# The Computational Complexity of Single-Player Imperfect-Recall Games[*]

**Emanuel Tewolde**[1] , **Caspar Oesterheld**[1] , **Vincent Conitzer**[1,2] and **Paul W. Goldberg**[2]

[1] Foundations of Cooperative AI Lab (FOCAL), Carnegie Mellon University
[2] University of Oxford

emanueltewolde@cmu.edu, oesterheld@cmu.edu, conitzer@cs.cmu.edu, paul.goldberg@cs.ox.ac.uk

## 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 alternative solution concepts 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 relate those three solution concepts of a game to solution concepts of a polynomial maximization problem: 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 such strategies. For ex-ante optimality and (EDT,GDH)-equilibria, we obtain NP-hardness and inapproximability, and for (CDT,GT)-equilibria we obtain CLS-completeness results.

## 1 Introduction

Most formalisms that are used for reasoning under uncertainty and for 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 various desirable 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 use 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. Conitzer and Oesterheld [2023] give the example of an AI driver assistant that can take over whenever the human car driver makes a major error. When that happens, the AI needs to reason about how good the human driver is in general, about whom it is not allowed to store information. 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 [Kovarik *et al.*, 2023]. 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.

Notably, we need to model this phenomenon as imperfect recall rather than merely as imperfect information. In single-agent *perfect-recall* imperfect-information games (say, a POMDP), there is never a (strict) reason to randomize, whereas in imperfect-recall games the agent might have to

randomize in order to perform well overall; cf. the Absent-minded Driver example [Piccione and Rubinstein, 1997] (Appendix C). For example, suppose we deploy a content recommendation system to many people's phones, in an edge-computing sort of setup: We are not in constant communication with the phones, so the nodes of our system have to act independently each day before getting back in touch with us. Over the next day, we would like to experiment (in an optimal way) what kind of content to recommend. With a pure strategy, we would show all users the same content and learn very little from it. Instead, we would prefer to randomize the content shown on each phone, that is, use a mixed strategy. Therefore, this situation cannot be a perfect-recall game (even of imperfect information).

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 becomes 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] (Appendix C) 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 $1/3$, and others (Halvers) believe it is $1/2$, 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]; here we build on that recent conceptual work 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, game-theoretic phenomena such as notions of equilibrium naturally come up in the presence of imperfect recall even when there is just a single agent. Intuitively that is because it is more challenging for that agent to coordinate its actions with those it takes at other times. Moreover, 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: (1) *ex ante* optimality, where the behavior strategy is one that maximizes expected utility at the outset; (2) 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 (3) 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. This makes gradient descent methods applicable to the computation of such an equilibrium, and, relatedly, we derive that problems of finding such an equilibrium – up to an inverse exponential precision – are complete for the class CLS (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. Naturally, all these complexity results also have implications for learning or dynamics that converge to these solutions.

## 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 *(denoted $\Gamma$), sometimes also called an extensive decision problem with imperfect recall, consists of:*

1. *A rooted tree, with nodes $\mathcal{H}$ and 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 $h \in \mathcal{H} \setminus \mathcal{Z}$.*

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 the player's* decision nodes *$\mathcal{H}_*$ and a set of* chance nodes *$\mathcal{H}_c$. This partition indicates whether the single player or exogenous stochasticity determines the action at any given 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 $\mathcal{H}_* = \sqcup_{I \in \mathcal{I}_*} I$ of the player's decision nodes into* information sets *("info sets" for short). We require $A_h = A_{h'}$ for any nodes $h, h'$ in the same info set $I$.*

Throughout this paper, we let $\ell := |\mathcal{I}_*|$ denote the number of info sets. For computational purposes, we assume that a game $\Gamma$ is represented by its game tree structure of size $\Theta(|\mathcal{H}|)$ (which includes the info set partition), and by a binary encoding of its chance node probabilities and its utility payoffs. The last two shall take on rational values only.

Any node $h \in \mathcal{H}$ uniquely corresponds to a (node,action)-history $\mathrm{hist}(h)$ from root $h_0$ to $h$ in the game tree. Define functions $d$, $\nu$, $a$ such that the
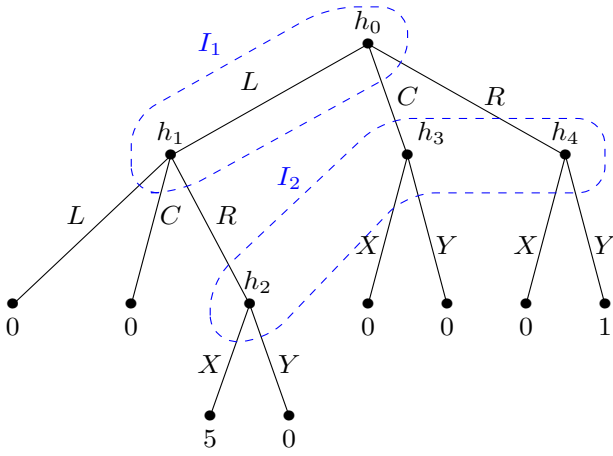
Figure 1: Running example of a single-player extensive-form game with imperfect recall. It has two info sets $I_1$ and $I_2$, and five nonterminal nodes.

node history and action history from $h_0$ to $h$ consist of the sequences $\big(\nu(h,0), \nu(h,1), \ldots, \nu(h,d(h))\big)$ and $\big(a(h,0), a(h,1), \ldots, a(h,d(h)-1)\big)$ respectively. In other words, function $d : \mathcal{H} \to \mathbb{N}_0$ identifies the tree depth of a node, $\nu : \mathcal{H} \times \mathbb{N}_0 \to \mathcal{H}$ the node ancestor at a specified depth, and $a : \mathcal{H} \times \mathbb{N}_0 \to \sqcup_{h \in \mathcal{H}} A_h$ the action ancestor at a specified depth. In particular, for all $h \in \mathcal{H}$, we have $\nu(h,0) = h_0$ and $\nu(h,d(h)) = h$. We restrict the domain of functions $\nu$ and $a$ to inputs $(h,k)$ with $k \le d(h)$ and $k \le d(h)-1$ respectively, and note that $a$ maps $(h,k)$ into $A_{\nu(h,k)}$.

The depth of $\Gamma$ is defined to be the maximal depth of the leaf nodes. For notational convenience, we add a singleton info set to $\Gamma$ for each chance node in $\Gamma$. 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) 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 in info set $I_1$ 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 Absentminded Driver from Piccione and Rubinstein [1997] (Appendix C).

Let $\Delta(A_I)$ denote the set of probability distributions over

the actions 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 $\mu$ of the player to info sets $I_h \in \mathcal{I}_c$ of chance nodes $h \in \mathcal{H}_c$ by setting $\mu(\cdot \mid I_h) := \mathbb{P}_c(\cdot \mid h)$ there.

Given that the player is currently at node $\bar{h} \in \mathcal{H} \setminus \mathcal{Z}$ and that she plays according to strategy $\mu$, we can calculate the probability of reaching node $h \in \mathcal{H}$ by multiplying the probabilities of the actions 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 $\Gamma$. Naturally, the reach probability of the root is 1.

The expected utility payoff for being at node $h \in \mathcal{H} \setminus \mathcal{Z}$ and using strategy $\mu$ from then on can be determined by $U(\mu \mid h) := \sum_{z \in \mathcal{Z}} \mathbb{P}(z \mid \mu, h) \cdot u(z)$. Furthermore, let $U : \mu \mapsto U(\mu) := U(\mu \mid h_0)$ be the function that takes a strategy $\mu$ of $\Gamma$ and returns the expected utility payoff of the player from following $\mu$ from the game start to termination. $U(\mu)$ is also called the *ex-ante* (expected) utility of $\mu$.

## 2.2 Utility as a Polynomial Function

Fix an ordering $I_1, \ldots, I_\ell$ of the info sets in $\mathcal{I}_*$ and denote $m_i := |A_{I_i}|$ for all $i \in [\ell]$. Moreover, fix an ordering $a_1, \ldots, a_{m_i}$ of the actions in $A_{I_i}$ for all $i \in [\ell]$.

We can uniquely describe a strategy $\mu$ of $\Gamma$ by the probability values that it assigns to each action $a_j$ at info set $I_i$, for $i \in [\ell]$ and $j \in [m_i]$. A strategy $\mu$ is a vector $\mu = (\mu_{ij})_{i,j} \in \times_{i=1}^{\ell} \mathbb{R}^{m_i}$ such that each subvector $\mu_{i\cdot} = (\mu_{ij})_j$ lies in the simplex $\Delta(A_{I_i}) \equiv \Delta^{m_i-1} := \{y \in \mathbb{R}^{m_i} : y_j \ge 0 \,\forall j, \sum_{j=1}^{m_i} y_j = 1\}$. Therefore, the strategy space of $\Gamma$ is $\times_{i=1}^{\ell} \Delta(A_{I_i}) \equiv \times_{i=1}^{\ell} \Delta^{m_i-1}$.

The expected utility function $U$ of a strategy $\mu$ can be fully written out as

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

As noted by Piccione and Rubinstein [1997], this is a polynomial function in the variables $(\mu_{ij})_{i,j}$. Recall that at chance nodes, the probabilities are exogenously fixed constants. Thus, the degree of the polynomial function $U$ is upper-bounded by the maximum number of times the player of $\Gamma$ might have to take a decision in order to reach a terminal node. Note that polynomial $U$ can be constructed in polynomial time in the encoding size of $\Gamma$.

**Example 2.** *In the game of Figure 1, we get $\ell = 2$, $m_1 = 3$, $m_2 = 2$. 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 $U(\mu) = 5\mu_{11}\mu_{13}\mu_{21} + \mu_{13}\mu_{22}$.*

We show in Appendix A that one can also reduce any multivariate polynomial $p : \bigtimes_{i=1}^{\ell} \mathbb{R}^{m_i} \to \mathbb{R}$ to a single-player extensive-form game $\Gamma$ with imperfect recall such that its expected utility function satisfies $U(\mu) = p(\mu)$ on $\bigtimes_{i=1}^{\ell} \mathbb{R}^{m_i}$.

## 2.3 (Computing) Ex-ante Optimal Strategies

Suppose we want to *solve* a given game $\Gamma$. 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 $\mu^*$ is ex-ante optimal for $\Gamma$ if it solves*

$$\max_{\mu} U(\mu) \quad \text{s.t.} \quad \mu \in \bigtimes_{i=1}^{\ell} \Delta(A_{I_i}). \tag{1}$$

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 contains all those decision problems that can be solved in expected polynomial time by a randomized (Las Vegas) algorithm. Let OPT be 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 $\Gamma$ and target value $t \in \mathbb{Q}$ (encoded in binary) as inputs and asks whether there is a strategy $\mu$ for $\Gamma$ with ex-ante expected utility $U(\mu) \geq t$. 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 instance $\Gamma$ has a tree depth of $3$ and only one info set.*
*(2.) NP-hardness holds even if $\Gamma$ has no absentmindedness, a tree depth of $4$ and the player has $2$ actions per info set.*
*(3.) NP-hardness holds even if $\Gamma$ has no absentmindedness, a tree depth of $3$ and the player has $3$ actions per info set.*

Note that finding the ex-ante optimal strategy of $\Gamma$ is at least as hard as this decision problem of whether a target value can be achieved in $\Gamma$. On the other hand, with an efficient solver of the decision version, one can recover the optimal ex-ante utility $U^* := \max_{\mu} U(\mu)$ through binary search.

An early such NP-hardness result was given by Koller and Megiddo [1992]. Gimbert *et al.* [2020] leverage the correspondence to polynomial optimization (Section 2.2) to further show that, in fact, this decision problem of whether a target value can be achieved in $\Gamma$ is $\exists \mathbb{R}$-complete. The complexity class $\exists \mathbb{R}$, called *the existential theory of the reals* [Renegar, 1992; Schaefer and Stefankovic, 2017], consists of all those problems that reduce to deciding whether a sentence of the following form is true: $\exists x_1 \ldots \exists x_n F(x_1, \ldots x_n)$, where

the $x_i$ are real-valued variables and where $F$ is a quantifier-free formula that may contain equalities and inequalities of real polynomials. $\exists \mathbb{R}$ lies in between NP and PSPACE [Shor, 1990; Canny, 1988]. The decision problem of Proposition 4 can also be shown to be a member of NP in an approximate sense; namely, when it is allowed to incorrectly return "yes" to the problem instance $(\Gamma, t, \epsilon)$ if there exists a strategy profile $\mu$ with $U(\mu) \geq t - \epsilon$. Here, $\epsilon > 0$ represents an inverse-exponential precision parameter.

A proof of Proposition 4 can be found in Appendix B. Result (1.) is based on 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]. The proofs reveal that NP-hardness remains even if the encoding size of the chance node probabilities and utility values are in $O(|\mathcal{H}|)$.

## 3 Equilibria in Imperfect-Recall Games

Proposition 4 shows a strong obstacle to finding or approximating ex-ante optimal strategies 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 has no influence over which actions she chooses at other info sets. Therefore, at an equilibrium strategy $\mu$, the player will play the best action at each info set, assuming that she has been playing according to $\mu$ 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 where an ex-ante optimal strategy also constitutes an equilibrium. In games without absentmindedness, these two equilibrium concepts coincide. In games with absentmindedness, the concepts differ in how expected utilities are evaluated for an action $a$ at a current info set $I$, given that the player plays according to strategy $\mu$ anywhere "else". Computing such expected utilities requires

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

In the sequel, let $I$ be the player's current info set at which she finds herself during play while playing $\mu$ in game $\Gamma$.

### 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 $\mu$ at past arrivals at $I$, or that she will be playing according to $\mu$ at future arrivals at $I$. The intuition behind CDT is that the player's choice to deviate away from $\mu$ 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 $\alpha$ whenever she arrived in $I$ in past play, and she will be deviating to $\alpha$ whenever she arrives in $I$ again in future play. Indeed, EDT argues that the choice to deviate to $\alpha$ now is evidence for the player taking the same deviation choice in the past and future.

Denote with $\mu_{I \mapsto \alpha}$ an EDT deviation, i.e., the strategy of $\Gamma$ that plays according to $\mu$ at every info set except at the info set $I \in \mathcal{I}_*$ where it plays according to $\alpha \in A_I$. By contrast, a CDT deviation may result in different actions taken at the same info set. This might not constitute 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 $\mu$-prescribed action $R$ to the action $L$ this one time only. Then, CDT argues that the player will stick to her $\mu$-prescribed action $R$ at the other node of $I_1$, leading to one of the two 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 history to enter info set $I$. Define the reach probability and (expected) visit frequency of $I$ under $\mu$ as $\mathbb{P}(I \mid \mu) := \sum_{h \in I^{1\text{st}}} \mathbb{P}(h \mid \mu)$ and $\text{Fr}(I \mid \mu) := \sum_{h \in I} \mathbb{P}(h \mid \mu)$. 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. However, we have in general that $\mathbb{P}(I \mid \mu) > 0$ if and only if $\text{Fr}(I \mid \mu) > 0$. Finally, denote with $\chi : P \to \{0, 1\}$ the function that takes a Boolean property $P$ as input and evaluate 1 if and only if $P$ is true.

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

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

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

The second belief system argues that one should rather focus on the reach probabilities. Note that the statement $I \cap \text{hist}(z) \neq \emptyset$ evaluates as true if and only if $I$ occurs in history $\text{hist}(z)$ of terminal node $z \in \mathcal{Z}$ at least once.

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

$$\mathbb{P}_{\text{GDH}}(\text{hist}(z) \mid \mu, I) := \chi(I \cap \text{hist}(z) \neq \emptyset) \cdot \frac{\mathbb{P}(z \mid \mu)}{\mathbb{P}(I \mid \mu)}.$$

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 type of events (to be at player node versus to be in the history of a terminal node). In Appendix D, we phrase GT and GDH in each other's language. In the language of GDH, GT assigns the event of being in history $\text{hist}(z)$ of a terminal node $z$ a higher probability if the reach probability of $z$ under $\mu$ is higher (same as GDH) and if $\text{hist}(z)$ visits info set $I$ very often (whereas GDH only cares about $I$ being visited at least once by $\text{hist}(z)$).

