Discrete Applied Mathematics 4 (1982) 247–267
North-Holland Publishing Company

SELECTING THE TOP THREE ELEMENTS

Martin AIGNER

II. Mathematisches Institut, Freie Universität Berlin, Königin-Luise Str. 24-26, D-1000 Berlin 33

Received 20 June 1980
Revised 2 October 1981

The paper contains the complete results for the complexity of the following three selection problems: (A) Find the 3 top elements (ordered) out of a linear order on n elements. (B) Find the 3rd element out of a linear order on n elements. (C) Find the 3 top elements (unordered) out of a linear order on n elements.

1. Introduction

Over the past years a considerable amount of research has been devoted to selection algorithms using binary comparisons and their complexity (see Knuth [5, Ch. 5.3]. The following two problems received particular attention:

(A) Find the t largest elements out of a linear order on n elements.

(B) find the t-th largest element out of a linear order on n elements.

A variant of problem A is:

(C) Find the t largest elements without regard to order out of a linear order on n elements.

Let us denote by $W_t(n)$, $V_t(n)$ and $U_t(n)$ the minimum number of comparisons needed for (A), (B) and (C), respectively. Then we have

$$U_t(n) \le V_t(n) \le W_t(n) \quad \text{for every } n \text{ and } t. \tag{1.1}$$

For a given t, the only complete results are

$$U_1(n) = V_1(n) = W_1(n) = n-1, \tag{1.2}$$

$$W_2(n) = V_2(n) = (n-2) + \lceil \log n \rceil \qquad \text{(Kislitsyn [4])}, \tag{1.3}$$

$$U_2(n) = (n-2) + \lceil \log (n-1) \rceil \qquad \text{(Sobel [6])}.^1 \tag{1.4}$$

Upper bounds were established by Kislitsyn [4], Hadian–Sobel [2], lower bounds by Hyafil [3] and others (see 2.7 to 2.9 below). For related results see Aigner [1].

In this paper the complete results are given for $t=3$.

[1] log always means $\log_2$.

Theorem. *Let $n \geq 6$, $n = 2^k + r$ where $0 \leq r < 2^k$. Then*

$$W_3(n) = \begin{cases} (n-3)+2k & r=0, \\ (n-3)+2k+1 & \text{for } 1 \leq r \leq 2^{k-2}, \\ (n-3)+2k+2 & \text{otherwise.} \end{cases} \tag{1.5}$$

$$V_3(n) = \begin{cases} (n-3)+2k & r=0,1, \\ (n-3)+2k+1 & \text{for } 2 \leq r \leq 2^{k-2}+1, \\ (n-3)+2k+2 & \text{otherwise.} \end{cases} \tag{1.6}$$

$$U_3(n) = \begin{cases} (n-3)+2k & r=0,1,2, \\ (n-3)+2k+1 & \text{for } 3 \leq r \leq 2^{k-2}+2, \\ (n-3)+2k+2 & \text{otherwise.} \end{cases} \tag{1.7}$$

The small values are: $W_3(3)=3$, $W_3(4)=5$, $W_3(5)=7$; $V_3(3)=2$, $V_3(4)=4$, $V_3(5)=6$; $U_3(3)=0$, $U_3(4)=3$, $U_3(5)=5$.

2. Preliminary results

The numbers $U_3(n)$, $V_3(n)$ and $W_3(n)$ are related by a few well-known inequalities (see e.g. [1]): For $n \geq 4$,

$$W_3(n) < W_3(n+1) \leq W_3(n)+2, \tag{2.1}$$

$$V_3(n) < V_3(n+1) \leq V_3(n)+2, \tag{2.2}$$

$$U_3(n) < U_3(n+1) \leq U_3(n)+2, \tag{2.3}$$

and further

$$U_3(n+1) \leq V_3(n)+1 \leq W_3(n)+1. \tag{2.4}$$

We will need the following generalizations of (1.3) and (1.4). Let P be the poset of Fig. 1:

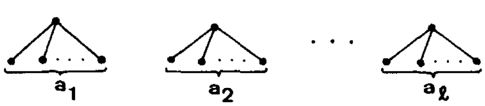


Fig. 1.

If, starting from P, $V_2(a_1,\dots,a_l) = W_2(a_1,\dots,a_l)$ and $U_2(a_1,\dots,a_l)$ denote the minimum number of comparisons needed to determine the first two elements in order and without regard to order, respectively, then

$$V_2(a_1,\dots,a_l) = W_2(a_1,\dots,a_l) = (l-2) + \left\lceil \log\left(\sum_{i=1}^{l} 2^{a_i}\right) \right\rceil \tag{2.5}$$

$$U_2(a_1,\dots,a_l) = (l-2) + \left\lceil \log\left(\sum_{i=1}^{l} 2^{a_i} - 1\right) \right\rceil. \tag{2.6}$$

(2.5) is an exercise in Knuth [5,p.219], (2.6) can be proved by an analogous argument. (1.3) and (1.4) become (2.5) and (2.6) in the special case $a_1 = a_2 = \cdots = a_l = 0$.

Knuth [5,p.212,214] gives upper bounds for $W_t(n)$ and $V_t(n)$ (due to Kislitsyn and Hadian–Sobel). By a variation of the tree selection method of Hadian–Sobel one can establish an upper bound for $U_t(n)$ as well. For $t = 3$ these bounds read as follows:

$$W_3(n) \le (n-3) + \lceil \log n \rceil + \lceil \log (n-1) \rceil, \tag{2.7}$$

$$V_3(n) \le (n-3) + 2\lceil \log (n-1) \rceil, \tag{2.8}$$

$$U_3(n) \le (n-3) + 2\lceil \log (n-2) \rceil. \tag{2.9}$$

A comparison of (1.5) to (1.7) with (2.7) to (2.9) shows that the algorithms giving rise to (2.7) to (2.9) are optimal except when r is in the range given in line 2 of (1.5) to (1.7).

3. The algorithms

The small values are easily disposed of, so let us assume $n \ge 8$, $n = 2^k + r$ where $0 \le r < 2^k$. We are going to design three algorithms whose respective lengths are given by the formulae (1.5), (1.6) and (1.7) thus proving these numbers to be upper bounds.

Let us denote the algorithms by W, V and U. By the remark at the end of the last section we may assume that r is in the range given in line 2 of (1.5), (1.6) and (1.7), respectively; in fact we may assume (by the inequalities (2.1) to (2.3)) that r attains the right-hand bound in each range.

Algorithm W. $n = 2^k + 2^{k-2}$.

Step 1. Split the elements into two disjoint sets S_1 and S_2 with $|S_1| = 2^k$, $|S_2| = 2^{k-2}$. By setting up balanced tournaments for S_1 and S_2 separately we then determine the largest element in each set. This takes

$$C_1 = n - 2 \tag{3.1}$$

comparisons. Let a denote the largest element of S_1 and b the element with which a was compared last; v shall denote the largest element of S_2. Hence after Step 1 we arrive at the poset shown in Fig. 2.

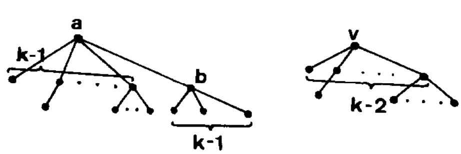


Fig. 2.

Step 2. Compare b with v. Thus

$$C_2 = 1. \tag{3.2}$$

If $b > v$, then a is the largest element. By (2.5) we can now determine the 2nd and 3rd element using at most

$$C_3' = (k-2) + \lceil \log(2^0 + 2^1 + \cdots + 2^{k-2} + 2^k) \rceil = 2k - 1 \tag{3.3'}$$

comparisons. If $b < v$, then:

Step 3. Compare a with v. If $a > v$, then a is the largest element. Counting the comparison $a \div v$ and using (2.5) we need at most

$$C_3'' = (k-1) + \lceil \log(2^0 + \cdots + 2^{k-2} + 2^{k-1}) \rceil = 2k - 1 \tag{3.3''}$$

comparisons to determine the other two.

If $a < v$, then v is the largest element. Counting the comparison $a \div v$ and using (2.5) again we need at most

$$C_3''' = (k-2) + \lceil \log(2^k + 2^0 + \cdots + 2^{k-3}) \rceil = 2k - 1 \tag{3.3'''}$$

further comparisons.

