid \mu, I)$.

EDT Deviation does not affect Reach Probabilities. Observe that a node $h \in I^{1\text{st}}$ in Γ satisfies $\mathbb{P}(h \mid \mu) = \mathbb{P}(h \mid \mu_{I \mapsto \alpha})$ because info set I - in which μ and $\mu_{I \mapsto \alpha}$ differ - never appeared in the history of h. Therefore,

$$\begin{aligned} \mathbb{P}(I \mid \mu) = \sum_{h \in I^{1\text{st}}} \mathbb{P}(h \mid \mu) &= \sum_{h \in I^{1\text{st}}} \mathbb{P}(h \mid \mu_{I \mapsto \alpha}) \\ &= \mathbb{P}(I \mid \mu_{I \mapsto \alpha}) \,. \end{aligned} \tag{7}$$

Without absentmindedness, (CDT,GT) equals (EDT,GDH). First, we shall show two facts that hold in games with imperfect recall (with or without absentmindedness). For any strategy μ' of Γ, node $h \in \mathcal{H} \setminus \mathcal{Z}$, and terminal node $z \in \mathcal{Z}$, we have

$$\begin{aligned} &\mathbb{P}(z \mid \mu', h) \\ &= \chi(h \in \text{hist}(z)) \cdot \prod_{k=d(h)}^{d(z)-1} \mu'\big(a(z,k) \mid I_{\nu(z,k)}\big) \\ &= \chi(h \in \text{hist}(z)) \cdot \mu'\big(a(z,d(h)) \mid I_{\nu(z,d(h))}\big) \cdot \\ &\qquad \prod_{k=d(h)+1}^{d(z)-1} \mu'\big(a(z,k) \mid I_{\nu(z,k)}\big) \\ &= \sum_{a' \in A_{I_h}} \chi(h \in \text{hist}(z)) \cdot \chi\Big(a' = a(z,d(h))\Big) \cdot \mu'\big(a' \mid I_h\big) \cdot \\ &\qquad \prod_{k=d(h)+1}^{d(z)-1} \mu'\big(a(z,k) \mid I_{\nu(z,k)}\big) \\ &= \sum_{a' \in A_{I_h}} \mu'\big(a' \mid I_h\big) \cdot \chi(h \circ a' \in \text{hist}(z)) \cdot \\ &\qquad \prod_{k=d(h \circ a)}^{d(z)-1} \mu'\big(a(z,k) \mid I_{\nu(z,k)}\big) \\ &= \sum_{a' \in A_{I_h}} \mu'\big(a' \mid I_h\big) \cdot \mathbb{P}(z \mid \mu', h \circ a') \,. \end{aligned}$$

Moreover, we can obtain

$$\begin{aligned} Q(\mu' \mid h) &= \sum_{z \in \mathcal{Z}} \mathbb{P}(z \mid \mu', h) \cdot u(z) \\ &= \sum_{z \in \mathcal{Z}} \sum_{a' \in A_{I_h}} \mu'\big(a' \mid I_h\big) \cdot \mathbb{P}(z \mid \mu', h \circ a') \cdot u(z) \\ &= \sum_{a' \in A_{I_h}} \mu'\big(a' \mid I_h\big) \cdot \sum_{z \in \mathcal{Z}} \mathbb{P}(z \mid \mu', h \circ a') \cdot u(z) \\ &= \sum_{a' \in A_{I_h}} \mu'\big(a' \mid I_h\big) \cdot Q(\mu' \mid h \circ a') \,. \end{aligned}$$

Now let Γ be a single-player extensive-form game with imperfect recall and without absentmindedness. Let μ be the strategy with which the player entered the game, and let the player find herself at info set $I \in \mathcal{I}_*$ with $\mathbb{P}(I \mid \mu) > 0 \iff \text{Fr}(I \mid \mu) > 0$.

CDT and EDT address how a players choices at the current node affect the players choice at other nodes of the same path and of the same info set I. Without absentmindedness, this consideration becomes obsolete because for any root-to-end path in Γ, the player can arrive in I at most once on that path. Thus, EDT deviations and CDT deviations have the same effect.

Moreover, that lack of absentmindedness means $I^{1\text{st}} = I$. Hence, $\mathbb{P}(I \mid \mu) = \text{Fr}(I \mid \mu)$ and $\chi(I \text{ in hist}(z)) = \#(I \text{ in hist}(z))$ for any terminal node $z \in \mathcal{Z}$. Therefore, by Appendix C, we have for any terminal node z:

$$\begin{aligned} \mathbb{P}_{\text{GT}}(\text{hist}(z) \mid \mu, I) &= \#(I \text{ in hist}(z)) \cdot \frac{\mathbb{P}(z \mid \mu)}{\text{Fr}(I \mid \mu)} \\ &= \chi(I \text{ in hist}(z)) \cdot \frac{\mathbb{P}(z \mid \mu)}{\mathbb{P}(I \mid \mu)} \\ &= \mathbb{P}_{\text{GDH}}(\text{hist}(z) \mid \mu, I) \end{aligned}$$

and for any node $h \in \mathcal{H}_*$:

$$\begin{aligned} &\mathbb{P}_{\text{GDH}}(h \mid \mu, I) \\ &\quad = \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\mathbb{P}(I \mid \mu)} \cdot \sum_{\substack{z \in \mathcal{Z} \\ h \in \text{hist}(z)}} \frac{\mathbb{P}(z \mid \mu, h)}{\#(I \text{ in hist}(z))} \\ &\quad = \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I \mid \mu)} \cdot \sum_{\substack{z \in \mathcal{Z} \\ h \in \text{hist}(z)}} \mathbb{P}(z \mid \mu, h) \\ &\quad = \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I \mid \mu)} \cdot 1 \\ &\quad = \mathbb{P}_{\text{GT}}(h \mid \mu, I) \end{aligned}$$

Finally, we show that (CDT,GT) equals (EDT,GDH) compute the same expected utilities from a deviation to a mixed action $\alpha \in \Delta(A_I)$. Since $I^{1\text{st}} = I$, we have $\mathbb{P}(h \mid \mu) = \mathbb{P}(h \mid \mu_{I \mapsto \alpha})$ for all $h \in I$. With the general facts showed in the beginning, we can derive for games without absentmindedness:

$$
\begin{aligned}
&\mathrm{EU}_{\mathrm{CDT,GT}}(\alpha \mid \mu, I) \\
&= \sum_{h\in I} \mathbb{P}_{\mathrm{GT}}(h \mid \mu, I) \cdot \Big(\sum_{a\in A_I} \alpha(a) \cdot Q(\mu \mid h \circ a) \Big) \\
&= \sum_{h\in I} \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\mathrm{Fr}(I \mid \mu)} \cdot \sum_{a\in A_I} \alpha(a) \cdot Q(\mu \mid h \circ a) \\
&= \sum_{h\in I} 1 \cdot \frac{\mathbb{P}(h \mid \mu)}{\mathbb{P}(I \mid \mu)} \cdot \sum_{a\in A_I} \mu_{I\mapsto\alpha}(a \mid I) \cdot Q(\mu_{I\mapsto\alpha} \mid h \circ a) \\
&= \sum_{h\in I} \frac{\mathbb{P}(h \mid \mu_{I\mapsto\alpha})}{\mathbb{P}(I \mid \mu_{I\mapsto\alpha})} \cdot Q(\mu_{I\mapsto\alpha} \mid h) \\
&= \sum_{h\in I} \frac{\mathbb{P}(h \mid \mu_{I\mapsto\alpha})}{\mathbb{P}(I \mid \mu_{I\mapsto\alpha})} \cdot \\
&\qquad \sum_{z\in\mathcal{Z}} \mathbb{P}(z \mid \mu_{I\mapsto\alpha}, h) \cdot \chi(h \in \mathrm{hist}(z)) \cdot u(z) \\
&= \sum_{z\in\mathcal{Z}} u(z) \cdot \sum_{h\in I\cap\mathrm{hist}(z)} \frac{\mathbb{P}(h \mid \mu_{I\mapsto\alpha})}{\mathbb{P}(I \mid \mu_{I\mapsto\alpha})} \cdot \mathbb{P}(z \mid \mu_{I\mapsto\alpha}, h) \\
&= \sum_{z\in\mathcal{Z}} u(z) \cdot \sum_{h\in I\cap\mathrm{hist}(z)} \frac{\mathbb{P}(z \mid \mu_{I\mapsto\alpha})}{\mathbb{P}(I \mid \mu_{I\mapsto\alpha})} \\
&= \sum_{z\in\mathcal{Z}} u(z) \cdot \#(I \text{ in } \mathrm{hist}(z)) \cdot \frac{\mathbb{P}(z \mid \mu_{I\mapsto\alpha})}{\mathbb{P}(I \mid \mu_{I\mapsto\alpha})} \\
&= \sum_{z\in\mathcal{Z}} u(z) \cdot \chi(I \text{ in } \mathrm{hist}(z)) \cdot \frac{\mathbb{P}(z \mid \mu_{I\mapsto\alpha})}{\mathbb{P}(I \mid \mu_{I\mapsto\alpha})} \\
&= \sum_{z\in\mathcal{Z}} u(z) \cdot \mathbb{P}_{\mathrm{GDH}}(\, \mathrm{hist}(z) \mid \mu_{I\mapsto\alpha}, I\,) \\
&= \mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha \mid \mu, I)
\end{aligned}
$$

In particular, the notions of (CDT,GT)-Equilibrium and (EDT,GDH)-Equilibrium coincide in games without absentmindedness.

