Appendix A DETAILED PROOFS

In this section, we include the full proofs that are omitted in the main manuscript.

A.1 Unbiased Estimate of The Zero-One Loss

We prove the unbaisedness of our loss estimator presented in (4).

Lemma A.1. *The estimator $\hat{l}_t^{0-1}$ in (4) is an unbiased estimator of the zero-one loss l_t^{0-1}:*

$$\mathbb{E}_{\tilde{y}_t \sim p_t} \hat{l}_t^{0-1} = l_t^{0-1}.$$

Proof. Since $\tilde{y}$ is drawn with respect to p_t, we can write

$$\begin{aligned}
\mathbb{E}_{\tilde{y}_t \sim p_t} \hat{l}_{t,i}^{0-1} &= \mathbb{I}(y_t \neq i)\mathbb{I}(\hat{y}_t \neq i) + \mathbb{I}(\hat{y}_t \neq y_t)\mathbb{I}(\hat{y}_t = i) \\
&= \mathbb{I}(y_t \neq i)\mathbb{I}(\hat{y}_t \neq i) + \mathbb{I}(i \neq y_t)\mathbb{I}(\hat{y}_t = i) \\
&= \mathbb{I}(y_t \neq i)(\mathbb{I}(\hat{y}_t \neq i) + \mathbb{I}(\hat{y}_t = i)) \\
&= \mathbb{I}(y_t \neq i),
\end{aligned}$$

where the last term is $l_{t,i}^{0-1}$, which completes the proof. □

A.2 Proof of Theorem 3.1

Since the agnostic learnability condition is given with deterministic cost vectors, we need to bridge the deterministic costs with the randomized ones. Observe that $\hat{c}_t^i$ relies solely on the random draw of $\tilde{y}_t$ at each round. Therefore, the partial sum of random vectors $S_j = \sum_{t=1}^{j} \hat{c}_t^i - c_t^i$ has the martingale property. Then we can prove the following lemma using the Azuma-Hoeffding inequality.

Lemma A.2. *Suppose the random cost vectors $\hat{c}_t$ are b-bounded. Let p_t be a probability vector in Δ_k. Then the following inequality holds with probability $1-\delta$:*

$$|\sum_{t=1}^{T} (\hat{c}_t - c_t) \cdot p_t| \leq b\sqrt{2T \log \frac{2}{\delta}}.$$

Proof. Since $\hat{c}_t$ is b-bounded and unbiased, we have

$$|(\hat{c}_t - c_t) \cdot p_t| \leq b \text{ a.s. and } \mathbb{E}(\hat{c}_t - c_t) \cdot p_t = 0.$$

Therefore the Azuma-Hoeffding's inequality implies

$$\mathbb{P}(|\sum_{t=1}^{T} (\hat{c}_t - c_t) \cdot p_t| \geq \epsilon) \leq 2e^{-\frac{\epsilon^2}{2b^2 T}}.$$

Putting $\epsilon = b\sqrt{2T \log \frac{2}{\delta}}$ finishes the proof. □

We now go into the main proof of Theorem 3.1.

Proof. Fix the sequence of cost vectors $w_t c_t$. From the richness condition with edge 2γ, we know

$$\inf_{h \in \mathcal{H}} \sum_{t=1}^{T} w_t c_{t,h(x_t)} \leq \sum_{t=1}^{T} w_t c_t \cdot u_{2\gamma}^{y_t}.$$

By applying Lemma A.2 with $p_t = e_{h(x_t)}$, we get with probability $1-\delta$,

$$\inf_{h \in \mathcal{H}} \sum_{t=1}^{T} w_t \hat{c}_{t,h(x_t)} \leq \sum_{t=1}^{T} w_t c_t \cdot u_{2\gamma}^{y_t} + b\sqrt{2T \log \frac{2}{\delta}}. \quad (13)$$

Then by the online learnability condition, the online learner based on $\mathcal{H}$ can generate predictions $\hat{y}_t$ that satisfies the following inequality with probability $1-\delta$:

$$\sum_{t=1}^{T} w_t \hat{c}_{t,\hat{y}_t} \leq \inf_{h \in \mathcal{H}} \sum_{t=1}^{T} w_t \hat{c}_{t,h(x_t)} + R_\delta(T),$$

