A Technical Appendix

Proof of Theorem 1.

Proof. First we need the following simple Lemma:

Lemma 2. *Let D be a discrete distribution with probabilities $p_1 \geq p_2 \geq \dots$. $\forall \gamma \in (0,1), z \in \mathbb{N}$, it holds that if $z \geq \gamma e^{H(D)/(1-\gamma)}$, then $\sum_{i=1}^{z} p_i \geq \gamma$.*

Order the histories $h_1, \dots \in S^{\sigma}(v)$ descending by probability with respect to $\pi^{(\sigma_{pre},\sigma_2)}$, with probabilities $p_1 \geq p_2 \geq \dots$ respectively. Then setting $\gamma = 1-\epsilon$ in Lemma 2 gives us

$$\sum_{i=1}^{z} \pi^{(\sigma_{pre},\sigma_2)}(h_i) \geq (1-\epsilon) P_{norm}(v)$$

Where $z = \lceil (1-\epsilon) e^{H(\mathcal{S}^{\sigma,\sigma_{pre}}(v))/\epsilon} \rceil$. Now consider the precomputation strategy $\tilde{\sigma}_1$ which memorizes all histories in $S := \{h \in H : \exists i \in [z] \text{ s.t. } h < h_i\}$, where $|S| \leq Lz$. Then $\forall i \in [z]$, we have $\pi^{(\sigma_{pre},\sigma_2)}(h_i) = \pi^{(\tilde{\sigma}_1,\sigma_2)}(h_i)$. Since $h_i \not\leq h_j$ for $i \neq j$, and $\forall i \in [z], u^{(\sigma_1,\sigma_2)}(h_i) \geq v$, we have

$$u^{(\tilde{\sigma}_1,\sigma_2)} \geq \sum_{i=1}^{z} \pi^{(\sigma_{pre},\sigma_2)}(h_i) u^{(\sigma_1,\sigma_2)}(h_i) \geq (1-\epsilon) v P_{norm}(v)$$

$\square$

Proof of Lemma 2.

Proof. Suppose $\sum_{i=1}^{z} p_i < \gamma$. Then we have

$$\begin{aligned} H(D) &\geq \sum_{i=z+1}^{|D|} p_i \log\left(\frac{1}{p_i}\right) \\ &\geq \sum_{i=z+1}^{|D|} p_i \log\left(\frac{1}{p_z}\right) \\ &> (1-\gamma)\log\left(\frac{1}{p_z}\right) \\ &\geq (1-\gamma)\log\left(\frac{1}{\gamma/z}\right) \end{aligned}$$

Where the last inequality follows because $p_z \leq \gamma/z$. Solving gives

$$z < \gamma e^{H(D)/(1-\gamma)}$$

and the result follows.

$\square$

Proof of Theorem 2.

Proof. Since $u_1^{(\sigma_{pre},\Delta\tilde{\sigma}_2^*)} = 1$, $P_{norm}(v') = 1$ for any $v' \leq 1$. Since the Nash equilibrium value is v, we must have $v \geq \tilde{u}^{(\Delta\tilde{\sigma}_1,\Delta\tilde{\sigma}_2^*)}$ for any mixture of precomputation strategies $\Delta\tilde{\sigma}_1$ for player 1. Write $H(v') := H\left(\mathcal{S}^{(\sigma_1,\Delta\tilde{\sigma}_2^*),\sigma_{pre}}(v')\right)$ as a shorthand, where we associate the mixed strategy $\Delta\tilde{\sigma}_2^*$ with a corresponding behavior strategy in the original game by Kuhn's theorem. Because the Nash equilibrium value of the game is v, any precomputation strategy for player 1 must have utility $\leq v$ in the meta game. Theorem 1 provides the existence of a precomputation strategy for player 1, and gives us that $\forall \epsilon \in (0,1), v' \in [0,1]$,

$$v \geq (1-\epsilon)v' - (1-\epsilon)L\lambda_1 e^{H(v')/\epsilon}$$

Requiring $v' > v + e^{-k}$, we can set $\epsilon = \frac{v'-v-e^{-k}}{v'-e^{-k}}$. Rearranging, simplifying, and assuming $e^{-k} < \lambda_1 L$ gives