E On Section 3.4: Proofs

The proofs of this section stay conceptually close to Piccione and Rubinstein [1997] and Oesterheld and Conitzer [2022].

E.1 Proof of Lemma 13

We established that the strategy space of a game Γ is $\bigtimes_{i=1}^{l} \Delta^{m_i}$ and that the strategy utility function Q extends to all $\bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1}$. In order to discuss differentiabity, we view elements $\mu \in \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1}$ as flattened vectors in $\mathbb{R}^{\sum_{i=1}^{l}(m_i+1)}$. However, for notational convenience, we keep the vector description $\mu = (\mu_{ij})_{i,j=1}^{l,m_i+1} \in \bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1}$ because it is indexed by a (info set, action)-pair for our game-theoretic perspective. Therefore, the basis vectors that span $\bigtimes_{i=1}^{l} \mathbb{R}^{m_i+1}$ are the vectors e_{ij}, for $i \in [l]$ and $j \in [m_i]$, with (i', j') entries

$$
(e_{ij})_{i'j'} := \begin{cases} 1 & \text{if } i' = i \text{ and } j' = j \\ 0 & \text{otherwise.} \end{cases}
$$

Then, polynomial Q has one partial derivative $\nabla_{ij}\, Q$ for each coordinate direction e_{ij} which is defined as

$$
\nabla_{ij}\, Q(\mu) := \lim_{\epsilon\to 0} \frac{1}{\epsilon} \cdot \Big(Q(\mu + \epsilon \cdot e_{ij}) - Q(\mu) \Big). \qquad (8)
$$

We are ready to prove Lemma 13.

Lemma. *Let I_i be an info set, $a_j \in A_{I_i}$ an action, and $\mu \in \bigtimes_{i=1}^{l} \Delta^{m_i}$ a strategy of the game. Then:*

1. $\nabla_{ij}\, Q(\mu) = 0$ *if* $\mathrm{Fr}(I_i \mid \mu) = 0$, *and*
2. $\nabla_{ij}\, Q(\mu) = \mathrm{Fr}(I_i \mid \mu) \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i)$ *otherwise.*

Proof. Using (8), let us first get an expression for $Q(\mu + \epsilon \cdot e_{ij})$. By Section 2.2, we have

$$
\begin{aligned}
&Q(\mu + \epsilon \cdot e_{ij}) \\
&= \sum_{z\in\mathcal{Z}} u(z) \cdot \qquad (9) \\
&\qquad \prod_{k=0}^{d(z)-1} \Big(\mu\big(a(z,k) \mid I_{\nu(z,k)}\big) + \epsilon \cdot e_{ij}\big(a(z,k) \mid I_{\nu(z,k)}\big) \Big)
\end{aligned}
$$

where we use $e_{ij}\big(a(z,k) \mid I_{\nu(z,k)}\big)$ to indicate the entry of coordinate direction e_{ij} at the info set index of $I_{\nu(z,k)}$ and the action index of $a(z,k)$. We continue the equation chain of (9) by sorting the product of sums by their order in ϵ:

$$
\begin{aligned}
&= \sum_{z\in\mathcal{Z}} \Bigg(u(z) \prod_{k=0}^{d(z)-1} \mu\big(a(z,k) \mid I_{\nu(z,k)}\big) \Bigg) + \\
&\qquad \sum_{z\in\mathcal{Z}} \Bigg(u(z) \sum_{h\in\mathrm{hist}(z)} \Big[\epsilon \cdot e_{ij}\Big(a(z,d(h)) \mid I_h\Big) \cdot \\
&\qquad \prod_{\substack{k=0 \\ k\neq d(h)}}^{d(z)-1} \mu\big(a(z,k) \mid I_{\nu(z,k)}\big) \Big] \Bigg) + \mathcal{O}(\epsilon^2) \\
&= Q(\mu) + \epsilon \cdot \sum_{z\in\mathcal{Z}} \Big[u(z) \cdot \sum_{h\in\mathrm{hist}(z)} e_{ij}\Big(a(z,d(h)) \mid I_h\Big) \cdot \\
&\qquad \mathbb{P}(h \mid \mu) \cdot \mathbb{P}\Big(z \mid \mu, h \circ a(z,d(h))\Big) \Big] + \mathcal{O}(\epsilon^2) \\
&\overset{(*)}{=} Q(\mu) + \mathcal{O}(\epsilon^2) + \\
&\qquad \epsilon \cdot \sum_{z\in\mathcal{Z}} \sum_{h\in\mathrm{hist}(z)\cap I_i} \chi\Big(a_j \neq a(z,d(h))\Big) \cdot u(z) \cdot \\
&\qquad \mathbb{P}(h \mid \mu) \cdot \mathbb{P}\Big(z \mid \mu, h \circ a(z,d(h))\Big) \\
&\overset{(**)}{=} Q(\mu) + \mathcal{O}(\epsilon^2) + \\
&\qquad \epsilon \cdot \underbrace{\sum_{z\in\mathcal{Z}} \sum_{h\in I_i} u(z) \cdot \mathbb{P}(h \mid \mu) \cdot \mathbb{P}\Big(z \mid \mu, h \circ a_j\Big)}_{\text{Denote this term as } (\dagger)}
\end{aligned}
$$

In equation line $(*)$, we use that e_{ij} is zero in any info set $\neq I_i$ or any action $\neq a_j \in A_{I_i}$. In equation line $(*)$, we use that $\mathbb{P}\Big(z \mid \mu, h \circ a_j\Big)$ is zero if $h \neq \mathrm{hist}(z)$ or $a_j \neq a(z, d(h))$.

Term (†) is constant in ϵ. Once we derived a better expression for (†), we can obtain the statement of the lemma from

$$\begin{aligned}\nabla_{ij} Q(\mu) &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\big(Q(\mu + \epsilon\cdot e_{ij}) - Q(\mu)\big)\\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(Q(\mu) + \epsilon\cdot(\dagger) + \mathcal{O}(\epsilon^2) - Q(\mu)\Big)\\ &= (\dagger)\,.\end{aligned}$$

Consider the case where $0 = \mathrm{Fr}(I_i \mid \mu) = \sum_{h\in I_i}\mathbb{P}(h \mid \mu)$. Recall that $\mu \in \bigtimes_{i=1}^{l}\Delta^{m_i}$. Therefore, all reach probabilities are non-negative. In particular, we obtain $\mathbb{P}(h \mid \mu) = 0$ for all $h \in I_i$. Hence, $(\dagger) = 0$.

Consider the other case, namely, where $\mathrm{Fr}(I_i \mid \mu) > 0$. Then, we can simplify:

$$\begin{aligned}(\dagger) &= \sum_{h\in I_i}\sum_{z\in\mathcal{Z}} u(z)\cdot\mathbb{P}(h\mid\mu)\cdot\mathbb{P}\Big(z \mid \mu, h\circ a_j)\Big)\\ &= \frac{\mathrm{Fr}(I_i\mid\mu)}{\mathrm{Fr}(I_i\mid\mu)}\cdot\sum_{h\in I_i}\mathbb{P}(h\mid\mu)\cdot\sum_{z\in\mathcal{Z}}u(z)\cdot\mathbb{P}\Big(z\mid\mu,h\circ a_j)\Big)\\ &= \mathrm{Fr}(I_i\mid\mu)\cdot\sum_{h\in I_i}\chi(h\in I_i)\cdot\frac{\mathbb{P}(h\mid\mu)}{\mathrm{Fr}(I_i\mid\mu)}\cdot Q(\mu\mid h\circ a_j)\\ &= \mathrm{Fr}(I_i\mid\mu)\cdot\sum_{h\in I_i}\mathbb{P}_{\mathrm{GT}}(h\mid\mu,I_i)\cdot Q(\mu\mid h\circ a_j)\\ &= \mathrm{Fr}(I_i\mid\mu)\cdot\mathrm{EU}_{\mathrm{CDT,GT}}(a_j\mid\mu,I_i)\end{aligned}$$

□

E.2 Proof of Lemma 14

Lemma. *Strategy $\mu \in \bigtimes_{i=1}^{l}\Delta^{m_i}$ of a game Γ is an (EDT,GDH)-Equilibrium if and only if for all $i\in[l]$:*

$$\mu^*_{i\cdot} \in \operatorname*{argmax}_{y\in\Delta^{m_i}} \; Q(\mu^*_{1\cdot},\ldots,\mu^*_{i-1\cdot},y,\mu^*_{i+1\cdot},\ldots,\mu^*_{l\cdot})\,.$$

Proof. We start with the definition (EDT,GDH)-expected utilities (Definition 11). Say, the player entered the game with strategy $\mu\in\bigtimes_{i=1}^{l}\Delta^{m_i}$, and arrived at an info set I_i with $\mathbb{P}(I_i\mid\mu)>0$. Let $\alpha\in\Delta(A_{I_i})$. Then

$$\begin{aligned}&\mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha\mid\mu,I_i)\\ &= \sum_{z\in\mathcal{Z}}\mathbb{P}_{\mathrm{GDH}}(\,\mathrm{hist}(z)\mid\mu_{I_i\mapsto\alpha},I_i\,)\cdot u(z)\\ &= \sum_{z\in\mathcal{Z}}\chi(I_i \text{ in hist}(z))\cdot\frac{\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})}{\mathbb{P}(I_i\mid\mu_{I_i\mapsto\alpha})}\cdot u(z)\\ &= \frac{1}{\mathbb{P}(I_i\mid\mu_{I_i\mapsto\alpha})}\cdot\sum_{\substack{z\in\mathcal{Z}\\(I_i \text{ in hist}(z))}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z) \quad (10)\end{aligned}$$

