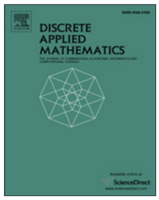
Discrete Applied Mathematics 156 (2008) 3305–3311

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam



Randomized strategies for the plurality problem

Daniel Král' [a], Jiří Sgall [b,*], Tomáš Tichý [b]

[a] Department of Applied Mathematics and Institute for Theoretical Computer Science,[1] Faculty of Mathematics and Physics, Charles University, Malostranské náměstí 25, 118 00 Prague, Czech Republic

[b] Mathematical Institute of the Academy of Sciences of the Czech Republic, Žitná 25, CZ-11567 Prague, Czech Republic

ARTICLE INFO

Article history:
Received 7 October 2005
Received in revised form 17 October 2006
Accepted 21 May 2008
Available online 1 July 2008

Keywords:
Concrete complexity
Randomized algorithms
Plurality game
Majority game

ABSTRACT

We consider a game played by two players, Paul and Carol. At the beginning of the game, Carol fixes a coloring of n balls. At each turn, Paul chooses a pair of the balls and asks Carol whether the balls have the same color. Carol truthfully answers his question. Paul's goal is to determine the most frequent (plurality) color in the coloring by asking as few questions as possible. The game is studied in the probabilistic setting when Paul is allowed to choose his next question randomly.

We give asymptotically tight bounds both for the case of two colors and many colors. For the balls colored by k colors, we prove a lower bound $\Omega(kn)$ on the expected number of questions; this is asymptotically optimal. For the balls colored by two colors, we provide a strategy for Paul to determine the plurality color with the expected number of $2n/3 + O(\sqrt{n \log n})$ questions; this almost matches the lower bound $2n/3 - O(\sqrt{n})$.

1. Introduction

We study a two-player game played by Paul and Carol. Paul wants to determine a property of the input based on Carol's answers. At the beginning of the game, Carol fixes a coloring of n balls by k colors. Paul does not know the coloring. At each step, he chooses two balls and asks Carol whether the balls have the same color. Carol truthfully answers. Paul wants to ask the least number of questions (in the worst case) to determine the desired property of the coloring.

In the probabilistic setting, Paul can choose his next question randomly, depending on the previous course of the game. As in the previous literature, we consider Las Vegas strategies, which means that at the end of the game Paul is not allowed to make any error. His goal is to minimize the expected number of questions asked for the worst case (most difficult) coloring.

1.1. Previous work

Several problems depending on the property that Paul is interested in have been studied. The goal of the *Majority problem* is to find a ball b such that the number of balls with the same color is greater than $n/2$, or to declare that there is no such ball. If the balls are colored with only two colors and no randomness is allowed, Paul always has to ask at least $n - \nu(n)$ questions for some coloring, and Paul has a strategy such that he asks at most $n - \nu(n)$ questions [14,4] for any coloring, where $\nu(n)$ is the number of 1's in the binary representation of n. If the number of colors is unrestricted, then the necessary and sufficient number of questions is $\lceil 3n/2 \rceil - 2$, see [9]. In the average case setting (the average is taken over all possible input colorings for a fixed deterministic strategy), the upper bound of $2n/3 - \sqrt{8n/(9\pi)} + \Theta(1)$ and the lower bound of $2n/3 - \sqrt{8n/(9\pi)} - O(\log n)$ were established [5].

* Corresponding author.
E-mail addresses: kral@kam.mff.cuni.cz (D. Král'), sgall@math.cas.cz (J. Sgall), tichy@math.cas.cz (T. Tichý).

[1] Institute for Theoretical Computer Science is supported by the Ministry of Education of the Czech Republic as project 1M0545.

0166-218X/$ – see front matter
doi:10.1016/j.dam.2008.05.014

Table 1
The summary of the previous and new results

Problem	Deterministic strategies		Probabilistic strategies	
	Lower bound	Upper bound	Lower bound	Upper bound
Plurality				
2 colors	$n - \nu(n)$	$n - \nu(n)$	$2n/3 - o(n)$	$2n/3 + o(n)$[∗]
3 colors	$3n/2 - O(1)$	$5n/3 + O(1)$	$3n/2 - o(n)$	$3n/2 + O(1)$
k colors	$kn/40$	$O(kn)$	$\Theta(kn)$[∗]	$kn/2 + o(n)$
Partition				
2 colors	$n - 1$	$n - 1$	$n - 1$	$n - 1$
3 colors	$2n - 3$	$2n - 3$	$5n/3 - 8/3$	$5n/3 - 8/3 + o(1)$
k colors	$(k-1)n - \binom{k}{2}$	$(k-1)n - \binom{k}{2}$	$\frac{(k-1)(n-k)}{4}$[∗]	$kn/2 + o(n)$

The results obtained in this paper are marked [∗].