**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 the 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 contrast but for a different example called Sleeping Beauty [Elga, 2000] (Appendix C).*

## 3.3 Two Equilibrium Concepts

Any claims made in this section are proven in Appendix E.

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

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

$$\text{EU}_{\text{CDT,GT}}(\alpha \mid \mu, I) := \sum_{h \in I} \mathbb{P}_{\text{GT}}(h \mid \mu, I)$$
$$\cdot \left( \sum_{a \in A_I} \alpha(a) \cdot U(\mu \mid h \circ a) \right).$$

Note that the inner sum collapses to $U(\mu \mid h \circ a)$ if the considered mixed action $\alpha$ is a pure action $a$.

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

$$\mu^*(\cdot \mid I) \in \underset{\alpha \in \Delta(A_I)}{\arg\max} \, \text{EU}_{\text{CDT,GT}}(\alpha \mid \mu^*, I).$$

Alternatively, we can use the easier-to-check condition that for all info sets $I \in \mathcal{I}_*$ with $\text{Fr}(I \mid \mu^*) > 0$ and all pure actions $a \in A_I$ with $\mu^*(a \mid I) > 0$, we have

$$a \in \underset{a' \in A_I}{\arg\max} \, \text{EU}_{\text{CDT,GT}}(a' \mid \mu^*, I). \qquad (2)$$

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 $\alpha$ now and according to $\mu$ otherwise is*

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

The GDH belief 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 $\mu^*$ of $\Gamma$ is a (EDT,GDH)-equilibrium if for all info sets $I \in \mathcal{I}_*$ with $\mathbb{P}(I \mid \mu^*) > 0$ under $\mu^*$, we have*

$$\mu^*(\cdot \mid I) \in \underset{\alpha \in \Delta(A_I)}{\arg\max} \operatorname{EU}_{\mathrm{EDT,GDH}}(\alpha \mid \mu^*, I). \quad (3)$$

For (EDT,GDH), it is not sufficient to only check for optimality of pure actions that are in the support of $\mu^*$. 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 pure actions $\{L, C, R\}$. But the player would strictly benefit from deviating to mixed action $\frac{1}{2}L + \frac{1}{2}R$.

Finally, observe that in games without absentmindedness, the following notions coincide: CDT and EDT, GT and GDH, and Definitions 9 and 11. In particular, both equilibrium concepts coincide:

**Lemma 13.** *In games without absentmindedness, a strategy $\mu$ is a (CDT,GT)-equilibrium if and only if it is an (EDT,GDH)-equilibrium.*

### 3.4 Equilibria from the Ex-Ante Perspective

Recall from Section 2.2 that the (ex-ante) strategy utility function $U$ of $\Gamma$ is a polynomial function from $\times_{i=1}^{\ell} \mathbb{R}^{m_i}$ to $\mathbb{R}$. In this section, we give characterizations for (CDT,GT)- and (EDT,GDH)-equilibria in terms of $U$, 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 $U$ is continuously differentiable in $\mu \in \times_{i=1}^{\ell} \mathbb{R}^{m_i}$. For $i \in [\ell]$ and $j \in [m_i]$, let $\nabla_{ij} U$ stand for the partial derivative in direction $(i, j)$, that is, the linear change of $U$ at a point $\mu$ if you infinitesimally increase its $\mu(a_j \mid I_i)$ value.

**Lemma 14.** *Let $I_i$ be an info set, $a_j \in A_{I_i}$ an action, and $\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})$ a strategy. Then:*
*1. $\nabla_{ij} U(\mu) = 0$ if $\operatorname{Fr}(I_i \mid \mu) = 0$, and*
*2. $\nabla_{ij} U(\mu) = \operatorname{Fr}(I_i \mid \mu) \cdot \operatorname{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i)$ otherwise.*

Note that an infinitesimal increase of $\mu(a_j \mid I_i)$ means in a game-theoretic sense that the decision for action $a_j$ is made slightly more probable at *every* node of info set $I_i$. This resembles an EDT type of deviation power but restricted to small deviations that stay close to the current action profile $\mu$. Then, Lemma 14 says that a CDT deviation – rescaled by $\operatorname{Fr}(I_i \mid \mu)$ – accurately captures the linear (=dominant) effect of such a "local EDT deviation".

**Lemma 15.** *Strategy $\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})$ of $\Gamma$ is an (EDT,GDH)-equilibrium if and only if for all $i \in [\ell]$:*

$$\mu_{i \cdot} \in \underset{y \in \Delta(A_{I_i})}{\arg\max} \ U(\mu_{1 \cdot}, \ldots, \mu_{i-1 \cdot}, y, \mu_{i+1 \cdot}, \ldots, \mu_{\ell \cdot}).$$

One possible interpretation of Lemma 15 is that (EDT,GDH)-equilibria of $\Gamma$ are exactly the Nash equilibria of an $\ell$-player simultaneous and identical-interest game $G$: Each player $i$ shall have the continuous action space $\Delta(A_{I_i})$ and the (single) utility payoff, a function of the chosen action profile $(\mu_{i \cdot})_{i=1}^{\ell} \in \times_{i=1}^{\ell} \Delta(A_{I_i})$, shall be the polynomial $U$.

### 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 (Appendix G). Therefore, we relax our search to $\epsilon$-approximate (CDT,GT)-equilibria where $\epsilon$ is an inverse-exponential numerical precision parameter.

**Definition 16.** *An instance of the problem* (CDT,GT)-EQUILIBRIUM *consists of a single-player extensive-form game $\Gamma$ with imperfect recall and a precision parameter $\epsilon > 0$ encoded in binary. A solution consists of a strategy $\mu$ for $\Gamma$ that satisfies for all $I \in \mathcal{I}_*$ with $\operatorname{Fr}(I \mid \mu) > 0$:*

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

There is also an alternative notion of being close to an equilibrium, called $\epsilon$-well-supported (CDT,GT)-equilibrium. It instead requires condition (2) to be satisfied up to $\epsilon$ precision. An analysis of when both approximation concepts are polynomial-time related can be found in Appendix H.

We will give hardness results and restricted membership results for (CDT,GT)-EQUILIBRIUM for the class CLS (Continuous Local Search). CLS was introduced by Daskalakis and Papadimitriou [2011] who noted that it contains various important problems of continuous local optimization that belong both to PPAD and PLS. PPAD [Papadimitriou, 1994] is well-known as the class that captures the complexity of many problems of Nash equilibrium computation ([Daskalakis *et al.*, 2009; Chen *et al.*, 2009] and much subsequent work), while PLS (Polynomial Local Search [Johnson *et al.*, 1988]) represents the complexity of many problems of discrete local optimization. Recently, Fearnley *et al.* [2023] showed that CLS is *equal* to the intersection of PPAD and PLS, indicating that CLS-hardness is quite a reliable notion of computational difficulty. In addition, the hardness of CLS can also be based on the cryptographic assumption of indistinguishability obfuscation [Hubáček and Yogev, 2017]. Fearnley *et al.* [2023] also showed that a version of the KKT point search problem is CLS-complete. Subsequent work has established CLS-completeness of mixed Nash equilibria of congestion games [Babichenko and Rubinstein, 2021] and solutions to a certain class of contests [Elkind *et al.*, 2022]. We can characterize CLS through any of its complete problems.

We will mainly be interested in KKT points. Consider a general non-linear maximization problem

$$\max_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Bx + b \leq 0, \, Cx + c = 0 \quad (4)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, $B \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n}$, and $c \in \mathbb{R}^{\ell}$, and the domain is bounded. A point $x \in \mathbb{R}^n$ is then said to be a KKT point for (4) if there exist KKT multipliers $\tau_1, \ldots, \tau_m, \kappa_1, \ldots, \kappa_{\ell} \in \mathbb{R}$ such that $Bx + b \leq 0$ and $Cx + c = 0$, and $\forall j \in [m] : \tau_j \geq 0, \forall j \in [m] : \tau_j = 0$ or $B_j . x + b_j = 0$, and

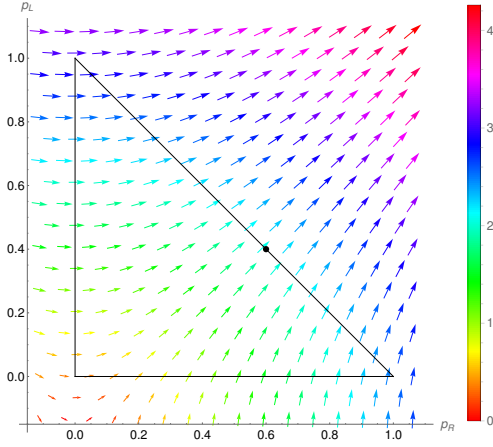$$\nabla f(x) = \sum_{j=1}^{m} \tau_j \cdot (B_j .)^T + \sum_{i=1}^{\ell} \kappa_i \cdot (C_i .)^T = 0.$$

Figure 2: A plot of the gradient vector field $\nabla U$ of the strategy utility function $U$. The underlying game is the one from Figure 1 except where the player nodes of info set $I_2$ are replaced by chance nodes that choose actions $X$ and $Y$ with chance probabilities $\frac{1}{2}$ each. Then the player only has to choose a mixed action for info set $I_1$. The plot is in 2D for visualization convenience: The x-axis and y-axis represent the probabilities put on action $R$ and $L$ respectively. The action simplex $\Delta(\{L, C, R\})$ becomes a right triangle with the point $(0, 0)$ corresponding to pure action $C$. The gradient coloring represents the vector length.
There is no KKT point in the interior of the (projected) simplex because the gradient does not vanish there. The KKT points on the boundary of the simplex are those where the gradient is directed perpendicularly outwards of the boundary constraint (except corner points, whose gradient only needs to lie in the positive cone of the boundary constraint directions). Thus $(0.6 \cdot L + 0.4 \cdot R)$ is the only KKT point. A game-theoretic analysis of the underlying game $\Gamma$ also yields $(0.6 \cdot L + 0.4 \cdot R)$ as the only (CDT,GT)-equilibrium.

The KKT conditions are necessary first-order conditions for a point to be a local optimum of (4). Furthermore, feasible stationary points satisfy the KKT conditions.

## 4 Main Results

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

### 4.1 Complexity of the Search Problems

First, we use Lemma 14 to give a characterization of (CDT,GT)-equilibria in terms of ex-ante utility. For that, recall the ex-ante maximization problem (1).

**Theorem 1.** *Strategy* $\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})$ *of* $\Gamma$ *is a (CDT,GT)-equilibrium if and only if* $\mu$ *is a KKT point of (1).*

We visualize Theorem 1 in Figure 2. This result also reveals a method to find (CDT,GT)-equilibria, namely by applying Gradient Descent on $U$. 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 because for any $\epsilon > 0$ a shift from $\mu(\cdot \mid I_1) = C$ to

$\epsilon \cdot (\frac{1}{2} L + \frac{1}{2} R) + (1 - \epsilon) \cdot C$ would yield the player ex-ante utility $5 \cdot \frac{\epsilon}{2} \cdot \frac{\epsilon}{2} > 0$. However, from a (CDT,GT) standpoint, the player should be 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 14 and Theorem 1, the issue is that the first order effect of increasing the probabilities of $R$ and $L$ is 0.

The three solution concepts considered in this paper form an inclusion hierarchy, a result shown by Oesterheld and Conitzer [2022] [cf. Piccione and Rubinstein, 1997]:

**Lemma 17.** *An ex-ante optimal strategy of a game* $\Gamma$ *is also an (EDT,GDH)-equilibrium. An (EDT,GDH)-equilibrium is also a (CDT,GT)-equilibrium.*

*In particular, any single-player extensive-form game* $\Gamma$ *with imperfect recall admits an (EDT,GDH)-equilibrium and a (CDT,GT)-equilibrium.*

The implication chain of Lemma 17 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 $\mu'' = (\frac{1}{2} L + \frac{1}{2} R, X)$ achieves the (optimal) ex-ante utility $5/4$.

The second part of Lemma 17 holds because ex-ante optimal strategies always exist. This is in contrast to the multi-player setting where Nash equilibria may not exist in the presence of imperfect recall. Moreover, Lemma 17 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-hard. CLS-hardness holds even for games restricted to:*
*(1.) a tree depth of* 6 *and the player has* 2 *actions per info set,*
*(2.) no absentmindedness and a tree depth of* 6, *and*
*(3.) no chance nodes, a tree depth of* 5, *and only one info set.*
*The problem is in CLS for the subclass of problem instances of* (CDT,GT)-EQUILIBRIUM *where a lower bound on positive visit frequencies in* $\Gamma$ *is easily obtainable.*

All CLS results of Theorem 2 also hold analogously for the search problem of an approximate well-supported (CDT,GT)-equilibrium. 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. Both search problems we reduce from were shown to be CLS-complete by Babichenko and Rubinstein [2021].

The CLS membership in Theorem 2 implies that, unless NP = co-NP, the considered problem cannot be hard for the class NP [Megiddo and Papadimitriou, 1991]. We only prove CLS membership for those games $\Gamma$ where we can construct a lower bound value $\lambda > 0$ that satisfies $\text{Fr}(I \mid \mu) = 0$ or $\text{Fr}(I \mid \mu) \geq \lambda$ for all strategies $\mu$ and info sets $I$ in $\Gamma$. That is because if $\text{Fr}(I \mid \mu) > 0$ is too small, the approximation error may explode when transitioning from the ex-ante perspective $\nabla_{ij} U(\mu)$ to the *de se* perspective $\text{EU}_{\text{CDT,GT}}(a_j \mid \mu, I_i)$. Fortunately, such a lower bound exists and is polynomial-time

computable for many well-known imperfect-recall games, such as the Absentminded Driver, game variants of the Sleepy Beauty problem, and all the games used in the CLS-hardness results of Theorem 2. Thus, the computation of an approximate (CDT,GT)-equilibrium is CLS-complete in those games that admit such a lower bound on positive visit frequencies. Statement (2.) shows in particular that absentmindedness is not the reason for CLS hardness. With Lemma 13, this implies

**Corollary 18.** *In games without absentmindedness where a lower bound on positive visit frequencies is easily obtainable, it is CLS-complete to find an $\epsilon$-(EDT,GDH)-equilibrium.*

The authors are not aware of any complexity classification for the problem of *finding* an approximate (EDT,GDH)-equilibrium in games that may have absentmindedness – even though Lemma 15 gives a nice optimization characterization of (EDT,GDH)-equilibria. Nonetheless, we are able to give conditional inapproximability results for (EDT,GDH)-equilibria with the next theorem.

## 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 $\Gamma$ and $t \in \mathbb{Q}$, is there a (CDT,GT)-equilibrium of $\Gamma$ with ex-ante utility $\geq t$?*

*(1b.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there a (CDT,GT)-equilibrium $\mu$ with $\text{Fr}(I \mid \mu) > 0$, and such that the player has a (CDT,GT)-expected utility $\geq t$ upon reaching $I$?*

*(1c.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there a strategy $\mu$ of $\Gamma$ with $\text{Fr}(I \mid \mu) > 0$, and such that the player has a (CDT,GT)-expected utility $\geq t$ upon reaching $I$?*

*(2a.) Given $\Gamma$ and $t \in \mathbb{Q}$, is there an (EDT,GDH)-equilibrium of $\Gamma$ with ex-ante utility $\geq t$?*

*(2b.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there an (EDT,GDH)-equilibrium $\mu$ with $\mathbb{P}(I \mid \mu) > 0$, and such that the player has an (EDT,GDH)-expected utility $\geq t$ upon reaching $I$?*

*(2c.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there a strategy $\mu$ of $\Gamma$ with $\mathbb{P}(I \mid \mu) > 0$, and such that the player has an (EDT,GDH)-expected utility $\geq t$ upon reaching $I$?*

*(3a.) Given $\Gamma$ and $t \in \mathbb{Q}$, do all (EDT,GDH)-equilibria of $\Gamma$ have ex-ante utility $\geq t$?*

*(3b.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, do all (EDT,GDH)-equilibria $\mu$ with $\mathbb{P}(I \mid \mu) > 0$ yield the player an (EDT,GDH)-expected utility $\geq t$ upon reaching $I$?*

All the results of Theorem 3 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.) also relies on

the observation that in games with one info set only, any (EDT,GDH)-equilibrium is also ex-ante optimal (cf. Lemma 15).