Adding (3.1), (3.2) and (3.3) we see that in each case the total number of comparisons is at most

$$(n-2) + 1 + (2k-1) = (n-3) + (2k+1).$$

The algorithms V and U are very similar whence we just list the steps.

Algorithm V. $n = 2^k + 2^{k-2} + 1$.

Step 1. Split the elements into 3 disjoint sets S_1, S_2 and S_3 with $|S_1| = 2^k$, $|S_2| = 2^{k-2}$ and $|S_3| = 1$, and determine the largest elements in S_1 and S_2. We use the notation a, b, v as in Fig. 2.

Step 2. $b \div v$. If $b > v$, then a is one of the two top elements. Now determine the 3rd using (2.5). If $b < v$, then:

Step 3. $a \div v$. The larger of a and v is one of the two top elements, and the 3rd element can now be determined, using (2.5).

Algorithm U. $n = 2^k + 2^{k-2} + 2$.

Step 1. Split the elements into 3 disjoint sets S_1, S_2 and S_3 with $|S_1| = 2^k$, $|S_2| = 2^{k-2}$ and $|S_3| = 2$, and determine the largest element of each set. We use the notation as in Fig. 2.

Step 2. $b \div v$. If $b > v$, then a is one of the three top elements and the other two can be determined, using (2.6). If $b < v$, then:

Step 3. $a \div v$. The larger of a and v is one of the three top elements, to determine the other two, we again use (2.6).

4. The basic oracle

In order to prove the theorem it remains to show that in any optimal binary tree (=algorithm) there is always a path from the root to an end-node of at least the length given on the right-hand side of (1.5) to (1.7). To do this we set up a suitable oracle in each case, a method described in Knuth [5,p.200,212].

The small cases $n \leq 8$ can be easily checked directly, hence we assume $n = 2^k + r$, $n \geq 9$. Using (1.1) and (2.1) to (2.4) we have the following:

Lemma 1. *Let $n = 2^k + r$, $n \geq 9$. In order to prove the theorem it suffices to show*

$$W_3(2^k+1) \geq (n-3)+(2k+1), \tag{4.1}$$

$$W_3(2^k+2^{k-2}+1) \geq (n-3)+(2k+2), \tag{4.2}$$

$$U_3(2^k+3) \geq (n-3)+(2k+1), \tag{4.3}$$

$$U_3(2^k+2^{k-2}+3) \geq (n-3)+(2k+2). \tag{4.4}$$

Let $\mathscr{T}$ be an optimal binary comparison tree for determining the first three elements (in order or without order). In order to prove (4.1), (4.2) and (4.3), (4.4), respectively, we must find a path $\mathscr{P}$ in $\mathscr{T}$ whose length $l(\mathscr{P})$ is given by the right hand side of (4.1) to (4.4). To do this we construct an oracle $\mathscr{O}$ (which is the same in either case) on $\mathscr{T}$. When there is no danger of confusion we identify $\mathscr{O}$ with the path $\mathscr{P}$ in $\mathscr{T}$ determined by $\mathscr{O}$ and call $l(\mathscr{O}) = l(\mathscr{P})$ the *length* of the oracle.

For the remainder of this section let $\mathscr{T}$ be a fixed optimal tree for determining the first three elements (in order or without order) out of a linear order N, $|N| = n$. To make the language more expressive we freely use the terms "players, games, wins, losses, etc.".

Some notation and terminology

Let $\mathscr{P}$ be a path in $\mathscr{T}$ (from the root to an end-node). All the terms that follow are defined with respect to $\mathscr{P}$.

(i) $l(\mathscr{P})$ is the length of $\mathscr{P}$.

(ii) $x \div y$ means that x and y are being compared.

(iii) Before the h-th comparison, $\mathscr{T}$ is said to be in *stage* h.

(iv) P_h is the poset determined by $\mathscr{T}$ in stage h; $<_h$ is the order relation in P_h and if $x \geq_h y$, then x is said to *dominate* y (in stage h). In a picture of P_h we usually only draw the upper part, deleting all elements which have already been found not to belong to the top three.

(v) Dropping the subscript means the final stage, i.e., after the last comparison. For example, P is the final poset, $<$ is the order relation in P, etc.

(vi) $x>y$ means that x defeated y; hence $x \succ y$ iff there exist $z_1, \ldots, z_t$ with $x > z_1 > \cdots > z_t = y$.

We come to the construction of the oracle $\mathcal{O}$. The idea of using a 'dominance function' is implicit in all oracle proofs of Kislitsyn's result (1.3). (See, in particular, Hyafil [3]). As before we set $n = 2^k + r$, $0 \le r < 2^k$, $n \ge 9$.

$\mathcal{O}$ is defined inductively on $\mathcal{T}$. To each stage h of $\mathcal{T}$ we associate a pair (T_h, φ_h) where $T_h \subseteq N$ and φ_h: $N \to \mathbb{N}_0$ are defined inductively as follows:

(S) $\qquad T_1 = \emptyset, \quad \varphi_1(x) = 1 \quad$ for all $x \in N$.

Suppose $x \div y$ is the h-th comparison. Then the oracle $\mathcal{O}$ says:

$x>y$ if (O1) $x, y \in T_h$ and x entered T_h before y,
or if (O2) $x \in T_h$, $y \notin T_h$,
or if (O3) $x \notin T_h$, $y \notin T_h$ and
(O3i) $\lceil \log \varphi_h(x) \rceil > \lceil \log \varphi_h(y) \rceil$,
or (O3ii) $\lceil \log \varphi_h(x) \rceil = \lceil \log \varphi_h(y) \rceil$ and x has not lost but y has already lost,
or (O3iii) $\lceil \log \varphi_h(x) \rceil = \lceil \log \varphi_h(y) \rceil$ and $\varphi_h(x) \ge \varphi_h(y)$.

That is, if in case (O3iii) $\varphi_h(x) = \varphi_h(y)$, then $\mathcal{O}$ makes an arbitrary decision compatible with the existing order. We use the convention $\lceil \log 0 \rceil = -1$.

After the decision $x>y$ the pair (T_h, φ_h) changes to:

(I1) If (O1) or (O2), then $T_{h+1} = T_h$, $\varphi_{h+1} = \varphi_h$.
If (O3), then

$$\text{(I2)} \quad T_{h+1} = T_h \text{ and } \begin{cases} \varphi_{h+1}(x) = \varphi_h(x) + \varphi_h(y), \\ \varphi_{h+1}(y) = 0, \\ \varphi_{h+1}(z) = \varphi_h(z) \text{ for } z \neq x, y, \end{cases} \quad \text{if } \varphi_h(x) + \varphi_h(y) \le 2^{k-1}.$$

$$\text{(I3)} \quad T_{h+1} = T_h \cup \{x\} \quad \text{and} \quad \varphi_{h+1} = \varphi_h \qquad \text{if } \varphi_h(x) + \varphi_h(y) > 2^{k-1}.$$

If (I3) applies, then $x \div y$ is called the *entry game* for x. In the case (I2) we call $x \div y$ an *essential game* if $\varphi_h(x) > 0$, $\varphi_h(y) > 0$. We say $x>y$ is an *essential win* for x if $x \div y$ is an essential game. Thus it is precisely the essential wins of x for which $\varphi_h(x) < \varphi_{h+1}(x)$.

Notation. (i) $D_h(x)$ is the set of players that have lost to x (in stage h), $D'_h(x)$ is the set of players that have lost to x in essential games; $d_h(x) = |D_h(x)|$, $d'_h(x) = |D'_h(x)|$.

(ii) $M_h = \{x \in N: x \notin T_h, \varphi_h(x) > 0\}$.

The following lemma lists the basic properties of the oracle $\mathcal{O}$. In particular, it says that T_h roughly corresponds to the top elements, M_h to the undecided elements and $\varphi_h(x)$ to the number of players dominated by x. Recall the convention that dropping the subscript means the final stage.

