ON SELECTING THE $t = 2$ BEST OF n ITEMS

USING BINARY COMPARISONS

by

Milton Sobel§
Technical Report No. 114
University of Minnesota, Minneapolis, Minnesota
and
Imperial College, London, England

§ Supported partially by a National Institutes of Health Special Fellowship while at Imperial College, London and partially by National Science Foundation Grant Number Gp-9018 through the University of Minnesota.

1. Introduction.

In a recent paper [5] the author has considered the problem of ranking the $t = 2$ best (i.e., the largest two) of n unequal numbers when only binary errorless comparisons are made; this paper considers the analogous problem of selecting the $t = 2$ best without ordering them. We are interested in two criteria: one is to minimize the expected number of comparisons required (called the E-criterion) and the other is to minimize the maximum number of comparisons required (called the M-criterion). Unlike the ranking problem which was considered by different authors, the selection problem appears not to have been previously considered; hence all the procedures discussed are new. The ideas behind some of the procedures and one method for obtaining an E-lower bound are similar to those used in [5]. The E- and M-efficiences of our procedures are numerically investigated.

In order to evaluate efficiency or prove optimality we need to develop an attainable lower bound over all possible procedures. The best M-lower bound is obtained; the E-lower bound obtained is only over a certain class of procedures. Our results (see table in Section 2) are optimal for $n \leq 5$. With the help of the above M-lower bound, one of the procedures R_M is shown to be M-optimal for all n.

This formulation is directly applicable to tournament problems and we use the associated terminology, i.e., the best player corresponds to the largest number, etc. The better player always wins and, since no two players have the same ability, a draw cannot occur.

2. Procedures for the Selection Problem for $t = 2$.

Six procedures for the selection problem with $t = 2$ are defined. Three of them use the concepts of pairing and one-step or higher-step expected

1. Introduction.

In a recent paper [illegible] the author has considered the problem of ranking the $t = 2$ best (i.e., the largest two) of n unequal numbers when only binary errorless comparisons are made; this paper considers the analogous problem of selecting the $t = 2$ best without ordering them. We are interested in two criteria; one is to minimize the expected number of comparisons required (called the E-criterion) and the other is to minimize the maximum number of comparisons required (called the M-criterion). Unlike the ranking problem which was considered by different authors, the selection problem appears not to have been previously considered; hence all the procedures discussed are new. The ideas behind some of the procedures and one method for obtaining an E-lower bound are similar to those used in [5]. The E- and M-efficiency of our procedures are numerically investigated.

In order to evaluate efficiency or prove optimality we need to develop an obtainable lower bound over all possible procedures. The best M-lower bound is obtained; the E-lower bound obtained is only over a certain class of procedures. Our results (see table in section 5) are optimal for $n \leq$ [illegible]. With the help of the above M-lower bound, one of the procedures [illegible] is shown to be M-optimal for all n.

This formulation is directly applicable to tournament problems and we use the associated terminology, i.e., the best player corresponds to the largest number, etc. The better player always wins and, since no two players have the same ability, a draw cannot occur.

2. Procedures for the Selection Problem for $t = 2$.

Six procedures for the selection problem with $t = 2$ are defined. Three of them use the concepts of pairing and one-step or higher-step expected entropy (the adjective expected is later deleted) and one is related to a procedure suggested by Picard [2] for the corresponding ordering problem. After some preliminary definitions, we briefly describe the procedures and give a table comparing the numerical results obtained for $n = 1(1)10$.

A state of nature (or case) is any one of the $n!$ possible ordered arrangements of the n players. There are, of course, $\binom{n}{2}$ possible decisions. At any stage of the procedure, we are concerned with the number of cases that are consistent with the results of comparisons already made for each of the $\binom{n}{2}$ decisions; these $J=\binom{n}{2}$ integers are proportional to the conditional probabilities that each of the $\binom{n}{2}$ decisions is the correct one given the results of the comparisons already made. Hence these J integers describe the state of our system, say S_α. Let the integers be $n_i^{(\alpha)}$ with sum $N^{(\alpha)}$ and let $p_i^{(\alpha)} = n_i^{(\alpha)}/N^{(\alpha)}$ $(i = 1, 2, \ldots, J)$ be the conditional probabilities given the system state S_α. The entropy (or uncertainty) associated with S_α is given by

