Three Optimal Algorithms for Balls of Three Colors*

Zdeněk Dvořák Vít Jelínek Daniel Král' Jan Kynčl
Department of Applied Mathematics and
Institute for Theoretical Computer Science (ITI)†
Charles University
Malostranské náměstí 25, 118 00 Prague, Czech Republic
E-mail: {rakdver,jelinek,kral,kyncl}@kam.mff.cuni.cz

Michael Saks‡
Department of Mathematics
Rutgers, the State University of NJ
110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA
E-mail: saks@math.rutgers.edu

Abstract

We consider a game played by two players, Paul and Carol. Carol fixes a coloring of n balls with three colors. At each step, Paul chooses a pair of balls and asks Carol whether the balls have the same color. Carol truthfully answers yes or no. In the Plurality problem, Paul wants to find a ball with the most common color. In the Partition problem, Paul wants to partition the balls according to their colors. He wants to ask Carol the least number of questions to reach his goal. We find optimal deterministic and probabilistic strategies for the Partition problem and an asymptotically optimal probabilistic strategy for the Plurality problem.

*The REU program, during which the research was conducted, is supported by grant KONTAKT ME 521.

†Institute for Theoretical Computer Science is supported by Ministry of Education of Czech Republic as project LN00A056.

‡Research supported in part by NSF grant CCR-9988526.

1 Introduction

We study a game played by two players, Paul and Carol, in which Paul wants to determine a certain property of the input based on Carol's answers. Carol fixes a coloring of n balls by k colors. Paul does not know the coloring of the balls. At each step, he chooses two balls and asks Carol whether they have the same color. Carol truthfully answers `YES` or `NO`. Paul wants to ask the least number of questions in the worst case to determine the desired property of the coloring.

The first problem of this kind which was considered is the *Majority problem*, in which Paul wants to find a ball b such that the number of balls colored with the same color as b is greater than $n/2$, or to declare that there is no such ball. Saks and Werman [10], later Alonso, Reingold and Schott [4], showed that $n - \nu(n)$ questions are necessary and sufficient for Paul to resolve the Majority problem with n balls of two colors, where $\nu(n)$ is the number of 1's in the binary representation of n. Fisher and Salzberg [6] showed that $\lceil 3n/2 \rceil - 2$ questions are necessary and sufficient to solve the Majority problem with n balls and an unrestricted number of colors. Some variants of the Majority problem were also considered in [1, 7].

In this paper, we consider the Plurality problem, introduced by Aigner et al. [2], and the Partition problem. In the *Plurality problem*, Paul seeks for a ball such that the number of balls with the same color exceeds the number of balls of any other color (or he finds out that there is a tie between two or more different colors). In the *Partition problem*, Paul wants to partition the balls according to their colors. Aigner et al. [2] found a strategy to solve the Plurality problem with n balls of three colors such that Paul asks at most $\lfloor \frac{5n}{3} \rfloor - 2$ questions. On the other hand, Carol can force Paul to ask at least $3\lfloor \frac{n}{2} \rfloor - 2$ questions. In addition, if the number k of colors of the balls contained in the input is not fixed to be three, Aigner et al. [3] provide lower and upper bounds of order $\Theta(kn)$. A problem similar to the Plurality problem was also studied by Srivastava [11].

We focus on the case when the balls are colored with three colors and present a probabilistic strategy for the Plurality problem and both deterministic and probabilistic strategies for the Partition problem. In the probabilistic setting, Paul may flip coins and use the outcome to choose his questions. The quality of a probabilistic strategy is measured as the maximum of the expected number of Paul's questions over all inputs. For both the deterministic and probabilistic strategies, we assume that Carol knows the strategy and chooses the worst coloring. In the deterministic setting,

The problem	Lower bound	Upper bound
The Plurality Problem		
Deterministic strategy [2]	$3\lfloor \frac{n}{2} \rfloor - 2$	$\lfloor \frac{5n}{3} \rfloor - 2$
Probabilistic strategy	$\frac{3}{2}n - O\left(\sqrt{n \log n}\right)$	$\frac{3}{2}n + O(1)$
The Partition Problem		
Deterministic strategy	$2n - 3$	$2n - 3$
Probabilistic strategy	$\frac{5}{3}n - \frac{8}{3}$	$\frac{5}{3}n - \frac{8}{3} + o(1)$

Table 1: Bounds for the Plurality and Partition problems with n balls of three colors.

Carol can also choose the coloring on-line in response to the questions. This is not appropriate for probabilistic strategies, since we assume that Carol does not know outcome of the coin flips.

Our results are summarized in Table 1. In the case of deterministic strategy of the Partition problem, we provide matching lower and upper bounds on the number of Paul's questions. The result can be generalized for balls with arbitrary number of colors (see Section 6). In the probabilistic setting, our bounds for the Partition problem match up to the $o(1)$ term. For the Plurality problem, we managed to prove a lower bound in the probabilistic setting which is close to the lower bound proved by Aigner et al. [2] in the (more constrained) deterministic setting and we show that the lower bound is asymptotically tight by providing a matching upper bound.

2 Notation

In this section, we introduce a compact way of representing a deterministic strategy for Paul, and the state of his information about the colors of the balls as the strategy proceeds.

The game of Paul and Carol can be viewed as a game on a graph whose vertices are the balls. Initially the graph is empty. At each turn, Paul chooses a pair of nonadjacent vertices and adds that edge to the graph. Carol then colors the edge by red if the two vertices have the same color, or by blue if the two vertices have a different color. This edge-colored graph represents the state of Paul's knowledge and is referred to as *Paul's graph*. Notice that each connected component of red edges consists of vertices corresponding to balls with the same color. *The reduced graph* has as its vertex set each of these red connected components, with two components joined if there is at least one blue edge between them. In the Partition problem with k colors, the game ends when the reduced graph is uniquely vertex k-colorable (up to a permutation of the colors). In the Plurality problem with k colors, the game ends when there is a vertex v in the reduced graph with the property that in every vertex k-coloring, v belongs to a largest color class, where the size of a color class is the sum of orders of the contracted components.

A deterministic strategy for Paul can be represented by a rooted binary tree in which the left edge from each internal vertex is colored with red and the right edge with blue. The root is associated with Paul's first question. The left subtree represents Paul's strategy for the case when Carol answers that the colors of the balls are the same, and the right one for the case when she answers that they are different. At each node, Paul's information can be represented by a graph as described above. For a given coloring of the balls, there is a unique path from the root to a leaf in the tree. This path in the tree is called the *computation path.*

3 Yao's Principle

Yao [12] proposed a technique for proving lower bounds on probabilistic algorithms which is based on the minimax principle from game theory. Informally, to prove such a lower bound, instead of constructing a hard coloring of the balls for every probabilistic algorithm, it is enough to find a probability distribution on colorings which is hard for every deterministic algorithm. Yao's technique applies to our setting, too. We formulate the principle formally using our notation as a proposition. Since the proof follows the same line as in the original setting, we do not include it and refer the reader, e.g., to [9, Subsection 2.2.2] if necessary.