Lemma 2. *Let $n = 2^k + r$, $0 \le r < 2^k$, $n \ge 9$. Let s be the length of the oracle $\mathcal{O}$ on the* *optimal comparison tree $\mathcal{T}$. For all $x, y \in N$ and $h = 1, 2, \ldots, s, s+1$:*

$$\emptyset = T_1 \subseteq T_2 \subseteq \cdots \subseteq T, \qquad N = M_1 \supseteq M_2 \supseteq \cdots \supseteq M,$$

$$d_h(x) \le d_{h+1}(x), \qquad d'_h(x) \le d'_{h+1}(x), \tag{4.5}$$

$$\sum_{x \in N} \varphi_h(x) = n, \tag{4.6}$$

$$\varphi_h(x) \le 2^{k-1}, \tag{4.7}$$

$$d_h(x) \ge d'_h(x) \ge \lceil \log \varphi_h(x) \rceil, \tag{4.8}$$

$$D'_h(x) \cap D'_h(y) = \emptyset, \tag{4.9}$$

$$x \in T_{h+1} \Rightarrow \varphi_h(x) > 2^{k-2},\ d'_h(x) \ge k-1,\ d_{h+1}(x) \ge k, \tag{4.10}$$

$$\varphi_h(x) > 2^{k-2} \Rightarrow x \in T \cup M, \tag{4.11}$$

$$x \in M_h,\ x \prec_h y \Rightarrow y \in T_h, \tag{4.12}$$

$$|\{y \in N: y \preceq_h x\}| \ge \varphi_h(x). \tag{4.13}$$

Proof. (4.5) is obvious from the definition. (4.6) is clear from (S) and (I1) to (I3). (4.7) follows from (I2) and (I3), (4.8) from (O3) and (I2). (4.9) is clear from the observation that for the loser of an essential game φ_h drops to 0 and stays at 0 thereafter (by (O3)). To prove (4.10) suppose the i-th game $x \div y$, $i \le h$, is the entry game for x. Then, by (I3), $\varphi_i(x) + \varphi_i(y) > 2^{k-1}$, hence $\varphi_i(x) > 2^{k-2}$ by (O3) and thus $\lceil \log \varphi_i(x) \rceil \ge k-1$. Now, by (4.8), $d'_h(x) \ge d'_i(x) \ge k-1$ and counting the win $x > y$ we have $d_{h+1}(x) \ge d_{i+1}(x) \ge k$. (4.11) follows from (O3) and (I3), (4.12) from the definition of M_h and (I2). (4.13), finally, is easily established by induction. □

5. Proof of the theorem

To prove (4.1) to (4.4) we consider separately the cases whether the algorithm $\mathcal{T}$ determines the top three elements in order or without regard to order.

Let $\mathcal{T}$ be an optimal algorithm for determining the top three elements without regard to order and $\mathcal{O}$ the oracle of the previous section. By (4.3) and (4.4) we may assume $n = 2^k + 3$ or $n = 2^k + 2^{k-2} + 3$, $n \ge 11$.

Notation. Let $\mathcal{P}$ be a path in $\mathcal{T}$ from the root to an end-node. Then $A(i) = \{x \in N$: x loses at least i times (in $\mathcal{P}$)$\}$, $a_i = |A(i)|$, $i = 1, 2, \ldots$, $\mathrm{Top}_h = \{x \in N$: x dominates at least $n-2$ elements in stage h (including itself)$\}$, $\mathrm{Bot}_h = \{x \in N$: x is dominated by at least 4 elements in stage h (including itself)$\}$.

Since $\mathcal{T}$ is optimal we clearly have $a_i = 0$ for $i \ge 4$.

Lemma 3. *Let $\mathcal{P}$ be a path in $\mathcal{T}$, then $l(\mathcal{P}) = a_1 + a_2 + a_3$.*

Proof. $a_1+a_2+a_3$ counts the total number of defeats in $\mathscr{P}$. $\square$

When $\mathscr{P}$ stops, then the three top elements must form one of the following posets in Fig. 3.

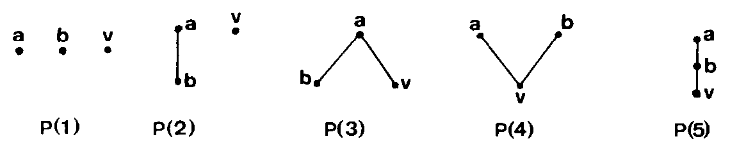


Fig. 3.

In the following lemma we use the letters a, b and v as in Fig. 3.

Lemma 4. *Let $\mathscr{P}$ be a path in $\mathscr{T}$ with $l(\mathscr{P})=s$. If the final poset is*

$P(1)$, *then* $s\geq(n-3)+|D(a)|+|D(b)|+|D(v)|-|D(a)\cap D(b)\cap D(v)|$. (5.1)

$P(2)$, *then* $s\geq(n-3)+|D(a)|+|D(b)|+|D(v)|-|D(b)\cap D(v)|$. (5.2)

$P(3)$, *then* $s\geq(n-3)+|D(a)|+|D(b)|+|D(v)|-|D(b)\cap D(v)|$. (5.3)

$P(4)$, *then* $A(1)=N-\{a,b\}$, $A(2)\supseteq D(a)\cup D(b)$,
$A(3)\supseteq(D(a)\cap D(b))-v$ *and* $s\geq(n-3)+|D(a)|+|D(b)|$. (5.4)

If we have equality for s, then we have set equality for $A(2)$ and $A(3)$.

$P(5)$, *then* $A(1)=N-\{a\}$, $A(2)\supseteq(D(a)-b)\cup(D(b)-v)$,
$A(3)\supseteq(D(a)\cap D(b))-v$ *and* $s\geq(n-3)+|D(a)|+|D(b)|$. (5.5)

If we have equality for s, then we have set equality for $A(2)$ and $A(3)$ and $v\notin D(a)$.

Proof. Let us just verify (5.1). Clearly, $A(1)=N-\{a,b,v\}$, hence $a_1=n-3$. Any player that loses to a, b or v must lose at least a second time, hence $A(2)\supseteq D(a)\cup D(b)\cup D(v)$. Any player that loses to two out of a,b,v must lose a third time, hence

$$A(3)\supseteq(D(a)\cap D(b))\cup(D(a)\cap D(v))\cup(D(b)\cap D(v)).$$

It follows that

$$a_2+a_3=|A(2)|+|A(3)|\geq|D(a)|+|D(b)|+|D(v)|-|D(a)\cap D(b)\cap D(v)|$$

and thus (5.1) by Lemma 3. $\square$

Top_h is by definition the set of elements which, in stage h, have already been found to belong to the top three; similarly for Bot_h. In particular, $\mathrm{Top}_1=\mathrm{Bot}_1=\emptyset$ and $|\mathrm{Top}|=3$, $|\mathrm{Bot}|=n-3$.

Lemma 5. *Suppose $l(\mathscr{P})=s$ and let $x \div y$ be the s-th (last) comparison.*

$$|\mathrm{Top}_s|=2, \qquad |\mathrm{Bot}_s|=n-4. \tag{5.6}$$

Let $\mathrm{Top}_s=\{a,b\}$. *Then $a>_s x$ or $a>_s y$ and similarly $b>_s x$ or $b>_s y$.* (5.7)

$z\in \mathrm{Bot}_s \Rightarrow z<_s a,\ z<_s b$ *and* $z<_s x$ *or* $z<_s y$. (5.8)

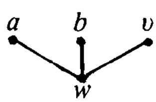
Proof. Since $x\div y$ is the final game the winner must go into Top and the loser into Bot. Hence $|\mathrm{Top}_s|\le 2$, $|\mathrm{Bot}_s|\le n-4$. Suppose there exists $w\in N-(\mathrm{Top}_s\cup \mathrm{Bot}_s \cup \{x,y\})$. If $w<_s x$, then in case $x>y$ no further relation is added involving w and hence w must have been in Bot_s. A similar reasoning shows $w\not<_s y$, $x\not<_s w$, $y\not<_s w$. Thus w is unrelated to x and y and could therefore not be determined by the comparison $x\div y$. (5.7) and (5.8) are proved by analogous arguments. □

The following propositions deal separately with each case whether the final poset determined by the oracle $\mathscr{O}$ is $P(1), P(2), \ldots$ or $P(5)$ in Fig. 3.