Continue with the definition of an (EDT,GDH)-Equilibrium (Definition 12). When taking the argmax of (10) over $\alpha\in\Delta(A_{I_i})$, we can rescale (10) by strictly positive factors and add terms to it without changing the solution set to the argmax (as long as the factors and terms are independent of α). First, multiply (10) by α-independent factor $\mathbb{P}(I_i\mid\mu_{I_i\mapsto\alpha})\overset{(7)}{=}\mathbb{P}(I_i\mid\mu)>0$. Second, observe that for any $z\in\mathcal{Z}$ for which info set I does not occur in hist(z), we have $\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha}) = \mathbb{P}(z\mid\mu)$. Hence, we can add α-independent term

$$\sum_{\substack{z\in\mathcal{Z}\\(I \text{ not in hist}(z))}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z)$$

to (10) afterwards. These two steps yield

$$\begin{aligned}&\operatorname*{argmax}_{\alpha\in\Delta(A_{I_i})}\mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha\mid\mu,I_i)\\ &\overset{(10)}{=}\operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}}\frac{1}{\mathbb{P}(I_i\mid\mu_{I_i\mapsto\alpha})}\cdot\sum_{\substack{z\in\mathcal{Z}\\(I \text{ in hist}(z))}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z)\\ &= \operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}}\sum_{\substack{z\in\mathcal{Z}\\(I \text{ in hist}(z))}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z)\\ &= \operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}}\sum_{z\in\mathcal{Z}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z) = \operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}} Q(\mu_{I_i\mapsto\alpha})\\ &= \operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}} Q(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{l\cdot})\,.\end{aligned}$$

where we use that strategy $\mu_{I_i\mapsto\alpha}$ has vector description $(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{l\cdot})$.

All in all, we obtain that μ is an (EDT,GDH)-Equilibrium

$$\begin{aligned}\iff\; &\forall i\in[l] \text{ with } \mathbb{P}(I_i\mid\mu)>0\;:\\ &\mu(\cdot\mid I_i)\in\operatorname*{argmax}_{\alpha\in\Delta(A_{I_i})}\mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha\mid\mu,I_i)\\ \iff\; &\forall i\in[l] \text{ with } \mathbb{P}(I_i\mid\mu)>0\;:\\ &\mu_{i\cdot}\in\operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}} Q(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{l\cdot})\\ \overset{(*)}{\iff}\; &\forall i\in[l]\;:\\ &\mu_{i\cdot}\in\operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}} Q(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{l\cdot})\end{aligned}$$

The last equivalence is based on the following argument: If info set I_i has reach probability $\mathbb{P}(I_i\mid\mu)=0$, then function $Q(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{l\cdot})$ is constant in the action choice α at info set I_i. Therefore, any $\alpha'\in\Delta^{m_i+1}$ satisfies

$$\alpha'\in\operatorname*{argmax}_{\alpha\in\Delta^{m_i+1}} Q(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{l\cdot})\,.$$

This is why we can add or remove this generally true statement as a condition requirement. □

F On Section 3.5: Solutions to Simple Games can be Irrational

We here show that the solutions (*ex ante* optimal policy, and (EDT,GDH)- and (CDT,GT)- equilibria) to simple games can be irrational. In fact, we will show that they are sometimes are not even expressible in radicals.

As a starting point, consider the polynomial equation

$$x^5 - x - 1 = 0.$$

This equation has a unique real-valued solution $x^*\approx 1.1673$ that cannot be expressed in radicals, i.e., that cannot be expressed using sums, products, divisions and roots [Lang, 1994][P.121]. In particular, it is irrational.

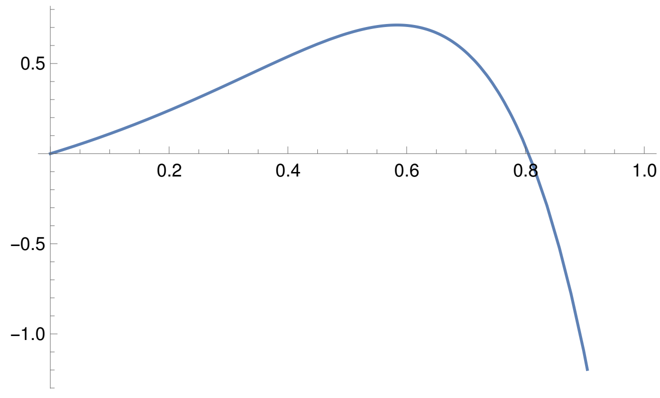

Figure 2: A plot of the function $p(x) = -\frac{16}{3}x^6 + x^2 + x$.

From the above polynomial equation we will construct a polynomial whose unique KKT point on $[0, 1]$ is $x^*/2$. First, note that the equation $32x^5 - 2x - 1 = (2x)^5 - 2x - 1 = 0$, obtained by substituting $2x$ for x in the above, has only one solution, $\tilde{x} = x^*/2 \approx 0.58365$. Now consider the polynomial

$$p(x) = -\frac{16}{3}x^6 + x^2 + x.$$

Notice that the derivative of this is exactly $-32x^5 + 2x + 1$. Thus p's only critical point is $\tilde{x}$. The function is plotted in Figure 2. This point is a local maximum, and the function has no other local maximum in the compact interval $[0, 1]$ (or elsewhere).

Now consider the game in Figure 3. Let $Q: \mathbb{R}^2 \to \mathbb{R}: (x, y) \mapsto -\frac{16}{3}x^6 + x^2 + x = p(x)$ be the *ex ante* expected utility function extended to $\mathbb{R}^2$. Clearly, the unique optimal policy is $x = \tilde{x}, y = 1 - \tilde{x}$. Next we will show that the unique (CDT,GT) equilibrium is also $x = \tilde{x}$. It is easy to see that neither $x = 0$ nor $x = 1$ induces a (CDT,GT) equilibrium. So for any (CDT,GT) equilibrium it has to be the case that $\mathrm{EU}_{\mathrm{CDT,GT}}(a_1 \mid x) = \mathrm{EU}_{\mathrm{CDT,GT}}(a_2 \mid x)$. By Lemma 13, this means that we must have $\frac{d}{dx}Q(x, y) = \frac{d}{dy}Q(x, y) = 0$. Therefore, the only (CDT,GT) equilibrium is at $x = \tilde{x}$. By Lemma 16, this means that the only (EDT,GDH) equilibrium is also at $x = \tilde{x}$.

All in all, we have given an easy to represent game for which the only solution $(\tilde{x}, 1 - \tilde{x})$ has entries which are irrational (or even more, cannot be expressed in radicals).

G Proofs of the Main Results

G.1 Proof of Theorem 1

Recall the problem of maximizing the ex-ante utility in a game Γ:

$$\begin{aligned} \max_{\mu \in \times_{i=1}^{l} \mathbb{R}^{m_i+1}} \quad & Q(\mu) \\ \text{s.t.} \quad & \mu_{ij} \geq 0 \quad \forall i \in [l], \forall j \in [m_i + 1] \\ & \sum_{j=1}^{m_i+1} \mu_{ij} = 1 \quad \forall i \in [l] \end{aligned} \tag{11}$$

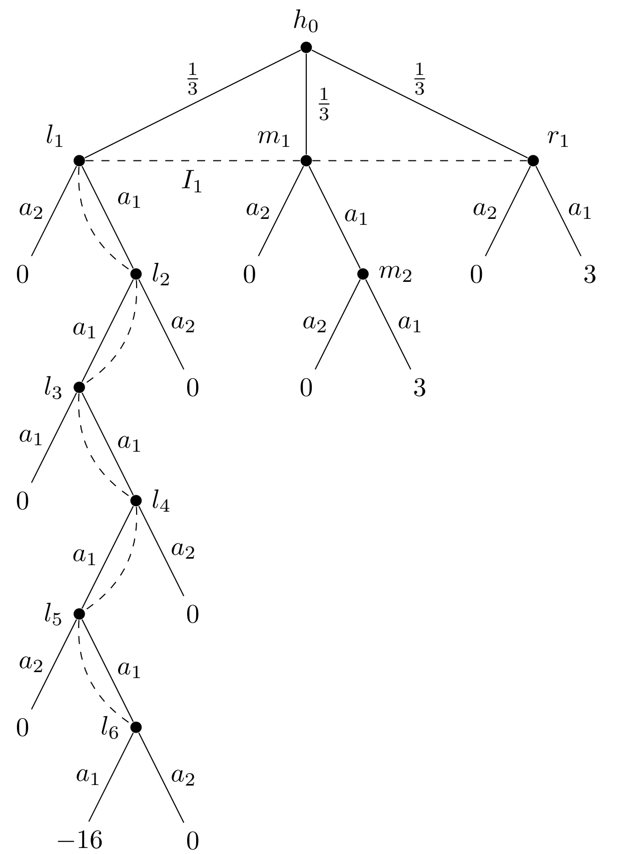

Figure 3: A single-player game of imperfect recall whose only solution cannot be expressed in radicals.