The *Plurality problem* was introduced by Aigner et al. [2]. Paul seeks a ball with the *plurality color*, i.e., the color such that the number of balls with this color exceeds the number of balls of any other color; if no such ball exists, Paul declares that there is a tie. Paul's goal is only to point to a single ball with this color, he does not want to find all such balls. Note that if the balls are colored just with two colors, the Majority and the Plurality problems are the same. Aigner et al. [2,3] give a deterministic strategy to solve the Plurality problem with n balls of three colors asking at most $\lfloor 5n/3 \rfloor - 2$ questions and showed that Carol can force Paul to ask at least $3\lfloor n/2 \rfloor - 2$ questions. In addition, if the number k of colors of the balls is not fixed to be three, they give lower and upper bounds of order $\Theta(kn)$ (the lower bound is $kn/40$ but the authors admit not trying to tune the constant). In the probabilistic setting, the necessary and sufficient expected number of questions to resolve the problem with n balls of three colors is $3n/2 + o(n)$, see [8].

Yet another problem is the *Partition problem* in which Paul wants to partition the balls according to their colors. In the deterministic case, the necessary and sufficient number of questions to resolve the Partition problem with n balls of k colors is $(k-1)n - \binom{k}{2}$, see [8]. In the probabilistic setting, the necessary and sufficient expected number of questions to resolve the problem for n balls of three colors is $5n/3 - 8/3 + o(1)$, see [8].

Other related problems were also addressed, e.g., determining a set of at least k balls with same color in [7,10], and many other variants in [1,6,12].

1.2. Our results

If balls are colored with k colors for a constant k, we show a lower bound $\Omega(kn)$ on the expected number of questions to resolve the Plurality problem with n balls; the constant at the linear term in Ω is approximately $2/27$. This extends the lower bound $\Omega(kn)$ of Aigner et al. [3] from deterministic to probabilistic strategies. Since the number of questions necessary to resolve the Partition problem is at least the number of questions for the Plurality problem, this yields a lower bound $\Omega(kn)$ on the expected number of questions for the Partition problem, too. However, for this problem, we provide a simpler and better lower bound of $(k-1)(n-k)/4$ on the number of questions.

We establish a new upper bound for the Plurality problem in the probabilistic setting (see Table 1). For n balls of two colors, we show that the sufficient expected number of questions to resolve the Plurality problem is $2n/3 + o(n)$. Note that for two colors, the Majority and Plurality problems coincide. A matching lower bound for this problem follows from the result of [5] on the average case complexity using Yao's principle: Yao's principle [13,15] asserts that the expected number of questions in the best probabilistic strategy (for the worst possible input, averaging over the random choices of the algorithm) is equal to the average number of questions needed by any deterministic strategy for the "most difficult" probabilistic distribution on the inputs (averaging over the choice of input instance, according to the distribution). Alonso et al. [5] showed that for a coloring chosen uniformly at random, any deterministic strategy requires at least $2n/3 - \sqrt{8n/(9\pi)} - O(\log n)$ questions on average; thus the same lower bound for probabilistic strategies follows.

2. Preliminaries

All the logarithms have base two. We use a way of representing Paul's information about the colors of the balls from [8]: the game is viewed as a game on a graph whose vertices represent the balls. At the beginning, the graph has no edges. At each turn, Paul chooses a pair of nonadjacent vertices and joins them by an edge. Carol colors the edge red if the balls corresponding to the end-vertices have the same color, or blue if their colors are different. The graph represents the state of Paul's knowledge and is referred to as *Paul's graph*. A coloring of the balls is *consistent* with Paul's graph if the colors of every pair of balls corresponding to the vertices joined by a red (blue) edge are the same (different). In the Partition problem with k colors, the game ends when there is a unique coloring (up to a permutation of the colors) consistent with the graph. In the Plurality problem with k colors, the game ends when there is a vertex v such that in any consistent coloring the ball corresponding to v has the plurality color, or when in any consistent coloring there is a tie (no plurality color).