Proposition 1. *Let $\mathscr{T}$ be an optimal comparison tree for determining the top three elements (without regard to order) out of a linear order N, $|N|=n$, $n\ge 11$. Suppose the basic oracle $\mathscr{O}$ determines $P(1)$, $P(2)$ or $P(3)$ as final poset. Let $s=l(\mathscr{O})$ be the length of $\mathscr{O}$. Then*

$$\begin{aligned} &s\ge (n-3)+2k+1 \quad \text{if } n=2^k+3,\\ &s\ge (n-3)+2k+2 \quad \text{if } n=2^k+2^{k-2}+3. \end{aligned} \tag{5.9}$$

Proof. Suppose the final poset is $P(1)$ and set $n=2^k+r$ where r is either 3 or $2^{k-2}+3$. We use the notation of Fig. 3. By (5.6), the last game must involve one of the top elements a, b or v. Suppose the last game is $v\div w$. Then, by (5.7), we have



as a subposet of P and further, by (5.8), $z<_s a,b$ for all $z\ne a,b,v$. We use the notation T,M as in Section 4.

Case 0. $|T|=0$. Then $M=\{a,b,v\}$ by (4.12). Both a and b defeated w; let a be the first to have beaten w. Since $\varphi(a)+\varphi(b)+\varphi(v)=n>2^k$ and $\varphi(a),\varphi(b),\varphi(v)\le 2^{k-1}$ by (4.6) and (4.7), two of $\varphi(a),\varphi(b),\varphi(v)$ must be $>2^{k-2}$.

Assume $\varphi(a)>2^{k-2}$, $\varphi(b)>2^{k-2}$. Then (by (4.8)) $d'(a)\ge k-1$, $d'(b)\ge k-1$ and $d(b)\ge k$ counting the non-essential win $b>w$. Since $D'(v)\subseteq D(v)$ we have

$$|D(v)|-|D(v)\cap D(a)\cap D(b)| \ge |D'(v)|-|D'(v)\cap D(a)\cap D(b)|.$$

Setting $u=|D'(v)\cap D(a)\cap D(b)|$ we note $d(a)\ge k-1+u$, $d(b)\ge k+u$ by (4.9) and the fact that $w\notin D'(v)$ and $d'(v)\ge \lceil \log r\rceil$ since $\varphi(v)\ge r$ by (4.7). Using (5.1) we thus obtain

$$s \geq (n-3)+(2k-1)+2u+d'(v)-u \geq (n-3)+2k+(\lceil \log r \rceil -1)$$

and hence (5.9).

The case $\varphi(a)>2^{k-2}$, $\varphi(v)>2^{k-2}$ is the same as before with the roles of b and v interchanged.

Assume, finally, $\varphi(b)>2^{k-2}$, $\varphi(v)>2^{k-2}$. Then $d'(b)\geq k-1$, $d'(v)\geq k-1$. Setting $u=|D'(a)\cap D(b)\cap D(v)|$ we have $u\geq 1$ since $w\in D'(a)\cap D(b)\cap D(v)$ and further $d(b)\geq d'(b)+u$, $d(v)\geq d'(v)+u$ by (4.9). Since by the same reasoning as before $\varphi(a)\geq r$, we obtain by (5.1)

$$s \geq (n-3)+2k-2+2u+d'(a)-u \geq (n-3)+2k+(\lceil \log r \rceil -1)$$

and hence (5.9).

Case 1. $|T|=1$. Then $T\cup M=\{a,b,v\}$ by (4.12). Let us again assume that a was the first to have beaten w. Assume $T=\{a\}$. Then $\varphi(a)>2^{k-2}$ and $d(a)\geq k$ by (4.10). One of $\varphi(b)$ and $\varphi(v)$ must be $>2^{k-2}$. Suppose w.l.o.g. $\varphi(b)>2^{k-2}$ and thus $\varphi(v)\geq r$. Setting $u=|D'(v)\cap D(a)\cap D(b)|$ we have as before $d(b)\geq d'(b)+u\geq (k-1)+u$. By (5.1), this yields

$$s \geq (n-3)+(2k-1)+u+d'(v)-u \geq (n-3)+2k+(\lceil \log r \rceil -1)$$

and thus (5.9). The case where $a\notin T$ and, say, $T=\{b\}$ is settled by an analogous argument.

Case 2. $|T|=2$. Then $T\cup M=\{a,b,v\}$ by (4.12). Assume $T=\{a,b\}$. Then $\varphi(a)>2^{k-2}$, $\varphi(b)>2^{k-2}$, $\varphi(v)\geq r$. Setting $u=|D'(v)\cap D(a)\cap D(b)|$ we have $d(a)\geq d'(a)+u\geq (k-1)+u$ and $d(b)\geq k$ by (4.10). Using (5.1) again, we obtain

$$s \geq (n-3)+2k-1+u+d'(v)-u \geq (n-3)+2k+(\lceil \log r \rceil -1)$$

and thus (5.9). The same reasoning goes through for $v\in T$.

Case 3. $|T|\geq 3$. Then $\{a,b,v\}\subseteq T$ by (4.12), and thus $d(a)\geq k$, $d(b)\geq k$ and $d(v)\geq k$. W.l.o.g. let a be the first of the three to have entered T. Setting $u=|D'(a)\cap D(b)\cap D(v)|$ we have $d(b)\geq k-1+u$, $d(v)\geq k$. Thus again

$$s \geq (n-3)+2k-1+u+d'(a)-u \geq (n-3)+2k+(\lceil \log r \rceil -1)$$

and hence (5.9).

The cases when the final poset of the top three is $P(2)$ or $P(3)$ in Fig. 3 are dealt with by entirely analogous arguments. □

The remaining two cases $P(4)$ and $P(5)$ are considerably more difficult. Note that these are the posets which are relevant to the V_3- and W_3-problem. Here the oracle will, in general, not produce a path of the length required in the theorem. Our method will thus consist in changing the oracle at a suitable stage in order to produce a required path.

To shorten the argument the following notation is useful: Suppose $\mathcal{O}$ is an oracle defined on the comparison tree $\mathcal{T}$, and let A, B be disjoint subsets of N. By setting $A <_{\mathcal{O}} B$ or just $A < B$ if there is no danger of confusion we mean that whenever $\mathcal{T}$ performs a comparison $a \div b$ with $a \in A$ and $b \in B$, then $\mathcal{O}$ decides $a < b$. Whenever an outcome is not explicitly stated in the definition of $\mathcal{O}$, then the oracle is assumed to make an arbitrary decision compatible with the existing order. The following two propositions sum up the results for the ordered and unordered case in the presence of $P(4)$ or $P(5)$ as final poset. We confine ourselves to a proof of the somewhat subtler ordered case. Here the final poset is $P(5)$, of course.

Proposition 2. *Let $\mathcal{T}$ be an optimal comparison tree for determining the top three elements (without regard to other) out of a linear order N, $|N| = n \geq 11$. Suppose the oracle $\mathcal{O}$ determines $P(4)$ or $P(5)$ as final poset. Then there exists a path $\mathcal{P}$ in $\mathcal{T}$ of length s^* where*

$$\begin{aligned} s^* &\geq (n-3) + 2k + 1 \quad \text{if } n = 2^k + 3, \\ s^* &\geq (n-3) + 2k + 2 \quad \text{if } n = 2^k + 2^{k-2} + 3. \end{aligned} \tag{5.10}$$

Proposition 3. *Let $\mathcal{T}$ be an optimal comparison tree for determining the top three elements in order out of a linear order N, $|N| = n \geq 9$. Then there exists a path $\mathcal{P}$ in $\mathcal{T}$ of length s^* where*

$$\begin{aligned} s^* &\geq (n-3) + 2k + 1 \quad \text{if } n = 2^k + 1, \\ s^* &\geq (n-3) + 2k + 2 \quad \text{if } n = 2^k + 2^{k-2} + 1. \end{aligned} \tag{5.11}$$

Proof. We use the notation a, b, v in the final poset $P(5)$ as in Fig. 3. Let $\mathcal{O}$ be the basic oracle of Section 4 and $s = l(\mathcal{O})$ its length. We have $a, b \in T$ and $v \in M \cup T$ by (4.12) and (4.7) and hence $s \geq (n-3) + 2k$ by (5.5) and (4.10).

Case 1. $n = 2^k + 1$. If $s > (n-3) + 2k$ we are finished, so assume $s = (n-3) + 2k$. It follows that $d(a) = d(b) = k$, $D(a) \cap D(b) = \emptyset$ and that $a \div b$, $b \div v$ are the entry games for a and b, respectively. Furthermore, any game in $\mathcal{O}$ not involving a or b must be between top elements of $N - \{a, b\}$ at that stage. Let $a \div b$ be the i-th game and $b \div v$ the u-th game where $i < u$ by (O1).

