Appendix

In the appendices we present the proofs, and additional lemmas that are used in the proofs.

A. Lemma 1

Lemma 1 proves that if (5) is satisfied for some action $a \in A(I)$ on iteration T, then the value of action a and all its descendants on every iteration played so far can be set to the T-near counterfactual best response value. The same lemma holds if one replaces the T-near counterfactual best response values with exact counterfactual best response values. The proof for Lemma 1 draws from recent work on warm starting CFR using only an average strategy profile (Brown & Sandholm, 2016).

Lemma 1. *Assume T iterations of CFR with RM have been played in a two-player zero-sum game. If $T\big(\psi^{\bar{\sigma}^T_{-i},T}(I,a)\big) \le \sum_{t=1}^{T} v^{\sigma^t}(I)$ and one sets $v^{\sigma^t}(I,a) = \psi^{\bar{\sigma}^T_{-i},T}(I,a)$ for each $t \le T$ and for each $I' \in D(I,a)$ sets $v^{\sigma^t}(I',a') = \psi^{\bar{\sigma}^T_{-i},T}(I',a')$ and $v^{\sigma^t}(I') = \psi^{\bar{\sigma}^T_{-i},T}(I')$ then after T' additional iterations of CFR with RM, the bound on exploitability of $\bar{\sigma}^{T+T'}$ is no worse than having played $T+T'$ iterations of CFR with RM unaltered.*

Proof. The proof builds upon Theorem 2 in (Brown & Sandholm, 2016). Assume $T\big(\psi^{\bar{\sigma}^T_{-i},T}(I,a)\big) \le \sum_{t=1}^{T} v^{\sigma^t}(I)$. We wish to warm start to T iterations. For each $I' \in D(I,a)$ set $v^{\sigma^t}(I',a') = \psi^{\bar{\sigma}^T_{-i},T}(I',a')$ and $v^{\sigma^t}(I') = \psi^{\bar{\sigma}^T_{-i},T}(I')$ and set $v^{\sigma^t}(I,a) = \psi^{\bar{\sigma}^T_{-i},T}(I,a)$ for all $t \le T$. For every other action, leave regret unchanged. For each $I' \in D(I,a)$ we know by construction that $\Phi(R^T(I'))$ is within the CFR bound $y^T_{I'}$ after changing regret. By assumption $T\big(\psi^{\bar{\sigma}^T_{-i},T}(I,a)\big) \le \sum_{t=1}^{T} v^{\sigma^t}(I)$, so $R^T(I,a) \le 0$ and therefore $\Phi(R^T(I))$ is unchanged. Finally, since the T iterations were played according to CFR with RM and regret is unchanged for every other information set I'', so the conditions for Theorem 2 in (Brown & Sandholm, 2016) hold for every information set, and therefore we can warm start to T iterations of CFR with RM with no penalty to the convergence bound. □

B. Proof of Theorem 1

Proof. From Lemma 1 we can immediately set regret for $a \in A(I)$ to $v^{\sigma^t}(I,a) = \psi^{\bar{\sigma}^T_{-i},T}(I,a)$. By construction of T', $R^t(I,a)$ is guaranteed to be nonpositive for $T \le t \le T+T'$ and therefore $\sigma^t(I,a) = 0$. Thus, $\bar{\sigma}_i^{T+T'}(I')$ for $I' \in D(I,a)$ is identical regardless of what is played in $D(I,a)$ during $T \le t \le T+T'$.

Since $(T+T')\big(\psi^{\bar{\sigma}^{T+T'}_{-i},T+T'}(I,a)\big) \le T\big(\psi^{\bar{\sigma}^T_{-i},T}(I,a)\big) + T'\big(U(I,a)\big)$ and $\sum_{t=1}^{T+T'} v^{\sigma^t}(I) \ge \sum_{t=1}^{T} v^{\sigma^t}(I) + T'\big(L(I)\big)$, so by the definition of T', $(T+T')\big(\psi^{\bar{\sigma}^{T+T'}_{-i},T+T'}(I,a)\big) \le \sum_{t=1}^{T+T'} v^{\sigma^t}(I)$. So if regrets in $D(I,a)$ and $R^{T+T'}(I,a)$ are set according to Lemma 1, then after T'' additional iterations of CFR with RM, the bound on exploitability of $\bar{\sigma}^{T+T'+T''}$ is no worse than having played $T+T'+T''$ iterations of CFR with RM from scratch. □