$$\begin{aligned} H(v') &\geq \frac{v'-v-e^{-k}}{v'-e^{-k}} \log\left(\frac{e^{-k}}{\lambda_1 L}\right) \\ &\geq (v'-v-e^{-k})\left(\log\left(\frac{1}{\lambda_1 L}\right) - k\right) \\ &\geq (v'-v-0.01)\left(\log\left(\frac{1}{\lambda_1 L}\right) - 5\right) \end{aligned}$$

for $k = 5$. When $e^{-5} \geq \lambda_1 L$ the final inequality is nonpositive and therefore automatically a lower bound on $H(v')$. $\square$

Proof of Theorem 3.

Proof. Let S be the memorization set of the precomputation strategy $\tilde{\sigma}_1$. Then without loss of generality we can assume $h \in S \implies \pi^{(\sigma_{pre},\sigma_2)}(h) \geq \lambda_1$; otherwise, we could strictly increase $\tilde{u}_1^{(\tilde{\sigma}_1,\sigma_2)}$ by removing h and any descendants of h from S. The second observation is that we can conclude $|S| \leq \frac{L+1}{\lambda_1}$ from Lemma 1.

Algorithm 1 Compute best precomputation response.

Input: $(\sigma_1,\sigma_2), \sigma_{pre}, \lambda_1, \epsilon, \delta$
Output: A precomputation strategy $\tilde{\sigma}_1'$ for player 1 which approximately maximizes $\tilde{u}^{(\tilde{\sigma}_1',\sigma_2)}$

1: Define $est(h, K)$ to be an estimate of $u_1^{\sigma_1,\sigma_2}(h)$ using the mean value of K games sampled from history h played according to $\sigma = (\sigma_1,\sigma_2)$. This can be computed in time $\mathcal{O}(LK)$.
2: Initialize BESTCHOICES, $S = \emptyset$.
3: Let bestvalue$(h) = est(h,K) \times \pi^{(\sigma_{pre},\sigma_2)}(h)$ if $\pi^{(\sigma_{pre},\sigma_2)}(h) < \lambda_1$. Otherwise, bestvalue$(h) = \max[\sum_{h' \in Succ(h)}$ bestvalue$(h') - \lambda_1$, $est(h,K)\pi^{(\sigma_{pre},\sigma_2)}(h)]$. If the left argument of the $\max$ is larger, we add h to BESTCHOICES. Here we set $K = \mathcal{O}\left(\ln\left(\frac{A^2(L)}{\lambda_1}\right)/\epsilon^2\right)$.
4: If the starting history $\emptyset \in$ BESTCHOICES, then depth first search from $\emptyset$, where there is an edge $h \to h' \in Succ(h)$ iff $h' \in$ BESTCHOICES. Add each history visited to S.
5: **Return** $\tilde{\sigma}_1'$, the precomputation strategy which memorizes σ_{pre} on set S.

Algorithm 1 recursively computes an optimal precomputation memorization set S, but uses the approximation $est(h,K)$ in place of $u_1^{(\sigma_1,\sigma_2)}(h)$. Let $W_1 = \{h \in H | \pi^{(\sigma_{pre},\sigma_2)}(h) \geq \lambda_1\}$, $W_2 = Succ(W_1)$ and $W = W_1 \cup W_2$. If we have $\forall h \in W, |est(h,K) - u_1^{(\sigma_1,\sigma_2)}(h)| < \frac{\epsilon}{2}$, then if there exists some $\tilde{\sigma}_1$ with true value $\geq v$, the precomputation strategy $\tilde{\sigma}_1$ with precomputation set S has value at least $v - \frac{\epsilon}{2}$ when we use the approximated utilities, and so Algorithm 1 finds some precomputation strategy with approximated value at least $v - \frac{\epsilon}{2}$. Conversely, this precomputation strategy has real value at least $v - \epsilon$.