Suppose $\varphi_u(v) > 2^{k-3}$. Then $d_u'(v) = l \geq k-2$ and P_u looks as depicted in Fig. 4.

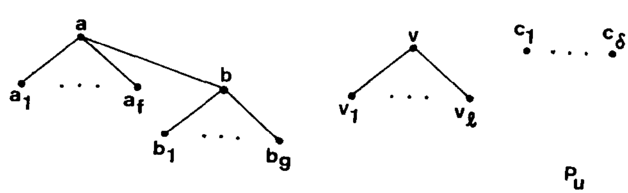


Fig. 4.

$A = \{a_1, \ldots, a_f\}$ is the set of players who have lost only to a (in stage u); similarly for $B = \{b_1, \ldots, b_g\}$ and $V = \{v_1, \ldots, v_l\}$. $C = \{c_1, \ldots, c_\delta\}$ is the set of unbeaten players different from a and v. Hence

$$s-u=f+g+\delta \quad \text{with} f \leq k-1,\ g \leq k-1. \tag{5.12}$$

If, after the outcome $b>v$, we change the oracle to $\mathcal{O}'$: $B \cup \{v\} < A < C < b < a$ we see that another game will involve b unless $f=\delta=0$. Thus, by (5.12), we may assume $s-u=g \leq k-1=d_u(b)$. But now, consider the outcome $b<v$ with the resulting poset P_{u+1} depicted in Fig. 5.

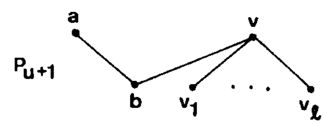


Fig. 5.

By choosing the oracle $\mathcal{O}''$: $b<V<a<v$ we need at least $l+2 \geq k$ games whence $u+l(\mathcal{O}'') \geq u+k \geq s+1$.

Let us then assume $\varphi_u(v) \leq 2^{k-3}$. By using the oracle $\mathcal{O}'$ from above we may assume that $\delta=0$ and hence

$$\varphi_u(a)+\varphi_u(b) \geq 2^k+1-2^{k-3}=2^{k-1}+2^{k-2}+2^{k-3}+1$$

by (4.6), and thus

$$\varphi_u(a), \varphi_u(b) \geq 2^{k-2}+2^{k-3}+1$$

by (4.7) and (O3). Clearly, $d_i'(a)=k-1$, but also $d_i'(b)=k-1$ since $d_i'(b)=h \leq k-2$, then $\varphi_i(b) \leq 2^h$ by (4.8) and thus

$$\varphi_u(b) \leq 2^{k-2}+2^{h-1} \leq 2^{k-2}+2^{k-3}$$

by (O3ii). Let $x_1, \ldots, x_{k-1}$ be the players defeated by a in this order before game i. From $\varphi_i(a) \geq 2^{k-2}+2^{k-3}+1$ it follows that $d'(x_1) \geq 0, d'(x_2) \geq 1, \ldots, d'(x_{k-1}) \geq k-2$. Similarly, let $y_1, \ldots, y_{k-1}$ be the players defeated by b in this order then $d'(y_1) \geq 0$, $\ldots, d'(y_{k-1}) \geq k-2$. P_i is thus of the form depicted in Fig. 6 where $X=\{x_{i_1}, \ldots, x_{i_\alpha}\}$, $Y=\{y_{j_1}, \ldots, y_{j_\beta}\}$, are again the sets of players that have only lost to a and b, respectively, and $C=\{c_1, \ldots, c_\gamma\}$ is the set of unbeaten players apart from a and b.

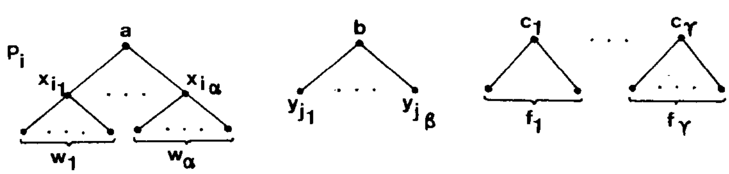


Fig. 6.

We have

$$s-i=\alpha+\beta+\gamma \quad \text{with } \gamma\geq 1. \tag{5.13}$$

Assuming $w_1\leq w_2\leq \cdots \leq w_\alpha$ we note $w_i\geq i-1$ for $i=1,\dots,\alpha$. Let the oracle decide $a>b$ if $\alpha\geq\beta$. If, after the outcome $a>b$, the oracle declares a to be the best player we need (by (2.5)) at least

$$\begin{aligned}&(\alpha+\gamma-1)+\lceil\log(2^{w_1}+\cdots+2^{w_\alpha}+2^\beta+2^{f_1}+\cdots+2^{f_\gamma})\rceil\\ &\quad\geq(\alpha+\gamma-1)+\lceil\log(2^\alpha+2^\beta+2^{f_1}+\cdots+2^{f_\gamma}-1)\rceil.\end{aligned}$$

For this expression to be $\leq\alpha+\beta+\gamma$ we must have $\alpha=\beta$, $w_1=0,\dots,w_\alpha=\alpha-1$, $\gamma=1$, $f_1=0$ and thus

$$s-i=2\alpha+1. \tag{5.14}$$

Hence P_{i+1} looks as in Fig. 7.

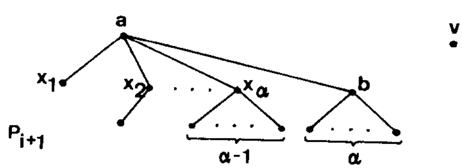


Fig. 7.

After game i the oracle $\mathcal{O}^*$ is defined very much like the basic oracle. First of all, $\mathcal{O}^*$ stipulates $N-(X\cup\{a,b,v\})<X\cup\{b,v\}<a$. Within $X\cup\{b,v\}$ the decisions are made according to the number of elements covered at stage t. To formalize this we define functions ψ_t: $X\cup\{b,v\}\to\mathbb{N}_0, t\geq i+1$, as follows. At the start

$$\begin{aligned}&\psi_{i+1}(x_j)=2^{j-1} \quad (j=1,\dots,\alpha),\\ &\psi_{i+1}(b)=2^\alpha, \qquad \psi_{i+1}(v)=1.\end{aligned}$$

Let $D_t=\{x\in X\cup\{b,v\}:\ \psi_t(x)>0\}$. If $x\div y$ is the t-th game with $x,y\in X\cup\{b,v\}$, then $\mathcal{O}^*$ decides $x>y$ iff $\psi_t(x)\geq\psi_t(y)$. If $\psi_t(x)=\psi_t(y)$ and $x=v$, then $\mathcal{O}^*$ says $v>y$. After the outcome $x>y$ with $x,y\in D_t$, ψ_t changes to

$$\begin{aligned}&\psi_{t+1}(x)=2\psi_t(x), \quad \psi_{t+1}(y)=0,\\ &\psi_{t+1}(z)=\psi_t(z) \quad \text{for } z\in X\cup\{b,v\},\ z\neq x,y.\end{aligned}$$

Hence in this case $D_{t+1}=D_t-\{y\}$. If the t-th game has as its loser an element covered by some $w\in D_t$ then we set $\psi_{t+1}(w)=\frac{1}{2}\psi_t(w)$ and $\psi_{t+1}(z)=\psi_t(z)$ for all other z. In all other cases we stipulate $\psi_{t+1}=\psi_t$.

It follows from the definition that, for $x\in D_t$, $\log\psi_t(x)$ is the number of elements, covered by x in stage t, which are still possible candidates for the third ranked element. If the t-th game matches two elements in D_t, then, clearly, $\sum_z\psi_{t+1}(z)\geq$

$\sum_z \psi_t(z)$ with strict inequality if v was beaten for the first time in the t-th game. Suppose there is no further game involving a. Then there must be a game, say the t-th game, which puts v below a. Suppose a covers f elements in P_{t+1} and suppose that there have been p games up to this stage which involved elements not in $X \cup \{b, v\}$. Then $t-i=\alpha+2-f+p$. From the set-up of $\mathcal{O}^*$ we have

$$\sum \psi_{t+1}(z) \geq (\sum \psi_{i+1}(z))2^{-p}+1 = \left(\sum_{j=1}^{\alpha} 2^j+1\right)2^{-p}+1 = 2^{\alpha+1-p}+1.$$

Applying (2.5) we thus need at least

$$(f-2)+\lceil \log(2^{\alpha+1-p}+1)\rceil = (f-2)+(\alpha+2-p) = f+\alpha-p$$

more games after the t-th game. Hence, by (5.14)

$$\begin{aligned} i+l(\mathcal{O}^*) &= i+(t-i)+(f+\alpha-p) \\ &= (s-2\alpha-1)+(\alpha+2-f+p)+(f+\alpha-p) = s+1. \end{aligned}$$

Suppose, finally, the game $a \div v$ is played in $\mathcal{O}^*$. Let it be the e-th game. Then by using the symbols f and p as before we have $e-i=\alpha+3-f+p$, $\sum \psi_{e+1}(z) \geq 2^{\alpha+1-p}$ and thus $i+l(\mathcal{O}^*) \geq s+1$.