Then, the KKT conditions for (11) and a point $\mu \in \times_{i=1}^{l} \mathbb{R}^{m_i+1}$ are: There exist KKT multipliers $\{\tau_{ij} \in \mathbb{R}\}_{i,j=1}^{l,m_i+1}$ and $\{\kappa_i \in \mathbb{R}\}_{i=1}^{l}$ such that

$$\begin{aligned} & \mu_{ij} \geq 0 \quad \forall i \in [l], \forall j \in [m_i + 1] \\ & \sum_{j=1}^{m_i+1} \mu_{ij} = 1 \quad \forall i \in [l] \\ & \tau_{ij} \geq 0 \quad \forall i \in [l], \forall j \in [m_i + 1] \\ & \tau_{ij} = 0 \quad \text{or} \quad \mu_{ij} = 0 \quad \forall i \in [l], \forall j \in [m_i + 1] \\ & \nabla_{ij} Q(\mu) = -\tau_{ij} + \kappa_i \quad \forall i \in [l], \forall j \in [m_i + 1]\,. \end{aligned} \tag{12}$$

We are ready to prove Theorem 1.

Theorem. *Strategy $\mu \in \times_{i=1}^{l} \Delta^{m_i}$ of Γ is a (CDT,GT)-Equilibrium if and only μ is a KKT point of (11).*

Proof. " $\implies$ ":
Suppose $\mu \in \times_{i=1}^{l} \Delta^{m_i}$ is a (CDT,GT)-Equilibrium of Γ. Then the first two KKT conditions of (12) are satisfied by assumption. Let $i \in [l], \forall j \in [m_i + 1]$.

If $\mathrm{Fr}(I_i \mid \mu) = 0$, then by Lemma 13 we have $\nabla_{ij} Q(\mu) = 0$. Therefore, choose $\tau_{ij} = 0$ and $\kappa_i = 0$ to satisfy the last three KKT conditions for the respective i and j.

Suppose $\mathrm{Fr}(I_i \mid \mu) > 0$. Choose

$$\kappa_i = \max_{j' \in [m_i+1]} \nabla_{ij'} Q(\mu) .$$

The choice of τ_{ij} depends on μ_{ij}. If $\mu_{ij} > 0$, then by Lemma 13 and by the equivalent characterization of a (CDT,GT)-Equilibrium right below Definition 10, we have

$$\begin{aligned} \nabla_{ij} Q(\mu) &= \mathrm{Fr}(I_i \mid \mu) \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i) \\ &= \max_{j' \in [m_i+1]} \left\{ \mathrm{Fr}(I_i \mid \mu) \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_{j'} \mid \mu, I_i) \right\} \\ &= \max_{j' \in [m_i+1]} \nabla_{ij'} Q(\mu) = \kappa_i . \end{aligned}$$

Therefore choose $\tau_{ij} = 0$. If $\mu_{ij} = 0$, on the other hand, then the above equation chain would yield inequality $\nabla_{ij} Q(\mu) \leq \kappa_i$ instead (when transitioning to the max). Thus, choose $\tau_{ij} = \kappa_i - \nabla_{ij} Q(\mu)$. These choices of κ_i and τ_{ij} satisfy the last three KKT conditions for the respective i and j.

“ $\Longleftarrow$ ”:

Suppose μ is a KKT point of problem (11), that is, it satisfies the KKT conditions (12) for some KKT multipliers $\{\tau_{ij} \in \mathbb{R}\}_{i,j=1}^{l,m_i+1}$ and $\{\kappa_i \in \mathbb{R}\}_{i=1}^{l}$. Then, the first two KKT conditions ensure that μ makes a valid strategy for Γ (recall that we always identify $\mu(a_j \mid I_i) := \mu_{ij}$ for info set $I_i \in \mathcal{I}_*$ and action $a_j \in I_i$). So let us check equivalent characterization of a (CDT,GT)-Equilibrium right below Definition 10. Let $i \in [l]$ and $j \in [m_i + 1]$ be such that $\mathrm{Fr}(I_i \mid \mu) > 0$ and $\mu_{ij} > 0$. Then, Lemma 13 and the last three KKT conditions give us for all $j' \in [m_i + 1]$:

$$\begin{aligned} &\mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i) = \frac{\mathrm{Fr}(I_i \mid \mu)}{\mathrm{Fr}(I_i \mid \mu)} \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i) \\ &= \frac{1}{\mathrm{Fr}(I_i \mid \mu)} \cdot \nabla_{ij} Q(\mu) = \frac{1}{\mathrm{Fr}(I_i \mid \mu)} \cdot (-\tau_{ij} + \kappa_i) \\ &\overset{\tau_{ij}=0}{=} \frac{1}{\mathrm{Fr}(I_i \mid \mu)} \cdot \kappa_i \geq \frac{1}{\mathrm{Fr}(I_i \mid \mu)} \cdot (-\tau_{ij'} + \kappa_i) \\ &= \frac{1}{\mathrm{Fr}(I_i \mid \mu)} \cdot \nabla_{ij'} Q(\mu) = \mathrm{EU}_{\mathrm{CDT,GT}}(a_{j'} \mid \mu, I_i) \end{aligned}$$

This implies $\mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i) = \max_{a \in A_{I_i}} \mathrm{EU}_{\mathrm{CDT,GT}}(a \mid \mu, I_i)$. Overall, we get that μ is a (CDT,GT)-Equilibrium for Γ □

G.2 Visualization of Theorem 1

Consider the game in Figure 1, except replace the player nodes of info set I_2 by chance nodes that choose actions X and Y with with chance probabilities $\frac{1}{2}$ each. Then the player only has to choose a mixed action for info set I_1. The only (CDT,GT)-Equilibrium of the underlying game is $(0.6 \cdot L + 0.4 \cdot R)$.

The gradient vector field ∇Q of the strategy utility function Q of this game is plotted in Figure 4. We plotted this in 2D for visualization convenience. The x-axis represents the probability put on action R, the y-axis represents the probability put on action L, and the probability put on action C can be determined as $1 - p_L - p_R$. Then the simplex $\Delta(\{L, C, R\})$

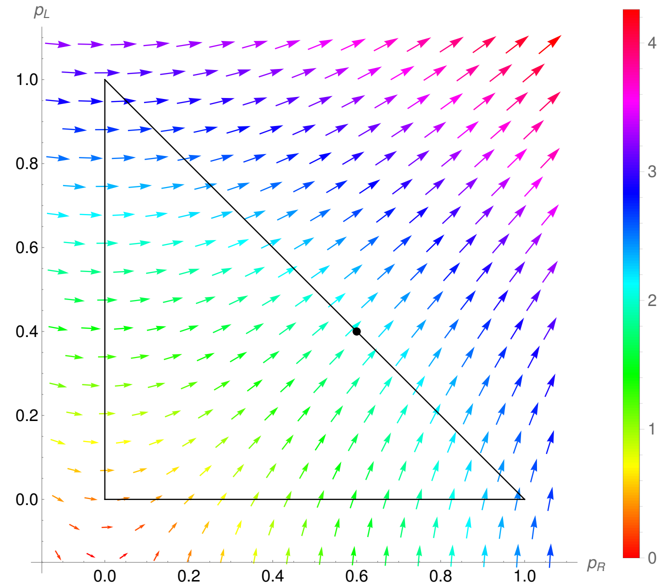

Figure 4: A visualization of Theorem 1.

becomes a triangle in Figure 4 with the point $(0, 0)$ corresponding to pure action C. Figure 4 shows the statement of Theorem 1: The only (CDT,GT)-Equilibrium of the underlying game is also the only KKT point of Q. This is because point in the interior of the (projected) simplex has a vanishing gradient, and because $(0.6 \cdot L + 0.4 \cdot R)$ is the only point of the boundary at which the gradient vector points outwards perpendicular to the boundary constraint.

This visual also reveals a method to find (CDT,GT)-Equilibria, namely by applying Gradient Descent on Q.

G.3 Reproof of Lemma 16

The next lemma is due to Oesterheld and Conitzer [2022] [cf. Piccione and Rubinstein, 1997].

Lemma. *Ex-ante optimal strategies μ of a game Γ are also (EDT,GDH)-Equilibria of Γ. (EDT,GDH)-Equilibria μ of a game Γ are also (CDT,GT)-Equilibria of Γ. It follows that a game Γ always admits an (EDT,GDH)-Equilibrium and a (CDT,GT)-Equilibrium.*

Proof. An ex-ante optimal strategy is defined as a maximum to the maximization problem (2). Therefore, it will in particular satisfy the characterization of an (EDT,GDH)-Equilibrium according to Lemma 14.