From (3a.) we obtain in particular that, unless NP = ZPP, there is no FPTAS for the search problem of an (EDT,GDH)-equilibrium in games with imperfect recall[1]. To compare this to Theorem 2, we remark that this conditional inapproximability result for (EDT,GDH)-equilibria (and ex-ante optimal strategies) is obtained even for games where a lower bound on positive visit frequencies is easily obtainable.

The decision problems of the form (1-.), (2a.), and (2c.) are all members of the complexity class $\exists \mathbb{R}$, and therefore, in particular, in PSPACE. On one hand, this is because (CDT,GT)-expected utilities and (EDT,GDH)-expected utilities can be described as rational functions (fractions of polynomial functions). Furthermore, this is because the alternative definition (2) of a (CDT,GT)-equilibrium gives rise to polynomially many comparisons of polynomial functions, and, for (2a.), because ex-ante optimal strategies are (EDT,GDH)-equilibria.

## 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 in the presence of imperfect recall. 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, to the extent that (CDT,GT)-equilibria are (under reasonable complexity assumptions) easier to compute than (EDT,GDH)-equilibria, does that provide support for using the former solution concept, at least for certain purposes? We hope that the work we

_____

[1]Note that FPTAS even allow for an approximation up to an inverse-*polynomial* precision $\epsilon$.

have done in this paper can serve as a springboard for further research into this fascinating and important topic.

## Acknowledgements

## 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.

[Canny, 1988] John Canny. Some algebraic and geometric computations in pspace. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, STOC '88, page 460–467, New York, NY, USA, 1988. Association for Computing Machinery.

[Chen et al., 2006] Xi Chen, Xiaotie Deng, and Shang-hua Teng. Computing nash equilibria: Approximation and smoothed complexity. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 603–612, 2006.

[Chen et al., 2009] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *J. ACM*, 56(3):14:1–14:57, 2009.

[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 and Oesterheld, 2023] Conitzer and Oesterheld. Foundations of cooperative ai. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, February 7 - 14, 2022*. AAAI Press, 2023.

[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.

[Daskalakis et al., 2009] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM J. Comput.*, 39(1):195–259, 2009.

[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.

[Elkind et al., 2022] Edith Elkind, Abheek Ghosh, and Paul W. Goldberg. Simultaneous contests with equal sharing allocation of prizes: Computational complexity and price of anarchy. In *Algorithmic Game Theory - 15th International Symposium, SAGT*, volume 13584 of *Lecture Notes in Computer Science*, pages 133–150. Springer, 2022.

[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.

[Fearnley et al., 2023] John Fearnley, Paul Goldberg, Alexandros Hollender, and Rahul Savani. The complexity of gradient descent: CLS = PPAD ∩ PLS. *J. ACM*, 70(1):7:1–7:74, 2023.

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

[Gimbert et al., 2020] Hugo Gimbert, Soumyajit Paul, and B. Srivathsan. A bridge between polynomial optimization and games with imperfect recall. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, page 456–464, Richland, SC, 2020. International Foundation for Autonomous Agents and Multiagent Systems.

[Håstad, 1996] Johan Håstad. Clique is hard to approximate within $n^{1\text{-epsilon}}$. In *37th Annual Symposium on Foundations of Computer Science, FOCS*, pages 627–636. IEEE Computer Society, 1996.

[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*, pages 1352–1371. SIAM, 2017.

[Johnson *et al.*, 1988] David S. Johnson, Christos H. Papadimitriou, and Mihalis Yannakakis. How easy is local search? *Journal of Computer and System Sciences*, 37(1):79–100, 1988.

[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.

[Kovarik *et al.*, 2023] Vojtech Kovarik, Caspar Oesterheld, and Vincent ConitzerYuta. Game theory with simulation of other players. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, Macao, 9-25 August 2023*. ijcai.org, 2023.

[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.

[Papadimitriou, 1994] Christos H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and System Sciences*, 48(3):498–532, 1994.

[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.

[Renegar, 1992] James Renegar. On the computational complexity and geometry of the first-order theory of the reals. part i: Introduction. preliminaries. the geometry of semi-algebraic sets. the decision problem for the existential theory of the reals. *Journal of Symbolic Computation*, 13(3):255–299, 1992.

[Schaefer and Stefankovic, 2017] Marcus Schaefer and Daniel Stefankovic. Fixed points, nash equilibria, and the existential theory of the reals. *Theory Comput. Syst.*, 60(2):172–193, 2017.

[Shor, 1990] Peter W. Shor. Stretchability of pseudolines is NP-hard. In Peter Gritzmann and Bernd Sturmfels, editors, *Applied Geometry And Discrete Mathematics, Proceedings of a DIMACS Workshop, Providence, Rhode Island, USA, September 18, 1990*, volume 4 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 531–554. DIMACS/AMS, 1990.

[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.

# A On Section 2.2: Reductions from Polynomial Maximization to Games

We consider problems of optimizing polynomials $p : \times_{i=1}^{\ell} \mathbb{R}^{m_i} \to \mathbb{R}$, where each subset of $m_i$ variables are constrained to lie in the standard $(m_i - 1)$-simplex.

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

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

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

of all variations to draw up to $d \in \mathbb{N}$ elements out of the set $\{(i,j)\}_{i,j=1}^{\ell, m_i}$, with replacement and without regard to draw order. Each element $D \in \mathrm{MB}(d, \boldsymbol{m})$ captures the variable degrees of the monomial basis element $\prod_{i,j=1}^{\ell, m_i} x_{ij}^{D_{ij}}$ that it represents.

Throughout this paper (specifically, this appendix), if the instance to a problem contains a polynomial function $p : \times_{i=1}^{\ell} \mathbb{R}^{m_i} \to \mathbb{R}$, then we assume it to be represented as a binary encoding of $\ell, (m_i)_{i=1}^{\ell}$, and the polynomials coefficients $(\lambda_D)_{D \in \mathrm{MB}(d, \boldsymbol{m})}$, which need to be rational. This is also called the Turing (bit) model. Denote $\mathrm{supp}(p) := \{D \in \mathrm{MB}(d, \boldsymbol{m}) : \lambda_D \neq 0\})$ and, for each $D \in \mathrm{MB}(d, \boldsymbol{m})$, $\mathrm{supp}(D) := \{(i,j) \text{ with } i \in [\ell], j \in [m_i] : D_{ij} > 0\})$ as well as $|D| := \sum_{i,j=1}^{\ell, m_i} D_{ij}$. Let $\mathrm{supp}(D)^{\mathrm{ms}}$ be the multiset that contains $D_{ij}$ many copies of the element $(i,j)$ in it for each $(i,j) \in \mathrm{supp}(D)$ (in multisets, duplicate elements are allowed). Then $|\mathrm{supp}(D)^{\mathrm{ms}}| = |D|$.

Given such a polynomial function, let us construct a corresponding single-player extensive-form game $\Gamma$ with imperfect recall. It shall have info sets $I_i$ for $i \in [\ell]$ with action sets $A_{I_i} := \{a_1, \ldots, a_{m_i}\}$, 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 uniformly at random. 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 $\prod_{i,j=1}^{\ell, m_i} x_{ij}^{D_{ij}}$. We will now build $T_D$ depth layer by depth layer, until a tree of depth $|D|$, by lexicographically going through the set $\mathrm{supp}(D)^{\mathrm{ms}}$. Depth $k - 1$ of $T_D$ corresponds to the $k$-th element of $\mathrm{supp}(D)^{\mathrm{ms}}$, referred to as $(i(k), j(k))$. We present two variants with how to continue building $\Gamma$:

**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 $m_{i(k)}$ outgoing edges out

of $h$ into a new node of depth $k$, and the edges shall be labeled with $\{a_1, \ldots, a_{m_{i(k)}}\}$ 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 $\left(a_{j(k)} \in A_{I_{i(k)}}\right)_{k \in [|D|]}$ 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 $\lambda_D \cdot |\mathrm{supp}(p)|$.

**Variant 2:** This variant does not have terminal nodes in depth layers $0, \ldots, |D| - 1$. 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 $m_{i(k)}$ outgoing edges out of each $h$ into a new node of depth $k$. The edges shall be labeled with $\{a_1, \ldots, a_{m_{i(k)}}\}$ 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 $\left(a_{j(k)} \in A_{I_{i(k)}}\right)_{k \in [|D|]}$ in $T_D$, which we will refer to as $z_D$. Assign it a utility of $\lambda_D \cdot |\mathrm{supp}(p)|$. Assign all other terminal nodes of $T_D$ a utility of 0.

In both variants, we have that any point $x \in \times_{i=1}^{\ell} \mathbb{R}^{m_i}$ of $p$ that is also in $\times_{i=1}^{\ell} \Delta^{m_i-1}$ naturally comprises a strategy $\mu$ in $\Gamma$ with probabilities $\mu(a_j \mid I_i) = x_{ij}$. Moreover, the strategy utility function $U$ of both variants of $\Gamma$ satisfies

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

for corresponding $x$ and $\mu$. This extends to $U = p$ on all $\times_{i=1}^{\ell} \mathbb{R}^{m_i}$.

The construction of the first variant of $\Gamma$ takes polynomial time in the encoding size of $p : \times_{i=1}^{\ell} \mathbb{R}^{m_i} \to \mathbb{R}$. That is, because the game tree has size $\leq |\mathrm{supp}(p)| \cdot \deg(p) \cdot (\max_i m_i)$. The second variant of $\Gamma$ 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 $\leq |\mathrm{supp}(p)| \cdot (\max_i m_i)^d$ which makes the whole construction of $\Gamma$ polynomial time in the size of the input instances again. An advantage of the second variant is that in that game, the reach probability $\mathbb{P}(I \mid \mu)$ and visit frequency $\mathrm{Fr}(I \mid \mu)$ of an info set $I \in \mathcal{I}$ are independent of the used strategy $\mu$. That is because a subtree $T_D$ is chosen by nature at the beginning, and within $T_D$, any outcome path visits the exact same info sets with the same multiplicity. The reach probability of info set $I_i$ becomes

$$\mathbb{P}(I_i) = \mathbb{P}(I_i \mid \mu) = \sum_{D \in \mathrm{supp}(p)} \frac{1}{|\mathrm{supp}(p)|} \cdot \chi\left( \sum_{j \in [m_i]} D_{ij} \geq 1 \right)$$

and its visit frequency becomes

$$\mathrm{Fr}(I_i) = \mathrm{Fr}(I_i \mid \mu) = \sum_{D \in \mathrm{supp}(p)} \frac{1}{|\mathrm{supp}(p)|} \cdot \sum_{j \in [m_i]} D_{ij} \,. \tag{5}$$

We will make use of this observation later when it comes to games that admit an easy-to-compute lower bound on positive visit frequencies.

## B   On Section 2.3: Proof of Proposition 4

We recall and prove Proposition 4:

**Proposition.** *Consider the problem that takes a game $\Gamma$ and target value $t \in \mathbb{Q}$ (encoded in binary) as inputs and asks whether there is a strategy $\mu$ for $\Gamma$ with ex-ante expected utility $U(\mu) \geq t$. 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 instance $\Gamma$ has a tree depth of 3 and only one info set.*
*(2.) NP-hardness holds even if $\Gamma$ has no absentmindedness, a tree depth of 4 and the player has 2 actions per info set.*
*(3.) NP-hardness holds even if $\Gamma$ 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 $p : \times_{i=1}^{\ell} \mathbb{R}^{m_i} \to \mathbb{R}$ over the product of simplices to maximizing ex-ante utility in a single-player extensive-form game $\Gamma$ 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 $p : \Delta^{m-1} \to \mathbb{R}$ and a target value $t \in \mathbb{Q}$, and that has to decide whether there exists a point $x \in \Delta^{m-1}$ 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 [1996]. 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 $x_1, \ldots, x_\ell$ be the variables of a 3CNF formula $\phi$ having $n$ clauses. We construct the corresponding game instance $\Gamma$ 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 $x_i, x_j, x_k$. 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, $\phi$ 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 $\mu$ is a behavioural (=randomizing) strategy with an ex-ante utility of 1. Take any pure-action strategy $\mu'$ that only uses actions that are in the support of $\mu$. Then $\mu'$ makes a satisfying truth assignment for $\phi$.
(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 $(\mu_1, \mu_2)$ that yields a payoff $\geq 1$ 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 all the 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 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 $\Gamma$ with imperfect recall in the following way: $\Gamma$ has depth 3 and information sets $\mathcal{I}_* \cup \mathcal{J}_*$. The root of $\Gamma$ 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 $\Gamma$ 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 $\Gamma$ be 1. A strategy $(\mu_1, \mu_2)$ of $G$ shall correspond to the strategy $\mu$ of $\Gamma$ 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 $J \in \mathcal{J}_* \subset \mathcal{I}_* \cup \mathcal{J}_*$

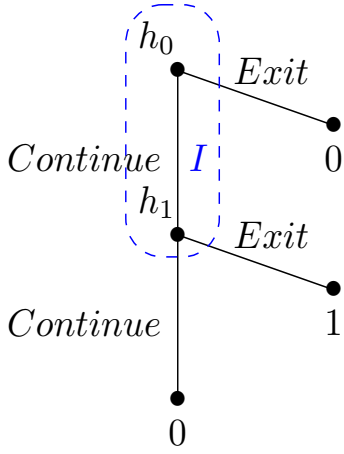Then, there is a strategy profile $(\mu_1, \mu_2)$ of $G$ that yields

Figure 3: Absentminded driver game. The driver reaches his destination by taking the second of two exits from the road he is driving on. Unfortunately, when approaching an exit, he forgets whether he has passed another exit before. Any pure strategy yields payoff 0. His best strategy is to exit with probability $\frac{1}{2}$ which gives him a payoff of $\frac{1}{4}$.
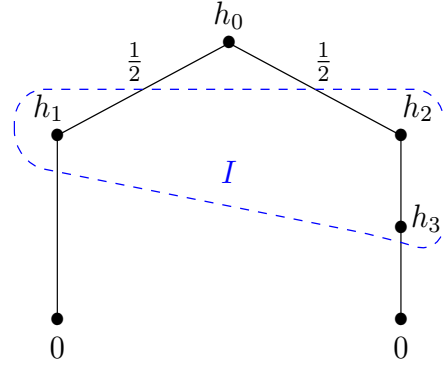


Figure 4: Sleeping Beauty. In the beginning, a coin is flipped in secret from the game participant, the sleeping beauty. Next, she is put to sleep, and woken up. If the coin flip turned out heads, the game ends. If the coin flip turned out tails, she will be put to sleep again, and woken up again, after which the game ends. At any point she wakes up, she forgot whether she has been awake before. When woken up (= arriving in $I$), what should her belief probabilities be on the coin flip having turned out heads (= her being in $h_1$)? Generalized Thirding argues it should be $\frac{1}{3}$ whereas Generalized Double Halving argues it should be $\frac{1}{2}$.

a payoff $\geq 1$ if and only if its corresponding strategy $\mu$ in $\Gamma$ yields an ex-ante utility $\geq t$.

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

□

## C   Popular Game Examples of Imperfect Recall

In Figure 3, we describe the Absentminded Driver game. In Figure 4, we describe the Sleeping Beauty problem, which is less of a decision problem, but rather a probability puzzle.

## D   On Section 3.2: Comparing Generalized Thirding and Generalized Double Halving

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

Let $\Gamma$ be a single-player extensive-form game with imperfect recall, $\mu$ 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 $\mathbb{P}(I \mid \mu) > 0 \iff \mathrm{Fr}(I \mid \mu) > 0$.