Case 2. $n=2^k+2^{k-2}+1$. Again $s=l(\mathcal{O}) \geq (n-3)+2k$.

Case 2a. $s=(n-3)+2k$. As in Case 1 we have $d(a)=d(b)=k$, $D(a) \cap D(b)=\emptyset$ with $a \div b$ and $b \div v$ being entry games. Let as before be $a \div b$ the i-th game and $b \div v$ the u-th game in $\mathcal{O}$. Consider the u-th game where we use the notation of Fig. 4. Suppose first $\sum_{j=1}^{\delta} \varphi_u(c_j) \geq 2^{k-3}+1$. After the outcome $b>v$ we define $\mathcal{O}'$ by stipulating

$$N-(A \cup B \cup C \cup \{a,b,v\}) < B \cup \{v\} < A < C < b < a$$

and keeping $\mathcal{O}$ within A and C. Since $\delta \geq 1$ there must be another game involving b. Let it be the t-th game. P_t looks as in Fig. 8.

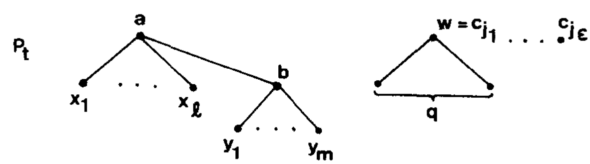


Fig. 8.

Then $t-1-u \geq (f+g+1+\delta)-(l+m+\varepsilon)$ with $\varepsilon \geq 1$, $m \leq k$, and thus, by (5.12), $s-t+1 \leq l+m+\varepsilon-1$. b must play some x_i or c_{j_i} in the t-th game. Now, if $l+\varepsilon \geq 2$, then b would have to play at least one more game whence

$$\begin{aligned} u+l(\mathcal{O}') &\geq u+(t-u)+l+m+\varepsilon \\ &\geq (s-f-g-\delta)+(f+g+\delta+2-l-m-\varepsilon)+(l+m+\varepsilon) = s+2. \end{aligned}$$

Hence we must have $l+\varepsilon=1$ and thus $l=0$, $\varepsilon=1$. Since $\sum_{j=1}^{\delta}\varphi_u(c_j)>2^{k-3}$ this implies $q=d_l'(w)\geq k-2$. The t-th game must therefore be $b\div w$. After changing the outcome to $b<w$ let the oracle $\mathcal{O}''$ declare w to be the best player. By (2.5), we need at least $q-1+\lceil\log(2+2^{k-3})\rceil\geq 2k-5$ more games to determine the other two. Hence

$$\begin{aligned}t+l(\mathcal{O}'')&=u+(t-u)+2k-5\geq(f+g+2+\delta)-(m+1)+u+2k-5\\&\geq(s-u+2)+u+k-6\geq s+2\quad\text{for } k\geq 6.\end{aligned}$$

The small cases $k=3,4,5$ are easily disposed of by considering a refined oracle $\mathcal{O}^*$ as in Case 1.

Suppose then $\sum_{j=1}^{\delta}\varphi_u(c_j)\leq 2^{k-3}$. Then $\varphi_u(v)>2^{k-3}$ and since $\varphi_u(v)\leq 2^{k-2}$ by (O3ii) we conclude

$$\varphi_u(a)+\varphi_u(b)\geq 2^k+2^{k-2}+1-2^{k-2}-2^{k-3}=2^{k-1}+2^{k-2}+2^{k-3}+1.$$

By the same argument as in Case 1 it follows that $d_i'(a)=d_i'(b)=k-1$ and, further, that $d'(x_1)\geq 0, d'(x_2)\geq 1,\ldots,d'(x_{k-1})\geq k-2$ where $X=\{x_1,\ldots,x_{k-1}\}$ are the players defeated by a (in that order) before game i and similarly $d'(y_1)\geq 0,\ldots,d'(y_{k-1})\geq k-2$ where $Y=\{y_1,\ldots,y_{k-1}\}$ are the players defeated by b up to game i. Let P_i be as in Fig. 6 and suppose $\mathcal{O}$ makes the decision $a>b$ iff $\alpha\geq\beta$. Then P_u is of the form depicted in Fig. 4 with $l=d'(v)\geq k-2$ since $\varphi_u(v)>2^{k-3}$, and $\delta\geq 1$ since $\varphi_u(a)+\varphi_u(b)+\varphi_u(v)\leq 2^k+2^{k-2}<n$. Let the outcome of the u-th game be $b<v$. Then by using the notation as in Fig. 4 and employing the oracle $\mathcal{O}'$ as given there we conclude that there must be at least $f+l+\delta+2\geq f+\delta+k$ more games to determine the top three. By (5.12), this gives

$$u+l(\mathcal{O}')\geq(s-f-g-\delta)+(\mathrm{f}+\delta+k)=s+(k-g).$$

Hence in order that $u+l(\mathcal{O}')\leq s+1$ we must have $g=k-1$, and thus $\alpha\geq\beta\geq g=k-1$, i.e., $\alpha=\beta=k-1$, in the notation of Fig. 6. Hence we may assume that P_{i+1} is as in Fig. 9.

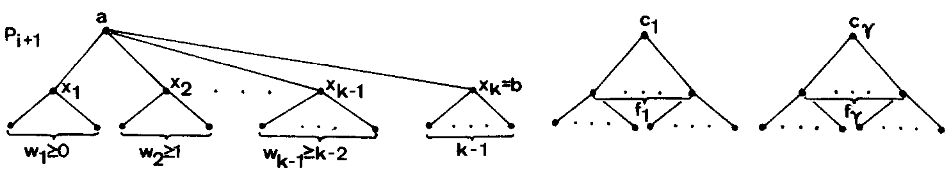


Fig. 9.

with

$$s-i=2k-2+\gamma,\quad \gamma\geq 1. \tag{5.15}$$

The following argument is the subtlest part of the proof which, of course, is to be expected since the situation of Fig. 9 corresponds precisely to the set-up of the (optimal) algorithm W of Section 3. After the i-th game we change to the following oracle $\tilde{\mathcal{O}}$. Set $X=\{x_1,\dots,x_k\}$, $C=\{c_1,\dots,c_\gamma\}$. First of all, $\tilde{\mathcal{O}}$ stipulates that

$$N-(X\cup C\cup\{a\})<X\cup C<a.$$

Within $X\cup C$ the decisions are made according to the functions $\chi_t: X\cup C\to\mathbb{N}_0$, $t\geq i+1$, in a similar way as in Case 1. At the start

$$\begin{aligned}
&\chi_{i+1}(x_j)=2^{w_j} && (j=1,\dots,k-1),\\
&\chi_{i+1}(x_k)=2^{k-1},\\
&\chi_{i+1}(c_j)=2^{f_j} && (j=1,\dots,\gamma).
\end{aligned}$$

Let $D_t=\{x\in X\cup C\colon \chi_t(x)>0\}$, $X_t=D_t\cap X$, $C_t=D_t\cap C$. If $y\div z$ is the t-th game with $y,z\in X\cup C$, then $\tilde{\mathcal{O}}$ says

$$\begin{aligned}
y>z\quad &\text{if}\quad y,z\in X \text{ and } \chi_t(y)\geq\chi_t(z),\\
&\text{or if}\quad y,z\in C \text{ and } \chi_t(y)\geq\chi_t(z),\\
&\text{or if}\quad y\in C,\ z\in X \text{ and } \chi_t(y)\geq\tfrac{1}{2}\chi_t(z).
\end{aligned}$$

After the outcome $y>z$, with $y,z\in D_t$, χ_t changes to

$$\begin{aligned}
&\chi_{t+1}(y)=2\chi_t(y), \qquad \chi_{t+1}(z)=0,\\
&\chi_{t+1}(w)=\chi_t(w)\quad \text{for } w\in X\cup C,\ w\neq y,z.
\end{aligned}$$

As in Case 1, if the t-th game has as its loser an element covered by some $w\in D_t$, then $\chi_{t+1}(w)=\frac{1}{2}\chi_t(w)$ and $\chi_{t+1}(z)=\chi_t(z)$ for all other $z\in D_t$. In all other cases, $\chi_{t+1}=\chi_t$.