Proposition 1 *If for the Plurality or Partition problem with n balls of k colors, there exists a probability distribution on colorings of the balls such that the expected number of Paul's questions is at least K for each deterministic strategy, then for each probabilistic strategy there exists a coloring $\mathcal{I}$ of the balls such that the expected number of Paul's questions for the coloring $\mathcal{I}$ is at least K.*

4 Probabilistic Strategy for the Plurality Problem

We are now ready to present the first of our results, an asymptotically optimal probabilistic strategy for the Plurality problem:

Theorem 2 *There is a probabilistic strategy for the Plurality problem with n balls of three colors such that the expected number of Paul's questions does not exceed $\frac{3}{2}n + O(1)$ for any coloring of the balls.*

Proof: Fix a coloring of the balls and choose any subset $\mathcal{B}_0$ of $3n'$ balls from the input, where $n' = \lfloor \frac{n}{3} \rfloor$. Partition randomly the set $\mathcal{B}_0$ into n' ordered triples (a_i, b_i, c_i), $1 \le i \le n'$. For each i, $1 \le i \le n'$, Paul asks Carol whether the balls a_i and b_i have the same color and whether the balls b_i and c_i have the same color. If Carol answers in both the cases that the balls have different colors, Paul asks, in addition, whether the colors of the balls a_i and c_i are the same.

Based on Carol's answers, Paul is able to classify the triples into three types:

type A All the three balls of the triple have the same color. This is the case when Carol answers both the initial questions positively.

type B Two balls of the triple have the same color, but the remaining one has a different color. This is the case when Carol answers one of the initial questions positively and the other one negatively, or both the initial questions negatively and the additional question positively.

type C All the three balls have different colors. This is the case when Carol answers the initial questions and the additional question negatively.

Paul now chooses randomly and independently a representative ball from each triple of type A or B. Let $\mathcal{B}$ be the set of (at most n') chosen balls. In addition, he chooses randomly a ball d from $\mathcal{B}_0$.

For each ball from the set $\mathcal{B}$, Paul asks Carol whether its color is the same as the color of d. Let $\mathcal{B}' \subseteq \mathcal{B}$ be the set of the balls whose colors are different. Paul chooses arbitrarily a ball $d' \in \mathcal{B}'$ and compares the ball d' with the remaining balls of $\mathcal{B}'$. Paul is able to determine the partition of the balls of $\mathcal{B}$ according to their colors: the balls of $\mathcal{B} \setminus \mathcal{B}'$, the balls of $\mathcal{B}'$ which have the same color as the ball d' and the balls of $\mathcal{B}'$ whose color is different from the color of d'.

Finally, Paul determines the partition of all the balls from the triples of type A or B. The balls contained in a triple of type A have the same color as its representative. In the case of a triple of type B, Paul asks Carol whether the colors of balls d_1 and d_2 are the same, where d_1 is a ball of the triple whose color is different from the color of the representative, and d_2 is a ball of $\mathcal{B}$ whose color is different from the color of the representative. In this way, Paul obtains the partition of all the balls from triples of type A or B according to their colors.

After at most $2(n-3n')$ additional questions, Paul knows the partition of the balls of $\overline{\mathcal{B}}$ according to their colors, where $\overline{\mathcal{B}}$ is the set of all the n balls except for the balls of $\mathcal{B}_0$ which are contained in triples of type C. Since each triple of type C contains one ball of each of the three colors, the plurality color of the balls of $\overline{\mathcal{B}}$ is also the plurality color of all the balls. If there is no plurality color in $\overline{\mathcal{B}}$, then there is no plurality color in the original problem either.

Before we formally analyze the described strategy, we explain some ideas behind it. Let α, β and γ be the fractions of the balls of each of the colors among the balls of $\mathcal{B}_0$. If the ratios α, β and γ are close to $1/3$, then a lot of the balls belong to the triples of type C. Clearly, such balls can be removed from the problem and we solve the Plurality problem for the remaining balls (this reduces the size of the problem). However, if the ratios α, β and γ are unbalanced, then the previous fails to work. But in this case, with high probability, the ball d has the same color as a lot of the balls of $\mathcal{B}$ and Paul does not need to compare too many balls of $\mathcal{B}$ with the ball d'.

We are now ready to start estimating the expected number of Paul's questions. The expected numbers of triples of each type are the following:

- $\left(\alpha^3+\beta^3+\gamma^3+O\left(\frac{1}{n'}\right)\right)n'$ triples of type A,
- $\left(3(\alpha^2\beta+\alpha^2\gamma+\beta^2\alpha+\beta^2\gamma+\gamma^2\alpha+\gamma^2\beta)+O\left(\frac{1}{n'}\right)\right)n'$ triples of type B, and
- $\left(6\alpha\beta\gamma+O\left(\frac{1}{n'}\right)\right)n'$ triples of type C.

The expected numbers of Paul's questions to determine the type of the triple are 2, $\frac{7}{3}$ and 3 for types A, B and C, respectively. Therefore, the expected number of Paul's questions to determine the types of all the triples is:

$$\left(2(\alpha^3+\beta^3+\gamma^3)+7(\alpha^2\beta+\alpha^2\gamma+\beta^2\alpha+\beta^2\gamma+\gamma^2\alpha+\gamma^2\beta)+18\alpha\beta\gamma\right)n'+O(1) \tag{1}$$

Next, we compute the expected number of Paul's questions to determine the partition of the balls of $\mathcal{B}$. Fix a single ball b_0 out of the $3n'$ balls. Assume that the color of b_0 is the color with the fraction α. Since the probability that the ball b_0 is in a triple of type C is $2\beta\gamma + O(\frac{1}{n'})$, the ball b_0 is contained in the set $\mathcal{B}$ with the probability $\frac{1}{3} - \frac{2}{3}\beta\gamma + O\left(\frac{1}{n'}\right)$. If $b_0 \in \mathcal{B}$ and the colors of the balls b_0 and d are the same, Paul asks Carol a single question; if $b_0 \in \mathcal{B}$, the colors of b_0 and d are different, and $b_0 \neq d'$, he asks two questions. The former is the case with the probability α. Hence, if $b_0 \in \mathcal{B}$, Paul asks $2 - \alpha$ questions on average. We may conclude that in this stage, the expected number of Paul's questions involving balls of the color with the fraction α does not exceed:

$$3\alpha n' \left(\frac{1}{3} - \frac{2}{3}\beta\gamma + O\left(\frac{1}{n'}\right)\right)(2-\alpha)$$