$$(2.1) \qquad \mathcal{E}(S_\alpha) = -\sum_{i=1}^{J} p_i^{(\alpha)} \log p_i^{(\alpha)} ;$$

all logs in this paper are to the base 2 unless stated otherwise. If we start from S_α and a comparison $C = C(a \text{ v } b)$ (where v means versus) leads to states S_1 (resp., S_2) with probabilities $q_1^{(\alpha)}$ (resp., $q_2^{(\alpha)} = 1 - q_1^{(\alpha)}$), then the expected one-step reduction in entropy due to the comparison C, applied to the system state S_α, is given by

$$(2.2) \qquad E(\Delta \mathcal{E} | C, S_\alpha) = \mathcal{E}(S_\alpha) - \{q_1^{(\alpha)}\mathcal{E}(S_1) + q_2^{(\alpha)}\mathcal{E}(S_2)\}.$$

If we look s steps ahead then the expected reduction is again given by $\mathcal{E}(S_\alpha)$ minus the appropriate average of (at most) 2^s uncertainties. Our basic idea is to fix an s and find the comparison that maximizes the expected entropy (the adjective expected is later deleted) and the fourth is related to a procedure suggested by Picard [2] for the corresponding ordering problem. After some preliminary definitions, we briefly describe the four procedures and give a table comparing the numerical results obtained for $n = 1(1)11$.

A state of nature (or case) is any one of the $n!$ possible ordered arrangements of the n players. There are, of course, $\binom{n}{2}$ possible decisions. At any stage of the procedure, we are concerned with the number of cases that are consistent with the results of comparisons already made for each of the $\binom{n}{2}$ decisions; these $\binom{n}{2}$ integers are proportional to the conditional probabilities that each of the $\binom{n}{2}$ decisions is the correct one given the results of the comparisons already made. These $J = \binom{n}{2}$ integers describe the state of our system, say S_α. Let the integers be $n_i^{(\alpha)}$ with sum $N^{(\alpha)}$ and let $p_i^{(\alpha)} = n_i^{(\alpha)}/N^{(\alpha)}$ $(i = 1, 2, \ldots, J)$ be the conditional probabilities given the system state S_α. The entropy (or uncertainty) associated with S_α is given by

$$(2.1) \qquad \mathfrak{H}(S_\alpha) = -\sum_{i=1}^{J} p_i^{(\alpha)} \log p_i^{(\alpha)} ;$$

all logs in this paper are to the base 2 unless stated otherwise. If we start from S_α and a comparison $C = C(u\,v\,w)$ (where v means versus) leads to states S_1 (resp., S_2) with probabilities $p_1^{(\alpha)}$ (resp., $p_2^{(\alpha)} = 1 - p_1^{(\alpha)}$) then the expected one-step reduction in entropy due to the comparison C, applied to the system state S_α is given by

$$(2.2) \qquad R(C\,|\,S_\alpha) = \mathfrak{H}(S_\alpha) - p_1^{(\alpha)}\mathfrak{H}(S_1) - p_2^{(\alpha)}\mathfrak{H}(S_2) .$$

If we look s steps ahead then the expected reduction is again given by $\mathfrak{H}(S_\alpha)$ minus the appropriate average of (at most) 2^s uncertainties. Our basic idea is to fix an s and find the comparison that maximizes the expected s-step reduction in entropy at each stage. When the one-step plan (our procedure R_E) does not give optimal results, we investigate the improvement of a two-step plan by allowing the use of two-step reduction in a non-systematic manner (see the procedure R_E* below). It is conjectured that a systematic two-step plan would do at least as well, but this has not been proved.

The use of the expected reduction in entropy as a tool for search problems was used by Sobel and Groll [6] for group-testing, by F. Dubail [1] for other search problems and also by the present author in [5].

Another point of interest is the distinction between cycle pairing and complete pairing. For any n, let the binary structure of n be

(2.3) $$n = 2^{r_1} + 2^{r_2} + \ldots + 2^{r_s} \quad (r_1 > r_2 > \ldots > r_s \geq 0),$$

so that s is the number of ones in the binary notation for n. Let p be the highest power of 2 that factors into n!. Then it is easy to prove (see [5]) that $p = n - s$.