Let us show that an (EDT,GDH)-Equilibrium is a (CDT,GT)-Equilibrium by using their respective ex-ante characterizations from Lemma 14 and Theorem 1. So let $\mu = \in \bigtimes_{i=1}^{l} \Delta^{m_i}$ satisfy for all $i \in [l]$

$$\mu_{i \cdot}^* \in \operatorname*{argmax}_{y \in \Delta^{m_i}} Q(\mu_{1\cdot}^*, \ldots, \mu_{i-1\cdot}^*, y, \mu_{i+1\cdot}^*, \ldots, \mu_{l\cdot}^*) .$$

For $i \in [l]$ denote the function

$$\begin{aligned} q_{i,\mu} : \Delta^{m_i} &\to \mathbb{R} \\ y &\mapsto Q(\mu_{1\cdot}, \ldots, \mu_{i-1\cdot}, y, \mu_{i+1\cdot}, \ldots, \mu_{l\cdot}) . \end{aligned}$$

Then, for $j \in [m_i + 1]$, we have

$$\nabla_j\, q_{i,\mu}(y) = \nabla_{ij}\, Q(\mu_{1\cdot}, \ldots, \mu_{i-1\cdot}, y, \mu_{i+1\cdot}, \ldots, \mu_{l\cdot})\,.$$

By assumption on μ, each $\mu_{i\cdot}$ is a solution to the maximum problem

$$\max_{y \in \Delta^{m_i}} \; q_{i,\mu}(y)\,.$$

Global optimum $\mu_{i\cdot}$ is therefore in particular a KKT point of that problem. So there exist KKT multipliers $\{\tau_{ij} \in \mathbb{R}\}_{j=1}^{m_i+1}$ and κ_i such that

$$\begin{aligned}
&\mu_{ij} \geq 0 \quad \forall j \in [m_i + 1] \\
&\sum_{j=1}^{m_i+1} \mu_{ij} = 1 \\
&\tau_{ij} \geq 0 \quad \forall j \in [m_i + 1] \\
&\tau_{ij} = 0 \quad \text{or} \quad \mu_{ij} = 0 \quad \forall j \in [m_i + 1] \\
&\nabla_{ij}\, Q(\mu) = \nabla_j\, q_{i,\mu}(\mu_{i\cdot}) = -\tau_{ij} + \kappa_i \quad \forall j \in [m_i + 1]\,.
\end{aligned}$$

Since $i \in [l]$ was arbitrary, we can use these multipliers to obtain that μ satisfies KKT conditions (12).

Lastly, since there always exists an ex-ante optimal strategy (as a maximum (2) of a continuous polynomial function over a compact domain), there also always exists a (EDT,GDH)-Equilibrium and (CDT,GT)-Equilibrium. □

G.4 Proof of Theorem 2

Theorem. (CDT,GT)-EQUILIBRIUM *is CLS-complete. CLS-hardness holds even for game instances that*

(1.) have a tree depth of 5 *and the player has* 2 *actions per info set*

(2.) have no absentmindedness and a tree depth of 6

(3.) have no chance nodes, a tree depth of 5*, and only one info set*

We prove this result in four parts.

CLS Membership of (CDT,GT)-EQUILIBRIUM

By Theorem 1, (CDT,GT)-Equilibria of Γ coincide with the KKT points of the strategy utility function Q of Γ. Finding an ϵ-KKT point of a polynomial function over a compact domain was shown to be CLS-complete by Fearnley *et al.* [2021]. It is easy to see that the following can be constructed in polynomial time

1. describing the domain condition $\mu \in \times_{i=1}^{l} \Delta^{m_i}$ as a linear inequality $A\mu \leq b$ for when μ is viewed as a flattened vector
2. the polynomial Q
3. its derivative ∇Q
4. a Lipschitz constant L for Q and ∇Q

First CLS Hardness Result of (CDT,GT)-EQUILIBRIUM

We will derive our first CLS hardness result from a KKT problem studied by Babichenko and Rubinstein [2021]. Consider the maximization of a polynomial function $p : \mathbb{R}^l \to \mathbb{R}$ over the hypercube:

$$\begin{aligned}
\max_{x \in \mathbb{R}^l} \quad & p(x) \\
\text{s.t.} \quad & x \in [0,1]^l\,.
\end{aligned} \tag{13}$$

Here, p is again assumed to be represented in the Turing (bit) model $p(x) = \sum_{D \in \mathrm{MB}(d,l)} \lambda_D \cdot \prod_{i \in [l]} x_i^{D_i}$.

Definition 18. *An instance of the problem* KKTPOLYOVERCUBE *consists of a polynomial function* $p : \mathbb{R}^l \to \mathbb{R}$ *together with a precision value* $\epsilon > 0$. *A solution consists of a point* $x \in [0,1]^l$ *that is an* ϵ*-KKT point of the maximization problem (13):*

$$\begin{aligned}
\forall i \in [l] : \quad & x_i > 0 \implies \nabla_i\, p(x) \geq -\epsilon \\
\forall i \in [l] : \quad & x_i < 1 \implies \nabla_i\, p(x) \leq \epsilon
\end{aligned} \tag{14}$$

Babichenko and Rubinstein [2021] call this problem GD-FIXEDPOINT.

Lemma 19 (Babichenko and Rubinstein [2021])**.** KKTPOLYOVERCUBE *is CLS-complete. Hardness holds even*[3] *for polynomials in which every monomial has degree* 5*.*

We can now derive CLS-hardness of (CDT,GT)-EQUILIBRIUM by reducing from KKTPOLYOVERCUBE.

Take an instance $(p : \mathbb{R}^l \to \mathbb{R},\, \epsilon)$ of KKTPOLYOVERCUBE where p is known to only have degree 5 monomials. There are at most $\binom{l+5-1}{5} = \mathcal{O}(l^5)$ many such monomials. Consider the modified polynomial

$$\begin{aligned}
\hat{p} : \times_{i=1}^{l} \mathbb{R}^2 \quad & \to \mathbb{R} \\
\Big((x_{i1}, x_{i2})\Big)_{i=1}^{l} \quad & \mapsto p(x_{11}, x_{21}, \ldots, x_{l1})\,.
\end{aligned} \tag{15}$$

Create a game Γ out of $\hat{p}$ as described in Appendix A with the first variant, which is constructed in polynomial time in the encoding of p. Let $n := |\mathcal{H}|$ be the number of nodes in Γ. Set $\delta := \frac{\epsilon}{n}$. This is also constructable in polynomial time in the encoding of p and ϵ.

Recall from Appendix A that the strategy utility function Q of Γ has the property $Q(\mu) = \hat{p}(\mu) = p(\mu_{11}, \ldots, \mu_{l1})$ for any point $\mu \in \times_{i=1}^{l} \mathbb{R}^2$. Therefore, $\nabla_{i1}\, Q(\mu) = \nabla_i\, p(\mu_{11}, \ldots, \mu_{l1})$ and $\nabla_{i2}\, Q(\mu) = 0$ for all $\mu \in \times_{i=1}^{l} \mathbb{R}^2$ and $i \in [l]$. Finally, note that for any information set I of Γ and any strategy μ of Γ, we have $\mathrm{Fr}(I \mid \mu) = \sum_{h \in I} \mathbb{P}(h \mid \mu) \leq \sum_{h \in I} 1 \leq n$.

Let $\mu^* = (\mu^*_{ij})_{i,j=1}^{l,2} \in \times_{i=1}^{l} \Delta^1$ be a a solution to the (CDT,GT)-EQUILIBRIUM-instance (Γ, δ). Then we can show that the point $x^* := (\mu^*_{i1})_{i=1}^{l} \in [0,1]^l$ is a solution to the KKTPOLYOVERCUBE-instance $(p : \mathbb{R}^l \to \mathbb{R},\, \epsilon)$.

Let us first consider any index $i \in [l]$ with $\mathrm{Fr}(I_i \mid \mu^*) = 0$. Then, by Lemma 13, we get $0 = \nabla_{i1}\, Q(\mu^*) = \nabla_i\, p(x^*)$. Thus, such indices i always satisfy the conditions (14) independent of the value x^*_i.

Now consider an index $i \in [l]$ with $\mathrm{Fr}(I_i \mid \mu^*) > 0$ and $0 < x^*_i = \mu^*_{i1}$. Then, due to μ^* being a solution to the (CDT,GT)-EQUILIBRIUM-instance, we get

$$\begin{aligned}
& \mathrm{EU}_{\mathrm{CDT,GT}}(a_1 \mid \mu^*, I_i) \geq \max_{j=1,2} \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu^*, I_i) - \delta \\
\iff\; & \mathrm{EU}_{\mathrm{CDT,GT}}(a_1 \mid \mu^*, I_i) \geq \mathrm{EU}_{\mathrm{CDT,GT}}(a_2 \mid \mu^*, I_i) - \delta
\end{aligned}$$

[3] Hardness even holds if each summand $\lambda_D \cdot \prod_{i \in [l]} x_i^{D_i}$ for $\sum_{D \in \mathrm{MB}(5,l)}$ has degree 5 and is component-wise concave (and therefore the polynomial is also component-wise concave).