Hence, the expected number of the questions to determine the partition of $\mathcal{B}$ is:

$$\begin{aligned}
& 3\alpha n' \left(\frac{1}{3} - \frac{2}{3}\beta\gamma + O\left(\frac{1}{n'}\right)\right)(2-\alpha) \\
+ \; & 3\beta n' \left(\frac{1}{3} - \frac{2}{3}\alpha\gamma + O\left(\frac{1}{n'}\right)\right)(2-\beta) \\
+ \; & 3\gamma n' \left(\frac{1}{3} - \frac{2}{3}\alpha\beta + O\left(\frac{1}{n'}\right)\right)(2-\gamma) \\
= \; & (2 - \alpha^2 - \beta^2 - \gamma^2)n' - 2\alpha\beta\gamma n'(6 - \alpha - \beta - \gamma) + O(1) \\
= \; & (2 - \alpha^2 - \beta^2 - \gamma^2)n' - 10\alpha\beta\gamma n' + O(1) \qquad (2)
\end{aligned}$$

Next, Paul asks Carol a single question for each triple of type B. The expected number of such questions is:

$$3(\alpha^2\beta + \alpha^2\gamma + \beta^2\alpha + \beta^2\gamma + \gamma^2\alpha + \gamma^2\beta)n' + O(1) \qquad (3)$$

Finally, Paul asks at most $2(n - 3n') \leq 4$ questions to find the partition of $\overline{\mathcal{B}}$.

The expected number of all the questions asked by Paul is given by the sum of (1), (2) and (3), which is equal to the following expression:

$$\begin{aligned}
& \big(2 + 2(\alpha^3 + \beta^3 + \gamma^3) + 10(\alpha^2\beta + \alpha^2\gamma + \beta^2\alpha + \beta^2\gamma + \gamma^2\alpha + \gamma^2\beta) \\
+ \; & 8\alpha\beta\gamma - \alpha^2 - \beta^2 - \gamma^2\big)n' + O(1) \qquad (4)
\end{aligned}$$

It can be verified, see Proposition 3 following the proof of this theorem, that the maximum of (4) for $\alpha, \beta, \gamma \in [0,1]$ with $\alpha+\beta+\gamma=1$ is attained for $\alpha=\beta=1/2$ and $\gamma=0$ (and the two other symmetric permutations of the values of the variables α, β and γ). Therefore, the expected number of Paul's questions does not exceed:

$$\left(2+\frac{2\cdot 2+10\cdot 2+8\cdot 0}{8}-\frac{2}{4}\right)n'+O(1)=\frac{9}{2}n'+O(1)=\frac{3}{2}n+O(1)\ .$$

■

We now prove technical Proposition 3 that is referred in the proof of Theorem 2:

Proposition 3 *The maximum, which is equal to $\frac{9}{2}$, of the function*

$$2+2(\alpha^3+\beta^3+\gamma^3)+10(\alpha^2\beta+\alpha^2\gamma+\beta^2\alpha+\beta^2\gamma+\gamma^2\alpha+\gamma^2\beta)+8\alpha\beta\gamma-\alpha^2-\beta^2-\gamma^2$$

with the variables $\alpha, \beta, \gamma \in [0,1]$ and with the additional constraint $\alpha+\beta+\gamma=1$ is attained for the following combinations of the values of α, β and γ:

- $\alpha=\beta=\frac{1}{2}$ *and* $\gamma=0$,
- $\alpha=\gamma=\frac{1}{2}$ *and* $\beta=0$, *or*
- $\beta=\gamma=\frac{1}{2}$ *and* $\alpha=0$.

Proof: First, we apply several substitutions to the function. Since $\alpha+\beta+\gamma=1$, we have the following:

$$\alpha^2\beta+\alpha^2\gamma+\beta^2\alpha+\beta^2\gamma+\gamma^2\alpha+\gamma^2\beta=\alpha^2+\beta^2+\gamma^2-\alpha^3-\beta^3-\gamma^3 \quad (5)$$

Plug the equality (5) to the examined function:

$$2-8(\alpha^3+\beta^3+\gamma^3)+8\alpha\beta\gamma+9(\alpha^2+\beta^2+\gamma^2) \quad (6)$$

Similarly, we establish the following equality using the constraint $\alpha+\beta+\gamma=1$ and (5):

$$\begin{aligned}
1 &= (\alpha+\beta+\gamma)^3\\
1 &= \alpha^3+\beta^3+\gamma^3+3(\alpha^2\beta+\alpha^2\gamma+\beta^2\alpha+\beta^2\gamma+\gamma^2\alpha+\gamma^2\beta)+6\alpha\beta\gamma\\
1 &= -2(\alpha^3+\beta^3+\gamma^3)+3(\alpha^2+\beta^2+\gamma^2)+6\alpha\beta\gamma\\
-6\alpha\beta\gamma &= -1-2(\alpha^3+\beta^3+\gamma^3)+3(\alpha^2+\beta^2+\gamma^2)\\
8\alpha\beta\gamma &= \frac{4}{3}+\frac{8}{3}(\alpha^3+\beta^3+\gamma^3)-4(\alpha^2+\beta^2+\gamma^2) \quad (7)
\end{aligned}$$

We combine (6) and (7) and obtain that the examined function is equal to:

$$\frac{10}{3} - \frac{16}{3}(\alpha^3 + \beta^3 + \gamma^3) + 5(\alpha^2 + \beta^2 + \gamma^2) \tag{8}$$

Let $S \subset \mathbb{R}^3$ be the set of $\alpha, \beta, \gamma \in [0,1]$ satisfying $\alpha + \beta + \gamma = 1$. Clearly, the set S is an equilateral triangle in the 3-dimensional space which is contained in the plane with the norm vector equal to $(1,1,1)$. The maximum of the function is attained on the boundary of the set S or in an internal point of S where the gradient of the function is a multiple of the norm vector of the support plane.

We first examine the function on the boundary of the set S. In such case, one of the variables α, β and γ is equal to zero. We can assume $\gamma = 0$ by symmetry. We substitute $\beta = 1 - \alpha$ and $\gamma = 0$ to (8):

$$\frac{10}{3} - \frac{16}{3}(1 - 3\alpha + 3\alpha^2) + 5(1 - 2\alpha + 2\alpha^2) \tag{9}$$

Since the first derivative of (9) is $16(1-2\alpha) - 10(1-2\alpha) = 6(1-2\alpha)$ and it is equal to zero only for $\alpha = 1/2$, the extremes of (9) for $\alpha \in [0,1]$ can be attained only for $\alpha = 0$, $\alpha = 1/2$, and $\alpha = 1$. For these values of α, the expression (9) is equal to 3, $9/2$ and 3, respectively.