Using (13) and the union bound, we have with probability $1-2\delta$,

$$\sum_{t=1}^{T} w_t \hat{c}_{t,\hat{y}_t} \leq \sum_{t=1}^{T} w_t c_t \cdot u_{2\gamma}^{y_t} + b\sqrt{2T \log \frac{2}{\delta}} + R_\delta(T).$$

Then by the definition of $\mathcal{C}_1^{eor}$ in (6), we can compute

$$c_t \cdot u_\gamma^y = \frac{1-\gamma}{k}.$$

Therefore, using the assumption $w_t \geq m$ for all t, we can bound

$$\sum_{t=1}^{T} w_t c_t \cdot (u_\gamma^{y_t} - u_{2\gamma}^{y_t}) = \frac{\gamma}{k} \sum_{t=1}^{T} w_t \geq \frac{\gamma}{k} mT.$$

Then by taking

$$S = \sup_T -\frac{\gamma}{k} mT + b\sqrt{2T \log \frac{1}{\delta}} + R_\delta(T),$$

we prove that with probability $1-2\delta$,

$$\sum_{t=1}^{T} w_t \hat{c}_{t,\hat{y}_t} \leq \sum_{t=1}^{T} w_t c_t \cdot u_\gamma^{y_t} + S,$$

which shows the learner and the adversary satisfy BanditWLC$(\gamma, 2\delta, S)$. □

A.3 Proof of Theorem 3.2

Note that the cost vectors defined in (8) does not put zero cost on the correct label. In order to apply the BanditWLC, we transform the cost vector. Since the zero-one loss vector has the minimal loss on the true

label, we can inductively check that $\operatorname{argmin}_l c^i_{t,l} = y_t$. Then we define $d^i_t \in \mathbb{R}^k$ as below:

$$d^i_{t,l} = c^i_{t,l} - c^i_{t,y_t}.$$

The minimal entry of d^i_t is zero. Let $w^i_t = ||d^i_t||_1$, which plays a similar role of the sample weight in that $\frac{d^i_t}{w^i_t} \in \mathcal{C}^{eor}_1$. We also define $w^{i*} = \sup_t w^i_t$.

We bound the cumulative potential functions by following the modified proof of Theorem 2 from Jung et al. (2017).

Lemma A.3. *With probability $1 - N\delta$, we have*

$$\sum_{t=1}^T \phi^{y_t}_0(s^N_t) \le \phi^1_N(\boldsymbol{0}) \cdot T + S\sum_{i=1}^N w^{i*}.$$

Proof. In the proof of Theorem 2 by Jung et al. (2017), the authors write

$$\sum_{t=1}^T \phi^{y_t}_{N-i+1}(s^{i-1}_t) = \frac{1-\gamma}{k}\sum_{t=1}^T w^i_t - \sum_{t=1}^T d^i_{t,h^i_t} + \sum_{t=1}^T \phi^{y_t}_{N-i}(s^i_t).$$

Using the fact that our weak learners satisfy BanditWLC(γ, δ, S), we have with probability $1-\delta$

$$\frac{1}{w^{i*}}\sum_{t=1}^T d^i_{t,h^i_t} \le \frac{1-\gamma}{kw^{i*}}\sum_{t=1}^T w^i_t + S,$$

from which we deduce

$$\sum_{t=1}^T \phi^{y_t}_{N-i+1}(s^{i-1}_t) + w^{i*}S \ge \sum_{t=1}^T \phi^{y_t}_{N-i}(s^i_t).$$

Summing this over i and using the union bound, we have with probability $1 - N\delta$,

$$\sum_{t=1}^T \phi^{y_t}_0(s^N_t) \le \sum_{t=1}^T \phi^{y_t}_N(\boldsymbol{0}) + S\sum_{i=1}^N w^{i*}.$$

By symmetry, we can check $\phi^l_N(\boldsymbol{0}) = \phi^1_N(\boldsymbol{0})$ for any label $l \in [k]$, which completes the proof. □

We now prove Theorem 3.2.