With Lemma 13, we can therefore derive

$$\begin{aligned}
&\nabla_i\, p(x^*) = \nabla_{i1}\, Q(\mu^*) \\
&= \mathrm{Fr}(I_i \mid \mu^*) \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_1 \mid \mu^*, I_i) \\
&\geq \mathrm{Fr}(I_i \mid \mu^*) \cdot \Big(\mathrm{EU}_{\mathrm{CDT,GT}}(a_2 \mid \mu^*, I_i) - \delta\Big) \\
&= \mathrm{Fr}(I_i \mid \mu^*) \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_2 \mid \mu^*, I_i) - \mathrm{Fr}(I_i \mid \mu^*) \cdot \delta \\
&= \nabla_{i2}\, Q(\mu) - \mathrm{Fr}(I_i \mid \mu^*) \cdot \delta = -\mathrm{Fr}(I_i \mid \mu^*) \cdot \delta \\
&\geq n \cdot \delta = -\epsilon
\end{aligned}$$

Now consider an index $i \in [l]$ with $\mathrm{Fr}(I_i \mid \mu^*) > 0$ and $1 > x_i^* = \mu_{i1}^* = 1 - \mu_{i2}^*$, i.e., $\mu_{i2}^* > 0$. Then, due to μ^* being a solution to the (CDT,GT)-EQUILIBRIUM-instance, we get

$$\begin{aligned}
&\mathrm{EU}_{\mathrm{CDT,GT}}(a_2 \mid \mu^*, I_i) \geq \max_{j=1,2} \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu^*, I_i) - \delta \\
&\iff \mathrm{EU}_{\mathrm{CDT,GT}}(a_1 \mid \mu^*, I_i) \leq \mathrm{EU}_{\mathrm{CDT,GT}}(a_2 \mid \mu^*, I_i) + \delta
\end{aligned}$$

Similarly to the other case, we can therefore derive

$$\begin{aligned}
&\nabla_i\, p(x^*) = \nabla_{i1}\, Q(\mu^*) \\
&= \mathrm{Fr}(I_i \mid \mu^*) \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(a_1 \mid \mu^*, I_i) \\
&\leq \mathrm{Fr}(I_i \mid \mu^*) \cdot \Big(\mathrm{EU}_{\mathrm{CDT,GT}}(a_2 \mid \mu^*, I_i) + \delta\Big) \\
&= \nabla_{i2}\, Q(\mu) + \mathrm{Fr}(I_i \mid \mu^*) \cdot \delta = \mathrm{Fr}(I_i \mid \mu^*) \cdot \delta \\
&\leq n \cdot \delta = \epsilon
\end{aligned}$$

Therefore, all in all, point x^* makes a solution to the KKTPOLYOVERCUBE-instance ($p : \mathbb{R}^l \to \mathbb{R}, \epsilon$). This finishes the polynomial time reduction from search problem of DEGREE-5-KKTPOLYOVERCUBE to search problem (CDT,GT)-EQUILIBRIUM.

Note that in the above reduction, the constructed game Γ has a tree depth of 6. The root is a chance node with a number of outgoing edges that equals the number of monomials in p. Any other node of Γ will have two outgoing actions. Therefore, CLS-hardness of (CDT,GT)-EQUILIBRIUM remains even if the game instance has a tree depth of 6 and the player only has 2 actions per info set.

Second CLS Hardness Result of (CDT,GT)-EQUILIBRIUM

The second and third CLS hardness results will both rely on the following game-theoretic problem studied by Babichenko and Rubinstein [2021] (again).

Let G be a n-player simultaneous game with action sets $\{A_i\}_{i\in[n]}$ of size $m_i := |A_i|$ and with utility functions $U_i : \times_{i'=1}^{n} \Delta(A_{i'}) \to \mathbb{R}$ for each player i. A mixed action profile $x = (x_{ij})_{i,j=1}^{n,m_i} \in \times_{i=1}^{n} \Delta^{m_i}$ is called an ϵ-Nash equilibrium of G if for every player $i \in [n]$ and every action $a_j \in A_i$ of player i, we have $U_i(x) \geq U_i(x_{1\cdot}, \dots, x_{i-1\cdot}, a_j, x_{i+1\cdot}, \dots, x_{n\cdot}) - \epsilon$. In standard game theory notation, this would be phrased as $U_i(x) \geq U_i(a_j, x_{-i\cdot}) - \epsilon$. For us, it will be helpful to use the notation of a EDT deviation, in which this condition becomes $U_i(x) \geq U_i(x_{I_i \mapsto a_j}) - \epsilon$.

A c-polytensor game is a multiplayer simultaneous game presented in terms of payoff tables, one for each subset of c players. Given a pure-strategy action profile of all players, a player's payoff consists of the sum of payoffs they get from the payoff tables of the c-subsets they belong to. For any constant c, such games with $n \geq c$ players and up to m pure actions per player have a polynomial-sized representation in n and m because there are $\leq n \cdot \binom{n}{c} \cdot m^c = \mathcal{O}(n^{c+1} \cdot m^5)$ many payoff entries overall. A c-polytensor identical interest game is a c-polytensor game in which every subgame associated with a c-subset yields the participating player the same utility.

Definition 20. *An instance of the problem c-POLYTENSOR-IDENTICALINTEREST consists of a c-polytensor identical interest game together with a precision value $\epsilon > 0$. A solution consist of an ϵ-Nash equilibrium of the game.*

Lemma 21 (Babichenko and Rubinstein [2021])**.** *5-POLYTENSOR-IDENTICALINTEREST is CLS-complete.*

By shifting and rescaling utilities, we may assume (without changing the complexity) that the game instances to 5-POLYTENSOR-IDENTICALINTEREST have payoffs in the interval $[0, 1]$.

We can now proof the second CLS hardness of (CDT,GT)-EQUILIBRIUM by reducing from 5-POLYTENSOR-IDENTICALINTEREST.

Take an instance (G, ϵ) of we 5-POLYTENSOR-IDENTICALINTEREST. Let n be the number of players of G, and m_i be the number of pure strategies of player i. Denote $m := \max_i m_i$. Then, each subset of 5 players of G has a table of up to m^5 payoffs in $[0, 1]$ one for each of their possible pure profiles. Moreover, there will be $\binom{n}{5}$ many subsets of 5 players among n players. Collect all those subsets to $\Lambda := \{X \subset [n] : |X| = 5\}$. If $X \in \Lambda$ is a 5-subset and the player of X play strategy profile $\mu_X \in \times_{i\in X} \Delta^{m_i}$, then denote the payoff that these player get in that subgame by $u^X(\mu_X)$. Then, the overall utility function U_i of a player $i \in [n]$ takes as input a strategy profile μ for the game Γ and returns the payoff:

$$U_i(\mu) = \sum_{X \in \Lambda:\, i \in X} u^X(\mu_X)$$

Construct an instance (Γ, δ) of (CDT,GT)-EQUILIBRIUM as follows. There will be n info sets $I_1, \dots, I_n$. The root of the tree of Γ is a chance node having $\binom{n}{5}$ subtrees, one for each set X of 5 players of G. The actions towards these subtrees are chosen uniform randomly, that is, with the same probability $1/\binom{n}{5}$. Consider a subtree T_X associated with a set $X = \{i_1, \dots, i_5\} \in \Lambda$. Then T_X shall have depth 5 with depth layer $k-1$ for $k \in [5]$ corresponding to player i_k. Every node of T_X of depth k_1 shall be assigned to info set I_{i_k} and have m_{i_k} outgoing edges labelled by the pure strategies of player i_k. Nodes of depth 5 shall be terminal nodes. If terminal node z of Γ has action history $(X, j_1, \dots, j_5)$, then it shall yield a payoff equal to $u^X(j_1, \dots, j_5)$, that is, the payoff of the subgame associated with X from the pure action profile $(j_1, \dots, j_5)$. Define the precision parameter as $\delta := \epsilon / \binom{n-1}{4}$. The instance (Γ, δ) can be constructed in polynomial time in the encoding of (G, ϵ). Note that the game Γ has tree depth 6 and no absentmindedness.

Take an info set I_i of Γ. The player reaches info set I_i with probability one in every subtree T_X with $i \in X$, independent of her strategy choice μ. Thus, the overall reach probability of each info set I_i of Γ is exactly the number subsets $X \in \Lambda$ that contain i divided by the number of 5-subsets overall, which is $\binom{n-1}{4}/\binom{n}{5} = \frac{5}{n}$. In particular, this reach probability is non-zero, so an approximate (CDT,GT)-Equilibrium must be approximately optimal in every information set I_i of Γ.

Since there is no absentmindedness in Γ, we get by Appendix D for all info sets I_i, all strategies μ, and all mixed actions $\alpha \in \Delta(A_{I_i})$:

$$\begin{aligned}
&\mathrm{EU}_{\mathrm{CDT,GT}}(\alpha \mid \mu, I_i) \\
&= \mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha \mid \mu, I_i) \\
&\overset{(10)}{=} \frac{1}{\mathbb{P}(I_i \mid \mu_{I_i \mapsto \alpha})} \cdot \sum_{\substack{z \in \mathcal{Z} \\ (I_i \text{ in hist}(z))}} \mathbb{P}(z \mid \mu_{I_i \mapsto \alpha}) \cdot u(z) \\
&= \frac{n}{5} \cdot \sum_{X \in \Lambda : \, i \in X} \sum_{z \in T_X} \mathbb{P}(z \mid \mu_{I_i \mapsto \alpha}) \cdot u(z) \\
&= \frac{n}{5} \cdot \sum_{X \in \Lambda : \, i \in X} \sum_{\mathbf{j} \in \times_{i' \in X} A_{I_{i'}}} \mathbb{P}(z_{X,\mathbf{j}} \mid \mu_{I_i \mapsto \alpha}) \cdot u^X(\mathbf{j}) \\
&\overset{(*)}{=} \frac{n}{5} \cdot \sum_{X \in \Lambda : \, i \in X} \frac{1}{\binom{n}{5}} \cdot u^X\Big((\mu_{I_i \mapsto \alpha})_X\Big) \\
&= \frac{n}{5} \cdot \frac{1}{\binom{n}{5}} \cdot U_i(\mu_{I_i \mapsto \alpha}) \\
&= \frac{1}{\binom{n-1}{4}} \cdot U_i(\mu_{I_i \mapsto \alpha})
\end{aligned}$$

where in $(*)$ we used that for any $X = \{i_1, \dots, i_5\} \in \Lambda$ and strategy μ that

$$\begin{aligned}
u^X(\mu_X) &= \sum_{(j_k)_{k=1}^5 \in \times_{k=1}^5 A_{I_{i_k}}} u^X(j_1, \dots, j_5) \cdot \prod_{k=1}^{5} \mu_X(j_k \mid I_{i_k}) \\
&= \sum_{\mathbf{j} \in \times_{i' \in X} A_{I_{i'}}} \binom{n}{5} \cdot \mathbb{P}(z_{X,\mathbf{j}} \mid \mu_{I_i \mapsto \alpha}) \cdot u^X(\mathbf{j}) \, .
\end{aligned}$$