We now examine the possibility that the extreme of (8) is attained at an internal point of S. In such case, the gradient determined by partial derivatives of (8) is a multiple of the vector $(1,1,1)$. The partial derivatives of (8) are the following:

$$\begin{array}{rcll}
\frac{\partial}{\partial\alpha}\left(\frac{10}{3} - \frac{16}{3}(\alpha^3+\beta^3+\gamma^3) + 5(\alpha^2+\beta^2+\gamma^2)\right) & = & -16\alpha^2 + 10\alpha & , \\
\frac{\partial}{\partial\beta}\left(\frac{10}{3} - \frac{16}{3}(\alpha^3+\beta^3+\gamma^3) + 5(\alpha^2+\beta^2+\gamma^2)\right) & = & -16\beta^2 + 10\beta & , \text{and} \\
\frac{\partial}{\partial\gamma}\left(\frac{10}{3} - \frac{16}{3}(\alpha^3+\beta^3+\gamma^3) + 5(\alpha^2+\beta^2+\gamma^2)\right) & = & -16\gamma^2 + 10\gamma & .
\end{array}$$

Since the gradient $(-16\alpha^2 + 10\alpha, -16\beta^2 + 10\beta, -16\gamma^2 + 10\gamma)$ is a multiple of the vector $(1,1,1)$ only if all the right hand sides are equal, it holds that $\alpha = \beta$ or $\alpha = \frac{5}{8} - \beta$. Similarly, $\alpha = \gamma$ or $\alpha = \frac{5}{8} - \gamma$, and $\beta = \gamma$ or $\beta = \frac{5}{8} - \gamma$. Hence, $\alpha = \beta = \gamma = 1/3$, or two of the variables are equal to $3/8$ and the remaining one to $1/4$. The value of (8) for $\alpha = \beta = \gamma = 1/3$ is equal to $\frac{119}{27} < \frac{9}{2}$. The value for $\alpha = \beta = 3/8$ and $\gamma = 1/4$ is equal to $\frac{141}{32} < \frac{9}{2}$.

We conclude that the maximum of the function from the statement of the proposition is equal to $\frac{9}{2}$ and it is attained for $\alpha = \beta = 1/2$ and $\gamma = 0$ and the other two permutation of the values of the variables α, β and γ.

∎

5 Lower Bound for the Plurality Problem

We first survey some basic results from information theory that we use in the proof of Lemma 6. The reader is welcome to see [5] for further details. If X is a random variable taking values in the finite set S, the *entropy* $H(X)$ of X is $\sum_{s\in S} p_s \log_2 1/p_s$ where p_s is the probability that X is equal to $s \in S$. In particular, if X is a zero-one random variable such that $X = 1$ with probability p, then the entropy of X is equal to $h(p) := p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$. We need the following elementary result on entropies of random variables taking values in finite sets:

Proposition 4 *Let $X_1, \ldots, X_k$ be (not necessarily independent) random variables and let $X = (X_1, \ldots, X_k)$ be the vector random variable whose entries are equal to random variables $X_1, \ldots, X_k$. The following holds:*

$$H(X) \le \sum_{i=1}^{k} H(X_i) \ .$$

In addition to the previous proposition, we state the following well-known combinatorial estimate on the middle binomial coefficient which can be found, e.g., in [8]:

Proposition 5 *Let n be a positive even integer. The following bounds hold:*

$$\frac{2^n}{\sqrt{2n}} \le \binom{n}{n/2} \le \frac{2^n}{\sqrt{n}} \ .$$

We are now ready to prove the following lemma on binary trees of bounded depths:

Lemma 6 *Let n be an even positive integer, and let T be a rooted binary tree of depth at most $n-1$ with $\binom{n}{n/2}/2$ leaves. Assume that for each inner node w of T, the edge which leads to its left child is colored by red and the edge which leads to its right child by blue. The average number of blue edges on the paths from the root to the leaves in T is at least:*

$$\frac{n}{2} - O\left(\sqrt{n\log n}\right) \ .$$

Proof: We first assign each leaf w of T a string $\sigma(w)$ of R's and B's of length n. If $v_0, \ldots, v_k$ is the path in T from the root to a leaf $w = v_k$, then the i-th letter of the sequence $\sigma(w)$ is R if the edge $v_{i-1}v_i$ is red, and B otherwise. If $i > k$, the i-th letter of $\sigma(w)$ is R. Observe that the strings assigned to the leaves of T are mutually distinct. Moreover, the average number μ of blue edges on the paths from the root to the leaves is equal to the average number of letters B contained in the strings $\sigma(w)$.

Choose uniformly at random a leaf w of the tree T and let X_i be the random variable equal to the i-th letter of the string $\sigma(w)$ for $1 \le i \le n$. Let p_i be the probability that X_i is equal to B. Clearly, $\mu = \sum_{i=1}^{n} p_i$. Consider the random variable X equal to the vector $(X_1, \ldots, X_n)$. For every string $\sigma(w)$, X is equal to $\sigma(w)$ with probability $2\binom{n}{n/2}^{-1}$. Hence, we have the following by Proposition 5:

$$H(X) = \log \binom{n}{n/2}/2 \ge \log \frac{2^n}{2\sqrt{2n}} \tag{10}$$

On the other hand, we can infer from Proposition 4:

$$H(X) \le \sum_{i=1}^{n} h(p_i) \tag{11}$$

We combine the inequalities (10) and (11), and using the fact that $h(p)$ is a concave function on the interval $(0,1)$, together with the estimate $h(1/2 - x) \le 1 - x^2$ for $x \in (-1/2, 1/2)$, we obtain the required bound (recall that $\mu = \sum_{i=1}^{n} p_i$):

$$\begin{aligned}
\log \frac{2^n}{2\sqrt{2n}} \le \quad H(X) \quad &\le \sum_{i=1}^{n} h(p_i) \\
n - \frac{\log 8n}{2} \quad &\le \quad n \cdot h\left(\frac{\mu}{n}\right) \\
1 - \frac{\log 8n}{2n} \quad &\le \quad 1 - \left(\frac{1}{2} - \frac{\mu}{n}\right)^2 \\
\left|\frac{1}{2} - \frac{\mu}{n}\right| \quad &\le \quad \sqrt{\frac{\log 8n}{2n}} \\
\frac{1}{2} - \sqrt{\frac{\log 8n}{2n}} \quad &\le \quad \frac{\mu}{n} \\
\frac{n}{2} - O\left(\sqrt{n \log n}\right) \quad &\le \quad \mu
\end{aligned}$$

The statement of the lemma now follows.