Proof. Since $\hat{y}_t = \operatorname{argmax}_l s^N_{t,l}$, we obtain

$$\phi^{y_t}_0(s^N_t) = \mathbb{I}(\hat{y}_t \ne y_t).$$

Furthermore, Jung et al. (2017) bound the terms that appear in the previous lemma:

$$\phi^l_N(\boldsymbol{0}) \le (k-1)e^{-\frac{\gamma^2 N}{2}}$$
$$\sum_{i=1}^N w^{i*} = O(k^{5/2}\sqrt{N}).$$

Combining these, we get with probability $1 - N\delta$,

$$\sum_{t=1}^T \mathbb{I}(\hat{y}_t \ne y_t) \le (k-1)e^{-\frac{\gamma^2 N}{2}}T + O(k^{5/2}\sqrt{N}S) \le (k-1)e^{-\frac{\gamma^2 N}{2}}T + \tilde{O}(\frac{k^{7/2}\sqrt{N}}{\rho}),$$

where the last inequality holds by (7).

To bound the booster's loss $\mathbb{I}(\tilde{y}_t \ne y_t)$, observe

$$\mathbb{E}_{\tilde{y}_t}\mathbb{I}(\tilde{y}_t \ne y_t) \le \mathbb{I}(\hat{y}_t \ne y_t) + \rho.$$

Using the concentration inequality, we have with probability $1 - (N+1)\delta$,

$$\sum_{t=1}^T \mathbb{I}(\tilde{y}_t \ne y_t) \le (k-1)e^{-\frac{\gamma^2 N}{2}}T + \tilde{O}(\frac{k^{7/2}\sqrt{N}}{\rho}) + \rho T + \sqrt{T\log\frac{1}{\delta}} \le (k-1)e^{-\frac{\gamma^2 N}{2}}T + 2\rho T + \tilde{O}(\frac{k^{7/2}\sqrt{N}}{\rho}),$$

where we use the relation $\rho T + \frac{\log\frac{1}{\delta}}{\rho} \ge 2\sqrt{T\log\frac{1}{\delta}}$ to absorb the term $\sqrt{T\log\frac{1}{\delta}}$. This proves the main theorem. □

A.4 Proof of Theorem 3.3

We first recall a lemma from Jung et al. (2017) to aid the proof.

Lemma A.4 (Jung et al. (2017), Lemma 11)**.** *Suppose $A, B \ge 0$, $B - A = \gamma \in [-1, 1]$, and $A + B \le 1$. Then we have*

$$\min_{\alpha\in[-2,2]} A(e^\alpha - 1) + B(e^{-\alpha} - 1) \le -\frac{\gamma^2}{2}.$$

Now we proceed with a bound of the zero-one loss of AdaBandit. The main structure of the proof results from the mistake bound of Adaboost.OLM by Jung et al. (2017).

Proof. We let M_i denote the number of mistakes made by expert i: $M_i = \sum_{t=1}^T \mathbb{I}(\hat{y}^i_t \ne y_t)$. We also let $M_0 = T$ for convenience. As the booster uses the estimate $\hat{l}^{0-1}_t$ to run the Hedge algorithm, we define $\hat{M}_i = \sum_{t=1}^T \hat{l}^{0-1}_{t,\hat{y}^i_t}$ so that $\mathbb{E}_{\tilde{y}_1,\cdots,\tilde{y}_T}\hat{M}_i = M_i$. If we write $i^* = \operatorname{argmin}_i M_i$, then by the Azuma-Hoeffding inequality and the fact that $\hat{l}^{0-1}_t$ is $\frac{k}{\rho}$-bounded, we have with probability $1-\delta$,

$$\min_i \hat{M}_i \le \hat{M}_{i^*} \le \min_i M_i + \tilde{O}(\frac{k}{\rho}\sqrt{T}),$$

where $\tilde{O}$ suppresses dependence on $\log \frac{1}{\delta}$.

Then a standard analysis of the Hedge algorithm (see Corollary 2.3 by Cesa-Bianchi and Lugosi (2006)) and the Azuma-Hoeffding inequality provide that with probability $1-3\delta$,

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{I}(\hat{y}_t \neq y_t) &\leq \sum_{t=1}^{T}\hat{l}^{0-1}_{t,\hat{y}_t} + \tilde{O}(\frac{k}{\rho}\sqrt{T}) \\ &\leq 2\min_i \hat{M}_i + 2\log N + \tilde{O}(\frac{k}{\rho}\sqrt{T}) \quad (14)\\ &\leq 2\min_i M_i + 2\log N + \tilde{O}(\frac{k}{\rho}\sqrt{T}).\end{aligned}$$

Now define $w^i = -\sum_{t=1}^{T} c^i_{t,y_t}$. If the expert $i-1$ makes a mistake at round t, there is $l \neq y_t$ such that $s^{i-1}_{t,y_t} \leq s^{i-1}_{t,l}$. According to (10), this implies that $-c^i_{t,y_t} \geq \frac{1}{2}$. From this, we can deduce that

$$w^i \geq \frac{M_{i-1}}{2}. \quad (15)$$

By our convention $M_0 = T$, the above inequality still holds for $i = 1$.