We have $|W| \leq |A|^2 \frac{(L+1)}{\lambda_1}$. Chernoff and union bound give us that the probability of the event $\exists h \in W$ s.t. $|est(h,K) - u_1^{(\sigma_1,\sigma_2)}(h)| \geq \frac{\epsilon}{2}$ is

$$\leq |W| 2 \exp\left(-\frac{K}{2}(\epsilon/2)^2\right) \leq \delta$$

for $K = 16 \ln\left(\frac{|W|}{\delta}\right)/\epsilon^2 = 16 \ln\left(\frac{A^2(L+1)}{\delta\lambda_1}\right)/\epsilon^2$.

Thus with probability at least $1-\delta$, the estimates are all within $\frac{\epsilon}{2}$ of the true values and we find a precomputation policy $\tilde{\sigma}_1'$ with $\tilde{u}_1^{(\tilde{\sigma}_1',\sigma_2)} \geq v - \epsilon$. The algorithm considers at most $|W|$ histories, where each history takes time LK to sample, giving a total running time of $|W|LK = \mathcal{O}\left(|A|^2 \frac{L^2}{\lambda_1 \epsilon^2} \ln(\frac{A^2 L}{\lambda_1})\right)$. □

Proof of Lemma 1.

Proof. We have

$$L+1 \geq \sum_{l=0}^{L} \sum_{\substack{h \in H \\ |h|=l}} \pi^{\sigma_1,\sigma_2}(h) = \sum_{h \in H} \pi^{\sigma_1,\sigma_2}(h)$$

$$\geq \sum_{\substack{h \in H \\ \pi^{\sigma_1,\sigma_2}(h) \geq p}} \pi^{\sigma_1,\sigma_2}(h) \geq p |\{h \in H | \pi^{\sigma_1,\sigma_2}(h) \geq p\}|$$

so

$$|\{h \in H | \pi^{\sigma_1,\sigma_2}(h) \geq p\}| \leq \frac{L+1}{p}$$

□

Lemma 3. *Let*

$$W_1 = \left\{ h \in H | \sup_{\Delta\tilde{\sigma}_2 \in \Delta Pre(\sigma_{pre},\sigma_2)} \pi^{\sigma_{pre},\Delta\tilde{\sigma}_2}(h) \geq \lambda_1 \right\}$$

$$W_2 = \left\{ h \in H | \sup_{\Delta\tilde{\sigma}_1 \in \Delta Pre(\sigma_{pre},\sigma_1)} \pi^{\Delta\tilde{\sigma}_1,\sigma_{pre}}(h) \geq \lambda_2 \right\}$$

$$Succ(W_i) := \cup_{h \in W_i} Succ(h)$$

for $i \in \{1,2\}$

$$W := W_1 \cup W_2 \cup W$$

Then $|W| \leq \mathcal{O}\left(A^2(L+2)^2(\frac{1}{\lambda_1} + \frac{1}{\lambda_2})\right)$, *where* $L = \max_{h \in H} |h|$ *and* $A = \max_{h \in H} |A(h)|$.

Proof of Lemma 3.

Proof. Let $\tilde{\sigma}_{-i,l}$ be the precomputation strategy for player $-i$ (the player opposing player i) which plays σ_{pre} on every $h \in H$ such that $|h| < l$, and σ_{-i} otherwise. Let $l(h) = \max_{|h|} \pi^{\sigma_{pre},\tilde{\sigma}_{-i,l}}(h) \geq p$, and ∞ if no such l exists. Define

$$W_i = \left\{ h \in H | \max_{\Delta\tilde{\sigma}_{-i} \in \Delta Pre(\sigma_{pre},\sigma_{-i})} \pi^{\sigma_{pre},\Delta\tilde{\sigma}_{-i}}(h) \geq p \right\}$$