■

We are now ready to prove the desired lower bound:

Theorem 7 *For any probabilistic strategy for the Plurality problem with* n *balls of three colors, there is a coloring of the balls such that Paul asks at least* $\frac{3}{2}n - O\left(\sqrt{n\log n}\right)$ *questions on average.*

Proof: By Yao's principle (Proposition 1), it is enough to find a probability distribution on colorings of the balls such that if Paul uses any deterministic strategy, then he asks at least $\frac{3}{2}n - O\left(\sqrt{n\log n}\right)$ questions on average. If n is even, then the desired distribution is the uniform distribution on colorings in which half of balls have the first color, the other half have the second color, and there are no balls of the third color. If n is odd, then the desired distribution is the uniform distribution on colorings such that half of the first $n-1$ balls have the first color, the other half the second color, and the n-th ball has the third color. Clearly, there is no plurality color for any of these colorings. Let $\mathcal{I}$ be the set of all $\binom{n'}{n'/2}$ such colorings where $n' = n$, if n is even, and $n' = n-1$, otherwise.

Fix a deterministic strategy of Paul. Let G be Paul's final graph for one of the colorings from $\mathcal{I}$. We claim that each of the two subgraphs of G induced by the balls of the first or the second color is connected. Otherwise, let V_1 and V_2 be the sets of the vertices corresponding to the balls of these colors, and assume that $G[V_1]$ contains a component with a vertex set $W \subset V_1$. Based on G, Paul is unable to determine whether the balls of W have the same color as the balls of $V_1 \setminus W$, and thus he is unable to decide whether there is a plurality color (which would be the case if the balls of W had the third color). We conclude that Paul's final graph contains at least $n' - 2$ red edges.

Let T be the rooted binary tree corresponding to Paul's strategy and let V_{lf} be the set of the leaves of T to which there is a valid computation path for a coloring of $\mathcal{I}$ (note that T contains additional leaves corresponding to colorings not contained in $\mathcal{I}$). Since for each of the colorings of $\mathcal{I}$ the two subgraphs induced by the balls of the first two colors in Paul's final graph are connected, each leaf of T can correspond to two such colorings (which differ just by a permutation of the first two colors). Therefore, V_{lf} consists of $\binom{n'}{n'/2}/2$ leaves of T.

We now modify T to a tree T' and eventually to a tree T''. The tree T' is a subtree of T formed by the union of the paths from the root to the leaves contained in V_{lf}. In particular, T' has exactly $\binom{n'}{n'/2}/2$ leaves. Color the edges of T' by red and blue according to whether they join an inner vertex with its left or right child. For each inner vertex w with a single child in T', contract the edge leading from w to its only child. Let T'' be the obtained binary tree with $\binom{n'}{n'/2}/2$ leaves. The colors of the edges which have not been contracted are preserved.

The tree T'' corresponds to a deterministic strategy for distinguishing the colorings from the set $\mathcal{I}$ restricted to the first n' balls. At each inner node w of T'', the edge corresponding to Paul's question at the node w joins two different components of Paul's graph. Otherwise, the answer is uniquely determined by his graph. Consequently, the node w has a single child (there are no colorings of $\mathcal{I}$ consistent with the other answer) and the edge leading from w to its child should have been contracted.

Since all Paul's questions correspond to edges between different components, Paul's final graph (for his strategy determined by T'') is a forest for each coloring of $\mathcal{I}$. In particular, Paul's final graph contains at most $n'-1$ edges. Therefore, the depth of T'' does not exceed $n'-1$. By Lemma 6, the average number of blue edges on the path from the root to a leaf of T'' is at least $\frac{n'}{2} - O\left(\sqrt{n'\log n'}\right)$. Since the number of blue edges on such a path is equal to the number of blue edges in Paul's final graph G'' if Paul follows the strategy determined by T'', and $n = n' + O(1)$, the average number of blue edges in Paul's graphs is at least $\frac{n}{2} - O\left(\sqrt{n\log n}\right)$.

Observe that for each coloring of $\mathcal{I}$, the edges of the computation path in T'' form a subset of the edges of the computation path in T. Therefore, the average number of blue edges in Paul's final graphs with respect to the strategy corresponding to T is also at least $\frac{n}{2} - O\left(\sqrt{n\log n}\right)$. Since each final graph contains also (at least) $n'-2 \geq n-3$ red edges, the average number of Paul's questions, which is equal to the average number of edges in the final graph, is at least $\frac{3}{2}n - O\left(\sqrt{n\log n}\right)$.

■

6 Deterministic Strategy for the Partition Problem

We first describe Paul's strategy:

Proposition 8 *There is a deterministic strategy for the Partition problem with n balls of k colors such that Paul always asks at most $(k-1)n - \binom{k}{2}$ questions if $n \geq k$, and at most $\binom{n}{2}$ questions otherwise.*

Proof: Paul's strategy is divided into n steps. In the i-th step, Paul determines the color of the i-th ball.

If the first $i-1$ balls have only $k' < k$ different colors, then Paul compares the i-th ball with the representatives of all the k' colors found so far. In this case, Paul finds either that the i-th ball has the same color as one of the first $i-1$ balls, or that its color is different from the colors of all of these balls.

If the first $i-1$ balls have k different colors, then Paul compares the i-th ball with the representatives of $k-1$ colors. If Paul does not find a ball with the same color, then the color of the i-th ball is the color with no representative.

In this way, Paul determines the colors of all the balls. We estimate the number of comparisons in the worst case. Since the first $i-1$ balls have at most $i-1$ different colors, the number of comparisons in the i-th step is at most $\min\{i-1, k-1\}$. Therefore, if $n < k$, the number of questions does not exceed:

$$\sum_{i=1}^{n}(i-1) = \frac{n(n-1)}{2} = \binom{n}{2} .$$

In the general case $n \geq k$, we have the following bound:

$$\sum_{i=1}^{k}(i-1) + \sum_{i=k+1}^{n}(k-1) = \frac{k(k-1)}{2} + (n-k)(k-1) = n(k-1) - \binom{k}{2} .$$

■

Next, we show that Carol can force Paul to ask $(k-1)n - \binom{k}{2}$ questions:

Theorem 9 *If Paul is allowed to use only a deterministic strategy to solve the Partition problem with n balls of k colors, then Carol can force him to ask at least $(k-1)n - \binom{k}{2}$ questions if $n \geq k$, and at least $\binom{n}{2}$ questions otherwise.*

Proof: We can assume that Paul never asks a question whose answer is uniquely determined. Therefore, Carol can answer each question that the colors of the pair of the balls are different. Let G be Paul's graph at the end of the game. Note that all the edges of G are blue because of Carol's strategy.