Let $\mu^* \in \times_{i=1}^{l} \Delta^{m_i}$ be a solution to the (CDT,GT)-EQUILIBRIUM-instance (Γ, δ). Then we can show that the mixed action profile $(\mu^*_{i\cdot})_{i=1}^n$ is a solution to 5-POLYTENSOR-IDENTICALINTEREST instance (G, ϵ), that is, an ϵ-Nash equilibrium of G. Note that $\mu^*_{I_i \mapsto \mu^*_{i\cdot}} = \mu^*$. We obtain for all player $i \in [n]$ and all pure actions $j \in [m_i]$ of player i:

$$\begin{aligned}
&U_i(\mu^*) = \\
&= \frac{\binom{n-1}{4}}{\binom{n-1}{4}} \cdot U_i(\mu^*_{I_i \mapsto \mu^*_{i\cdot}}) \\
&= \binom{n-1}{4} \cdot \mathrm{EU}_{\mathrm{CDT,GT}}(\mu^*_{i\cdot} \mid \mu^*, I_i) \\
&\geq \binom{n-1}{4} \cdot \Big[\mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu^*, I_i) - \delta\Big] \\
&= \frac{\binom{n-1}{4}}{\binom{n-1}{4}} \cdot U_i(\mu^*_{I_i \mapsto a_j}) - \binom{n-1}{4} \cdot \delta \\
&= U_i(\mu^*_{I_i \mapsto a_j}) + \epsilon \, .
\end{aligned}$$

Therefore, a δ-(CDT,GT)-Equilibrium in Γ makes an ϵ-Nash equilibrium in G.

Third CLS Hardness Result of (CDT,GT)-EQUILIBRIUM We proof the third CLS hardness of (CDT,GT)-EQUILIBRIUM by reducing from 5-POLYTENSOR-IDENTICALINTEREST again.

Take an instance (G, ϵ_N) of we 5-POLYTENSOR-IDENTICALINTEREST. Use the same notation as in the last reduction proof from 5-POLYTENSOR-IDENTICALINTEREST. In particular, G has n players with up to m actions (for simplicity, extend the action set of every player to exactly m actions). They are assumed to lie in $[0, 1]$ again. Let M be an exponentially-large quantity (this will then still be polynomially large in the size of $1/\epsilon_N$).

Construct an instance (Γ, δ) of (CDT,GT)-EQUILIBRIUM as follows. Γ is a game tree of depth 5 and all nodes of Γ will be player nodes that belong to the same info set I. Starting from the root node, each node of of depth ≤ 4 shall have $n \cdot m$ outgoing edges, each associated with a (player, action)-pair $(i, j) \in [n] \times [m]$ of G. At depth 5, we have $(n \cdot m)^5$ many nodes that shall be terminal nodes. If the path to a terminal node z is $((i_1, j_1), \dots, (i_5, j_5))$, and if $k_z \in [5]$ is the number of distinct values in $\{i_1, \dots, i_5\}$, then z shall yield a payoff of

1. $k_z \cdot M$ if $k_z \leq 4$, and
2. $k_z \cdot M + u^{\{i_1, \dots, i_5\}}(j_1, \dots, j_5)$ if $k_z = 5$.

Finally, set $\delta := \epsilon_N$.

Notice that the exponential value $k_z \cdot M$ dominates $u^{\{i_1, \dots, i_5\}}(j_1, \dots, j_5) \in [0, 1]$. Moreover, no matter what strategy μ the player chooses, we have $\mathrm{Fr}(I \mid \mu) = 5$ and $\mathbb{P}_{GT}(h \mid \mu, I) = \frac{1}{5} \cdot \mathbb{P}(h \mid \mu)$ for all $h \in \mathcal{H} \setminus \mathcal{Z}$.

An ex-ante strategy $p = (p_{i,j})_{i,j=1}^{n,m} \in \Delta^{n \cdot m}$ of Γ shall correspond to the following strategy profile of G: Denote $p_i = \sum_j p_{i,j}$. Then, player $i \in [n]$ of G shall play action $j \in [m]$, with probability $p_{i,j}/p_i$.

We analyse what values $(p_{i,j})_{i,j}$ can take in a solution. We start by proving that p_i is very close to $1/n$.

Say that the player of Γ arrives in info set I and considers her best action $\alpha \in \Delta^{n \cdot m}$ to deviate to in a (CDT,GT) fashion. Then, she can split the task into choosing the values $(\alpha_i)_i$ that sum up to 1 followed by partitioning them into $(\alpha_{i,j})_{i,j}$ such that for each i, $(\alpha_{i,j})_j$ sum up to α_i. The choice of values $(\alpha_i)_i$ determines the agent's expected payoff from the values kM in the utility function. The choice of how to partition into $(\alpha_{i,j})_{i,j}$ determines the payoffs from 5-subset subgames of G but it does not affect the payoffs associated with the $(\alpha_i)_i$.

Suppose that p is an δ-(CDT,GT)-Equilibrium and that $\max_i p_i$ exceeds $\min_i p_i$ by more than δ. Then, upon arriving in info set I, the agent's best response is to allocate zero probability to any action (i, j) where $i \in \arg\max p_i$. This is because the agent aims to maximise the number of distinct values $i_1, \dots, i_5$ in the history of actions $(i_1, j_1), \dots, (i_5, j_5)$ because the values u^X are inverse exponentially smaller than the values $k \cdot M$. On the other hand, the agent should put weight on actions (i, j) for which p_i is small which will yield her an utility increase of $\geq \delta$. This contradicts the assumption that p is an δ-(CDT,GT)-Equilibrium. So assume henceforth that the values $(p_i)_i$ are all within ϵ_N of each other.

Aim: the values $p_{i,j}/p_i$ represent $\mathbb{P}(i \text{ plays } j)$ in a ϵ_N approximate solution of G.

The task of partitioning each p_i into $p_{i,j}$ affects the part of the agent's payoff that is attributable to the polytensor game payoffs from G.

A δ-(CDT,GT)-Equilibrium to Γ should be an δ-best response to itself once arriving in I. Hence, every action (i, j) with positive probability $p_{i,j}$ yields a payoff of at most δ less than any other action (i, j'). Any difference between their payoffs is associated with the payoffs in the 5-polytensor game G, and not with the kM portions of these payoffs.

The part of the payoff of action (i_1, j_1) that comes from the payoffs in G is

$$\sum_{i_2,i_3,i_4,i_5} u(i_{[5]}; j_{[5]}) \prod_{k=2}^{5} p_{i_k,j_k} \tag{16}$$

Using the fact that values p_i differ from $1/n$ by an inverse-exponential multiplicative factor from $1/n$, (16) differs by an inverse-exponential multiplicative factor from

$$n^4 \sum_{i_2,i_3,i_4,i_5} u(i_{[5]}; j_{[5]}) \prod_{k=2}^{5} p_{i_k,j_k}/p_{i_k} \tag{17}$$

So the values $p_{i,j}/p_i$ have to correspond to probabilities $\mathbb{P}(i \text{ plays } j)$ in an approximate solution of G.

G.5 Proof of Theorem 4

Theorem 4. *The following problems are all NP-hard. Unless NP = ZPP, there is also no FPTAS for these problems.*

- *(1a.) Given Γ and $t \in \mathbb{Q}$, is there an (CDT,GT)-Equilibrium of Γ with ex-ante utility $\geq t$?*
- *(1b.) Given Γ, an info set I of Γ and $t \in \mathbb{Q}$, is there a (CDT,GT)-Equilibrium μ such that $\mathrm{Fr}(I \mid \mu) > 0$, and such that the player has a (CDT,GT)-expected utility $\geq t$ upon reaching I?*
- *(1c.) Given Γ, an info set I of Γ and $t \in \mathbb{Q}$, is there a strategy μ of Γ such that $\mathrm{Fr}(I \mid \mu) > 0$, and such that the player has a (CDT,GT)-expected utility $\geq t$ upon reaching I?*
- *(2a.) Given Γ and $t \in \mathbb{Q}$, is there an (EDT,GDH)-Equilibrium of Γ with ex-ante utility $\geq t$?*
- *(2b.) Given Γ, an info set I of Γ and $t \in \mathbb{Q}$, is there a (EDT,GDH)-Equilibrium μ such that $\mathbb{P}(I \mid \mu) > 0$, and such that the player has a (EDT,GDH)-expected utility $\geq t$ upon reaching I?*
- *(2c.) Given Γ, an info set I of Γ and $t \in \mathbb{Q}$, is there a strategy μ of Γ such that $\mathbb{P}(I \mid \mu) > 0$, and such that the player has a (EDT,GDH)-expected utility $\geq t$ upon reaching I?*
- *(3a.) Given Γ and $t \in \mathbb{Q}$, do all (EDT,GDH)-Equilibria of Γ have ex-ante utility $\geq t$?*
- *(3b.) Given Γ, an info set I of Γ and $t \in \mathbb{Q}$, do all (EDT,GDH)-Equilibria μ with $\mathbb{P}(I \mid \mu) > 0$ promise the player a (EDT,GDH)-expected utility $\geq t$ upon reaching I?*

We reduce all those decision problems from the problem in Proposition 4, which we will henceforth call ExAnteOpt-D.