Then we have

$$|W_i| = \sum_{l'=0}^{L+1} \sum_{\substack{h \in H \\ l(h)=l'}} 1 \leq \sum_{l'=0}^{L+1} |h \in H : \pi^{\sigma_{pre},\tilde{\sigma}_{-i,l}}(h) \geq p|$$

$$\leq (L+2)\frac{(L+1)}{p} \leq \frac{(L+2)^2}{p}$$

The first equality follows because if $\pi_2^{\sigma_{pre},\Delta\tilde{\sigma}}(h) \geq p$ for some mixed precomputation strategy $\Delta\tilde{\sigma}_2$, then this must hold true for some pure precomputation strategy $\tilde{\sigma}_2$; if $\tilde{\sigma}_2$ precomputes on histories $h_{:0}, \ldots, h_{:l}$ for some l, then $\pi^{\sigma_{pre},\tilde{\sigma}_{-i,l}}(h) \geq p$. The last inequality follows from Lemma 1. Thus setting $p = \lambda_{-i}$, using that $|W_i \cup Succ(W_i)| \leq \mathcal{O}(A^2|W_i|)$, and summing for $i = 1,2$ gives the required bound. □

Lemma 4. *(Kuhn's Theorem applied to the meta precomputation game). Given a two player game G with policies $\sigma_1, \sigma_2, \sigma_{pre}$, define the extensive form imperfect information, perfect recall game G' with the following tree structure:*

1. *Each history $h' \in H'$ for game G' corresponds to a history $h = Q(h') \in H$ in the original game G. At history h', if h' is not a chance node, then $P(h') = P(h) \in \{1,2\}$. Instead of player $P(h')$ choosing an action in $A(h)$, $P(h')$ chooses whether to continue precomputing if they have not already stopped (action= 1) or stop precomputing (action=0). h' then transitions to a chance node h'' ($P(h'') = c$), which picks an action $a \in A(h)$ according to the distribution $\sigma_{pre}(h)$ if $P(h')$ chose to precompute in the previous step, or $\sigma_{P(h)}(h)$ otherwise. The history which is transitioned to after the chance node in game G' is then associated with history $\pi(h,a)$ in game G.*
2. *Players can see their own actions and the actions of the chance player, but not the actions of the opposing player. Formally, two histories $h', h'' \in H'$ for player $P(h) = P(h') = i \in \{1,2\}$ are in the same information set iff they have the same sequence of actions for player i and the chance player leading up to them.*

3. *Terminal histories in game G' are histories $h' \in H'$ where $P(h') \in \{1,2\}$ and the associated history $h \in H$ is a terminal history in game G. In each case, the utility of this terminal history h' is $u'_1(h') = u_1^{(\sigma_1,\sigma_2)}(h) - \lambda_1 z_1(h') + \lambda_2 z_2(h')$ for some functions z_i to be specified.*

Then for every pair of behavior strategies $(\tilde{\sigma}'_1, \tilde{\sigma}'_2)$ in G' with utility u, there is a pair of mixed strategies $(\tilde{\sigma}_1 = f(\tilde{\sigma}'_1), \tilde{\sigma}_2 = f(\tilde{\sigma}'_2))$ in the meta precomputation game with the same utility u, and vice versa (for some function f).

Proof of Lemma 4.

Proof. First, we argue that there is a bijective correspondence between pure strategies in the meta precomputation game G, and pure strategies in game G'. If we can establish this equivalence, then we know that there is also a bijective relationship between mixed strategies in G and mixed strategies in G'. Finally, by Kuhn's Theorem, we know that there is a correspondence between mixed strategies in game G', and behavior strategies in game G' (because G' is an extensive form perfect recall game), from which the conclusion of Theorem 4 follows. It remains to show the precise relationship between pure strategies in the meta precomputation game G, and pure strategies in G'.