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

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

The number of times info set $I$ occurs in the history of a terminal node $z$ can be described with $|I \cap \mathrm{hist}(z)|$. 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 $z \in \mathcal{Z}$:

$$\mathbb{P}_{\mathrm{GT}}(\mathrm{hist}(z) \mid \mu, I) = \sum_{h \in \mathcal{H}_*} \mathbb{P}_{\mathrm{GT}}(\text{at } h \text{ in hist}(z) \mid \mu, I)$$

$$:= \sum_{h \in \mathcal{H}_* \cap \mathrm{hist}(z)} \mathbb{P}_{\mathrm{GT}}(h \mid \mu, I) \cdot \mathbb{P}(z \mid \mu, h)$$

$$= \sum_{h \in I \cap \mathrm{hist}(z)} \frac{\mathbb{P}(h \mid \mu)}{\mathrm{Fr}(I \mid \mu)} \cdot \mathbb{P}(z \mid \mu, h)$$

$$= |I \cap \mathrm{hist}(z)| \cdot \frac{\mathbb{P}(z \mid \mu)}{\mathrm{Fr}(I \mid \mu)}$$

Compare this to the definition of GDH:

$$\mathbb{P}_{\mathrm{GDH}}(\mathrm{hist}(z) \mid \mu, I) = \chi(I \cap \mathrm{hist}(z) \neq \emptyset) \cdot \frac{\mathbb{P}(z \mid \mu)}{\mathbb{P}(I \mid \mu)}$$

So GT assigns a history hist$(z)$ of a terminal node $z$ a higher probability if the reach probability of $z$ under $\mu$ 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 $\mu$ 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

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

# E    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

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

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

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

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

**Proof of Lemma 13:    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 $\mu'$ of $\Gamma$, node $h \in \mathcal{H} \setminus \mathcal{Z}$,

and terminal node $z \in \mathcal{Z}$, we have

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

Moreover, we can obtain

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

Now let $\Gamma$ be a single-player extensive-form game with imperfect recall and without absentmindedness. Let $\mu$ 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 $\Gamma$, 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 $\chi(I \cap \text{hist}(z) \neq \emptyset) = |I \cap \text{hist}(z)|$ for any terminal node $z \in \mathcal{Z}$. Therefore, by

Appendix D, we have for any terminal node $z$:

$$\mathbb{P}_{\text{GT}}(\text{hist}(z) \mid \mu, I) = |I \cap \text{hist}(z)| \cdot \frac{\mathbb{P}(z \mid \mu)}{\text{Fr}(I \mid \mu)}$$

$$= \chi(I \cap \text{hist}(z) \neq \emptyset) \cdot \frac{\mathbb{P}(z \mid \mu)}{\mathbb{P}(I \mid \mu)}$$

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

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

$$\mathbb{P}_{\text{GDH}}(h \mid \mu, I)$$

$$= \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\mathbb{P}(I \mid \mu)} \cdot \sum_{\substack{z \in \mathcal{Z} \\ h \in \text{hist}(z)}} \frac{\mathbb{P}(z \mid \mu, h)}{|I \cap \text{hist}(z)|}$$

$$= \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I \mid \mu)} \cdot \sum_{\substack{z \in \mathcal{Z} \\ h \in \text{hist}(z)}} \mathbb{P}(z \mid \mu, h)$$

$$= \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I \mid \mu)} \cdot 1$$

$$= \mathbb{P}_{\text{GT}}(h \mid \mu, I)$$

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 $I^{\text{1st}} = I$, 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:

$$\text{EU}_{\text{CDT,GT}}(\alpha \mid \mu, I)$$

$$= \sum_{h \in I} \mathbb{P}_{\text{GT}}(h \mid \mu, I) \cdot \Big( \sum_{a \in A_I} \alpha(a) \cdot U(\mu \mid h \circ a) \Big)$$

$$= \sum_{h \in I} \chi(h \in I) \cdot \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I \mid \mu)} \cdot \sum_{a \in A_I} \alpha(a) \cdot U(\mu \mid h \circ a)$$

$$= \sum_{h \in I} 1 \cdot \frac{\mathbb{P}(h \mid \mu)}{\mathbb{P}(I \mid \mu)} \cdot \sum_{a \in A_I} \mu_{I \mapsto \alpha}(a \mid I) \cdot U(\mu_{I \mapsto \alpha} \mid h \circ a)$$

$$= \sum_{h \in I} \frac{\mathbb{P}(h \mid \mu_{I \mapsto \alpha})}{\mathbb{P}(I \mid \mu_{I \mapsto \alpha})} \cdot U(\mu_{I \mapsto \alpha} \mid h)$$

$$= \sum_{h \in I} \frac{\mathbb{P}(h \mid \mu_{I \mapsto \alpha})}{\mathbb{P}(I \mid \mu_{I \mapsto \alpha})} \cdot$$

$$\sum_{z \in \mathcal{Z}} \mathbb{P}(z \mid \mu_{I \mapsto \alpha}, h) \cdot \chi(h \in \text{hist}(z)) \cdot u(z)$$

$$= \sum_{z \in \mathcal{Z}} u(z) \cdot \sum_{h \in I \cap \text{hist}(z)} \frac{\mathbb{P}(h \mid \mu_{I \mapsto \alpha})}{\mathbb{P}(I \mid \mu_{I \mapsto \alpha})} \cdot \mathbb{P}(z \mid \mu_{I \mapsto \alpha}, h)$$

$$= \sum_{z \in \mathcal{Z}} u(z) \cdot \sum_{h \in I \cap \text{hist}(z)} \frac{\mathbb{P}(z \mid \mu_{I \mapsto \alpha})}{\mathbb{P}(I \mid \mu_{I \mapsto \alpha})}$$

$$= \sum_{z \in \mathcal{Z}} u(z) \cdot |I \cap \text{hist}(z)| \cdot \frac{\mathbb{P}(z \mid \mu_{I \mapsto \alpha})}{\mathbb{P}(I \mid \mu_{I \mapsto \alpha})}$$

$$= \sum_{z \in \mathcal{Z}} u(z) \cdot \chi(I \cap \text{hist}(z) \neq \emptyset) \cdot \frac{\mathbb{P}(z \mid \mu_{I \mapsto \alpha})}{\mathbb{P}(I \mid \mu_{I \mapsto \alpha})}$$

$$= \sum_{z \in \mathcal{Z}} u(z) \cdot \mathbb{P}_{\text{GDH}}(\text{hist}(z) \mid \mu_{I \mapsto \alpha}, I)$$

$$= \text{EU}_{\text{EDT,GDH}}(\alpha \mid \mu, I)$$

As a consequence, we get Lemma 13:

**Lemma.** *In games without absentmindedness, a strategy $\mu$ is a (CDT,GT)-equilibrium if and only if it is an (EDT,GDH)-equilibrium.*

## F    On Section 3.4: Proofs

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

### F.1    Proof of Lemma 14

We identified the strategy space $\times_{i=1}^{\ell} \Delta(A_{I_i})$ of a game $\Gamma$ with $\times_{i=1}^{\ell} \Delta^{m_i - 1}$ and established that the strategy utility function $U$ extends to all $\times_{i=1}^{\ell} \mathbb{R}^{m_i}$. In order to discuss differentiabity, we view elements $\mu \in \times_{i=1}^{\ell} \mathbb{R}^{m_i}$ as flattened vectors in $\mathbb{R}^{\sum_{i=1}^{\ell} m_i}$. However, for notational convenience, we keep the vector description $\mu = (\mu_{ij})_{i,j=1}^{\ell, m_i} \in \times_{i=1}^{\ell} \mathbb{R}^{m_i}$ because it is indexed by a (info set, action)-pair for our game-theoretic perspective. Therefore, the basis vectors that span $\times_{i=1}^{\ell} \mathbb{R}^{m_i}$ are the vectors $e_{ij}$, for $i \in [\ell]$ 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 $U$ has one partial derivative $\nabla_{ij} U$ for each coordinate direction $e_{ij}$ which is defined as

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

We are ready to prove Lemma 14.

**Lemma.** *Let $I_i$ be an info set, $a_j \in A_{I_i}$ an action, and $\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})$ a strategy of the game. Then:*
*1. $\nabla_{ij} U(\mu) = 0$  if $\text{Fr}(I_i \mid \mu) = 0$, and*
*2. $\nabla_{ij} U(\mu) = \text{Fr}(I_i \mid \mu) \cdot \text{EU}_{\text{CDT,GT}}(a_j \mid \mu, I_i)$ otherwise.*

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

$$U(\mu + \epsilon \cdot e_{ij})$$

$$= \sum_{z \in \mathcal{Z}} u(z) \cdot \qquad\qquad\qquad\qquad\qquad (9)$$

$$\prod_{k=0}^{d(z)-1} \Big( \mu\big(a(z,k) \mid I_{\nu(z,k)}\big) + \epsilon \cdot e_{ij}\big(a(z,k) \mid I_{\nu(z,k)}\big) \Big)$$

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 $\epsilon$:

$$= \sum_{z \in \mathcal{Z}} \left( u(z) \prod_{k=0}^{d(z)-1} \mu\big(a(z,k) \mid I_{\nu(z,k)}\big) \right) +$$

$$\sum_{z \in \mathcal{Z}} \left( u(z) \sum_{h \in \text{hist}(z)} \left[ \epsilon \cdot e_{ij}\big(a(z, d(h)) \mid I_h\big) \cdot \right. \right.$$

$$\left. \left. \prod_{\substack{k=0 \\ k \neq d(h)}}^{d(z)-1} \mu\big(a(z,k) \mid I_{\nu(z,k)}\big) \right] \right) + \mathcal{O}(\epsilon^2)$$

$$= U(\mu) + \epsilon \cdot \sum_{z \in \mathcal{Z}} \left[ u(z) \cdot \sum_{h \in \text{hist}(z)} e_{ij}\big(a(z, d(h)) \mid I_h\big) \cdot \right.$$

$$\left. \mathbb{P}(h \mid \mu) \cdot \mathbb{P}\big(z \mid \mu, h \circ a(z, d(h))\big) \right] + \mathcal{O}(\epsilon^2)$$

$$\overset{(*)}{=} U(\mu) + \mathcal{O}(\epsilon^2) +$$

$$\epsilon \cdot \sum_{z \in \mathcal{Z}} \sum_{h \in \text{hist}(z) \cap I_i} \chi\big(a_j \neq a(z, d(h))\big) \cdot u(z) \cdot$$

$$\mathbb{P}(h \mid \mu) \cdot \mathbb{P}\big(z \mid \mu, h \circ a(z, d(h))\big)$$

$$\overset{(**)}{=} U(\mu) + \mathcal{O}(\epsilon^2) +$$

$$\epsilon \cdot \underbrace{\sum_{z \in \mathcal{Z}} \sum_{h \in I_i} u(z) \cdot \mathbb{P}(h \mid \mu) \cdot \mathbb{P}\big(z \mid \mu, h \circ a_j\big)}_{\text{Denote this term as } (\dagger)}$$

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 $h \neq \text{hist}(z)$ or $a_j \neq a(z, d(h))$.

Term $(\dagger)$ is constant in $\epsilon$. Once we derived a better expression for $(\dagger)$, we can obtain the statement of the lemma from

$$\nabla_{ij} U(\mu) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \big( U(\mu + \epsilon \cdot e_{ij}) - U(\mu) \big)$$

$$= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \big( U(\mu) + \epsilon \cdot (\dagger) + \mathcal{O}(\epsilon^2) - U(\mu) \big)$$

$$= (\dagger).$$

Consider the case where $0 = \text{Fr}(I_i \mid \mu) = \sum_{h \in I_i} \mathbb{P}(h \mid \mu)$. Recall that $\mu \in \times_{i=1}^{\ell} \Delta^{m_i - 1}$. 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 $\text{Fr}(I_i \mid \mu) > 0$.

Then, we can simplify:

$$(\dagger) = \sum_{h \in I_i} \sum_{z \in \mathcal{Z}} u(z) \cdot \mathbb{P}(h \mid \mu) \cdot \mathbb{P}\big(z \mid \mu, h \circ a_j\big)$$

$$= \frac{\text{Fr}(I_i \mid \mu)}{\text{Fr}(I_i \mid \mu)} \cdot \sum_{h \in I_i} \mathbb{P}(h \mid \mu) \cdot \sum_{z \in \mathcal{Z}} u(z) \cdot \mathbb{P}\big(z \mid \mu, h \circ a_j\big)$$

$$= \text{Fr}(I_i \mid \mu) \cdot \sum_{h \in I_i} \chi(h \in I_i) \cdot \frac{\mathbb{P}(h \mid \mu)}{\text{Fr}(I_i \mid \mu)} \cdot U(\mu \mid h \circ a_j)$$

$$= \text{Fr}(I_i \mid \mu) \cdot \sum_{h \in I_i} \mathbb{P}_{\text{GT}}(h \mid \mu, I_i) \cdot U(\mu \mid h \circ a_j)$$

$$= \text{Fr}(I_i \mid \mu) \cdot \text{EU}_{\text{CDT,GT}}(a_j \mid \mu, I_i)$$

$$\square$$

### F.2 Proof of Lemma 15

**Lemma.** *Strategy* $\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})$ *of a game* $\Gamma$ *is an (EDT,GDH)-equilibrium if and only if for all* $i \in [\ell]$:

$$\mu_{i \cdot} \in \underset{y \in \Delta(A_{I_i})}{\text{argmax}} \; U(\mu_{1 \cdot}, \ldots, \mu_{i-1 \cdot}, y, \mu_{i+1 \cdot}, \ldots, \mu_{\ell \cdot}).$$

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

$$\text{EU}_{\text{EDT,GDH}}(\alpha \mid \mu, I_i)$$

$$= \sum_{z \in \mathcal{Z}} \mathbb{P}_{\text{GDH}}(\text{hist}(z) \mid \mu_{I_i \mapsto \alpha}, I_i) \cdot u(z)$$

$$= \sum_{z \in \mathcal{Z}} \chi(I \cap \text{hist}(z) \neq \emptyset) \cdot \frac{\mathbb{P}(z \mid \mu_{I_i \mapsto \alpha})}{\mathbb{P}(I_i \mid \mu_{I_i \mapsto \alpha})} \cdot u(z)$$

$$= \frac{1}{\mathbb{P}(I_i \mid \mu_{I_i \mapsto \alpha})} \cdot \sum_{\substack{z \in \mathcal{Z} \\ I_i \cap \text{hist}(z) \neq \emptyset}} \mathbb{P}(z \mid \mu_{I_i \mapsto \alpha}) \cdot u(z) \quad (10)$$

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 $\alpha$). First, multiply (10) by $\alpha$-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 $\text{hist}(z)$, we have $\mathbb{P}(z \mid \mu_{I_i \mapsto \alpha}) = \mathbb{P}(z \mid \mu)$. Hence, we can add $\alpha$-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

$$\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ \mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha\mid\mu,I_i)$$

$$\overset{(10)}{=}\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ \frac{1}{\mathbb{P}(I_i\mid\mu_{I_i\mapsto\alpha})}\cdot\sum_{\substack{z\in\mathcal{Z}\\ I\cap\mathrm{hist}(z)\neq\emptyset}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z)$$

$$=\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ \sum_{\substack{z\in\mathcal{Z}\\ I\cap\mathrm{hist}(z)\neq\emptyset}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z)$$

$$=\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ \sum_{z\in\mathcal{Z}}\mathbb{P}(z\mid\mu_{I_i\mapsto\alpha})\cdot u(z)=\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ U(\mu_{I_i\mapsto\alpha})$$

$$=\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ U(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{\ell\cdot}).$$

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

All in all, we obtain that $\mu$ is an (EDT,GDH)-equilibrium

$$\iff \forall i\in[\ell]\ \text{with}\ \mathbb{P}(I_i\mid\mu)>0:$$
$$\mu(\cdot\mid I_i)\in\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ \mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha\mid\mu,I_i)$$

$$\iff \forall i\in[\ell]\ \text{with}\ \mathbb{P}(I_i\mid\mu)>0:$$
$$\mu_{i\cdot}\in\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ U(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{\ell\cdot})$$

$$\overset{(*)}{\iff} \forall i\in[\ell]:$$
$$\mu_{i\cdot}\in\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ U(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{\ell\cdot})$$

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 $U(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{\ell\cdot})$ is constant in the action choice $\alpha$ at info set $I_i$. Therefore, any $\alpha'\in\Delta(A_{I_i})$ satisfies

$$\alpha'\in\underset{\alpha\in\Delta(A_{I_i})}{\mathrm{argmax}}\ U(\mu_{1\cdot},\ldots,\mu_{i-1\cdot},\alpha,\mu_{i+1\cdot},\ldots,\mu_{\ell\cdot}).$$

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

# G   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 $x^*\approx1.1673$ 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.