For $n = 2^r$ a knock-out tournament for finding the best one consists of r rounds where the number of contenders is halved at each round. Under complete pairing we start a procedure by randomly breaking up n into subsets of size 2^{r_i} as in (2.3) and doing a knock-out tournament within each of these subsets. After this, we use the comparison that maximizes the one-step expected reduction in entropy.

Under cycle pairing we start with a knock-out tournament only for one subset of size 2^s (usually $s = r_1$ defined in (2.3)) and then continue with the expected reduction in entropy. The procedure R_E* uses cycle pairing for n - 1 i.e., it uses cycle pairing with $s = r_1$ for $n \neq 2^r$ and for $n = 2^r$ we take $s = r-1$.

s-step reduction in entropy at each stage. Since the one-step plan (our procedure R_1) does not give optimal results, we investigate the improvement of a two-step plan by allowing the use of two-step reduction in a non-systematic manner (see the procedure $R*$ below). It is then presumed that a systematic two-step plan would do at least as well, but this has not been proven.

The use of the expected reduction in entropy as a tool for search problems was used by Sobel and Groll [6] for group-testing, by M. Sobel [1] for other search problems and also in [4]. [present author in [5]].

Another point of interest is the distinction between cycle and complete pairing. For any n, let the binary structure of n be

$$n = 2^{r_1} + 2^{r_2} + \dots + 2^{r_s} \quad (r_1 > r_2 > \dots > r_s \geq 0), \tag{2.3}$$

so that s is the number of ones in the binary notation for n. Let p be the highest power of 2 that factors into $n!$. Then it is easy to prove (see [3]) that $p = n - s$.

For $n = 2^r$ a knock-out tournament for finding the best one consists of r rounds where the number of contenders is halved at each round. Under complete pairing we start a procedure by randomly breaking up n into subsets of size 2^{r_i} as in (2.3) and doing a knock-out tournament within each of these subsets. After this, we use the comparison that maximizes the expected reduction in entropy.

Under cycle pairing, we start with a knock-out tournament only for one subset of size 2^{r} (usually $r = r_1$ defined in (2.3)) and then continue with the expected reduction in entropy. The procedure $R*$ uses cycle pairing for $n - 1$, i.e. it uses cycle pairing with $r = r_1$ for $n \neq 2^r$ and for $n = 2^r$ we take $r = r-1$.

- 7 -

Procedure R_A: Let $n = 2^r + c$ with $0 < c \leq 2^r$. This procedure uses cycle pairing with 2^s players with some $s \leq r$ (actually $s = r-1$ or r, the particular value to be determined later). After this we have one large connected set of size 2^s for the graph (or tree) and the remaining $n - 2^s$ players (called newcomers below) are unconnected, i.e., have not yet played. The connected set has a best player x_1 with i_1 inferiors and among these a contender for second best x_2 with the largest number of inferiors, say i_2 among the contenders for second best. Then $i_2 \leq i_1 - 1 \leq n - 2$ and our goal is to make $i_2 = n - 2$, which implies that $i_1 = n - 1$.

Each newcomer except for the last one (and possibly the one before that as explained in the ending E_1 below), comes up in turn (we assume they are in order) and plays x_2. If he loses he retires; if he wins he plays again, this time against x_1. If he loses to x_1 he takes over as the new x_2; if he wins against x_1 he becomes the new x_1 and the old x_1 becomes the new x_2.

Three different endings are used with this procedure, say E_1, E_2 and E_3, according to whether the original x_1 (after cycle pairing) is beaten by a newcomer not among the last two, he is beaten for the first time by the next-to-last newcomer n_1, or he is better than all the newcomers except possibly the last one n_0.

Under E_1 there is only 1 contender for second best (in the connected subset), namely x_2, when n_1 is ready to play. Again n_1 plays x_2 and retires if he loses, but if he wins he 'sits out' a game letting n_0 play x_2. Then if n_0 loses, he (i.e., n_0) retires and we are through. If n_0 wins, then x_2 retires and we need exactly 2 more games to find the best 2 of the 3 players, x_1, n_0 and n_1.