If $n < k$ and G is not a complete graph, then all the balls can have n distinct colors or the two balls which are not joined by an edge in G can have the same color and the remaining balls can have $n-2$ additional distinct colors. Therefore, Paul cannot determine the partition of the balls. We may conclude that if $n < k$, G must be a complete graph and Paul asked $\binom{n}{2}$ questions.

In the rest of the proof, we consider the general case $n \geq k$. Let $V_1, \ldots, V_k$ be the vertices of G corresponding to the sets of the balls of the same color. Each of the sets V_i, $1 \leq i \leq k$, is non-empty: otherwise, there exist an empty set V_i and a set $V_{i'}$ with at least two vertices (recall that $n \geq k$). Move a vertex from $V_{i'}$ to V_i. The new partition is also consistent with the graph G and therefore Paul is unable to uniquely determine the partition.

Assume now that the subgraph of G induced by $V_i \cup V_{i'}$ for some i and i', $1 \leq i < i' \leq k$, is disconnected. Let W be the vertices of one of the components of the subgraph. Move the vertices of $V_i \cap W$ from V_i to $V_{i'}$ and the vertices of $V_{i'} \cap W$ from $V_{i'}$ to V_i. Since the new partition is consistent with the graph G, Paul cannot uniquely determine the partition of the balls according to their colors. We may conclude that each set V_i is non-empty and the subgraph of G induced by any pair of V_i and $V_{i'}$ is connected.

Let n_i be the number of vertices of V_i. For every i and i', $1 \leq i < i' \leq k$, the subgraph of G induced by $V_i \cup V_{i'}$ contains at least $n_i + n_{i'} - 1$ edges because it is connected. Since the sets V_i are disjoint, the number of edges of G, which is the number of questions asked by Paul, is at least the following:

$$\sum_{1 \leq i < i' \leq k} (n_i + n_{i'} - 1) = \sum_{i=1}^{k} (k-1)n_i - \sum_{1 \leq i < i' \leq k} 1 = (k-1)n - \binom{k}{2} .$$

■

7 Probabilistic Strategy for the Partition Problem

We first state the following auxiliary lemma:

Lemma 10 *Consider a random ordering of n balls, out of which, ξn balls are white and $(1-\xi)n$ are black. The expected length of the initial segment comprised entirely of white balls in the random ordering is at least:*

$$\min\left\{9, \frac{\xi}{1-\xi} - O\left(\frac{\log n}{\sqrt{n}}\right)\right\} .$$

Proof: Because of the O-term, it is enough to prove the lemma for sufficiently large n, say $n \geq 1000$. We distinguish several cases according to whether ξ is close to one, to zero or to neither of them. Assume first that $\xi > \frac{99}{100}$. The probability that the initial segment of length 10 is comprised entirely by white balls is the following:

$$\frac{\xi n}{n} \cdot \frac{\xi n - 1}{n-1} \cdots \frac{\xi n - 9}{n-9} > \left(\frac{\xi n - 9}{n-9}\right)^{10} > \left(\frac{98}{99}\right)^{10} > \frac{9}{10}.$$

Hence, the expected length of the initial segment comprised entirely by white balls, is at least $\frac{9}{10} \cdot 10 = 9$.

In the rest, we assume that $\xi \leq \frac{99}{100}$. Let $m = \lfloor \sqrt{n} \log n \rfloor$. If $\xi < \frac{2m}{n}$, the expected length of the initial segment of white balls is at least ξ (the probability that the first ball is white). Since the difference between $\frac{\xi}{1-\xi}$ and the lower bound ξ is small:

$$\frac{\xi}{1-\xi} - \xi = \frac{\xi^2}{1-\xi} \leq O(\xi^2) = O\left(\frac{\log^2 n}{n}\right) \leq O\left(\frac{\log n}{\sqrt{n}}\right) ,$$

the expected length of the initial segment of white balls is at least $\frac{\xi}{1-\xi} - O\left(\frac{\log n}{\sqrt{n}}\right)$ as claimed in the statement of the lemma.

Therefore, we assume that $\frac{2m}{n} \leq \xi \leq \frac{99}{100}$ in what follows. The expected length of the initial segment comprised by white balls is at least:

$$\begin{aligned}
&\sum_{i=1}^{m} \mathbb{P}(\text{the first } i \text{ balls are white}) = \sum_{i=1}^{m} \prod_{j=0}^{i-1} \frac{\xi n - j}{n-j} \geq \sum_{i=1}^{m} \left(\frac{\xi n - m}{n-m}\right)^i \\
= \quad & \sum_{i=1}^{m} \left(\frac{\xi n - m}{n}\right)^i \geq \sum_{i=1}^{m} \left(\xi - \frac{m}{n}\right)^i = \frac{\left(\xi - \frac{m}{n}\right) - \left(\xi - \frac{m}{n}\right)^{m+1}}{1 - \left(\xi - \frac{m}{n}\right)} \\
\geq \quad & \frac{\xi}{1 - \left(\xi - \frac{m}{n}\right)} - O\left(\frac{m}{n}\right) - e^{-\Theta\left(\frac{m^2}{(n)^2}\right)} \geq \frac{\xi}{1-\xi} - O\left(\frac{\log n}{\sqrt{n}}\right) .
\end{aligned}$$

■

We can now describe Paul's strategy:

Theorem 11 *There is a probabilistic strategy for the Partition problem with* n *balls of three colors such that the expected number of Paul's questions does not exceed* $\frac{5}{3}n - \frac{8}{3} + O\left(\frac{\log n}{\sqrt{n}}\right)$ *for any coloring of the balls.*

Proof: Fix a coloring of the n balls. Let α, β and γ be fractions of the balls of each of the three colors in the coloring. Paul first chooses a random ordering of the balls. His strategy is divided into n steps, and in the i-th step Paul determines the color of the i-th ball (in the random ordering).

If the first $i-1$ balls have the same color, then Paul just compares the i-th ball with any of the first $i-1$ balls. If the first $i-1$ balls have two distinct colors, then Paul randomly chooses one of the two colors and compares the i-th ball with a representative of this color. If Carol answers that the balls have different colors, then he compares the i-th ball with a representative of the other color. Finally, if the first $i-1$ balls have three distinct colors, then he randomly chooses one of the colors and compares the i-th ball with a representative of this color. If Carol answers that the balls have different colors, Paul randomly chooses one of the two remaining colors and compares the i-th ball with its representative.

If the input does not contain balls of all three colors, then the expected number of Paul's questions in each step is at most $3/2$. Consequently, the expected number of all the questions asked by Paul does not exceed $\frac{3}{2}n$, and the statement of the lemma readily follows. Therefore, we assume in the rest that there is a ball of each of the three colors, i.e., $\alpha, \beta, \gamma > 0$.