3. Lower bounds for many colors

In this section, we prove the lower bound $\Omega(kn)$ for the Partition and the Plurality problems with n balls of k colors in the probabilistic setting. Since any strategy for the Partition problem also yields a solution of the Plurality Problem, it would be sufficient to consider only the Plurality problem. However, the proof of the lower bound for the Partition problem is simpler, we achieve a better constant, and the bound holds for all values of n and k. In the case of the Plurality problem, the bound holds only if k is a constant and n is sufficiently large.

3.1. Partition problem

Let G be Paul's graph. We define a potential of Paul's graph that roughly measures the progress achieved by Paul during the game. The *partition potential* of a vertex v of G is equal to 0, if v is adjacent to a red edge (recall that red edges correspond to equal comparisons), and to $\max\{0, k-1-\deg_G(v)\}$, otherwise. The partition potential of Paul's graph is the sum of the partition potentials of all the vertices of G.

Lemma 1. *Suppose that G is Paul's final graph for the Partition problem such that it is consistent with some coloring that uses all k colors. Then the partition potential of every vertex of G is zero.*

Proof. Suppose that the potential of a vertex v is non-zero, i.e., there is no red edge incident with v in G and $\deg_G(v) < k-1$. Consider a coloring c of the vertices of G that uses all the k colors (such a coloring exists by our assumption) and uncolor the vertex v. The coloring still uses at least $k-1$ colors. There are now at least two different colors that can be assigned to the vertex v, as at most $\deg_G(v) < k-1$ colors for v are forbidden. Since at least one of these colors is used by another vertex, the partitions of the vertices corresponding to the possible extensions are different. □

Theorem 2. *Any probabilistic strategy for the Partition problem with n balls of k colors requires asking at least $(k-1)(n-k)/4$ questions on average for some coloring of the balls.*

Proof. By Yao's principle, it is enough to show that every deterministic strategy for the Partition problem with n balls of k colors requires asking at least $(k-1)(n-k)/4$ questions on average for some fixed probability distribution on the colorings. The distribution that we consider is the following: color the first k balls by mutually different colors and use the uniform distribution on all the k^{n-k} colorings of the remaining $n-k$ balls. We reveal the information about the coloring of the first k balls at the beginning, thus Paul's graph initially contains blue edges among all the first k vertices and no other edges. The initial partition potential of Paul's graph is equal to $(k-1)(n-k)$. By Lemma 1 and the fact that any used coloring uses all k colors, the potential of Paul's final graph is zero. We show that after each question, on average, the partition potential of G decreases by at most four. The theorem then follows.

Assume that Paul asks for the comparison of the balls corresponding to the vertices v and w. Let p be the partition potential of v. If $p = 0$ (this includes the case that v corresponds to one of the first k balls), the potential of v cannot further decrease. If $p > 0$, the vertex v is not incident with a red edge. Given any precoloring of all the vertices of G except for v, which is consistent with G, there are at least $k - \deg_G(v)$ different colors that may be assigned to v. Therefore, by averaging over all possible precolorings, the probability that the edge vw will be red is at most $1/(k-\deg_G(v)) = 1/(p+1) \leq 1/p$. In such a case, the partition potential of v drops down from p to zero. Otherwise, the potential of v is decreased by one. Therefore, the potential of v decreases by at most 2 on average. Similarly, the potential of w decreases by at most 2 on average. Since the potentials of the remaining vertices do not change, the average decrease (after each question) of the partition potential of G is at most four. □

3.2. Plurality problem

We introduce notation used in this subsection: a coloring of n balls with k colors is *nice* if there exists the plurality color and the number of balls of any of the colors is at most $n/k + n^{2/3}$. If k is fixed, then Chernoff's bound implies that a random coloring of n balls with k colors is nice with probability $1 - o(1)$ as n tends to infinity. (This is true even if $n^{2/3}$ is replaced by any function in $\omega(n^{1/2})$; we use $n^{2/3}$ for simplicity.)

A vertex of Paul's graph is *free* if it is incident with no red edge and with less than $\lfloor k/3 \rfloor$ blue edges. The other vertices are said to be *non-free*. The *plurality potential* of a vertex v of Paul's graph is $\lfloor k/3 \rfloor - \deg_G(v)$ if v is free, and 0 otherwise. The plurality potential of Paul's graph is the sum of the plurality potentials of all the vertices of G.