Procedure R_a: Let $n = 2^s + c$ with $0 < c \leq 2^s$. This procedure uses cycle, starting with 2^s players with some a [illegible] (actually a = [illegible] or [illegible], the part particular to be determined later). After this we have one large connected set of size 2^s for the graph (or tree) and the remaining $n - 2^s$ players (called newcomers below) are unconnected, i.e., have not yet played. The connected set has a best player x_1 with s inferiors and among these a contender for second best x_2 with the largest number of inferiors, say i_2, among the contenders for second best. Then $i_2 \leq i_1 - 1 \leq n - 2$ and our goal is to make $i_2 = n - 2$ which implies that $i_1 = n - 1$.

Each newcomer except for the last one (and possibly the one before that as explained in the ending R_1 below) comes in in turn (we assume they are in order) and plays x_2. If he loses he is out; if he wins he plays again, this time against x_1. If he loses to x_1 he takes over as the new x_2; if he wins against x_1 he becomes the new x_1 and the old x_1 becomes the new x_2.

Three different endings are used with this procedure, say R_1, R_2 and R_3, according to whether the original x_1 (after cycle pairing) is beaten by a newcomer not among the last two or is beaten for the first time by the next-to-last newcomer n_1, or he is better than all the newcomers except possibly the last one n_0.

Under R_1 there is only 1 contender for second best (in the connected subset), namely x_2, when n_1 is ready to play. Again n_1 plays x_2 and retires if he loses, but if he wins he sits out a game letting n_0 play x_2. Then if [illegible] n_0 (x_2) retires and we are through. If n_0 wins, then x_2 retires and we need exactly 2 more games to find the best 2 of the players x_1, n_0 and n_1.

Under E_2 there is only 1 contender for second best (in the connected subset), namely x_2, when n_0 is ready to play (here n_1 plays as usual). Then n_0 plays x_2 and the procedure terminates.

Under E_3 there are s contenders (including x_2) for second best (in the connected subset) when n_0 is ready to play. Again n_0 plays x_2. If he loses, he retires and we need exactly $s - 1$ more games to find the 2 best players. If he wins, he continues to play the other $s - 1$ contenders, eliminating one at each step, until he loses (at which point he retires) or he wins over all of them. If he loses and there are still c contenders then we need exactly $c - 1$ more games to complete the procedure.

To determine the value for s for the initial cycle pairing we use the exact formula for the expectation derived in Section 3 for procedure R_A and use the s that gives the smaller expectation. Thus for $5 \leq n \leq 10$ it can be verified that $s = 2$, or cyle pairing with $2^s = 4$ units, is best and for $n = 11$ we start cycle pairing with 8 units (see equation (3.20) for the asymptotic equivalent).

<u>Procedure R_E*</u>: Start with cycle pairing for $n - 1$ and then maximize the 1 step (or 2 step) expected reduction in entropy. Two-step reductions are sometimes used but not systematically.

<u>Procedure R_E</u>: This consists of a 'pure' strategy for maximizing the one-step expected reduction in entropy. Here two-step reductions are used only when different comparisons give the same entropy reduction in 1 step.

<u>Procedure R_{CP}</u>: Start with complete pairing (for n) and then continue with the strategy of maximizing the one-step expected reduction in entropy.

Procedure R_M: For this procedure we use 'complete pairing for $n - 1$' and for $n \geq 3$ make use of the binary expansion of $n - 1$

$$n - 1 = 2^{r_1} + 2^{r_2} + \ldots + 2^{r_s} \quad (r_1 > r_2 > \ldots > r_s \geq 0) \tag{2.4}$$

to explain the different steps in the procedure, after putting one unit (or player) aside until the very last comparison.

1. Find the best one separately in each of the subsets for which $r_i > 0$.

2. Play the best one of the smallest subset of size 2^{r_s} against the best of the second smallest, the best of these two against the best of the third smallest subset, etc., until the best one of $n - 1$ is determined. Let c denote the number of contenders obtained for second best.