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
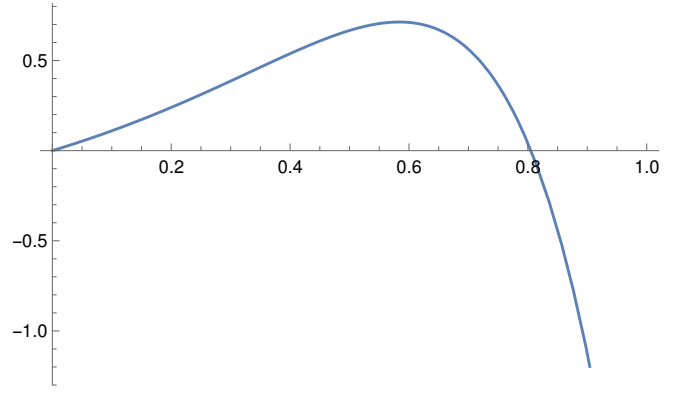


Figure 5: A plot of the function $p(x)=-\frac{16}{3}x^6+x^2+x$.

solution, $\tilde{x}=x^*/2\approx0.58365$. 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 5. 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 6. Let $U\colon\mathbb{R}^2\to\mathbb{R}\colon(x,y)\mapsto-\frac{16}{3}x^6+x^2+x=p(x)$ 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 14, this means that we must have $\frac{d}{dx}U(x,y)=\frac{d}{dy}U(x,y)=0$. Therefore, the only (CDT,GT)-equilibrium is at $x=\tilde{x}$. By Lemma 17, 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).

# H   Proofs of the Main Results

## H.1   Proof of Theorem 1

Recall the problem of maximizing the ex-ante utility in a game $\Gamma$:

$$\max_{\mu\in\times_{i=1}^{\ell}\mathbb{R}^{m_i}}\quad U(\mu)$$
$$\text{s.t.}\qquad \mu_{ij}\geq0\quad\forall i\in[\ell],\forall j\in[m_i]$$
$$\sum_{j=1}^{m_i}\mu_{ij}=1\quad\forall i\in[\ell]\tag{11}$$

Then, the KKT conditions for (11) and a point $\mu\in\times_{i=1}^{\ell}\mathbb{R}^{m_i}$ are: There exist KKT multipliers $\{\tau_{ij}\in\mathbb{R}\}_{i,j=1}^{\ell,m_i}$
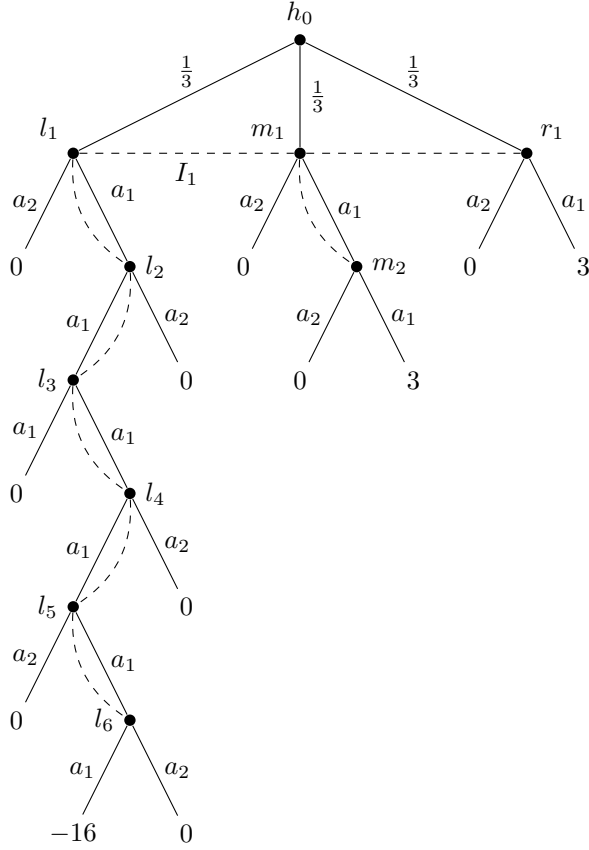
Figure 6: A single-player game of imperfect recall whose only solution cannot be expressed in radicals.

and $\{\kappa_i \in \mathbb{R}\}_{i=1}^{\ell}$ such that

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

We are ready to prove Theorem 1.

**Theorem.** *Strategy $\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})$ of $\Gamma$ is a (CDT,GT)-equilibrium if and only if $\mu$ is a KKT point of (11).*

*Proof.* " $\Longrightarrow$ ":
Suppose $\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})$ is a (CDT,GT)-equilibrium of $\Gamma$. Then the first two KKT conditions of (12) are satisfied by assumption. Let $i \in [\ell]$ and $\forall j \in [m_i]$.

If $\mathrm{Fr}(I_i \mid \mu) = 0$, then by Lemma 14 we have $\nabla_{ij} U(\mu) = 0$. 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

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

The choice of $\tau_{ij}$ depends on $\mu_{ij}$. If $\mu_{ij} > 0$, then by Lemma 14 and by the equivalent characterization of a (CDT,GT)-equilibrium right below Definition 10, we have

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

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

" $\Longleftarrow$ ":
Suppose $\mu$ is a KKT point of problem (11), that is, it satisfies the KKT conditions (12) for some KKT multipliers $\{\tau_{ij} \in \mathbb{R}\}_{i,j=1}^{\ell,m_i}$ and $\{\kappa_i \in \mathbb{R}\}_{i=1}^{\ell}$. Then, the first two KKT conditions ensure that $\mu$ makes a valid strategy for $\Gamma$ (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 $i \in [\ell]$ and $j \in [m_i]$ be such that $\mathrm{Fr}(I_i \mid \mu) > 0$ and $\mu_{ij} > 0$. Then, Lemma 14 and the last three KKT conditions give us for all $j' \in [m_i]$:

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

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 $\mu$ is a (CDT,GT)-equilibrium for $\Gamma$ $\qquad\square$

### H.2 Reproof of Lemma 17

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

**Lemma.** *An ex-ante optimal strategy of a game $\Gamma$ is also an (EDT,GDH)-equilibrium. An (EDT,GDH)-equilibrium is also a (CDT,GT)-equilibrium.*

*In particular, any single-player extensive-form game $\Gamma$ with imperfect recall 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 (1). Therefore, it will in particular satisfy the characterization of an (EDT,GDH)-equilibrium according to Lemma 15.

Let us show that an (EDT,GDH)-equilibrium is a (CDT,GT)-equilibrium by using their respective ex-ante characterizations from Lemma 15 and Theorem 1. So let $\mu = \in$

$\times_{i=1}^{\ell} \Delta(A_{I_i})$ satisfy for all $i \in [\ell]$

$$\mu_{i\cdot}^* \in \underset{y \in \Delta(A_{I_i})}{\text{argmax}} \; U(\mu_1^*, \ldots, \mu_{i-1\cdot}^*, y, \mu_{i+1\cdot}^*, \ldots, \mu_{\ell\cdot}^*).$$

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

$$q_{i,\mu} : \Delta^{m_i-1} \to \mathbb{R}$$
$$y \mapsto U(\mu_{1\cdot}, \ldots, \mu_{i-1\cdot}, y, \mu_{i+1\cdot}, \ldots, \mu_{\ell\cdot}).$$

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

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

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

$$\max_{y \in \Delta^{m_i-1}} \; q_{i,\mu}(y).$$

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

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

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

Lastly, since there always exists an ex-ante optimal strategy (as a maximum (1) of a continuous polynomial function over a compact domain), there also always exists an (EDT,GDH)-equilibrium and (CDT,GT)-equilibrium. $\qquad \square$

## H.3 Approximate and Well-Supported (CDT,GT)-equilibria

Before we get to the proof of Theorem 2, we should discuss two definitions of being close to a (CDT,GT)-equilibrium that are useful to computatational considerations:

**Definition 19.** *Let $\Gamma$ be a single-player extensive-form game with imperfect recall, $\mu$ be a strategy for $\Gamma$, and $\epsilon > 0$ be a precision parameter. Then:*

- *$\mu$ is called an $\epsilon$-approximate (CDT,GT)-equilibrium if for all $I \in \mathcal{I}_*$ with $\text{Fr}(I \mid \mu) > 0$:*

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

- *$\mu$ is called an $\epsilon$-well-supported (CDT,GT)-equilibrium if for all $I \in \mathcal{I}_*$ with $\text{Fr}(I \mid \mu) > 0$ and all $a \in A_I$ with $\mu(a \mid I) > 0$:*

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

Identity (6) directly implies the following relationship:

**Lemma 20.** *Let $\Gamma$ be a single-player extensive-form game with imperfect recall and $\epsilon > 0$. Then any $\epsilon$-well-supported (CDT,GT)-equilibrium $\mu$ is also an $\epsilon$-approximate (CDT,GT)-equilibrium.*

We can also prove a polynomial-time reduction in the reverse direction for all those games that have admit the following characteristic:

**Definition 21.** *Ia single-player extensive-form game $\Gamma$ (with imperfect recall), a value $\lambda > 0$ is said to be a lower bound on positive visit frequencies in $\Gamma$ if it satisfies for all info sets $I$ and all strategies $\mu$ in $\Gamma$*

$$\text{Fr}(I \mid \mu) = 0 \quad \text{or} \quad \text{Fr}(I \mid \mu) \geq \lambda.$$

Since $\text{Fr}(I \mid \mu)$ as a function in $\mu$ is a polynomial function, and therefore continuous over the connected strategy set $\times_{i=1}^{\ell} \Delta(A_{I_i})$, this definition becomes equivalent to

$$\forall I \in \mathcal{I}_* \; : \; \text{Fr}(I \mid \mu) = 0 \, \forall \mu \quad \text{or} \quad \text{Fr}(I \mid \mu) \geq \lambda \, \forall \mu. \quad (13)$$

While the next lemma seems to be novel for imperfect-recall games, the proof borrows its main ideas from the equivalency of $\epsilon$-approximate Nash equilibria and $\epsilon$-well-supported Nash equilibria, shown by Chen *et al.* [2006]. In the below formulation and proof, any roots $\sqrt{x}$ taken shall always refer to the positive valued root.

**Lemma 22.** *Let $\Gamma$ be a single-player extensive-form game with imperfect recall with a lower bound $\lambda \in \mathbb{Q}$ on its positive visit frequencies. Then we can compute a Lipschitz constant $L \geq 1$ w.r.t. the infinity norm $\|\cdot\|_\infty$ for all functions $\big(\text{EU}_{\text{CDT,GT}}(a \mid \cdot, I)\big)_{I \in \mathcal{I}, \text{Fr}(I \mid \cdot) \geq \lambda, a \in A_I}$ in $\mu$ over the strategy space $\times_{i=1}^{\ell} \Delta(A_{I_i})$. Let $\epsilon > 0$ such that $\epsilon < \frac{1}{m_i}$ for all info sets $I_i \in \mathcal{I}_*$.*

*Then, given an $\epsilon$-approximate (CDT,GT)-equilibrium $\mu$ of $\Gamma$, we can compute a $(3L|H|\sqrt{\epsilon})$-well-supported (CDT,GT)-equilibrium of $\Gamma$ in polynomial time.*

*Proof.* Let us observe how to compute such a Lipschitz constant $L$ in polynomial time. We first describe the general process of computing a Lipschitz constant $L_q$ of any polynomial function $q : \mathbb{R}^n \to \mathbb{R}^m$ w.r.t. the infinity norm and over the hypercube $[0,1]^n$. One possible Lipschitz constant is

$$\max_{x \in [0,1]^n} \{\|\nabla q(x)\|_\infty\} = \max_{x \in [0,1]^n} \max_{i \in [n]} |\nabla_i q(x)|$$
$$= \max_{i \in [n]} \max_{x \in [0,1]^n} |\nabla_i q(x)|.$$

Note that $\nabla_i q(x)$ is a polynomial function itself and that the factors $x_i^k$ in its monomials are $< 1$. Thus, $\max_{x \in [0,1]^n} |\nabla_i q(x)|$ can be upper bounded by the absolute value sum $\sum_{\lambda_D \in \nabla_i q} |\lambda_D|$ of all coefficients present in $\nabla_i q$. Taking the maximum over index $i$ over those upper bound sums gives us a Lipschitz value $L_q$ that can be computed in polynomial time in the description size of $q$.

Going back to the setup of this lemma, take an info set $I_i$ of $\Gamma$ with $\text{Fr}(I_i \mid \cdot) \geq \lambda$ and an action $a_j \in A_i$. Then we can

derive as possible Lipschitz constant for $\mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \cdot, I_i)$:

$$\max_{\mu \in \times_{i=1}^{\ell} \Delta(A_{I_i})} \{\|\nabla_\mu \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i)\|_\infty\}$$

$$\leq \max_{\mu \in [0,1]^{\Sigma_i m_i}} \{\|\nabla_\mu \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i)\|\}$$

$$= \max_{\mu \in [0,1]^{\Sigma_i m_i}} \left\{\left\|\nabla_\mu \left(\frac{\nabla_{ij} U(\mu)}{\mathrm{Fr}(I_i \mid \mu)}\right)\right\|\right\}$$

$$= \max_{\mu \in [0,1]^{\Sigma_i m_i}}$$

$$\left\{\left\|\frac{\mathrm{Fr}(I_i \mid \mu) \cdot \nabla_\mu \nabla_{ij} U(\mu) - \nabla_{ij} U(\mu) \cdot \nabla_\mu \mathrm{Fr}(I_i \mid \mu)}{\mathrm{Fr}(I_i \mid \mu)^2}\right\|\right\}$$

$$\leq \max_\mu \frac{1}{\mathrm{Fr}(I_i \mid \mu)^2} \cdot \left(\max_\mu \mathrm{Fr}(I_i \mid \mu) \cdot \max_\mu \|\nabla_\mu \nabla_{ij} U(\mu)\|\right.$$

$$\left. + \max_\mu |\nabla_{ij} U(\mu)| \cdot \max_\mu \|\nabla_\mu \mathrm{Fr}(I_i \mid \mu)\|\right)$$

$$\leq \frac{1}{\lambda^2} \cdot \left(|\mathcal{H}| \cdot \max_\mu \|\nabla_\mu \nabla_{ij} U(\mu)\|\right.$$

$$\left. + \max_\mu |\nabla_{ij} U(\mu)| \cdot \max_\mu \|\nabla_\mu \mathrm{Fr}(I_i \mid \mu)\|\right).$$

The terms $\max_\mu \|\nabla_\mu \nabla_{ij} U(\mu)\|$ and $\max_\mu \|\nabla_\mu \mathrm{Fr}(I_i \mid \mu)\|$ as well $\max_\mu |\nabla_{ij} U(\mu)|$ can be bounded above with the method described in the previous paragraph. The final value can be chosen as our Lipschitz constant $L_{ij}$ for $\mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \cdot, I_i)$. Finally, we can choose

$$L = \max_{I_i \in \mathcal{I}_*, \mathrm{Fr}(I_i \mid \cdot) \geq \lambda, a_j \in A_I} \{1, \max_{i,j} L_{ij}\}$$

as the desired mutual Lipschitz constant. This construction takes polynomial time in the encoding of $\Gamma$ and $\lambda$.