Lemma 3. *Suppose that G is Paul's final graph for the Plurality problem of n balls with $k \geq 3$ colors such that G is consistent with some nice coloring. Then G contains at most $n/3 + 2kn^{2/3}$ free vertices. In particular, the plurality potential of such Paul's final graph is at most $\lfloor k/3 \rfloor \cdot n/3 + o(n)$ for a fixed k.*

Proof. Assume for a contradiction that G has more than $n/3 + 2kn^{2/3}$ free vertices. Depending on Paul's answer, we construct a coloring that is consistent with G but for which Paul's answer is incorrect. We distinguish two cases.

In the first case, Paul claims that a ball b has the plurality color. Color the ball b and all the balls corresponding to non-free vertices in G as in the nice coloring that exists by our assumption. Since the coloring is nice, at most $n/k + n^{2/3}$ balls have the same color as b. Fix any $\lfloor k/3 \rfloor$ colors distinct from the color of b. Recolor all the balls corresponding to the free vertices greedily by the fixed $\lfloor k/3 \rfloor$ colors in such a way that the resulting coloring is consistent with G. In case that the vertex corresponding to b is free, the ball b keeps its original color. Note that such a recoloring is always possible because each free vertex is incident with at most $\lfloor k/3 \rfloor - 1$ blue edges (that forbid at most $\lfloor k/3 \rfloor - 1$ colors at each such vertex). Since there are at least $n/3 + 2kn^{2/3}$ balls corresponding to free vertices and they are colored with $\lfloor k/3 \rfloor$ colors, there exists a color used for more than $n/k + 2n^{2/3}$ such balls. Then, the color of b is not the plurality color in the constructed coloring.

In the second case, Paul claims that there is a tie. Color all the balls corresponding to non-free vertices in G as in the nice coloring. Fix any $\lfloor k/3 \rfloor$ colors and use them to color all the remaining balls greedily and consistently with G, as in the previous case. If there is no tie, we have a coloring that contradicts Paul's answer. Otherwise, similarly as in the previous case, there exists a color used for at least $n/k + 2n^{2/3}$ such balls and thus any color class that has the maximum number of balls contains a ball corresponding to a free vertex. Choose from all but one of the maximal color classes such a vertex and recolor them one by one by a color not used for any ball corresponding to a free vertex: there is always a choice of at least $\lfloor k/3 \rfloor$ colors, as at most $\lfloor k/3 \rfloor$ colors are used for the original coloring of the free vertices and at most $\lfloor k/3 \rfloor$ free vertices are recolored. This coloring has no tie, as no color used for recoloring can have more than $n/k + n^{2/3} + 1$ balls. □

Theorem 4. *Let $k \geq 3$ be a fixed integer. For any probabilistic strategy for the Plurality problem with n balls of k colors, there exists a coloring such that Paul asks at least $\lfloor k/3 \rfloor \cdot 2n/9 - o(n)$ questions on average.*

Proof. We again apply Yao's principle and show that every deterministic strategy for the Plurality problem requires asking at least $\lfloor k/3 \rfloor \cdot 2n/9 - o(n)$ questions on average for some distribution on colorings. We use the uniform distribution on all k^n colorings of n balls with k colors. Fix a deterministic strategy for the Plurality problem. Since a random coloring is nice with probability $1 - o(1)$, we may analyze the strategy only for nice colorings. The initial potential of Paul's graph is $\lfloor k/3 \rfloor n$. By Lemma 3, the plurality of Paul's final graph is at least $\lfloor k/3 \rfloor \cdot n/3 + o(n)$. We show that the expected decrease of this sum is at most 3 after every single question of Paul.

Consider a moment when Paul asked for the comparison of the balls corresponding to vertices v and w. If v is non-free, then its potential does not change after Carol's answer. Assume that v is free. Given any precoloring c of all the vertices of G except for v, there are at least $k - \lfloor k/3 \rfloor \geq 2k/3$ colors that may be assigned to v. Therefore, the probability that Carol answers that the balls have the same color (with respect to the uniform distribution on all the colorings) is at most $3/(2k)$. In such case, the plurality potential of v decreases by at most $k/3$. Otherwise, it decreases by one. In particular, the plurality potential of v decreases by at most $1 + 3/(2k) \cdot k/3 = 1.5$ on average. Similarly, the plurality potential of w decreases by at most 1.5 on average. Since the plurality potentials of the remaining vertices of G do not change, the average decrease of the plurality potential of G is at most three. □

4. Plurality problem with two colors

A natural idea for an upper bound is to compare random pairs of balls. If the colors are different, the pair can be discarded, as it does not influence the result. We then recurse on the pairs of balls of the same color. This strategy performs well if the numbers of balls of both the colors are about the same. However, on other instances the expected number of comparisons may be too large. To circumvent this problem, Paul applies the strategy only for a half of the balls, obtaining not only the plurality color, but also an estimate of the number of balls of each color. He then applies the same strategy to classify a fraction of the remaining balls so that only a small number of additional comparisons is needed (with high probability).