Fix any pure strategies $\tilde{\sigma}_i \in Pre(\sigma_{pre}, \sigma_i)$ for $i \in \{1,2\}$ in the meta-precomputation game G, where $\tilde{\sigma}_i$ has memorization set S_i. Recall that for every $h' \in H'$ for the game G' where $P(h') \neq c$, there is a unique associated history $Q(h') := h$ in game G. We map $\tilde{\sigma}_i$ to the pure strategy $\tilde{\sigma}'_i$ in game G', where player i chooses action 1 (to precompute) at history h' iff $Q(h') \in S_i$ and player 1 has not previously chosen to stop precomputing. Note that this is mapping is valid (i.e. respects information sets in game G') and is bijective: any information set I for player i in game G' where player i has more than one choice (i.e. has always chosen action 1 leading up to this history) is uniquely determined by the sequence of actions of the chance player, which corresponds to a unique history h in game G. Thus the inverse map, from any pure strategy $\tilde{\sigma}'_i$ in game G' to a pure strategy $\tilde{\sigma}_i$ in game G, where $h \in S_i$ if player i chooses to precompute at the information set corresponding to history h in game G', is well defined. Finally, we need to show that

$$\tilde{u}_1^{(\tilde{\sigma}_1,\tilde{\sigma}_2)} = (u')_1^{(\tilde{\sigma}'_1,\tilde{\sigma}'_2)}$$

which establishes the desired relationship. We begin by defining the functions z_i for $i \in \{1,2\}$. For $h' \in H'$, let $\pi'_c(h')$ be the probability of reaching h' where we only take into account the chance player (i.e. the product of the probabilities of all edges leading to h' where the chance player chooses an action). Now fix any terminal history $h' \in Z'$ for game G'. Without loss of generality, we assume $\pi'_c(h') > 0$ (otherwise we can remove it from the game without affecting the utility). Let $M_i(h') = \{h'_{i,1}, \ldots, h'_{i,j_i}\}$ be the set of histories $h'_{i,1} < h'_{i,2}, \ldots, h'_{i,j_i} < h'$ where player i chose to memorize (action 1) in order to reach history h'. Then we let

$$z_i(h') := \sum_{h'' \in M_i(h')} \frac{1}{\pi'_c(h'')}$$

Let $\tilde{\sigma} = (\tilde{\sigma}_1, \tilde{\sigma}_2)$ and $\tilde{\sigma}' = (\tilde{\sigma}'_1, \tilde{\sigma}'_2)$. If Z' is the set of terminal histories for G' where $\pi_c(h') \neq 0$, we can compute

$$\begin{aligned}(u')_1^{\tilde{\sigma}'} = \sum_{h' \in Z'} (\pi')^{\tilde{\sigma}'}(h')[u_1(Q(h')) \\ - \lambda_1 z_1(h') + \lambda_2 z_2(h')]\end{aligned}$$

Focusing on the first term gives

$$\sum_{h' \in Z'} (\pi')^{\tilde{\sigma}'}(h') u_1(Q(h')) = \sum_{h \in Z} \pi^{\tilde{\sigma}}(h) u_1(h)$$

by construction of the game G'. Focusing on the remaining terms, we have

$$\begin{aligned}
\sum_{h' \in Z'} (\pi')^{\tilde{\sigma}'}(h') z_i(h') &= \sum_{h' \in Z'} (\pi')^{\tilde{\sigma}'}(h') \sum_{h'' \in M_i(h')} \frac{1}{\pi'_c(h'')} \\
&= \sum_{h' \in Z'} (\pi')^{\tilde{\sigma}'}(h') \sum_{\substack{h'' < h' \\ Q(h'') \in S_i}} \frac{1}{\pi'_c(h'')} \\
&= \sum_{h \in S_i} \sum_{\substack{h' \in Q^{-1}(h) \\ (\pi')^{\tilde{\sigma}'}(h') > 0}} \pi'_c(h') \sum_{\substack{h'' \in Z' \\ h'' > h' \\ (\pi')^{\tilde{\sigma}'}(h'') > 0}} \frac{1}{\pi'_c(h'')} \\
&= \sum_{h \in S_i} \sum_{\substack{h' \in Q^{-1}(h) \\ (\pi')^{\tilde{\sigma}'}(h') > 0}} \sum_{\substack{h'' \in Z' \\ h'' > h' \\ (\pi')^{\tilde{\sigma}'}(h'') > 0}} \pi'_c(h', h'') \\
&= \sum_{h \in S_i} \sum_{\substack{h' \in Q^{-1}(h) \\ (\pi')^{\tilde{\sigma}'}(h') > 0}} 1 \\
&= \sum_{h \in S_i} 1 = |S_i|
\end{aligned}$$