Fix an ordering of the balls. Let j be the largest integer such that the first j balls have the same color, and j' the largest integer such that the first j' balls have at most two different colors. We compute the expected number of Paul's questions for the fixed ordering. In the first $j+1$ steps (except for the first step), Paul always asks a single question. At each of the next $j'-j-1$ steps, Paul asks $3/2$ questions on average. In the $(j'+1)$-th step, Paul asks two questions. At each of the $n-j'-1$ remaining steps, Paul asks $5/3$ questions on average. Hence, the expected number of Paul's questions for the fixed ordering is:

$$j + \frac{3}{2}(j'-j-1) + 2 + \frac{5}{3}(n-j'-1) = \frac{5}{3}n - \frac{1}{2}j - \frac{1}{6}j' - \frac{7}{6}\ .$$

The expected number of the questions (averaged through all the orderings) is:

$$\frac{5}{3}n - \frac{1}{2}\overline{j} - \frac{1}{6}\overline{j'} - \frac{7}{6} \tag{12}$$

where $\overline{j}$ and $\overline{j'}$ are the expected lengths of the initial segments in the random ordering comprised by balls of one color and two colors, respectively.

Let $\overline{j_A}$, $\overline{j_B}$ and $\overline{j_C}$ be the expected lengths of initial segments in the ordering formed entirely by balls of the color with the fractions α, β and γ, respectively. Similarly, $\overline{j_{AB}}$, $\overline{j_{AC}}$ and $\overline{j_{BC}}$ are the expected lengths of initial segments formed by balls of two indexed colors. Clearly, the following holds:

$$\begin{aligned}
\overline{j} &= \overline{j_A} + \overline{j_B} + \overline{j_C} \\
\overline{j'} &= \overline{j_{AB}} + \overline{j_{AC}} + \overline{j_{BC}} - \overline{j_A} - \overline{j_B} - \overline{j_C}
\end{aligned}$$

If one of the numbers $\overline{j_A}$, $\overline{j_B}$, $\overline{j_C}$, $\overline{j_{AB}}$, $\overline{j_{AC}}$ and $\overline{j_{BC}}$ is at least 9, then the sum $\overline{j} + \overline{j'}$ is at least 9, and the expected number of Paul's questions is at least $\frac{5}{3}n - \frac{9}{6} - \frac{7}{6} = \frac{5}{3}n - \frac{8}{3}$. If none of the numbers exceed 9, we have by Lemma 10:

$$\begin{aligned}
\overline{j} &= \tfrac{\alpha}{1-\alpha} + \tfrac{\beta}{1-\beta} + \tfrac{\gamma}{1-\gamma} + O\left(\tfrac{\log n}{\sqrt{n}}\right) \qquad \text{and} \\
\overline{j} + \overline{j'} &= \tfrac{\alpha+\beta}{1-\alpha-\beta} + \tfrac{\alpha+\gamma}{1-\alpha-\gamma} + \tfrac{\beta+\gamma}{1-\beta-\gamma} + O\left(\tfrac{\log n}{\sqrt{n}}\right) \quad .
\end{aligned}$$

Since the function $\frac{\xi}{1-\xi}$ is convex and $\alpha + \beta + \gamma = 1$, the minimum of both the expressions on the right hand side in the above equations is attained for $\alpha = \beta = \gamma = 1/3$. Therefore, we have the following:

$$\overline{j} \geq 3 \cdot \frac{1/3}{1-1/3} - O\left(\frac{\log n}{\sqrt{n}}\right) = \frac{3}{2} - O\left(\frac{\log n}{\sqrt{n}}\right) \tag{13}$$

$$\overline{j} + \overline{j'} \geq 3 \cdot \frac{2/3}{1-2/3} - O\left(\frac{\log n}{\sqrt{n}}\right) = 6 - O\left(\frac{\log n}{\sqrt{n}}\right) \tag{14}$$

Let us plug the estimates (13) and (14) to the expression (12) for the expected number of Paul's questions:

$$\frac{5}{3}n - \frac{1}{3}\overline{j} - \frac{1}{6}\left(\overline{j} + \overline{j'}\right) - \frac{7}{6} \leq \frac{5}{3}n - \frac{1}{2} - 1 - \frac{7}{6} + O\left(\frac{\log n}{\sqrt{n}}\right) = \frac{5}{3}n - \frac{8}{3} + O\left(\frac{\log n}{\sqrt{n}}\right) .$$

■

As the first step towards our lower bound, we state the following lemma on the average depths of leaves in binary trees with unbalanced inner vertices:

Lemma 12 *Let T be a rooted binary tree with N leaves. Assume that on each path from the root of T to a leaf, at least ℓ inner nodes have the following property ($\star$): the number of the leaves in the left subtree is precisely half of the number of the leaves in the right subtree. The average length of the path from the root to a leaf of T is at least the following:*

$$\frac{\log N}{\log 2} + \left(\frac{5}{3} - \frac{\log 3}{\log 2}\right)\ell$$

Proof: The proof proceeds by induction on N. If $N = 1$, the tree consists just of its root, ℓ must be equal to zero and the average length of the path from the root to a leaf is also zero. Assume that $N > 1$. In particular, the root has two children. Let N_1 and N_2 be the numbers of leaves in each of the two subtrees. Clearly, $N = N_1 + N_2$.

If the root does not have the property ($\star$), then there are at least ℓ inner nodes with the property ($\star$) on every path from the root to a leaf in each of the two subtrees. Using the induction, we infer that the average length of the path from the root to each of the leaves in the tree T is at least the following:

$$\begin{aligned}
&\frac{\left(1+\frac{\log N_1}{\log 2}+\left(\frac{5}{3}-\frac{\log 3}{\log 2}\right)\ell\right)N_1+\left(1+\frac{\log N_2}{\log 2}+\left(\frac{5}{3}-\frac{\log 3}{\log 2}\right)\ell\right)N_2}{N}\\
= \quad & 1+\frac{N_1\log N_1+N_2\log N_2}{N\log 2}+\left(\frac{5}{3}-\frac{\log 3}{\log 2}\right)\ell\\
\geq \quad & 1+\frac{\log\frac{N}{2}}{\log 2}+\left(\frac{5}{3}-\frac{\log 3}{\log 2}\right)\ell=\frac{\log N}{\log 2}+\left(\frac{5}{3}-\frac{\log 3}{\log 2}\right)\ell\,.
\end{aligned}$$