Algorithm Count determines the number of balls of each of the two colors and finds a representative of the larger color class (in case that there is no tie).

Algorithm Count.

(1) if there is no ball, return 0 and no representative;
(2) if there is a single ball, return 1 and set the representative r to this ball;
(3) randomly permute the n balls;
(4) for each $i \leq n/2$, compare the $(2i - 1)$th and $2i$th balls (in the order determined by the random permutation);
(5) let R be the set of balls containing one ball from each compared pair for which Carol answered that the balls have the same color;
(6) apply Algorithm Count recursively to the set R;
let r be the representative and m' be the number of balls of the plurality color (in case of a tie, no r is found);
(7) let $m = \lfloor n/2 \rfloor - |R| + 2m'$;
(8) if n is odd then
  if there is no representative r then
    set r to be the nth ball and let $m := m + 1$;
  else
    compare the nth ball and r,
    if they have the same color, then let $m := m + 1$;
(9) output m and r (if r was defined);

We summarize the properties of Algorithm COUNT in the next lemma:

Lemma 5. *Let n balls be colored so that there are a and b balls of each color and $a \geq b$. Then Algorithm* COUNT *correctly determines the number of balls with the plurality color and provides its representative (if there is no tie). The expected number of Paul's questions is at most $a + b/3 + \log n = n - 2b/3 + \log n$.*

Proof. The pairs where Carol answered that the balls have different colors contain equal number of balls of each color. Therefore, the plurality color for the original coloring is the plurality color among the balls of R, if n is even. If n is odd, additional adjustment performed in Step (8) is needed. Overall, Algorithm COUNT correctly determines the number of balls with the plurality color and finds its representative.

Let $x_{a,b}$ be the expected number of comparisons on an instance with a and b balls of each color. We show by induction on n that

$$x_{a,b} \leq a + \frac{b}{3} + \log(a+b).$$

The bound trivially holds if n is one or two. Let $n = a + b \geq 3$. The probability that a given ball of the second color is paired with a ball of the first color is at least 1/2: it is $a/(a+b-1) \geq 1/2$ for n even, and $a/(a+b) \geq 1/2$ for n odd (counting the probability that the ball is not matched). Thus the expected number of pairs of balls with different colors is at least $b/2$. Let a' and b' be the numbers of balls of each of the two colors in R. Since the expected number of pairs of balls with different colors is at least $b/2$, we have $\mathbf{Exp}[a'] \leq (a - b/2)/2 = a/2 - b/4$ and $\mathbf{Exp}[b'] \leq (b - b/2)/2 = b/4$. In particular, the expected number of balls with the plurality color in R is at most $a/2 - b/4$. Since the plurality color in R is the same as in the original problem, we conclude by induction:

$$\begin{aligned} x_{a,b} &\leq \frac{a+b}{2} + 1 + \mathbf{Exp}[x_{a',b'}] \\ &\leq \frac{a+b}{2} + 1 + \mathbf{Exp}[a'] + \frac{\mathbf{Exp}[b']}{3} + \log\frac{a+b}{2} \\ &\leq \frac{a+b}{2} + 1 + \frac{a}{2} - \frac{b}{4} + \frac{b}{12} + \log(a+b) - 1 \\ &= a + \frac{b}{3} + \log(a+b). \end{aligned}$$

The bound from the statement of the lemma now readily follows. □

We now analyze Algorithm PLURALITY for the Plurality problem:

Algorithm PLURALITY.