The first two equalities follow from the definitions of z_i and M_i. In the third equality, we reverse the order of the summation. Instead of fixing a terminal history $h' \in Z'$ and summing over all ancestors $h'' < h'$ where player i chose to precompute, in the third equality we sum over all histories h' where player i chose to precompute (i.e. $h' \in Q^{-1}(h)$ where $h \in S_i$ and h' has positive support on $\tilde{\sigma}'$), and then over the relevant terminal histories. Note that the summation $\sum_{\substack{h' \in Q^{-1}(h) \\ (\pi')^{\tilde{\sigma}'}(h') > 0}}$ is over exactly one element (using the fact that we are only considering pure strategies). We are able to replace $(\pi')^{\tilde{\sigma}'}(h')$ by $\pi'_c(h')$ because $\tilde{\sigma}'$ consists of pure strategies, i.e. for any h' in the support of $(\pi')^{\tilde{\sigma}'}$, the probabilities on the edges of the game tree leading up to h' where players 1 and 2 make moves are all 1, and so only the probability contribution from the chance player is relevant.

Collecting these results, we find that

$$\begin{aligned}(u')_1^{\tilde{\sigma}'} &= \sum_{h \in Z} \pi^{\tilde{\sigma}}(h) u_1(h) - \lambda_1 |S_1| + \lambda_2 |S_2| \\ &= \tilde{u}_1^{\tilde{\sigma}}\end{aligned}$$

Algorithm 2 Compute an approximate Nash equilibrium to the meta-precomputation game.

Input: $(\sigma_1, \sigma_2), \sigma_{pre}, \lambda_1, \lambda_2, \epsilon, \delta$
Output: An ϵ-Nash equilibrium to the meta precomputation game.

1: Define $est(h, K)$ to be an estimate of $u_1^{\sigma_1,\sigma_2}(h)$ using the mean value of K games sampled from history h played according to $\sigma = (\sigma_1, \sigma_2)$. This can be computed in time $\mathcal{O}(LK)$.
2: Define G to be the meta-precomputation game.
3: Compute the set of high probability histories W for G. Let G' be the equivalent extensive form game for G (Lemma 4), where histories $h' \in H'$ for game G' are now made to be terminal histories if they are associated with low probability histories $h = Q(h') \in H$ for game G (i.e. $h \notin W$ from Lemma 3). If h' is one of these terminal histories, we estimate the utility by $est(Q(h), K) + \lambda_1 z_1(h') - \lambda_2 z_2(h')$ (corresponding to Lemma 4) for $K = \mathcal{O}(\ln(\frac{|W|}{\delta})/\epsilon^2)$.
4: Run vanilla CFR from [Zinkevich *et al.*, 2007] on this approximate game G' for $\mathcal{O}\left(\frac{|W|^2}{\epsilon^2}\right)$ iterations.
5: Return the result of CFR on game G'.

showing the required equivalence.

□

Proof of Theorem 4

Proof. Suppose we pick $K = \mathcal{O}(\ln(\frac{|W|}{\delta})/\epsilon^2)$ samples for each call to $est(h, K)$, so that with probability at least $1-\delta$ we have $\forall h \in Succ(W_1) \cup Succ(W_2), |est(h, K) - u^{(\sigma_1,\sigma_2)}(h)| < \epsilon/x'$ for some constant x' (by applying Chernoff+Union bound, similarly to the proof of Theorem 3). Let $\tilde{u}_1'$ be the utility of the extensive form game where we use the estimated values to compute the terminal values of the game. Suppose we run counter factual regret minimization to get a $2\frac{\epsilon}{x}$ Nash equilibrium $(\Delta\tilde{\sigma}_1, \Delta\tilde{\sigma}_2)$ to this game using utility function $\tilde{u}_1'$. Then we have