Hence by the set-up of $\tilde{\mathcal{O}}$ a player $x\in X$ beats a player $c\in C$ iff $\chi_t(x)>0$ and $\chi_t(x)\geq 4\chi_t(c)$. Note that if $\chi_t(z)>0$ where $z\in X\cup C$, then it must be a power of 2. Define

$$S_t(X)=\sum_{j=1}^{k}\chi_t(x_j)\quad\text{and}\quad S_t(C)=\sum_{j=1}^{\gamma}\chi_t(c_j)\quad\text{for } t\geq i+1.$$

Thus $S_{i+1}(X)\geq 2^k-1$, $S_{i+1}(C)\geq 2^{k-2}+1$ (by (4.8) and (4.7)). The following observations are immediate from the definition of $\tilde{\mathcal{O}}$. Whenever $x\in X$ loses to $c\in C$ in, say, the t-th game, then $S_t(X)$ will decrease by $\chi_t(x)$ and $S_t(C)$ will increase by at least $\frac{1}{2}\chi_t(x)$. Whenever $c\in C$ loses to $x\in X$, then $S_t(C)$ will decrease by $\chi_t(c)$ and $S_t(X)$ will increase by at least $4\chi_t(c)$. Whenever the t-th game matches two members of X or two members of C then both sums $S_{t+1}(X)$ and $S_{t+1}(C)$ will be at least as large as $S_t(X)$ and $S_t(C)$, respectively. Set $L_t(X)=\sum_j\chi_j(x)$ where the summation is over all $\chi_j(x)$ such that in game j some $x\in X$ lost to some $c\in C$, for $i+1\leq j\leq t$. That is, $L_t(X)$ is the sum of the amounts that were substracted from $S_{i+1}(X)$ up to stage t due to losses of players of X to players of C. Similarly, define $L_t(C)$ by the losses of members of C to members of X. From what we just said the corresponding increments will be at least $\frac{1}{2}L_t(X)$ to $S_{i+1}(C)$ and at least $4L_t(C)$ to $S_{i+1}(X)$.

Suppose there is no more game involving a. Let the t-th game be the last game involving (unbeaten) players of C and suppose there have been p games (up to the t-th game) involving elements not in D_j, $j=i+1,\dots,t$. Then P_{t+1} is as in Fig. 10.

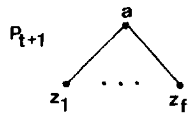


Fig. 10.

and we have $t-i=k+\gamma-f+p$. Now

$$0 \geq 2^{k-2}+1-L_{t+1}(C)+\tfrac{1}{2}L_{t+1}(X),$$

$$S_{t+1}(X) \geq (2^k-1-L_{t+1}(X)+4L_{t+1}(C))2^{-p}.$$

From this it follows that $2L_{t+1}(C)-L_{t+1}(X) \geq 2^{k-1}+2$ and thus

$$S_{t+1}(X) \geq (2^k-1+2^k+4+L_{t+1}(X))2^{-p} > 2^{k+1-p}.$$

Using (2.5), we conclude that we need at least another

$$(f-2)+\lceil \log(S_{t+1}(X)) \rceil \geq (f-2)+(k+2-p)=f+k-p$$

games. By (5.15), this gives

$$\begin{aligned} i+l(\tilde{\mathcal{O}}) &\geq i+(t-i)+(f+k-p) \\ &=(s-2k+2-\gamma)+(k+\gamma-f+p)+(f+k-p)=s+2. \end{aligned}$$

Suppose there are $m \geq 1$ games after the i-th game involving a and elements of C, say games $u_1, u_2, \ldots, u_m = u$ and suppose again there have been p games (up to the u-th game) involving elements not in D_j, $j=i+1, \ldots, u$. Let the opponents of a in these games be $v_1, \ldots, v_m$ with $d'_{u_j}(v_j)=l_j$. Then

$$X_{u_j+1}=X_{u_j} \cup \{v_j\}, \qquad C_{u_j+1}=C_{u_j}-v_j$$

and

$$S_{u_j+1}(X)=S_{u_j}(X)+\chi_{u_j}(v_j), \qquad S_{u_j+1}(C)=S_{u_j}(C)-\chi_{u_j}(v_j)$$

for $j=1, \ldots, m$. P_{u+1} looks therefore as in Fig. 11,

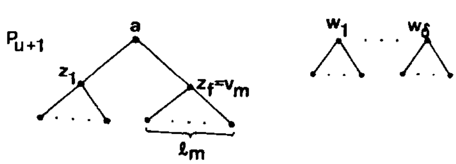


Fig. 11.

where $Z=\{z_1, \ldots, z_f\}=X_{u+1}$, $W=\{w_1, \ldots, w_\delta\}=C_{u+1}$ and

$$u-i=k+\gamma-f-\delta+m+p. \tag{5.16}$$

Claim: We may assume $l_j \le k-j$ for $j=1,\dots,m-1$. Suppose on the contrary $l_j > k-j$ for some j. Then by choosing after the u_j-th game the oracle $\mathcal{O}'$:

$$N-(D'_{u_j}(v_j)\cup X_{u_j}\cup C_{u_j}\cup\{a,v_j\})<D'_{u_j}(v_j)<X_{u_j}<C_{u_j}<v_j<a$$

we need at least $l_j+|X_{u_j}|+|C_{u_j}|-1$ more games (since $|C_{u_j+1}|\ge 1$) whence, by (5.15),

$$\begin{aligned}u_j+l(\mathcal{O}')&=i+(u_j-i)+l(\mathcal{O}')\\&\ge(s-2k+2-\gamma)+(k+\gamma-|X_{u_j}|-|C_{u_j}|+j)\\&\quad+(l_j+|X_{u_j}|+|C_{u_j}|-1)=s-k+1+j+l_j\ge s+2.\end{aligned}$$

Suppose first $\delta=0$, and hence $u-i=k+\gamma-f+m+p$.

If $f=1$, then

$$0=S_u(X)\ge(2^k-1-L_u(X)+4L_u(C)+2^{l_1}+\cdots+2^{l_{m-1}})2^{-p}$$

and thus

$$-L_u(C)+\tfrac{1}{2}L_u(X)\ge 2^{k-1}-\tfrac{1}{2}+2^{l_1-1}+\cdots+2^{l_{m-1}-1}+L_u(C)$$

whence

$$\begin{aligned}S_u(C)=2^{l_m}&\ge(2^{k-2}+1+2^{k-1}-\tfrac{1}{2}-2^{l_1-1}-\cdots-2^{l_{m-1}-1}+L_u(C))2^{-p}\\&\ge(2^{k-2}+2^{k-m}+\tfrac{1}{2}+L_u(C))2^{-p}.\end{aligned}\tag{5.17}$$

For $m\ge 2$ this last expression is $>2^{k-m+1-p}$. Hence in this case $l_m\ge k-m+2-p$ and therefore

$$\begin{aligned}i+l(\tilde{\mathcal{O}})&\ge i+(u-i)+(k-m+1-p)\\&=(s-2k+2-\gamma)+(k+\gamma-1+m+p)+(k-m+1-p)=s+2.\end{aligned}$$

If $m=1$, then $l_1\ge k-p$. If there exists $w\in D'_u(v_u)$ with $w \not<_u a$, then by choosing the oracle $(\tilde{\tilde{\mathcal{O}}})$: $D'_u(v_u)-w<w<a<v_u$ we need after the $(u-1)$-st game at least $l_1+1\ge k+1-p$ more games whence

$$(u-1)+l(\tilde{\tilde{\mathcal{O}}})\ge(s-k+1+p)+(k+1-p)=s+2.$$

If, on the other hand, $D'_u(v_u)<a$, i.e. $D'_u(v_u)\subseteq X_{i+1}$, then $L_u(C)\ge 2^{k-2}$ and thus, by (5.17),

$$S_u(C)=2^{l_1}\ge(2^k+\tfrac{1}{2})2^{-p}>2^{k-p},$$

i.e., $l_1\ge k-p+1$ from which again $i+l(\tilde{\mathcal{O}})\ge s+2$ results.