(1) randomly permute the n balls;
(2) apply Algorithm COUNT to the first $\lfloor n/2 \rfloor + 1$ balls;
let a' and b' be the numbers of balls of each color, $a' \geq b'$, and
let r' be the representative of the plurality color (if it exists);
(3) apply Algorithm COUNT to the next $2b'$ balls (or all the remaining balls if $2b' + \lfloor n/2 \rfloor + 1 > n$); let a'' and b'' be the numbers of balls of each color, $a'' \geq b''$, and let r'' be the representative of the plurality color (if it exists);
(4) if no representative (r' or r'') exists, output "TIE" and halt;
(5) if exactly one of the representatives r' and r'' exists, output this representative as the ball with the plurality color and halt;
(6) compare r' and r'' and if they have the same color,
output r' as the ball with the plurality color and halt;
(7) otherwise continue by comparing all the remaining balls to r';
this determines the color of all the balls relatively to r' and thus eventually determines the correct answer; stop and output the answer as soon as it is known.

Theorem 6. *Algorithm* PLURALITY *correctly determines the plurality color. For every coloring of the balls, the expected number of Paul's questions does not exceed $2n/3 + O(\sqrt{n \log n})$.*

Proof. We first show that Algorithm PLURALITY correctly determines the plurality color. If there is no representative r', then all the balls are examined. If the representative r' exists and either there is no r'' or the color of r'' is the same as the color of r', the number of balls with the same color as r' is at least $a' + a'' \geq a' + b' \geq \lfloor n/2 \rfloor + 1$ and thus the color of r' is the plurality color. Otherwise, all the balls are examined in the last step and the plurality color is correctly determined.

Next, we analyze the expected number of questions. Let a be the number of balls with the plurality color and b the number of balls with the other color. If $b \leq n/12$, then $b' \leq b$ and the entire number of examined balls is at most $n/2 + 2b' + 1 \leq 2n/3 + 1$. In the rest of the proof, we assume that $b \geq n/12$.

Since the number of Paul's questions never exceeds n, the contribution of the cases which happen with probability $O(1/n)$ is only a constant. Therefore, we can assume that such events do not happen. This excludes the cases when the sample of the first $n/2$ or the next $2b'$ balls is bad. More precisely, let B' be the number of balls of the smaller color among $\lfloor n/2 \rfloor + 1$ balls examined in Step (2). **Exp**$[B']$ is $b/2 + O(1)$ and by Chernoff's bound, the probability that $|B' - b/2| \geq \sqrt{n \log n}$ is at most $O(1/n)$ (see [11,13], for example). Thus, we assume that $|B' - b/2| < \sqrt{n \log n}$.

We now distinguish two cases. The first case is $b \geq n/2 - 25\sqrt{n \log n}$. Then $|B' - n/4| \leq 14\sqrt{n \log n}$ and, no matter if r' has the plurality color, we have $b' \geq n/4 - 14\sqrt{n \log n}$. Therefore, there are at most $O(\sqrt{n \log n})$ balls not examined in Steps (2) and (3), and thus $b'' \geq n/4 - O(\sqrt{n \log n})$. By Lemma 5, the expected number of questions in the two runs of Algorithm Count is at most:

$$\begin{aligned} \frac{n}{2} - \frac{2b'}{3} + 2b' - \frac{2b''}{3} + O(\log n) &< \frac{n}{2} - \frac{n}{6} + \frac{n}{2} - \frac{n}{6} + O(\sqrt{n \log n}) \\ &= \frac{2n}{3} + O\left(\sqrt{n \log n}\right). \end{aligned}$$

Since there are at most $O(\sqrt{n \log n})$ additional questions, the theorem follows.

The second case is $b < n/2 - 25\sqrt{n \log n}$. Then $B' \leq b/2 + \sqrt{n \log n} < n/4$ (using our previous assumption about B') and thus r' has the plurality color, $b' = B'$, and there are at most $2b' < n/2$ balls examined in Step (3). In addition, $b' = B' > b/2 - \sqrt{n \log n} > n/25$ for large n, using the fact that $b \geq n/12$.

Let B'' be the number of balls of the smaller color among the $2b'$ balls in Step (3). After removing the first $\lfloor n/2 \rfloor + 1$ balls, there are $b - B' = b/2 \pm \sqrt{n \log n}$ balls of the second color. Hence, **Exp**$[B'']$ is $2b'b/n \pm \sqrt{n \log n}$ and the probability that $|B'' - 2b'b/n| \geq 2\sqrt{n \log n}$ is at most the probability that $|B'' - \mathbf{Exp}[B'']| \geq \sqrt{n \log n}$ which is at most $O(1/n)$, by Chernoff's bound (using also $b' > n/25$). Similarly as before, the contribution of this case to the expected number of Paul's questions is a constant. Thus we assume for the rest of the proof that $|B'' - 2b'b/n| < 2\sqrt{n \log n}$.