Next we define the difference in the cumulative logistic loss between two consecutive experts as

$$\begin{aligned}\Delta_i &= \sum_{t=1}^{T} l^{\log}_{y_t}(s^i_t) - l^{\log}_{y_t}(s^{i-1}_t) \\ &= \sum_{t=1}^{T} l^{\log}_{y_t}(s^{i-1}_t + \alpha^i_t e_{h^i_t}) - l^{\log}_{y_t}(s^{i-1}_t).\end{aligned}$$

From (12), we have with probability $1-\delta$,

$$\begin{aligned}\Delta_i \leq \min_{\alpha\in[-2,2]} \sum_{t=1}^{T}[l^{\log}_{y_t}(s^{i-1}_t + \alpha e_{h^i_t}) - l^{\log}_{y_t}(s^{i-1}_t)] \\ + \tilde{O}(\frac{k^2}{\rho}\sqrt{T}). \quad (16)\end{aligned}$$

Let us record an inequality:

$$\begin{aligned}\log(1+e^{s+\alpha}) - \log(1+e^s) &= \log(1+\frac{e^\alpha - 1}{1+e^{-s}}) \\ &\leq \frac{e^\alpha - 1}{1+e^{-s}}.\end{aligned}$$

Using this, we can write

$$\begin{aligned}&l^{\log}_{y_t}(s^{i-1}_t + \alpha e_{h^i_t}) - l^{\log}_{y_t}(s^{i-1}_t) \\ &\qquad \leq \begin{cases} c^i_{t,h^i_t}(e^\alpha - 1) & \text{if } h^i_t \neq y_t \\ c^i_{t,h^i_t}(-e^{-\alpha}+1) & \text{if } h^i_t = y_t \end{cases}.\end{aligned}$$

Summing this over t, we get

$$\begin{aligned}&\sum_{t=1}^{T} l^{\log}_{y_t}(s^{i-1}_t + \alpha e_{h^i_t}) - l^{\log}_{y_t}(s^{i-1}_t) \\ &\qquad \leq w^i(A(e^\alpha - 1) + B(e^{-\alpha} - 1)),\end{aligned}$$

where

$$A = \sum_{t:h^i_t \neq y_t} c^i_{t,h^i_t}/w^i,\ B = -\sum_{t:h^i_t = y_t} c^i_{t,h^i_t}/w^i.$$

By (10), A and B are non-negative and $B - A = \gamma_i \in [-1,1]$, which is the empirical edge of WL^i. Then Lemma A.4 implies

$$\min_{\alpha\in[-2,2]} \sum_{t=1}^{T} l^{\log}_{y_t}(s^{i-1}_t + \alpha e_{h^i_t}) - l^{\log}_{y_t}(s^{i-1}_t) \leq -\frac{\gamma_i^2}{2} w^i.$$

Combining this result with (15) and (16), we have with probability $1-\delta$,

$$\Delta_i \leq -\frac{\gamma_i^2}{4} M_{i-1} + \tilde{O}(\frac{k^2}{\rho}\sqrt{T}).$$

Summing this over i and using the union bound, we have with probability $1-N\delta$,

$$\begin{aligned}&\sum_{t=1}^{T} l^{\log}_{y_t}(s^N_t) - l^{\log}_{y_t}(\mathbf{0}) \\ &\qquad \leq -\frac{\min_i M_i}{4}\sum_{i=1}^{N}\gamma_i^2 + \tilde{O}(\frac{k^2 N}{\rho}\sqrt{T}).\end{aligned}$$