C. Proof of Theorem 2

Proof. Consider an information set I and action $a \in A(I)$ where for every opponent Nash equilibrium strategy $\sigma^*_{-P(I)}$, $CBV^{\sigma^*_{-P(I)}}(I,a) < CBV^{\sigma^*_{-P(I)}}(I)$. Let $i = P(I)$. Let $\delta = \min_{\sigma_{-i} \in \Sigma^*}\big(CBV^{\sigma_{-i}}(I) - CBV^{\sigma_{-i}}(I,a)\big)$ where Σ^* is the set of Nash equilibria. Let $\sigma'_{-i} = \arg\max_{\sigma_{-i} \in \Sigma_{-i} | CBV^{\sigma_{-i}}(I) - CBV^{\sigma_{-i}}(I,a) \le \frac{3\delta}{4}} u_{-i}(\sigma_{-i}, BR(\sigma_{-i}))$ Since σ'_{-i} is not a Nash equilibrium strategy and CFR converges to a Nash equilibrium strategy for both players, so there exists a T_δ such that for all $T \ge T_\delta$, $CBV^{\bar{\sigma}^T_{-i}}(I) - CBV^{\bar{\sigma}^T_{-i}}(I,a) > \frac{3\delta}{4}$. Let $T'_{I,a} = \frac{4|\mathcal{I}|^2\Delta^2|A|}{\delta^2}$. For $T \ge T'_{I,a}$ since $R^T_i \le \sum_{I \in \mathcal{I}_i} R^T(I)$, so $CBV^{\bar{\sigma}^T_{-i}}(I) - \sum_{t=1}^{T} v^{\sigma^t}(I) \le \frac{\delta}{2}$. Let $T_{I,a} = \max(T'_{I,a}, T_\delta)$ and $\delta_{I,a} = \frac{\delta}{4}$. Then for $T \ge T_{I,a}$, $CBV^{\bar{\sigma}^T_{-i}}(I,a) - \frac{\sum_{t=1}^{T} v^{\sigma^t}(I)}{T} \le -\delta_{I,a}$. □

D. Proof of Corollary 1

Proof. Let $I \notin \mathcal{I}_S$. Then $I \in D(I',a')$ for some I' and $a' \in A(I')$ such that for every opponent Nash equilibrium strategy $\sigma^*_{-P(I')}$, $CBV^{\sigma^*_{-P(I')}}(I',a') < CBV^{\sigma^*_{-P(I')}}(I')$. Applying Theorem 2, this means there exists a $T_{I',a'}$ and $\delta_{I',a'} > 0$ such that for $T \ge T_{I',a'}$, $CBV^{\bar{\sigma}^T_{-i}}(I',a') - \frac{\sum_{t=1}^{T} v^{\sigma^t}(I')}{T} \le -\delta_{I',a'}$. So (5) always applies for $T \ge T_{I',a'}$ for I' and a' and I will always be pruned. Since (8) does not require knowledge of regret, it need not be stored for I.

Since $D(I',a')$ will always be pruned for $T \ge T_{I',a'}$, so for any $T \ge \frac{(T_{I',a'})^2}{C^2}$ iterations for some constant $C > 0$, $\pi_i^{\bar{\sigma}^T}(I) \le \frac{C}{\sqrt{T}}$, which satisfies the threshold of the average strategy. Thus, the average strategy in $D(I,a)$ can be discarded. □

E. Lemma 2

Lemma 2. *If for all $T \ge T'$ iterations of CFR with BRP, $T\big(CBV^{\bar{\sigma}^T}(I,a)\big) - \sum_{t=1}^{T} v^{\sigma^t}(I) \le -xT$ for some $x > 0$, then any history h' such that $h \cdot a \sqsubseteq h'$ for some $h \in I$ need only be traversed at most $O\big(\ln(T)\big)$ times.*

Proof. Let $a \in A(I)$ be an action such that for all $T \geq T'$, $T\big(CBV^{\bar{\sigma}^T}(I,a)\big) - \sum_{t=1}^{T} v^{\sigma^t}(I) \leq -xT$ for some $x > 0$. $\psi^{\bar{\sigma}^T_{-i},T}(I,a) \leq CBV^{\bar{\sigma}^T_{-i}}$, so from Theorem 1, $D(I,a)$ can be pruned for $m \geq \lfloor \frac{xT}{U(I,a)-L(I)} \rfloor$ iterations on iteration T. Thus, over iterations $T \leq t \leq T+m$, only a constant number of traversals must be done. So each iteration requires only $\frac{C}{m}$ work when amortized, where C is a constant. Since x, $U(I,a)$, and $L(I)$ are constants, so on each iteration $t \geq T'$, only an average of $\frac{C}{t}$ traversals of $D(I,a)$ is required. Summing over all $t \leq T$ for $T \geq T'$, and recognizing that T' is a constant, we get that action a is only taken $O\big(\ln(T)\big)$ over T iterations. Thus, any history h' such that $h \cdot a \sqsubseteq h'$ for some $h \in I$ need only be traversed at most $O\big(\ln(T)\big)$ times. □

F. Proof of Theorem 3

Proof. Consider an $h^* \notin S$. Then there exists some $h \cdot a \sqsubseteq h^*$ such that $h \in S$ but $h \cdot a \notin S$. Let $I = I(h)$ and $i = P(I)$. Since $h \cdot a \notin S$ but $h \in S$, so for every Nash equilibrium σ^*, $CBV^{\sigma^*}(I,a) < CBV^{\sigma^*}(I)$. From Theorem 2, there exists a $T_{I,a}$ and $\delta_{I,a} > 0$ such that after $T \geq T_{I,a}$ iterations of CFR, $CBV^{\bar{\sigma}^T_{-i}}(I,a) - \frac{\sum_{t=1}^{T} v^{\sigma^t(I)}}{T} \leq -\delta_{I,a}$. Thus from Lemma 2, h^* need only be traversed at most $O\big(\ln(T)\big)$ times. □