3. Use any knock-out tournament with exactly $c - 1$ games to determine the second best of $n - 1$.

4. Play the second best against the one set aside to complete the procedure.

Procedure R_P: Let the players in random order be denoted by $1, 2, \ldots, n$; we describe the scheme in three steps:

1. Play 1 v 2 and assume 1 loses to 2.

2. Play 3 v 1. If 3 loses then he is removed from contention; if 3 wins then 1 is removed and we play 3 v 2 to reestablish an ordering between the top two contenders.

3. Repeat this procedure, except that if the last player n wins then the extra game to reestablish the order is not played.

Although the procedure R_P appears to be inadmissible in the sense that one of the other procedures is at least as good or better for every n (for

Procedure K_{II}: For this procedure we use [illegible] late pairing for $n - 1$, and for $n \geq 2$ make use of the binary expansion of $n - 1$.

$$n - 1 = 2^{k_1} + 2^{k_2} + \dots + 2^{k_s} \quad (k_1 > k_2 > \dots > k_s \geq 0) \tag{2.5}$$

To explain the different steps in the procedure, after putting one man (or player) aside until the very last comparison:

1. Find the best one separately in each of the subsets for which $k_i > 0$.
2. Play the best one of the smallest subset of size 2^{k_s} against the best of the second smallest, the best of these two against the best of the third smallest subset, etc., until the best one of $n - 1$ is determined. Let c denote the number of contenders for second best place.
3. Use any knock-out tournament with exactly $c - 1$ games to determine the second best of $n - 1$.
4. Play the second best against the one set aside to complete the procedure.

Procedure K_I: Let the players in random order be denoted by 1, 2, ..., n; we describe the scheme in three steps:

1. Play 1 v 2 and assume 1 loses to 2.
2. Play 3 v 1. If 3 loses then he is removed from contention; if 3 wins then 1 is removed and we play 3 v 2 to establish an ordering between the top two contenders.
3. Repeat this procedure, except that if the last player n wins then the extra game to establish the order is not played.

Although the procedure K_I appears to be inadmissible in the sense that one of the other procedures is at least as good or better for every (for both the E-problem and the M-problem), the simplicity of the procedure enables us to get specific formulas, which throws light on the asymptotic properties of the other procedures.

Numerical Results for Six Procedures for Selecting the t = 2 Best Players Using Only Binary Errorless Comparisons

Bounds and Procedures	Expected Values $E\{T\|R\}$								
	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
LBE#	---	1.918	2.918	4.522	5.174	5.775	6.095	8.837	9.292
CLB##	---	---	3.386	4.685	5.899	7.066	8.197	9.307	10.401
R_A	---	2	3 $\underline{4}$	5 $\underline{0}$	6 $\underline{8}$	7 $\underline{309}$	8 $\underline{144}$	9 $\underline{121}$	11 $\underline{38}$
R_E*	---	2	3 $\underline{4}$	5 $\underline{0}$	6 $\underline{8}$	7 $\underline{315}$	8 $\underline{149}$	N.C.	N.C.
R_E	---	2	3 $\underline{5}$	5 $\underline{0}$	6 $\underline{10}$	7 $\underline{550}$	N.C.	N.C.	N.C.
R_{CP}	---	2	4 $\underline{0}$	5 $\underline{0}$	6 $\underline{10}$	7 $\underline{550}$	9 $\underline{0}$	10 $\underline{0}$	11 $\underline{336}$
R_M	---	2	3 $\underline{4}$	5 $\underline{0}$	6 $\underline{18}$	7 $\underline{420}$	8 $\underline{180}$	10 $\underline{0}$	11 $\underline{840}$
R_P	---	2	3 $\underline{4}$	5 $\underline{1}$	6 $\underline{17}$	7 $\underline{567}$	9 $\underline{39}$	10 $\underline{61}$	11 $\underline{829}$
D§§	---	---	6	6	30	630	210	140	1260
	Maximum Length $M\{T\|R\}$								
LBM§	---	2	4	5	7	8	9	10	12
R_A	---	2	4	5	7	9	11	13	15
R_E*	---	2	4	5	7	9	11	N.C.	N.C.
R_E	---	2	4	5	7	9	N.C.	N.C.	N.C.
R_{CP}	---	2	4	5	7	9	9	10	12
R_M	---	2	4	5	7	8	9	10	12
R_P	---	2	4	6	8	10	12	14	16

The LBE is the lower bound (3.14) for all procedures that use 'cycle pairing for n-1' defined in the text.