Using the case condition and $b' > n/25$ for large n, we have:

$$\begin{aligned} B'' &< 2b'\frac{b}{n} + 2\sqrt{n \log n} \\ &< b' \cdot \frac{n - 50\sqrt{n \log n}}{n} + 2\sqrt{n \log n} \\ &< b' - \frac{b' \cdot 50\sqrt{n \log n}}{n} + 2\sqrt{n \log n} \\ &< b' - \frac{n \cdot 50\sqrt{n \log n}}{25 \cdot n} + 2\sqrt{n \log n} = b'. \end{aligned}$$

Thus r'' is a ball of the plurality color, $b'' = B''$, and the algorithm terminates in Step (6). By Lemma 5, the expected number of questions is at most:

$$\begin{aligned} \frac{n}{2} - \frac{2b'}{3} + 2b' - \frac{2b''}{3} + O(\log n) &\leq \frac{n}{2} + b'\left(\frac{4}{3} - \frac{4b}{3n}\right) + O\left(\sqrt{n \log n}\right) \\ &\leq \frac{n}{2} + \frac{b}{2}\left(\frac{4}{3} - \frac{4b}{3n}\right) + O\left(\sqrt{n \log n}\right) \\ &\leq \frac{2n}{3} + O\left(\sqrt{n \log n}\right). \end{aligned}$$

The last inequality holds since the quadratic function of b is maximal for $b = n/2$. □

Acknowledgements

The first author was a postdoctoral fellow at Technical University Berlin within the framework of the European training network COMBSTRU from October 2004 to July 2005. The second and third authors were partially supported by Institutional Research Plan No. AV0Z10190503, Institute for Theoretical Computer Science, and grant 201/05/0124 of GA ČR.

References

[1] M. Aigner, Variants of the majority problem, Discrete Appl. Math. 137 (2004) 3–25.
[2] M. Aigner, G. De Marco, M. Montangero, The plurality problem with three colors, in: V. Diekert, M. Habib (Eds.), Proceedings of STACS 2004, in: LNCS, vol. 2996, Springer-Verlag, Berlin, 2004, pp. 513–521.
[3] M. Aigner, G. De Marco, M. Montangero, The plurality problem with three colors and more, Theoret. Comput. Sci. 337 (1–3) (2005) 319–330.
[4] L. Alonso, E.M. Reingold, R. Schott, Determining the majority, Inform. Process. Lett. 47 (1993) 253–255.
[5] L. Alonso, E.M. Reingold, R. Schott, The average-case complexity of determining the majority, SIAM J. Comput. 26 (1) (1997) 1–14.
[6] L. Alonso, P. Chassaing, E.M. Reingold, R. Schott, The worst-case chip problem, Inform. Process. Lett. 89 (2004) 303–308.

[7] P.G. Binderup, G.S. Frandsen, P.B. Miltersen, S. Skyum, The complexity of identifying large equivalence classes, Fund. Inform. 38 (1999) 25–37.
[8] Z. Dvořák, V. Jelínek, D. Král', J. Kynčl, M. Saks, Three optimal algorithms for balls of three colors, Random Struct. Algorithms 30 (2007) 63–77.
[9] M. Fisher, S. Salzberg, Finding a majority among *n* votes, J. Algorithms 3 (1982) 375–379.
[10] G.S. Frandsen, P.B. Miltersen, S. Skyum, The complexity of finding replicas using equality tests, in: A.M. Borzyszkowski, S. Sokolowski (Eds.), Proceedings of MFCS 1993, in: LNCS, vol. 711, Springer-Verlag, Berlin, 1993, pp. 463–472.
[11] T. Hagerup, C. Rüb, A guided tour Chernoff bounds, Inform. Process. Lett. 33 (1989) 305–308.
[12] G. de Marco G, A. Pelc, Randomized algorithms for determining the majority on graphs, Combin. Probab. Comput. 15 (2006) 823–834.
[13] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, Cambridge, 1995.
[14] M. Saks, M. Werman, On computing majority by comparisons, Combinatorica 11 (1991) 383–387.
[15] A.C.-C. Yao, Probabilistic computations: Towards a unified measure of complexity, in: Proceedings of FOCS, 1977, pp. 222–227.