It remains to prove the first result. Denote with $U_{ij}(\mu) := \mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \mu, I_i)$ the (CDT,GT)-expected utility function in $\mu$ from being in info set $I_i$ and using action $a_j$. Let $\mu$ be an $\epsilon$-approximate (CDT,GT)-equilibrium of $\Gamma$, and take an info set $I_i$ with $\mathrm{Fr}(I_i \mid \cdot) > 0$, thus $\mathrm{Fr}(I_i \mid \cdot) \geq \lambda$. As the first step, let us show that any action that sufficiently suboptimal will also be played with only very low probability. Assume $a_{j'} \in A_{I_i}$ is a suboptimal action, that is, there exists some $a_j \in A_{I_i}$ such that $U_{ij}(\mu) \geq U_{ij'}(\mu) + \sqrt{\epsilon}$. Consider the mixed action $\alpha \in \Delta(A_{I_i})$ defined as $\alpha_{j'} = 0$, $\alpha_j = \mu_{ij} + \mu_{ij'}$, and $\alpha_{\hat{\jmath}} = \mu_{i\hat{\jmath}}$ for all $\hat{\jmath} \neq j, j'$. Since $\mu$ is assumed to be an $\epsilon$-approximate (CDT,GT)-equilibrium, we get

$$\sum_{\hat{\jmath} \in [m_i]} \mu_{i\hat{\jmath}} \cdot U_{i\hat{\jmath}}(\mu) \overset{(6)}{=} \mathrm{EU}_{\mathrm{CDT,GT}}(\mu(\cdot \mid I_i) \mid \mu, I_i)$$

$$\geq \max_{k \in [m_i]} \mathrm{EU}_{\mathrm{CDT,GT}}(a_k \mid \mu, I_i) - \epsilon$$

$$= -\epsilon + \sum_{\hat{\jmath} \in [m_i]} \alpha_{\hat{\jmath}} \cdot \max_{k \in [m_i]} U_{ik}(\mu)$$

$$\geq -\epsilon + \sum_{\hat{\jmath} \in [m_i]} \alpha_{\hat{\jmath}} \cdot U_{ij}(\mu)$$

$$= -\epsilon + (\mu_{ij} + \mu_{ij'}) \cdot U_{ij}(\mu) + \sum_{\hat{\jmath} \neq j, j'} \mu_{i\hat{\jmath}} \cdot U_{i\hat{\jmath}}(\mu).$$

Rearranging yields

$$\mu_{ij'} \cdot U_{ij'}(\mu) \geq \mu_{ij'} \cdot U_{ij}(\mu) - \epsilon.$$

Using $U_{ij}(\mu) \geq U_{ij'}(\mu) + \sqrt{\epsilon}$, we therefore get

$$\mu_{ij'} \leq \frac{\epsilon}{U_{ij}(\mu) - U_{ij'}(\mu)} \leq \frac{\epsilon}{\sqrt{\epsilon}} = \sqrt{\epsilon}.$$

In other words, a suboptimal action $a_{j'} \in A_{I_i}$ in an $\epsilon$-approximate (CDT,GT)-equilibrium will be played with low probability. Denote with $\mathrm{low}_i \subseteq [m_i]$ the set of all those action indices $j'$ such that $a_{j'}$ is played with $\leq \sqrt{\epsilon}$ probability in $\mu$. Since we assumed $\sqrt{\epsilon} < \frac{1}{m_i}$, there will be at least one action index $j \in [m_i] \setminus \mathrm{low}_i$. We can therefore create a new candidate strategy $\pi$ from $\mu$ by redistributing the probability of the $\mathrm{low}_i$ actions to the others and achieve well-supportedness for $\pi$. More precisely, define $\pi \in \times_{i=1}^{\ell} \Delta^{m_i-1}$ as follows:

- for all $I_i$ with $\mathrm{Fr}(I_i \mid \cdot) = 0$ and all $a_j \in A_{I_i}$ set $\pi_{ij} := \mu_{ij}$,
- for all $I_i$ with $\mathrm{Fr}(I_i \mid \cdot) \geq \lambda$ and all $a_j \in A_{I_i}$ with $\mu_{ij} \leq \sqrt{\epsilon}$ set $\pi_{ij} = 0$, and
- for all $I_i$ with $\mathrm{Fr}(I_i \mid \cdot) \geq \lambda$ and all $a_j \in A_{I_i}$ with $\mu_{ij} > \sqrt{\epsilon}$ set

$$\pi_{ij} = \mu_{ij} + \frac{\sum_{\hat{\jmath} \in \mathrm{low}_i} \mu_{i\hat{\jmath}}}{m_i - |\mathrm{low}_i|}.$$

This is easily computable. Let us prove that $\pi$ is a $(3L|H|\sqrt{\epsilon})$-well-supported (CDT,GT)-equilibrium: First note $\|\pi - \mu\|_\infty \leq \sqrt{\epsilon} \cdot |H|$. Let $I_i \in \mathcal{I}_*$ again be an info set with $\mathrm{Fr}(I_i \mid \cdot) > 0$, hence $\mathrm{Fr}(I_i \mid \cdot) \geq \lambda$, and let an action $a_j \in A_{I_i}$ have $\pi(a_j \mid I_i) > 0$. Recall that $L$ serves as a Lipschitz constant on $U_{ij}$. By construction of $\pi$, we get $\mu_{ij} > \sqrt{\epsilon}$ which implies that $a_j$ could not have been suboptimal for $\mu$, that is,

$$U_{ij}(\mu) \geq \max_{\hat{\jmath} \in [m_i]} U_{i\hat{\jmath}}(\mu) - \sqrt{\epsilon}. \tag{14}$$

Therefore,

$$\mathrm{EU}_{\mathrm{CDT,GT}}(a_j \mid \pi, I_i)$$

$$= U_{ij}(\pi) = U_{ij}(\pi) - U_{ij}(\mu) + U_{ij}(\mu)$$

$$\geq -L \cdot \|\pi - \mu\| + U_{ij}(\mu)$$

$$\overset{(14)}{\geq} -L \cdot \sqrt{\epsilon} \cdot |H| + \max_{\hat{\jmath} \in [m_i]} U_{i\hat{\jmath}}(\mu) - \sqrt{\epsilon}$$

$$\geq -2L|H|\sqrt{\epsilon} + \max_{\hat{\jmath} \in [m_i]} \{U_{i\hat{\jmath}}(\mu) - U_{i\hat{\jmath}}(\pi) + U_{i\hat{\jmath}}(\pi)\}$$

$$\geq -2L|H|\sqrt{\epsilon} + \max_{\hat{\jmath} \in [m_i]} \{-L \cdot \|\pi - \mu\| + U_{i\hat{\jmath}}(\pi)\}$$

$$\geq -3L|H|\sqrt{\epsilon} + \max_{\hat{\jmath} \in [m_i]} U_{i\hat{\jmath}}(\pi)$$

$$= \max_{a_{\hat{\jmath}} \in A_{I_i}} \mathrm{EU}_{\mathrm{CDT,GT}}(a_{\hat{\jmath}} \mid \pi, I_i) - 3L|H|\sqrt{\epsilon}$$

$$\square$$

## H.4 Proof of Theorem 2

**Theorem.** (CDT,GT)-EQUILIBRIUM *is CLS-hard. CLS-hardness holds even for games restricted to:*

*(1.) a tree depth of* 6 *and the player has* 2 *actions per info set,*

*(2.) no absentmindedness and a tree depth of 6, and*
*(3.) no chance nodes, a tree depth of 5, and only one info set.*

*The problem is in CLS for the subclass of problem instances of* (CDT,GT)-EQUILIBRIUM *where a lower bound on positive visit frequencies in* $\Gamma$ *is easily obtainable.*

We prove this result in four parts.

**CLS Membership of restricted** (CDT,GT)-EQUILIBRIUM
By Theorem 1, (CDT,GT)-equilibria of $\Gamma$ coincide with the KKT points of the strategy utility function $U$ of $\Gamma$. Finding an approximate KKT point of a continuously differentiable function over a compact domain was shown to be CLS-complete by Fearnley *et al.* [2023]:

**Definition 23.** *An instance of the problem* KKT *consists of (1) a precision parameter $\epsilon > 0$ encoded in binary, (2) a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ defining a bounded non-empty domain $D = \{x \in \mathbb{R}^n : Ax \leq b\}$, (3) two well-behaved arithmetic circuits $f : \mathbb{R}^n \to \mathbb{R}$ and $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$, and (4) a Lipschitz constant $L > 0$.*

*A solution consists of a point $x \in D$ such that there exist $\mu_1, \ldots, \mu_m \geq 0$ such that $||\nabla f(x) + A^T \mu|| \leq \epsilon$ and $\mu^T(Ax - b) = 0$. Alternatively, we also accept points $x, y \in D$ as a solution that show that one of the following is true: (i) $f$ or $\nabla f$ is not $L$-Lipschitz, or (ii) $\nabla f$ is not the gradient of $f$.*

**Lemma 24** (Fearnley *et al.* [2021]). KKT *is CLS-complete.*

Fearnley *et al.* [2021] note that any norm can be used in Definition 23. We will henceforth use the infinity norm $||\cdot||_\infty$.

Let us outline the reduction from the restricted (CDT,GT)-EQUILIBRIUM to KKT. Let $(\Gamma, \epsilon)$ be an instance of (CDT,GT)-EQUILIBRIUM for which a lower bound $\lambda$ on positive visit frequencies in $\Gamma$ can be computed in polynomial time. Construct the strategy utility function $U$ of $\Gamma$ and its gradient $\nabla U$. Then $-U$ and $-\nabla U$ as polynomial functions will make well-behaved arithmetic circuits $f$ and $\nabla f$ (the negative sign comes in because KKT is about minimizing $f$ whereas (CDT,GT)-equilibrium is about utility maximization). The domain $D$ will be the Cartesian product of simplices $\times_{i=1}^l \Delta(A_{I_i})$ described with inequalities as in (11). Next, we can construct a Lipschitz constant $L$ for the polynomial functions $-U$ and $-\nabla U$ over $\times_{i=1}^l \Delta(A_{I_i})$ as described in the proof of Lemma 22. Finally, set the precision parameter $\delta$ of the KKT instance to $\delta = \frac{1}{2}\lambda\epsilon$, which is polynomial time computable in the encoding size of $(\Gamma, \epsilon)$.

Get a solution to the KKT instance. We have by construction that $L$ is a valid Lipschitz constant for $f$ and $\nabla f$, and that $\nabla f$ is indeed the gradient of $f$. Thus, the solution has to be a $\delta$-KKT point $\mu$, i.e. have KKT multipliers $\{\tau_{ij} \in \mathbb{R}\}_{i,j=1}^{\ell,m_i}$ and $\{\kappa_i \in \mathbb{R}\}_{i=1}^\ell$ such that

$$\mu_{ij} \geq 0 \quad \forall i \in [\ell], \forall j \in [m_i]$$
$$\sum_{j=1}^{m_i} \mu_{ij} = 1 \quad \forall i \in [\ell]$$
$$\tau_{ij} \geq 0 \quad \forall i \in [\ell], \forall j \in [m_i]$$
$$\tau_{ij} = 0 \quad \text{or} \quad \mu_{ij} = 0 \quad \forall i \in [\ell], \forall j \in [m_i]$$
$$|\nabla_{ij} U(\mu) - (-\tau_{ij} + \kappa_i)| \leq \delta \quad \forall i \in [\ell], \forall j \in [m_i].$$

The point $\mu$ forms a valid strategy of $\Gamma$. Let us show that it is also an $\epsilon$-well-supported (CDT,GT)-equilibrium of $\Gamma$. Let $I_i$ be an info set with $\text{Fr}(I_i \mid \mu) > 0$, hence $\text{Fr}(I_i \mid \mu) \geq \lambda$, and let $a_j \in A_{I_i}$ be an action with $\mu_{ij} = \mu(a_j \mid I_i) > 0$. Then, for any other action $a_{j'} \in A_{I_i}$, we have by Lemma 14 and the KKT conditions:

$$\text{EU}_{\text{CDT,GT}}(a_j \mid \mu, I_i) = \frac{\nabla_{ij} U(\mu)}{\text{Fr}(I_i \mid \mu)} \geq \frac{-\tau_{ij} + \kappa_i - \delta}{\text{Fr}(I_i \mid \mu)}$$
$$= \frac{\kappa_i - \delta}{\text{Fr}(I_i \mid \mu)} \geq \frac{-\tau_{ij'} + \kappa_i - \delta}{\text{Fr}(I_i \mid \mu)} \geq \frac{\nabla_{ij'} U(\mu) - 2\delta}{\text{Fr}(I_i \mid \mu)}$$
$$= \text{EU}_{\text{CDT,GT}}(a_{j'} \mid \mu, I_i) - \frac{2}{\text{Fr}(I_i \mid \mu)} \cdot \delta$$
$$\geq \text{EU}_{\text{CDT,GT}}(a_{j'} \mid \mu, I_i) - \frac{2}{\lambda} \cdot \delta$$
$$= \text{EU}_{\text{CDT,GT}}(a_{j'} \mid \mu, I_i) - \epsilon$$

Thus $\mu$ forms an $\epsilon$-well-supported (CDT,GT)-equilibrium for $\Gamma$, and therefore, also an $\epsilon$-(CDT,GT)-equilibrium for $\Gamma$.

**First CLS Hardness Result of Theorem 2**
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 $p : \mathbb{R}^\ell \to \mathbb{R}$ over the hypercube:

$$\max_{x \in \mathbb{R}^\ell} \quad p(x) \tag{15}$$
$$\text{s.t.} \quad x \in [0, 1]^\ell .$$

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

**Definition 25.** *An instance of the problem* KKTPOLYOVER-CUBE *consists of a polynomial function $p : \mathbb{R}^\ell \to \mathbb{R}$ together with a precision value $\epsilon > 0$. A solution consists of a point $x \in [0,1]^\ell$ that is an $\epsilon$-KKT point of the maximization problem (15):*

$$\forall i \in [\ell] : \quad x_i > 0 \implies \nabla_i p(x) \geq -\epsilon \tag{16}$$
$$\forall i \in [\ell] : \quad x_i < 1 \implies \nabla_i p(x) \leq \epsilon$$

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

**Lemma 26** (Babichenko and Rubinstein [2021]). KKT-POLYOVERCUBE *is CLS-complete. Hardness holds even[2] 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 $(p : \mathbb{R}^\ell \to \mathbb{R}, \epsilon)$ of KKTPOLYOVER-CUBE where $p$ is known to only have degree 5 monomials. There are at most $\binom{\ell+5-1}{5} = \mathcal{O}(\ell^5)$ many such monomials. Consider the modified polynomial

$$\hat{p} : \underset{i=1}{\overset{\ell}{\bigtimes}} \mathbb{R}^2 \qquad \to \mathbb{R} \tag{17}$$
$$\left((x_{i1}, x_{i2})\right)_{i=1}^\ell \qquad \mapsto p(x_{11}, x_{21}, \ldots, x_{\ell 1}) .$$

---

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

Create a game $\Gamma$ out of $\hat{p}$ as described in Appendix A with the second variant, which is constructed in polynomial time in the encoding of $p$ since its degree is known to be 5. Let $|\mathcal{H}|$ be the number of nodes in $\Gamma$. Recall that for any info set $I_i$ of $\Gamma$ its visit frequency $\text{Fr}(I_i \mid \mu)$ is constant in the used strategy $\mu$ and easily computable by (5). Thus, we are able to get a lower bound on the positive visit frequencies, which allows us to use Lemma 22 later on. Note moreover that the visit frequencies are always bounded above by $|H|$. Set

$$\delta := \min\{\frac{1}{3}, \frac{\epsilon^2}{(3L|\mathcal{H}|^2)^2}\}$$

which is computable in polynomial time in the encoding of $p$ and $\epsilon$.

Recall from Appendix A that the strategy utility function $U$ of $\Gamma$ has the property $U(\mu) = \hat{p}(\mu) = p(\mu_{11}, \ldots, \mu_{\ell 1})$ for any point $\mu \in \bigtimes_{i=1}^{\ell} \mathbb{R}^2$. Therefore, $\nabla_{i1} U(\mu) = \nabla_i p(\mu_{11}, \ldots, \mu_{\ell 1})$ and $\nabla_{i2} U(\mu) = 0$ for all $\mu \in \bigtimes_{i=1}^{\ell} \mathbb{R}^2$ and $i \in [\ell]$.