Since $l^{\log}_{y_t}(\mathbf{0}) = (k-1)\log 2$ and $l^{\log}_{y_t}(s^N_t) \geq 0$, we have with probability $1-N\delta$,

$$\min_i M_i \leq \frac{4(k-1)\log 2}{\sum_{i=1}^{N}\gamma_i^2} T + \tilde{O}(\frac{k^2 N}{\rho\sum_{i=1}^{N}\gamma_i^2}\sqrt{T}).$$

Using this to (14), we get with probability $1-(N+3)\delta$,

$$\sum_{t=1}^{T}\mathbb{I}(\hat{y}_t \neq y_t) \leq \frac{8(k-1)\log 2}{\sum_{i=1}^{N}\gamma_i^2} T + \tilde{O}(\frac{k^2 N}{\rho\sum_{i=1}^{N}\gamma_i^2}\sqrt{T}).$$

Then by the same argument as in Appendix A.3, we get with probability $1-(N+4)\delta$,

$$\begin{aligned}&\sum_{t=1}^{T}\mathbb{I}(\tilde{y}_t \neq y_t) \\ &\quad \leq \frac{8(k-1)\log 2}{\sum_{i=1}^{N}\gamma_i^2} T + 2\rho T + \tilde{O}(\frac{k^2 N}{\rho\sum_{i=1}^{N}\gamma_i^2}\sqrt{T}) \\ &\quad \leq \frac{8k}{\sum_{i=1}^{N}\gamma_i^2} T + 2\rho T + \tilde{O}(\frac{k^3 N^2}{\rho^2\sum_{i=1}^{N}\gamma_i^2}),\end{aligned}$$

where the last inequality comes from the arithmetic mean and geometric mean relation:

$$2ck^2 N\sqrt{T} \leq kT + c^2 k^3 N^2.$$

□

Table 2: Bandit and Full Information Total Accuracy

Data	k	StreamCnt	OptBandit	AdaBandit	BinBandit	OptFull	AdaFull
Balance	3	6250	0.83	0.91	0.73	0.71	0.85
Car	4	10368	0.82	0.93	0.78	0.80	0.93
Nursery	4	51840	0.89	0.92	0.89	0.91	0.95
Movement	5	165631	0.89	0.95	0.84	0.87	0.95
Mice	8	8640	0.53	0.71	0.62	0.65	0.84
Isolet	26	116955	0.32	0.51	0.52	0.74	0.78

Appendix B DETAILED DESCRIPTIONS OF EXPERIMENTS

We discuss the experimental results more in detail.

B.1 Data Set Details

We modified the data sets identically as in Jung et al. (2017). In particular, we replaced missing data values in Mice data with 0 and removed user information from Movement, leaving only sensor data in the latter. A single data point with missing values was also removed from Movement. Lastly, for Isolet the original 617 covariates were projected onto the top 50 principal components of the data set, retaining 80% of the variance. Table 3 summarizes the data information after preprocessing.

B.2 Parameter Tuning

Our boosting algorithms have a few parameters: the number of weak learners N, the edge γ for BanditBBM, and the exploration rate ρ. As the goal of the experiment is to compare the bandit algorithms with their full information counterparts, we did not optimize the parameters too hard. We optimized N up to multiples of 5 and fix $\gamma = 0.1$ for all data sets.

The only parameter we tried to fit is exploration rate. To obtain a reasonable ρ, we ran a grid search keeping all other parameters stable and observing accuracy on a random stream from each data set. The chosen ρ values reflect choices that made all the bandit algorithms perform well on each set. Table 4 shows the chosen ρ from each of these grid searches. How many data points were streamed to obtain the final accuracy is shown in column Count.

B.3 Updating Weak Learners

We used the VFDT algorithm designed by Domingos and Hulten (2000). The learner takes a label and an importance weight to be updated. When a cost vector $\hat{c}_t^i \in \mathbb{R}^k$ is passed to the learner, we used

$$l = \operatorname*{argmin}_j \hat{c}_{t,j}^i \text{ and } w = \sum_{j=1}^{k} (\hat{c}_{t,j}^i - \hat{c}_{t,l}^i)$$

as the label and the importance weight, respectively. If there were multiple minima in $\hat{c}_t^i$, we chose one of them randomly in a mistake round and selected the true label y_t in a correct round if it minimized the cost vector.