First note that any single-player extensive-form game with imperfect recall Γ has an ex-ante optimal strategy since the maximization problem (2) is about a continuous polynomial function over a compact domain. Moreover, observe that (Γ, t) is a yes instance for ExAnteOpt-D (by definition) if and only if there is a strategy μ in Γ with $Q(\mu) \geq t$ if and only if there is an ex-ante optimal strategy μ^* in Γ with $Q(\mu^*) \geq t$.

(1a.) and (2a.):
Let (Γ, t) be an instance for ExAnteOpt-D. Without transforming, we can choose (Γ, t) as the instance to the problems (1a.) and (2a.). Then, it will be a yes instance of (-a.) if and only if there is a (CDT,GT)-Equilbrium or, resp., (EDT,GDH)-Equilibrium in Γ with ex-ante utility $\geq t$ if and only if (by Lemma 16) there is an ex-ante optimal strategy in Γ with with ex-ante utility $\geq t$ if and only if (by the comment above) (Γ, t) is a yes instance for ExAnteOpt-D.

(1b.), (1c.), (2b.), and (2c.):
Let (Γ, t) be an instance for ExAnteOpt-D. Let us refer to the root of Γ with h_0. We construct a new game Γ' by adding a new artificial game start: Copy the game tree of Γ. Add a new node h_{-1} to it which shall represent root of Γ'. Connect h_{-1} to h_0 by one edge, called action a_{-1}), and assign h_{-1} to a new information I_{-1} that is added to $\mathcal{I}_*$. The corresponding instance shall be (Γ', I_{-1}, t).

Observe that any strategy μ for Γ corresponds to a strategy μ' in Γ' that behaves like μ at info sets I that were inherited from Γ and that takes the only viable action a_{-1} at info set I_{-1} with a 100% certainty. Moreover, info set I_{-1} has a visit frequency and reach probability of 1 for any strategy μ' of Γ' because they occur in the history of any terminal node exactly once in the beginning. Most crucially, we also get that the (CDT,GT)-expected utility (resp. (EDT,GDH)-expected utility) of μ' upon reaching info set I_{-1} is equal to the ex-ante expected utility of μ' in Γ' and of corresponding strategy μ in Γ.

We obtain:

- (Γ', I_{-1}, t) is a yes instance for problem (1b.) (resp. (2b.))

$\implies$ (Γ', I_{-1}, t) it is a yes instance for problem (1c.) (resp. (2c.))

$\implies$ there is a strategy μ' in Γ' with (CDT,GT)-expected utility (resp. (EDT,GDH)-expected utility) $\geq t$ upon reaching I_{-1}

$\implies$ corresponding strategy μ in Γ has ex-ante utility $\geq t$

$\implies$ (Γ, t) is a yes instance for ExAnteOpt-D.

and for the other direction:

- (Γ, t) is a yes instance for ExAnteOpt-D

$\implies$ there exists an ex-ante optimal strategy μ^* of (Γ, t) with ex-ante utility $\geq t$

$\implies$ corresponding strategy $(\mu^*)'$ in Γ' is ex-ante optimal for Γ', and it has (CDT,GT)-expected utility (resp. (EDT,GDH)-expected utility) $\geq t$ upon reaching I_{-1}

$\implies$ there exists a (CDT,GT)-Equilibrium in Γ' (resp. (EDT,GDH)-Equilibrium) with (CDT,GT)-expected utility (resp. (EDT,GDH)-expected utility) $\geq t$ upon reaching I_{-1}

$\implies$ (Γ', I_{-1}, t) is a yes instance for problem (1b.) (resp. (2b.))

$\implies$ (Γ', I_{-1}, t) is a yes instance for problem (1c.) (resp. (2c.)).

which completes the reduction. Note that G' has one additional info set and tree depth in comparison to G.

(3a.) and (3b.):

For (3a.), consider the reductions for (2a.) again and assume the starting instance (Γ, t) of ExAnteOpt-D has one info set only. By Proposition 4 we know that even with such instances ExAnteOpt-D is NP-hard and conditionally inapproximable. But an (EDT,GDH)-Equilibrium of such an instance must be an ex-ante optimal strategy (seen by Lemma 14). Therefore, each (EDT,GDH)-Equilibrium of Γ must have the same (ex-ante optimal) value in ex-ante utility. Thus, deciding whether all (EDT,GDH)-Equilibrium exceeds a target ex-ante utility (problem (3a.)) coincides with deciding whether one (EDT,GDH)-Equilibrium does that (problem (2a.)).

An analogous argument holds for (3b.) by considering the reductions for (2b.) for starting instances (Γ, t) with only one info set. There, we observe that (EDT,GDH)-Equilibria of Γ' correspond exactly to the ex-ante optimal strategies in Γ, and must therefore all promise the same (EDT,GDH)-expected utility upon reaching I_{-1}.

References

[Babichenko and Rubinstein, 2021] Yakov Babichenko and Aviad Rubinstein. Settling the complexity of nash equilibrium in congestion games. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1426–1437. ACM, 2021.

[Briggs, 2010] Rachael Briggs. Putting a value on Beauty. In Tamar Szabó Gendler and John Hawthorne, editor, *Oxford Studies in Epistemology: Volume 3*, pages 3–34. Oxford University Press, 2010.

[Camerer, 2003] Colin F. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, 2003.

[Chu and Halpern, 2001] Francis C. Chu and Joseph Y. Halpern. On the np-completeness of finding an optimal strategy in games with common payoffs. *Int. J. Game Theory*, 30(1):99–106, 2001.

[Conitzer, 2015] Vincent Conitzer. A Dutch book against sleeping beauties who are evidential decision theorists. *Synthese*, 192(9):2887–2899, 2015.

[Daskalakis and Papadimitriou, 2011] Constantinos Daskalakis and Christos Papadimitriou. Continuous local search. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, page 790–804, USA, 2011. Society for Industrial and Applied Mathematics.

[de Klerk, 2008] Etienne de Klerk. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *Central Eur. J. Oper. Res.*, 16(2):111–125, 2008.

[Draper and Pust, 2008] Kai Draper and Joel Pust. Diachronic Dutch Books and Sleeping Beauty. *Synthese*, 164(2):281–287, 2008.

[Elga, 2000] Adam Elga. Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60(2):143–147, 2000.

[Fearnley *et al.*, 2021] John Fearnley, Paul W. Goldberg, Alexandros Hollender, and Rahul Savani. The complexity of gradient descent: CLS = PPAD ∩ PLS. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 46–59. ACM, 2021.

[Fudenberg and Tirole, 1991] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, October 1991.

[Håstad, 1997] Johan Håstad. Clique is hard to approximate within $\mathrm{n}^{\text{1-epsilon}}$. *Electron. Colloquium Comput. Complex.*, TR97-038, 1997.

[Hitchcock, 2004] Christopher Hitchcock. Beauty and the bets. *Synthese*, 139(3):405–420, 2004.

[Hubácek and Yogev, 2017] Pavel Hubácek and Eylon Yogev. Hardness of continuous local search: Query complexity and cryptographic lower bounds. In Philip N. Klein, editor, *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 1352–1371. SIAM, 2017.

[Koller and Megiddo, 1992] Daphne Koller and Nimrod Megiddo. The complexity of two-person zero-sum games in extensive form. *Games and Economic Behavior*, 4(4):528–552, 1992.

[Kroer and Sandholm, 2016] Christian Kroer and Tuomas Sandholm. Imperfect-recall abstractions with bounds in games. In *Proceedings of the Seventeenth ACM Conference on Economics and Computation (EC)*, pages 459–476, Maastricht, the Netherlands, 2016.

[Lanctot *et al.*, 2012] Marc Lanctot, Richard G. Gibson, Neil Burch, and Michael Bowling. No-regret learning in extensive-form games with imperfect recall. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, Edinburgh, Scotland, UK, 2012.

[Lang, 1994] S. Lang. *Algebraic Number Theory*. Graduate Texts in Mathematics. Springer, 1994.

[Megiddo and Papadimitriou, 1991] Nimrod Megiddo and Christos H. Papadimitriou. On total functions, existence theorems and computational complexity. *Theor. Comput. Sci.*, 81(2):317–324, apr 1991.

[Nisan *et al.*, 2007] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.

[Oesterheld and Conitzer, 2022] Caspar Oesterheld and Vincent Conitzer. Can *de se* choice be *ex ante* reasonable in games of imperfect recall? https://www.andrew.cmu.edu/user/coesterh/DeSeVsExAnte.pdf, 2022. Working paper. Accessed: 2022-12-14.

[Piccione and Rubinstein, 1997] Michele Piccione and Ariel Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20(1):3–24, 1997.

[Waugh *et al.*, 2009] Kevin Waugh, Martin Zinkevich, Michael Johanson, Morgan Kan, David Schnizlein, and Michael H. Bowling. A practical use of imperfect recall. In *Eighth Symposium on Abstraction, Reformulation, and Approximation (SARA-09)*, Lake Arrowhead, CA, USA, 2009.