Let $\pi^* \in \bigtimes_{i=1}^{\ell} \Delta^1$ be a a solution to the (CDT,GT)-EQUILIBRIUM-instance $(\Gamma, \delta)$. By Lemma 22, we can compute from it a $(3L|\mathcal{H}|\sqrt{\delta})$-well-supported (CDT,GT)-equilibrium $\mu^* = (\mu_{ij}^*)_{i,j=1}^{\ell,2} \in \bigtimes_{i=1}^{\ell} \Delta^1$ of $\Gamma$ in polynomial time. For us $\mu^*$ is, in particular, a $\frac{\epsilon}{|\mathcal{H}|}$-well-supported (CDT,GT)-equilibrium because we set $\delta$ such that $3L|\mathcal{H}|\sqrt{\delta} \leq \frac{\epsilon}{|\mathcal{H}|}$. We can now show that the point $x^* := (\mu_{i1}^*)_{i=1}^{\ell} \in [0, 1]^{\ell}$ is a solution to the KKTPOLYOVERCUBE-instance $(p : \mathbb{R}^{\ell} \to \mathbb{R}, \epsilon)$.

Let us first consider any index $i \in [\ell]$ with $\text{Fr}(I_i \mid \mu^*) = 0$. Then, by Lemma 14, we get $0 = \nabla_{i1} U(\mu^*) = \nabla_i p(x^*)$. Thus, such indices $i$ always satisfy the conditions (16) independent of the value $x_i^*$.

Now consider an index $i \in [\ell]$ with $\text{Fr}(I_i \mid \mu^*) > 0$ and $0 < x_i^* = \mu_{i1}^*$. Then, due to $\mu^*$ being a $\frac{\epsilon}{|\mathcal{H}|}$-well-supported (CDT,GT)-equilibrium, we get

$$\text{EU}_{\text{CDT,GT}}(a_1 \mid \mu^*, I_i) \geq \max_{j=1,2} \text{EU}_{\text{CDT,GT}}(a_j \mid \mu^*, I_i) - \frac{\epsilon}{|\mathcal{H}|}$$

$$\Longleftrightarrow \text{EU}_{\text{CDT,GT}}(a_1 \mid \mu^*, I_i) \geq \text{EU}_{\text{CDT,GT}}(a_2 \mid \mu^*, I_i) - \frac{\epsilon}{|\mathcal{H}|}$$

With Lemma 14, we can therefore derive

$$\nabla_i p(x^*) = \nabla_{i1} U(\mu^*)$$
$$= \text{Fr}(I_i \mid \mu^*) \cdot \text{EU}_{\text{CDT,GT}}(a_1 \mid \mu^*, I_i)$$
$$\geq \text{Fr}(I_i \mid \mu^*) \cdot \left(\text{EU}_{\text{CDT,GT}}(a_2 \mid \mu^*, I_i) - \frac{\epsilon}{|\mathcal{H}|}\right)$$
$$= \text{Fr}(I_i \mid \mu^*) \cdot \text{EU}_{\text{CDT,GT}}(a_2 \mid \mu^*, I_i) - \text{Fr}(I_i \mid \mu^*) \cdot \frac{\epsilon}{|\mathcal{H}|}$$
$$= \nabla_{i2} U(\mu) - \text{Fr}(I_i \mid \mu^*) \cdot \frac{\epsilon}{|\mathcal{H}|} = -\text{Fr}(I_i \mid \mu^*) \cdot \frac{\epsilon}{|\mathcal{H}|}$$
$$\geq -|\mathcal{H}| \cdot \frac{\epsilon}{|\mathcal{H}|} = -\epsilon$$

Now consider an index $i \in [\ell]$ with $\text{Fr}(I_i \mid \mu^*) > 0$ and $1 > x_i^* = \mu_{i1}^* = 1 - \mu_{i2}^*$, i.e., $\mu_{i2}^* > 0$. Then with analogous

arguments, we get

$$\text{EU}_{\text{CDT,GT}}(a_2 \mid \mu^*, I_i) \geq \max_{j=1,2} \text{EU}_{\text{CDT,GT}}(a_j \mid \mu^*, I_i) - \frac{\epsilon}{|\mathcal{H}|}$$

$$\Longleftrightarrow \text{EU}_{\text{CDT,GT}}(a_1 \mid \mu^*, I_i) \leq \text{EU}_{\text{CDT,GT}}(a_2 \mid \mu^*, I_i) + \frac{\epsilon}{|\mathcal{H}|}$$

Similarly to the other case, we can therefore derive

$$\nabla_i p(x^*) = \nabla_{i1} U(\mu^*)$$
$$= \text{Fr}(I_i \mid \mu^*) \cdot \text{EU}_{\text{CDT,GT}}(a_1 \mid \mu^*, I_i)$$
$$\leq \text{Fr}(I_i \mid \mu^*) \cdot \left(\text{EU}_{\text{CDT,GT}}(a_2 \mid \mu^*, I_i) + \frac{\epsilon}{|\mathcal{H}|}\right)$$
$$= \nabla_{i2} U(\mu) + \text{Fr}(I_i \mid \mu^*) \cdot \frac{\epsilon}{|\mathcal{H}|} = \text{Fr}(I_i \mid \mu^*) \cdot \frac{\epsilon}{|\mathcal{H}|}$$
$$\leq |\mathcal{H}| \cdot \frac{\epsilon}{|\mathcal{H}|} = \epsilon$$

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

Note that in the above reduction, the constructed game $\Gamma$ 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 $\Gamma$ 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 Theorem 2**

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 $V_i : \bigtimes_{i=1}^{n} \Delta(A_i) \to \mathbb{R}$ for each player $i$. A mixed action profile $x = (x_{ij})_{i,j=1}^{n,m_i} \in \bigtimes_{i=1}^{n} \Delta(A_{I_i})$ is called an $\epsilon$-Nash equilibrium of $G$ if for every player $i \in [n]$ and every action $a_j \in A_i$ of player $i$, we have $V_i(x) \geq V_i(x_1, \ldots, x_{i-1}, a_j, x_{i+1}, \ldots, x_n.) - \epsilon$. In standard game theory notation, this would be phrased as $V_i(x) \geq V_i(a_j, x_{-i.}) - \epsilon$. For us, it will be helpful to use the notation of a EDT deviation, in which this condition becomes $V_i(x) \geq V_i(x_{I_i \mapsto a_j}) - \epsilon$.

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 27.** *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 $\epsilon$-Nash equilibrium of the game.*

**Lemma 28** (Babichenko and Rubinstein [2021]). 5-POLYTENSOR-IDENTICALINTEREST *is CLS-complete.*

Without changing the complexity, we may assume that in instances of 5-POLYTENSOR-IDENTICALINTEREST every player has exactly $m$ actions (by possibly copying actions) and that the payoffs lie in $[0, 1]$ (by shifting and rescaling utilities them).

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

Take an instance $(G, \epsilon)$ of 5-POLYTENSOR-IDENTICALINTEREST. Let $n$ be the number of players of $G$, and $m$ be the number of pure strategies of player $i$. Then, each subset of 5 players of $G$ has a table of $m^5$ 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 $\Lambda := \{i_{[5]} \subset [n] : |i_{[5]}| = 5\}$. If $i_{[5]} \in \Lambda$ is a 5-subset and the players of $i_{[5]}$ play strategy profile $\mu_{i_{[5]}} \in \times_{i \in i_{[5]}} \Delta(A_{I_i})$, then denote the payoff that these players get in that subgame by $u^{i_{[5]}}(\mu_{i_{[5]}})$. Then, the overall utility function $V_i$ of a player $i \in [n]$ takes as input a strategy profile $\mu$ for the game $\Gamma$ and returns the payoff:

$$V_i(\mu) = \sum_{i_{[5]} \in \Lambda:\, i \in i_{[5]}} u^{i_{[5]}}(\mu_{i_{[5]}})$$

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

Take an info set $I_i$ of $\Gamma$. The player reaches info set $I_i$ with probability one in every subtree $T_{i_{[5]}}$ with $i \in i_{[5]}$, independent of her strategy choice $\mu$. Thus, the overall reach probability of each info set $I_i$ of $\Gamma$ is exactly the number subsets $i_{[5]} \in \Lambda$ that contain $i$, divided by the number of 5-subsets overall, which is $\binom{n-1}{4}/\binom{n}{5} = \frac{5}{n}$.[3] In particular, this reach probability is non-zero, so an approximate (CDT,GT)-equilibrium must be approximately optimal in every info set $I_i$ of $\Gamma$.

Since there is no absentmindedness in $\Gamma$, we get by Appendix E for all info sets $I_i$, all strategies $\mu$, and all mixed

---

[3]Note that this value would also serve as a lower bound on positive frequencies.

actions $\alpha \in \Delta(A_{I_i})$:

$$
\begin{aligned}
&\mathrm{EU}_{\mathrm{CDT,GT}}(\alpha \mid \mu, I_i) \\
&= \mathrm{EU}_{\mathrm{EDT,GDH}}(\alpha \mid \mu, I_i) \\
&\overset{(10)}{=} \frac{1}{\mathbb{P}(I_i \mid \mu_{I_i \mapsto \alpha})} \cdot \sum_{\substack{z \in \mathcal{Z} \\ I_i \cap \mathrm{hist}(z) \neq \emptyset}} \mathbb{P}(z \mid \mu_{I_i \mapsto \alpha}) \cdot u(z) \\
&= \frac{n}{5} \cdot \sum_{i_{[5]} \in \Lambda:\, i \in i_{[5]}} \sum_{z \in T_{i_{[5]}}} \mathbb{P}(z \mid \mu_{I_i \mapsto \alpha}) \cdot u(z) \\
&= \frac{n}{5} \cdot \sum_{i_{[5]} \in \Lambda:\, i \in i_{[5]}} \sum_{j_{[5]} \in \times_{i \in i_{[5]}} A_{I_i}} \\
&\qquad\qquad \mathbb{P}(z_{i_{[5]}, j_{[5]}} \mid \mu_{I_i \mapsto \alpha}) \cdot u^{i_{[5]}}(j_{[5]}) \\
&\overset{(*)}{=} \frac{n}{5} \cdot \sum_{i_{[5]} \in \Lambda:\, i \in i_{[5]}} \frac{1}{\binom{n}{5}} \cdot u^{i_{[5]}}\Big((\mu_{I_i \mapsto \alpha})_{i_{[5]}}\Big) \\
&= \frac{n}{5} \cdot \frac{1}{\binom{n}{5}} \cdot V_i(\mu_{I_i \mapsto \alpha}) \\
&= \frac{1}{\binom{n-1}{4}} \cdot V_i(\mu_{I_i \mapsto \alpha})
\end{aligned}
$$

where in $(*)$ we used that for any $i_{[5]} = \{i_1, \ldots, i_5\} \in \Lambda$ and strategy $\mu$ that

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

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

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

Therefore, a $\delta$-(CDT,GT)-equilibrium in $\Gamma$ makes an $\epsilon$-Nash equilibrium in $G$.

## Third CLS Hardness Result of Theorem 2

We prove the third CLS hardness of Theorem 2 by again reducing from 5-POLYTENSOR-IDENTICALINTEREST.

Take an instance $(G, \epsilon)$ of 5-POLYTENSOR-IDENTICALINTEREST. Use the same notation as in the last reduction proof from 5-POLYTENSOR-IDENTICALINTEREST. In particular, $G$ has $n$ players with $m$ actions, and payoffs lie in $[0, 1]$. Moreover, we can assume $\epsilon < 1$.

**Corresponding Game** Construct an instance $(\Gamma, \delta)$ of (CDT,GT)-EQUILIBRIUM as follows. $\Gamma$ is a game tree of depth 5 and all nodes of $\Gamma$ will be player nodes that belong to the same info set $I$. The action set $A_I$ is defined as $\{(i, j) : i \in [n], j \in [m]\}$. Nodes of depth $\leq 4$ all have out-degree $n \cdot m$. 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 $(i_{[5]}, j_{[5]}) = ((i_1, j_1), \ldots, (i_5, j_5))$, and if $\eta(i_{[5]}) \in \{1, \ldots, 5\}$ is the number of distinct values present[4] in $i_{[5]}$ of $z$, then $z$ shall yield the player a payoff of

1. $M_2 \cdot \eta(i_{[5]})$ if $\eta(i_{[5]}) \leq 4$, and

2. $M_2 \cdot \eta(i_{[5]}) + u^{i_{[5]}}(j_{[5]})$ if $\eta(i_{[5]}) = 5$.

Note that we have constant visit frequency $\text{Fr}(I \mid \mu) = \text{Fr}(I) = 5$ independent of the strategy $\mu$ used in $\Gamma$. This makes Lemma 22 applicable to $\Gamma$. Let $L$ be the Lipschitz constant $L$ from Lemma 22. Then we can define the values

$$M_1 := 2 \cdot 100 \cdot n^9 m^4 \cdot \frac{1}{\epsilon}$$

$$\delta_1 := \frac{1}{5}\left(\frac{1}{n} - \frac{1}{M_1}\right)^4 \cdot \frac{\epsilon}{2}$$

$$\delta_2 := \left(\frac{\delta_1}{3L \sum_{k=0}^{5}(nm)^k}\right)^2$$

$$M_2 := (\delta_1 + n^4) \cdot M_1^4$$

where the sum $\sum_{k=0}^{5}(nm)^k$ represents the number of nodes $|H|$ in $\Gamma$. To give some intuition, $\delta_2$ is chosen such that Lemma 22 can be applied to get a $\delta_1$-well-supported (CDT,GT)-equilibrium. $M_2$ is chosen large enough (in comparison to the payoffs in $[0, 1]$) such that for equilibrium play, the player of $\Gamma$ mostly cares about mixing up the players $i$ from which she chooses an action in an approximate uniform fashion. $M_1$ is chosen large enough to counterbalance the errors in this approximate uniform mixing, and $\delta_1$ is chosen small enough to recover an $\epsilon$-Nash equilibrium at the end.

Let us show that a $\delta_2$-(CDT,GT)-equilibrium of $\Gamma$ gives rise to an $\epsilon$-Nash equilibrium of $G$. For that, we first have to collect some further observations about $\Gamma$.

**Ex-ante Utility in $\Gamma$** Let us describe the expected utility $U(\mu)$ of a strategy $\mu = (\mu_{ij})_{i,j=1}^{n,m} \in \Delta(A_I)$ of $\Gamma$. Split it into $U(\mu) = M_2\phi(\mu) + \psi(\mu)$ where the first part shall come from the $M_2 \cdot \eta(i_{[5]})$ and the second part from the $u^{i_{[5]}}(j_{[5]})$. The

latter is simply

$$\psi(\mu) = \sum_{(i_{[5]}, j_{[5]}):\eta(i_{[5]})=5} u^{i_{[5]}}(j_{[5]}) \prod_{k=1}^{5} \mu_{i_k j_k}. \quad (18)$$

For the former, denote with $p_i = \sum_j \mu_{ij}$ the total weight on actions that came from player $i$ of the original game. Consider the stochastic experiment of drawing 5 items out of the set of players $[n]$ according to the probability distribution $p = (p_i)_i$. Let $\chi_i \in \{0, 1\}$ be the random variable that denotes whether player $i$ was drawn at least once in the stochastic experiment. Then we can calculate the expected number of distinct players that are drawn in the experiment as

$$\phi(\mu) = \mathbb{E}[\sum_{i \in [n]} \chi_i] = \sum_{i \in [n]} \mathbb{E}[\chi_i]$$

$$= \sum_{i \in [n]} \mathbb{P}(\chi_i = 1) = \sum_{i \in [n]} \left(1 - \mathbb{P}(\chi_i = 0)\right)$$

$$= n - \sum_{i \in [n]} (1 - p_i)^5.$$

**De Se Utility in $\Gamma$** Next, we describe $\text{EU}_{\text{CDT,GT}}(a_{ij} \mid \mu, I)$ for an action $a_{ij}$ at $I$ with the help of Lemma 14.

$$\text{EU}_{\text{CDT,GT}}(a_{ij} \mid \mu, I) = \frac{1}{\text{Fr}(I)}\nabla_{ij} U(\mu) = \frac{1}{5}\nabla_{ij} U(\mu)$$

$$= \frac{M_2}{5} \cdot \nabla_{ij} \phi(\mu) + \frac{1}{5}\nabla_{ij} \psi(\mu)$$

$$= \frac{M_2}{5} \cdot 5(1 - p_i)^4$$

$$+ \frac{1}{5} \sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4, i \notin i_{[4]}}} u^{(i, i_{[4]})}(j, j_{[4]}) \prod_{k=2}^{5} \mu_{i_k j_k}$$

$$\overset{(*)}{\in} \left[M_2(1 - p_i)^4, M_2(1 - p_i)^4 + \frac{1}{5}\binom{n-1}{4}\right]$$

$$\subseteq \left[M_2(1 - p_i)^4, M_2(1 - p_i)^4 + n^4\right].$$

In $(*)$, we used that all factors in the indexed sum lie in $[0, 1]$.