One weakness of VFDT is that its performance is very sensitive to the range of importance weights. To address this, we added clipping in our implementation to prevent cost vectors from having excessively large entries. We introduced a magic number 100 and clipped the entry whenever it went outside the range $[-100, 100]$. Clipping was especially helpful in stabilizing results for data sets with large k. Additionally,

Table 3: Data Set Summary

Data	Size	Dimension	k
Balance	625	4	3
Car	1728	82	4
Nursery	12960	4	8
Mice	1080	82	8
Isolet	7797	50	26
Movement	165631	12	4

Table 4: Parameters for Bandit Algorithms

Data	Count	ρ	N_{Opt}	N_{Ada}	γ
Balance	2500	0.001	20	15	0.1
Car	6912	0.001	15	15	0.1
Nursery	12960	0.001	10	5	0.1
Movement	165631	0.001	10	20	0.1
Mice	4320	0.1	10	20	0.1
Isolet	38985	0.1	10	20	0.1

we scaled the importance weights for Isolet, as large k tends to create larger weights.

B.4 Total Accuracy

Table 2 shows the average accuracy of all five algorithms on each of the data sets, in contrast to the asymptotic performance in Table 1. The effect of increased k on the excess loss is noticeable, showing that the bandit algorithms learn less quickly. For data sets with smaller k, the total loss is a fairly large percentage of the asymptotic loss, indicating the the algorithms stay at their asymptotic accuracy for a larger fraction of rounds. For Mice and Isolet data, however, the total accuracy is significantly lower than the asympototic loss, indicating a much more linear improvement that is amortized over the whole data stream. Indeed, this corroborates Figure 1, where accuracy improvement of the bandit algorithms slows and tends towards a straight line for large k.

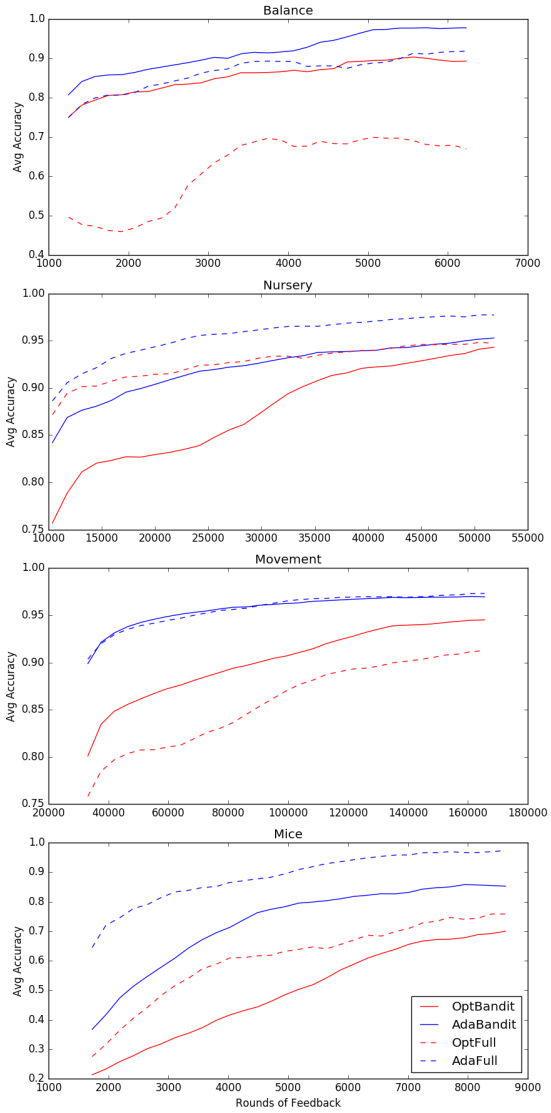

Figure 2: Learning Curves on Balance (Top), Nursery, Movement, and Mice (Bottom); Best Viewed in Color

B.5 Learning Curves

Figure 2 exhibits the learning curves of our bandit algorithms and the full information ones. Similar to the analysis in the main paper, the bandit algorithms learn slower than the full information algorithms in general, and the trend becomes obvious when k gets larger. However, the bandit algorithms, especially AdaBandit, become competitive in the end and sometimes outperform the full information algorithms. The underperformance of the optimal algorithms is partially because we did not optimize the edge γ, and this aspect makes adaptive algorithms more suitable in practice. It should be noted that the learning curves do not begin at round 0 because the window accuracy is not defined for a number of rounds less than 20% of the total rounds to be given.