$$\begin{aligned}
\tilde{u}_1^{(\Delta\tilde{\sigma}_1,\Delta\tilde{\sigma}_2)} + \frac{\epsilon}{x'} &\geq (\tilde{u}')_1^{(\Delta\tilde{\sigma}_1,\Delta\tilde{\sigma}_2)} \\
&\geq \sup_{\Delta\tilde{\sigma}_1 \in \Delta(Pre(\sigma_{pre},\sigma_1))} (\tilde{u}')_1^{(\Delta\tilde{\sigma}_1,\Delta\tilde{\sigma}_2)} - 2\frac{\epsilon}{x} \\
&\geq \sup_{\Delta\tilde{\sigma}_1 \in \Delta(Pre(\sigma_{pre},\sigma_1))} (\tilde{u})_1^{(\Delta\tilde{\sigma}_1,\Delta\tilde{\sigma}_2)} - 2\frac{\epsilon}{x} - \frac{\epsilon}{x'}
\end{aligned}$$

Thus

$$\tilde{u}_1^{(\Delta\tilde{\sigma}_1,\Delta\tilde{\sigma}_2)} + 2\left(\frac{\epsilon}{x'} + \frac{\epsilon}{x}\right) \geq \sup_{\Delta\tilde{\sigma}_1 \in \Delta(Pre(\sigma_{pre},\sigma_1))} (\tilde{u})_1^{(\Delta\tilde{\sigma}_1,\Delta\tilde{\sigma}_2)}$$

and the analogous inequality holds for player 2, giving a $2\left(\frac{\epsilon}{x'} + \frac{\epsilon}{x}\right)$ Nash equilibrium. Picking $x' = x = 4$ gives an ϵ Nash equilibrium.

Each round of Counter Factual Regret Minimization requires time $\mathcal{O}(|W|)$, and after T rounds the average regret for player i is

$$R_i \leq \frac{|W|}{\sqrt{T}}$$

Setting $R_i \leq \frac{\epsilon}{x}$ to give a $2\frac{\epsilon}{x}$ Nash equilibrium gives

$$T \geq \mathcal{O}\left(\frac{|W|^2}{\epsilon^2}\right)$$

The total runtime for the regret minimization component is $\mathcal{O}(T \times |W|) = \mathcal{O}\left(\frac{|W|^3}{\epsilon^2}\right)$, and the runtime for sampling utility estimates is $\mathcal{O}(K|Succ(W_1) \cup Succ(W_2)|L)$, giving a total runtime bound of

$$\frac{1}{\epsilon^2}\left(A^2L^2\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)\right)^3$$

$$+\frac{1}{\epsilon^2}A^2L^3\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)\ln\left(AL\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)/\delta\right)$$

□

Experimental Details:

Experiments were run on a single thread of a Intel Xeon E5-2678 v3 2.5GHz CPU for around 3 hours per plot. While it is feasible to run this experiment in full with more hardware, we make the following key simplifications to make the computation manageable on a desktop:

1. Instead of considering all possible actions, we limit policies to play the from the top $k = 2$ choices recommended by the Stockfish engine.
2. We set the maximum game length to $L = 100$ half moves; a draw with utility 0.5 is declared beyond this point.
3. Instead of explicitly estimating the utility of each board state, we use Stockfish to compute the centerpawn (cp) score of each board state. If there is a centerpawn advantage of ≥ 400 for any player, then this is calculated as a win for that player ($u = 1$ or 0). Otherwise the utility value is given as a draw ($u = 0.5$). This choice (as opposed to e.g. estimating the expected value as a linear function of the cp score) was made to give the qualitative interpretation of the results a more conservative lower bound. In particular, we can interpret the utility of the precomputing player as a qualitative indication of the fraction of the time they are able to reach an overwhelming advantage against their opponent (where we can be reasonably sure they would win if playing normally from that point onwards). In contrast, it isn't clear that e.g. a 50cp score for white (scored by Stockfish(50ms)) would translate to Stockfish(10ms) having a slightly higher probability of winning, because Stockfish(10ms) may not be powerful enough to make use of that advantage.