On the other hand, if the root has the property ($\star$), then $N_1 = N/3$, $N_2 = 2N/3$, and there are at least $\ell - 1$ inner nodes with the property ($\star$) on every path from the root to a leaf in each of the two subtrees. We again use the induction and derive the following lower bound on the average length of the paths:

$$\frac{\left(1+\frac{\log N_1}{\log 2}+\left(\frac{5}{3}-\frac{\log 3}{\log 2}\right)(\ell-1)\right)N_1+\left(1+\frac{\log N_2}{\log 2}+\left(\frac{5}{3}-\frac{\log 3}{\log 2}\right)(\ell-1)\right)N_2}{N}$$

$$
\begin{aligned}
&= 1 + \frac{\frac{N}{3}\log\frac{N}{3} + \frac{2N}{3}\log\frac{2N}{3}}{N\log 2} + \left(\frac{5}{3} - \frac{\log 3}{\log 2}\right)(\ell - 1) \\
&= 1 + \frac{\log N - \log 3 + \frac{2}{3}\log 2}{\log 2} + \left(\frac{5}{3} - \frac{\log 3}{\log 2}\right)(\ell - 1) \\
&= \frac{\log N}{\log 2} + \frac{5}{3} - \frac{\log 3}{\log 2} + \left(\frac{5}{3} - \frac{\log 3}{\log 2}\right)(\ell - 1) = \frac{\log N}{\log 2} + \left(\frac{5}{3} - \frac{\log 3}{\log 2}\right)\ell\,.
\end{aligned}
$$

■

We are now able to show that the number of Paul's questions in Theorem 11 is optimal:

Theorem 13 *For any probabilistic strategy for the Partition problem with n balls of three colors, there is a coloring of the balls which forces Paul to ask at least $\frac{5}{3}n - \frac{8}{3}$ questions on average.*

Proof: By Yao's principle (Proposition 1), it is enough to find a probability distribution on colorings such that if Paul uses any deterministic strategy, then he asks at least $\frac{5}{3}n - \frac{8}{3}$ questions on average. We claim that the uniform distribution on all 3^n possible colorings has this property.

Fix a deterministic strategy and let T be the corresponding binary tree. Since Paul is able to solve the Partition problem using this strategy, the computation paths may end in the same leaf for at most six different colorings (they can differ only by a permutation of the colors). Hence, T has at least $N = 3^n/6$ leaves.

Consider an inner node w of T at which Paul's question corresponds to an edge between two different components of Paul's graph. Let $\mathcal{I}$ be the set of colorings whose computation path reaches the node w. Since the edge corresponding to the question joins two different components, for exactly one third of the input colorings Carol answers that the balls have the same colors, and for exactly two thirds she answers that their colors are different: for each coloring $X \in \mathcal{I}$, the set $\mathcal{I}$ contains all the five other colorings obtained from X by permutation of the colors in one of the components. We conclude that if Paul's question at the node w corresponds to an edge between two different components of his graph, then the node w has the property $(\star)$ from the statement of Lemma 12.

Observe now that Paul's final graph is connected: otherwise, we could permute the colors of the balls in one of the components while keeping the colors of the remaining balls the same, which would yield a different partition consistent with Paul's final graph, and Paul would be unable to uniquely determine the partition of the balls. Since Paul's final graph is connected, on each path from the root to a leaf, there are at least $n-1$ nodes in which Paul asked a question which corresponds to an edge between two different components. Therefore, on each such path, at least $n-1$ nodes have the property $(\star)$ from Lemma 12.

By Lemma 12, the average length of the path from the root to a leaf of T, which is equal to the expected number of Paul's questions, is at least:

$$\frac{\log\frac{3^n}{6}}{\log 2} + \left(\frac{5}{3} - \frac{\log 3}{\log 2}\right)(n-1) = \frac{\log 3}{\log 2}n - \frac{\log 3}{\log 2} - 1 + \left(\frac{5}{3} - \frac{\log 3}{\log 2}\right)(n-1)$$
$$= \frac{5}{3}n - 1 - \frac{5}{3} = \frac{5}{3}n - \frac{8}{3}\,.$$

■

The arguments of this section can be generalized to inputs with balls of more colors. However, the obtained lower and upper bounds do not match: if Paul uses a probabilistic strategy similar to that in Proposition 11, he asks $\frac{k^2+k-2}{2k}n + O(1)$ questions on average for the Partition problem with n balls of k colors. On the other hand, the tree of any deterministic strategy must contain at least $n-1$ nodes w on any path from the root to a leaf such that the subtree of the left child of w contains exactly the fraction of $1/k$ of the leaves of the whole subtree of w. Based on this, one can establish the lower bound $\left(\frac{k-1}{k}\frac{\log(k-1)}{\log 2} + 1\right)n - \Theta(k\log k)$ on the expected number of questions asked by Paul (in the worst case).

Acknowledgements

The paper is based on the results obtained during the DIMACS/DIMATIA Research Experience for Undergraduates program in 2004. The REU program is a joint project of DIMACS at Rutgers University and DIMATIA at Charles University. The authors are also grateful to János Komlós, the director of the REU program at Rutgers, for his helpful comments on the considered problems. The authors would also like to thank Jiří Sgall for pointing out an easy argument how to extend the proof of Theorem 7 to the case when the number n of the balls is odd.

References

[1] Aigner M.: Variants of the majority problem, Discrete Applied Mathematics **137** (2004) 3–25.

[2] Aigner M., De Marco G., Montangero M.: The plurality problem with three colors, in: Diekert V., Habib M. (eds.): STACS 2004, LNCS 2296, Springer-Verlag, Berlin, Heidelberg, 2004, 513–521.

[3] Aigner M., De Marco G., Montangero M.: The plurality problem with three colors and more, submitted.

[4] Alonso L., Reingold E. M., Schott R.: Determining the majority, Information Processing Letters **47** (1993) 253–255.

[5] Cover T. M., Thomas J. A.: Elements of information theory, John Wiley & Sons, 1991.

[6] Fisher M., Salzberg S.: Finding a majority among n votes, Journal of Algorithms **3** (1982) 375–379.

[7] De Marco G., Pelc A.: Randomized algorithms for determining the majority on graphs, in: Rovan B., Vojtás P. (eds.): MFCS 2003, LNCS 2747, Springer-Verlag, Berlin, Heidelberg, 2003, 368–377.

[8] Matoušek J., Nešetřil J.: Invitation to discrete mathematics, Oxford University Press, 1998.

[9] Motwani R., Raghavan P.: Randomized Algorithms, Cambridge University Press, 1995.

[10] Saks M., Werman M.: On computing majority by comparisons, Combinatorica **11** (1991) 383–387.

[11] Srivastava N.: Tight bounds on plurality using equality tests, unpublished manuscript, 2003.

[12] Yao A. C.-C.: Probabilistic computations: Towards a unified measure of complexity, in: Proc. of the 17th Annual Symposium on Foundations of Computer Science (FOCS), 1977, 222–227.