If $f\ge 2$, then

$$0=S_{u+1}(C)\ge(2^{k-2}+1-L_{u+1}(C)+\tfrac{1}{2}L_{u+1}(X)-2^{l_1}-\cdots-2^{l_m})2^{-p},$$

hence

$$-L_{u+1}(X)+2L_{u+1}(C)\ge 2^{k-1}+2-2^{l_1+1}-\cdots-2^{l_m+1}.$$

It follows that

$$\begin{aligned} S_{u+1}(X) &\geq (2^k - 1 + 2^{l_1} + \cdots + 2^{l_m} - L_{u+1}(X) + 4L_{u+1}(C))2^{-p} \\ &\geq (2^k - 1 + 2^{k-1} + 2 - 2^{l_1} - \cdots - 2^{l_m} + 2L_{u+1}(C))2^{-p} \\ &\geq (2^{k-1} + 2^{k-m+1} - 2^{l_m} + 2L_{u+1}(C) + 1)2^{-p} \\ &> 2^{k-m+1-p} + (2^{k-1} - 2^{l_m})2^{-p}. \end{aligned}$$

Since $v_m \in X_{u+1}$ we also have $S_{u+1}(X) > 2^{l_m}$ because of $f \geq 2$. Now after the u-th game we need, by (2.5), at least

$$(f-2) + \lceil \log(S_{u+1}(X)) \rceil \geq (f-2) + (k - m + 2 - p)$$

more games whence (by (5.15) and (5.16))

$$\begin{aligned} i + l(\tilde{\mathcal{O}}) &\geq i + (u - i) + (f + k - m - p) \\ &\geq (s - 2k + 2 - \gamma) + (k + \gamma - f + m + p) + (f + k - m - p) = s + 2. \end{aligned}$$

Suppose, finally $\delta \geq 1$ in P_{u+1} of Fig. 11. In this case we may assume $l_m \leq k - m$ by employing the oracle $\mathcal{O}'$ of above. Let the e-th game be the last one to involve (unbeaten) members of C, and suppose there have been q games (between the u-th and e-th game) involving elements not in D_j, $j = u+1, \ldots, e$. Then P_{e+1} looks as in Fig. 12.

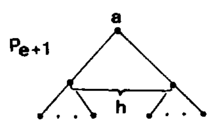


Fig. 12.

with

$$e - u = f + \delta - h + q, \quad h \geq 1. \tag{5.18}$$

We have

$$0 = S_{e+1}(C) \geq (2^{k-2} + 1 - L_{e+1}(C) + \tfrac{1}{2}L_{e+1}(X) - 2^{l_1} - \cdots - 2^{l_m})2^{-(p+q)},$$

thus

$$-L_{e+1}(X) + 2L_{e+1}(C) \geq 2^{k-1} + 2 - 2^{l_1+1} - \cdots - 2^{l_m+1}.$$

It follows that

$$\begin{aligned} S_{e+1}(X) &\geq (2^k - 1 + 2^{l_1} + \cdots + 2^{l_m} - L_{e+1}(X) + 4L_{e+1}(C))2^{-(p+q)} \\ &\geq (2^k - 1 + 2^{k-1} + 2 - 2^{l_1} - \cdots - 2^{l_m} + 2L_{e+1}(C))2^{-(p+q)} \\ &\geq (2^{k-1} + 2^{k-m} + 1 + 2L_{e+1}(X))2^{-(p+q)} \\ &> 2^{k-m+1-p-q}. \end{aligned}$$

By (2.5), we need as least another

$$(h-2)+\lceil\log(S_{e+1}(X))\rceil \geq (h-2)+(k-m+2-p-q)$$

games after the e-th game whence by (5.15), (5.16) and (5.18),

$$i+l(\tilde{\mathcal{O}}) \geq i+(u-i)+(e-u)+(h+k-m-p-q)=s+2.$$

Case 2b. $s=(n-3)+(2k+1)$. By (4.10), we have $d(a)\geq k$, $d(b)\geq k$. If $d(a)=d(b)=k$ then $a \div b$ and $b \div v$ are entry games (in the notation of Fig. 3) and we are back to Case 2a. Let us thus assume $d(a)+d(b)=2k+1$. There are two possibilities, $d(a)=k+1$, $d(b)=k$ or $d(a)=k$, $d(b)=k+1$.

Assume $d(a)=k+1$, $d(b)=k$. Then $b \div v$ is the entry game for b; let it be the u-th game. By (O1), we have $a \in T_u$. If $a >_u b$ and $d_u(a)=k$ then we are back in Case 2a. Let $a >_u b$ and $d_u(a)=k+1$. We use the notation of Fig. 4. By (5.12), we have $s-u=f+g+\delta$, i.e. all the remaining games in $\mathcal{O}$ must be between maximal elements of $N-\{a,b\}$. By employing the oracle $\mathcal{O}'$: $B\cup\{v\}<A<C<b<a$ we conclude that we may assume $f=\delta=0$, and thus $s-u=g\leq k-1$. Now, b lost to a in the k-th or $k+1$-st game involving a, let z be the other k-th or $k+1$-st loser. Then $\varphi_u(v)+\varphi_u(z)\geq 2^{k-2}+1$ and thus $\varphi_u(v)\geq 2^{k-3}+1$ from (O3) and the fact that $f=\delta=0$. Hence $l=d_u'(v)\geq k-2$ and by choosing the oracle $\mathcal{O}''$: $b<V<a<v$ in the notation of Fig. 5 we need at least $l+2\geq k$ games after the u-th game whence $u+l(\mathcal{O}'')\geq u+k\geq s+1$. If, on the other hand, $a \not>_u b$ (and therefore $d_u(a)=k$), then $a \div z$ must have been the entry game for a and P_{u+1} looks as in Fig. 13

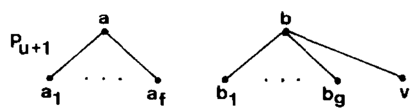


Fig. 13

where $s-u=f+g+\delta+1$. Note that $v \not<_u a$ by (5.5). By using the oracle $\mathcal{O}'''$:

$$N-(A\cup B\cup C\cup\{a,b,v\})<A<B<C<v<a<b$$

we need at least $f+g+\delta+2$ more games whence again $u+l(\mathcal{O}''')\geq s+1$.

Assume, finally, $d(a)=k$ and $d(b)=k+1$. Then $a \div b$ was the entry game for a. If $d_u(b)=k-1$, then $b \div v$ is the entry game for b and we are back in Case 2a. If $d_u(b)=k$, then by using the oracle $\mathcal{O}'$ after Fig. 4 we may again assume $f=\delta=0$ in Fig. 4. Let z be the loser to b in the k-th game of b. Since $\varphi_u(v)\leq 2^{k-2}$ by (O3ii) and $\delta=0$ we have

$$n=2^k+2^{k-2}+1=\varphi_u(a)+\varphi_u(b)+\varphi_u(v)+\varphi_u(z)$$

and thus $\varphi_u(z)\geq 1$. Hence $b \div z$ must have been the entry game for b and since v will eventually beat z we must have $\varphi_u(v) \geq 2^{k-3}+1$. But now we are back to Case 2a (Fig. 8) with z playing the role of v in the argument there. $\square$

This completes the proof of the theorem.

References

[1] M. Aigner, Producing posets, Discrete Math. 35 (1981) 1-15.
[2] A. Hadian and M. Sobel, Selecting the t-th largest using binary errorless comparisons, Coll. Math. Soc. János Bolyai (1969) 585-599.
[3] L. Hyafil, Bounds for selection, SIAM J. Comput. 5 (1976) 109-114.
[4] S.S. Kislitsyn, On the selection of the k-th element of an ordered set by pairwise comparisons, Sibirsk Math. Zk. 5 (1964) 557-564.
[5] D.E. Knuth, The Art of Computer Programming, Vol. 3, Sorting and Searching (Addison-Wesley, Reading, MA 1973).
[6] M. Sobel, On an optimal search for the t best using binary errorless comparisons: The selection problem, Technical Report No. 114, Dept. Stat. Univ. Minn. (1968).