**Solution Recovery** Let $\mu'$ be a $\delta_2$-(CDT,GT)-equilibrium of $\Gamma$. By Lemma 22, we can then compute a $\delta_1$-well-supported (CDT,GT)-equilibrium $\mu$ of $\Gamma$ from $\mu'$. In the remaining part of the proof, we will only work with $\mu$.

**In equilibrium, every player gets to play similarly often** Denote $p_i = \sum_j \mu_{ij}$ again. Take the two players $i^* \in \text{argmax}_i p_i$ and $i_* \in \text{argmax}_i p_i$ that are played the most and least often under $\mu$. Choose any actions $j^*, j_* \in [m]$ of players $i^*$ and $i_*$ respectively. Then we have

$$\delta_1 \geq \text{EU}_{\text{CDT,GT}}(a_{i_* j_*} \mid \mu, I) - \text{EU}_{\text{CDT,GT}}(a_{i^* j^*} \mid \mu, I)$$

$$\geq M_2(1 - p_{i_*})^4 - M_2(1 - p_{i^*})^4 - n^4$$

$$\geq M_2 \cdot \left[(1 - p_{i^*} + p_{i^*} - p_{i_*})^4 - (1 - p_{i^*})^4\right] - n^4$$

$$\overset{(\dagger)}{\geq} M_2 \cdot (p_{i^*} - p_{i_*})^4 - n^4.$$

---

[4]Note our shift in notation from the previous reduction proof. Now, tuple $i_{[5]}$ may contain duplicates from the set of players $[n]$.

In (†) we first used the binomial theorem for the components $1 - p_{i^*} \geq 0$ and $p_{i^*} - p_{i_*} \geq 0$ to split up the first exponent term. Next, the term $(1 - p_{i^*})^4$ canceled out and we dropped three other positive terms. From this, we conclude

$$0 \leq p_{i^*} - p_{i_*} \leq \left| \sqrt[4]{\frac{\delta_1 + n^4}{M_2}} \right| = \frac{1}{M_1} \, .$$

All in all we get that in a $\delta_1$-well-supported (CDT,GT)-equilibrium – such as $\mu$ – the actions $a_{ij}$ of each player $i \in [n]$ get played with summed probability

$$p_i \in \left[ p_{i_*} \, , \, p_{i^*} \right] \subseteq \left[ \frac{1}{n} - \frac{1}{M_1} \, , \, \frac{1}{n} + \frac{1}{M_1} \right] \, .$$

**Recovering an $\epsilon$-Nash equilibrium** We now have everything needed to complete the proof. Define the corresponding strategy profile $\pi = (\pi_i.)_{i=1}^n \in \times_{i=1}^n \Delta^{m-1}$ in $G$ to $\mu$ as $\pi_{ij} := \frac{\mu_{ij}}{p_i}$ for each player $i \in [n]$ and action $j \in [m]$. This is well-defined because $p_i \geq \frac{1}{n} - \frac{1}{M_1} > 0$. We show that $\pi$ is an $\epsilon$-Nash equilibrium of $G$ by contradiction. Assume $\pi$ is not an $\epsilon$-Nash equilibrium. Then, by Chen *et al.* [2006], it is also not an $\epsilon$-well-supported Nash equilibrium , that is, there exists a player $i \in [n]$ with actions $j, j' \in [m]$ such that action $j'$ is played with positive probability $\pi_{ij'} > 0$ and such that playing $j$ yields her more than $\epsilon$ more utility than playing $j'$, given that the other players play $\pi_{-i}.$. Formally, this means

$$
\sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4 \, , i \notin i_{[4]}}} u^{(i, i_{[4]})}(j, j_{[4]}) \prod_{k=2}^{5} \pi_{i_k j_k}
$$

$$
> \epsilon + \sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4 \, , i \notin i_{[4]}}} u^{(i, i_{[4]})}(j', j_{[4]}) \prod_{k=2}^{5} \pi_{i_k j_k} \, . \tag{19}
$$

Since $\frac{1}{n} - \frac{1}{M_1} \leq p_{\hat{i}} \leq \frac{1}{n} + \frac{1}{M_1}$ for all $\hat{i} \in [n]$, we can derive

$$
\sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4 \, , i \notin i_{[4]}}} u^{(i, i_{[4]})}(j, j_{[4]}) \prod_{k=2}^{5} \pi_{i_k j_k}
$$

$$
= \sum u^{(i, i_{[4]})}(j, j_{[4]}) \prod_{k=2}^{5} \frac{\mu_{i_k j_k}}{p_{i_k}}
$$

$$
\leq \sum u^{(i, i_{[4]})}(j, j_{[4]}) \prod_{k=2}^{5} \frac{\mu_{i_k j_k}}{\frac{1}{n} - \frac{1}{M_1}}
$$

$$
= \left( \frac{1}{\frac{1}{n} - \frac{1}{M_1}} \right)^4 \cdot \sum u^{(i, i_{[4]})}(j, j_{[4]}) \prod_{k=2}^{5} \mu_{i_k j_k} \, ,
$$

and analogously

$$
\sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4 \, , i \notin i_{[4]}}} u^{(i, i_{[4]})}(j', j_{[4]}) \prod_{k=2}^{5} \pi_{i_k j_k}
$$

$$
\geq \sum u^{(i, i_{[4]})}(j', j_{[4]}) \prod_{k=2}^{5} \frac{\mu_{i_k j_k}}{\frac{1}{n} + \frac{1}{M_1}}
$$

$$
= \left( \frac{1}{\frac{1}{n} + \frac{1}{M_1}} \right)^4 \cdot \sum u^{(i, i_{[4]})}(j', j_{[4]}) \prod_{k=2}^{5} \mu_{i_k j_k} \, .
$$

Observe the connection to the previously defined function $\psi$ here:

$$
\sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4 \, , i \notin i_{[4]}}} u^{(i, i_{[4]})}(j, j_{[4]}) \prod_{k=2}^{5} \mu_{i_k j_k} = \nabla_{ij} \psi(\mu) \, ,
$$

and

$$
\sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4 \, , i \notin i_{[4]}}} u^{(i, i_{[4]})}(j', j_{[4]}) \prod_{k=2}^{5} \mu_{i_k j_k} = \nabla_{ij'} \psi(\mu) \, .
$$

Inserting the bounds into (19) yields

$$
\left( \frac{1}{\frac{1}{n} - \frac{1}{M_1}} \right)^4 \nabla_{ij} \psi(\mu) > \epsilon + \left( \frac{1}{\frac{1}{n} + \frac{1}{M_1}} \right)^4 \nabla_{ij'} \psi(\mu)
$$

$$
= \epsilon + \left( \frac{1}{\frac{1}{n} - \frac{1}{M_1}} \right)^4 \nabla_{ij'} \psi(\mu)
$$

$$
- \left[ \left( \frac{1}{\frac{1}{n} - \frac{1}{M_1}} \right)^4 - \left( \frac{1}{\frac{1}{n} + \frac{1}{M_1}} \right)^4 \right] \cdot \nabla_{ij'} \psi(\mu)
$$

$$
= \epsilon + \left( \frac{1}{\frac{1}{n} - \frac{1}{M_1}} \right)^4 \nabla_{ij'} \psi(\mu)
$$

$$
- n^4 M_1^4 \underbrace{\left( \frac{1}{(M_1 - n)^4} - \frac{1}{(M_1 + n)^4} \right)}_{\text{Denote this term as } (\dagger)} \cdot \nabla_{ij'} \psi(\mu)
$$

$$
= \dots \text{ to be continued under } (\circ) \, .
$$

We have $(\dagger) \geq 0$ as well as

$$
(\dagger) = \frac{(M_1 + n)^4 - (M_1 - n)^4}{(M_1 - n)^4 (M_1 + n)^4}
$$

$$
= \frac{8 M_1^3 n + 8 M_1 n^3}{(M_1 - n)^4 (M_1 + n)^4}
$$

$$
\leq \frac{16 M_1^3 n}{(M_1 - n)^4 (M_1 + n)^4}
$$

$$
\leq \frac{16 M_1^3 n}{(M_1 - \frac{M_1}{2})^4 (M_1 + \frac{M_1}{2})^4}
$$

$$
= \frac{16}{\frac{3^4}{2^8}} \frac{M_1^3 n}{M_1^8} \leq 100 \frac{n}{M_1^5} \, .
$$

Moreover, we have

$$0 \leq \nabla_{ij'}\,\psi(\mu) \leq \sum_{\substack{(i_{[4]}, j_{[4]}): \\ \eta(i_{[4]})=4\,,\, i \notin i_{[4]}}} 1 \cdot 1 \leq (nm)^4\,.$$

Therefore, we can continue the inequality chain at $(\circ)$ with

$$(\circ) \geq \epsilon + \left(\frac{1}{\frac{1}{n} - \frac{1}{M_1}}\right)^4 \nabla_{ij'}\,\psi(\mu) - n^4 M_1^4 \cdot 100\frac{n}{M_1^5} \cdot (nm)^4$$

$$= \epsilon + \left(\frac{1}{\frac{1}{n} - \frac{1}{M_1}}\right)^4 \nabla_{ij'}\,\psi(\mu) - 100\frac{n^9 m^4}{M_1}$$

$$= \epsilon + \left(\frac{1}{\frac{1}{n} - \frac{1}{M_1}}\right)^4 \nabla_{ij'}\,\psi(\mu) - \frac{\epsilon}{2}\,,$$

where we inserted for the value of $M_1$. All in all, we derived

$$\left(\frac{1}{\frac{1}{n} - \frac{1}{M_1}}\right)^4 \nabla_{ij}\,\psi(\mu) > \left(\frac{1}{\frac{1}{n} - \frac{1}{M_1}}\right)^4 \nabla_{ij'}\,\psi(\mu) + \frac{\epsilon}{2}\,.$$

Thus, we can conclude with our previous observations about the de se utility in $\Gamma$ that

$$\mathrm{EU}_{\mathrm{CDT,GT}}(a_{ij} \mid \mu, I) = \frac{M_2}{5} \cdot \nabla_{ij}\,\phi(\mu) + \frac{1}{5}\nabla_{ij}\,\psi(\mu)$$

$$= \frac{M_2}{5} \cdot 5(1 - p_i)^4 + \frac{1}{5}\nabla_{ij}\,\psi(\mu)$$

$$= \frac{M_2}{5} \cdot \nabla_{ij'}\,\phi(\mu) + \frac{1}{5}\nabla_{ij}\,\psi(\mu)$$

$$> \frac{M_2}{5} \cdot \nabla_{ij'}\,\phi(\mu) + \frac{1}{5} \cdot \left[\nabla_{ij'}\,\psi(\mu) + \left(\frac{1}{n} - \frac{1}{M_1}\right)^4 \cdot \frac{\epsilon}{2}\right]$$

$$= \frac{M_2}{5} \cdot \nabla_{ij'}\,\phi(\mu) + \frac{1}{5}\nabla_{ij'}\,\psi(\mu) + \delta_1$$

$$= \mathrm{EU}_{\mathrm{CDT,GT}}(a_{ij'} \mid \mu, I) + \delta_1\,.$$

Hence, we derived that an action $(i, j')$ of $\Gamma$ has positive play probability

$$\mu_{ij'} = \pi_{ij'} \cdot p_i \geq \pi_{ij'} \cdot \left(\frac{1}{n} - \frac{1}{M_2}\right) > 0\,,$$

under $\mu$ while simultaneously being more than $\delta_1$ dominated by another action $(i, j)$. This contradicts $\mu$ being a $\delta_1$-well-supported (CDT,GT)-equilibrium. The contradiction completes our proof that $\pi$ is an $\epsilon$-Nash equilibrium.

## H.5 Proof of Theorem 3

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

*(1a.) Given $\Gamma$ and $t \in \mathbb{Q}$, is there a (CDT,GT)-equilibrium of $\Gamma$ with ex-ante utility $\geq t$?*

*(1b.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there a (CDT,GT)-equilibrium $\mu$ 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 $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there a strategy $\mu$ of $\Gamma$ 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 $\Gamma$ and $t \in \mathbb{Q}$, is there an (EDT,GDH)-equilibrium of $\Gamma$ with ex-ante utility $\geq t$?*

*(2b.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there an (EDT,GDH)-equilibrium $\mu$ such that $\mathbb{P}(I \mid \mu) > 0$, and such that the player has an (EDT,GDH)-expected utility $\geq t$ upon reaching $I$?*

*(2c.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, is there a strategy $\mu$ of $\Gamma$ such that $\mathbb{P}(I \mid \mu) > 0$, and such that the player has an (EDT,GDH)-expected utility $\geq t$ upon reaching $I$?*

*(3a.) Given $\Gamma$ and $t \in \mathbb{Q}$, do all (EDT,GDH)-equilibria of $\Gamma$ have ex-ante utility $\geq t$?*

*(3b.) Given $\Gamma$, an info set $I$ of $\Gamma$ and $t \in \mathbb{Q}$, do all (EDT,GDH)-equilibria $\mu$ with $\mathbb{P}(I \mid \mu) > 0$ yield the player an (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 $\Gamma$ has an ex-ante optimal strategy since the maximization problem (1) is about a continuous polynomial function over a compact domain. Moreover, observe that $(\Gamma, t)$ is a yes instance for ExAnteOpt-D (by definition) if and only if there is a strategy $\mu$ in $\Gamma$ with $U(\mu) \geq t$ if and only if there is an ex-ante optimal strategy $\mu^*$ in $\Gamma$ with $U(\mu^*) \geq t$.

(1a.) and (2a.):
Let $(\Gamma, t)$ be an instance for ExAnteOpt-D. Without transforming, we can choose $(\Gamma, t)$ 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 $\Gamma$ with ex-ante utility $\geq t$ if and only if (by Lemma 17) there is an ex-ante optimal strategy in $\Gamma$ with with ex-ante utility $\geq t$ if and only if (by the comment above) $(\Gamma, t)$ is a yes instance for ExAnteOpt-D.

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

Observe that any strategy $\mu$ for $\Gamma$ corresponds to a strategy $\mu'$ in $\Gamma'$ that behaves like $\mu$ at info sets $I$ that were inherited from $\Gamma$ 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 $\mu'$ of $\Gamma'$ 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 $\mu'$ upon reaching info set $I_{-1}$ is equal to the ex-ante expected utility of $\mu'$ in $\Gamma'$ and of corresponding strategy $\mu$ in $\Gamma$.

We obtain:

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

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

$\implies$ there is a strategy $\mu'$ in $\Gamma'$ with (CDT,GT)-expected

utility (resp. (EDT,GDH)-expected utility) $\geq t$ upon reaching $I_{-1}$

$\implies$ corresponding strategy $\mu$ in $\Gamma$ has ex-ante utility $\geq t$

$\implies$ $(\Gamma, t)$ is a yes instance for EXANTEOPT-D.

and for the other direction:

· $(\Gamma, t)$ is a yes instance for EXANTEOPT-D

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

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

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

$\implies$ $(\Gamma', I_{-1}, t)$ 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 $(\Gamma, t)$ 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 15). Therefore, each (EDT,GDH)-equilibrium of $\Gamma$ 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 $(\Gamma, t)$ with only one info set. There, we observe that (EDT,GDH)-equilibria of $\Gamma'$ correspond exactly to the ex-ante optimal strategies in $\Gamma$, and must therefore all promise the same (EDT,GDH)-expected utility upon reaching $I_{-1}$.