The CLB is the conjectured lower bound (3.22) for all procedures. Since the LBE-values are smaller, they are also conjectured to hold for all procedures.

§ The LBM is the best lower bound for the maximum branch length given on the right sides of (3.23) and (3.25).

§§ Each D is the common denominator for all the underlined numerators above it.

3. Properties and Bounds.

Since procedure R_P gives easy results, we consider it first. Let $f_P(n) = E\{T|R_P\}$ denote the expected number of comparisons and let $\overline{f}_P(n)$ denote the maximum length under R_P. It is easy to see that the expected number of games for the j^{th} player $(3 \le j \le n - 1)$ is $1 + \frac{2}{j}$, while for $j = n$, only one game is played. It follows that for $n \ge 3$

$$(3.1) \qquad f_P(n) = n - 1 + \sum_{j=3}^{n-1} 2/j = n - 4 + 2 \sum_{j=1}^{n-1} 1/j \approx n + 2 \ln n.$$

Clearly if players 3, 4,..., n-1 all win we obtain the maximum length; hence for $n \ge 3$

$$(3.2) \qquad \overline{f}_P(n) = 2n - 4.$$

Although R_P has a better expectation than R_E for small powers of 2 (see $n = 4$), the maximum length grows very rapidly compared to that of R_E.

Let $f_M(n) = E\{T|R_M\}$ denote the expected number of comparisons under R_M and let $\overline{f}_M(n)$ denote the maximum number under R_M. Using the result for R_M in [5] for the ordering problem with n replaced by n - 1, and adding 1, gives

$$(3.3) \qquad \overline{f}_M(n) = n - 1 + [\log (n - 2)]$$

and

$$(3.4) \qquad f_M(n) = n - 2 + \frac{1}{n-1} \sum_{j=1}^{s} (r_j + j - \delta_{js}) 2^{r_j}$$

where the r_j are now defined by (2.4) and $\delta_{js} = 1$ if $j = s$ and $= 0$ otherwise. The asymptotic analysis in [5] is also appropriate here. It is shown below that (3.3) is the smallest possible maximum branch length and hence the procedure R_M is M-optimal.

3. Properties and Bounds.

Since procedure P_T gives easy results, we consider it first. Let $e_T(n) = E[Z_T]$ denote the expected number of comparisons and let $\bar{L}_T(n)$ denote the maximum length under P_T. It is easy to see that the expected number of games for the j^{th} player $(3 \leq j \leq n - 1)$ is $1 + \frac{1}{j}$, while for $j = n$, only one game is played. It follows that for $n \geq 3$

$$(3.1) \qquad e_T(n) = n - 1 + \sum_{j=3}^{n-1} (1/j) = n - \frac{3}{2} + \sum_{j=1}^{n-1} 1/j \approx n + \ln n.$$

Clearly if players $3, 4, \ldots, n-1$ all win we obtain the maximum length; hence for $n \geq 3$

$$(3.2) \qquad \bar{L}_T(n) = 2n - 4.$$

Although P_T has a better expectation than P_K for small powers of 2 (see $n = 4$), the maximum length grows very rapidly compared to that of P_K.

Let $e_M(n) = E[Z_M]$ denote the expected number of comparisons under P_M and let $\bar{L}_M(n)$ denote the maximum number under P_M. Using the result for P_K in [5] for the ordering problem with n replaced by $n - 1$, and adding 1, gives

$$(3.3) \qquad \bar{L}_M(n) = n - 1 + [\log (n - 2)]$$

and

$$(3.4) \qquad e_M(n) = n - 2 + \frac{n-1}{2} \sum_{j=1}^{n} (\alpha_j + 1 - \delta_{jn})/x_j$$

where the α_j are now defined by (4.4) and $\delta_{jn} = 1$ if $j = n$ and $= 0$ otherwise. The asymptotic analysis in [5] is also appropriate here. It is shown below that (3.3) is the smallest possible maximum branch length and hence the procedure P_M is the best.

Since we have a special interest in procedures that use cycle pairing for $n - 1$ we let R_C denote any such procedure and study its properties.

For $n = 2^r + 1$ the cycle-pairing and complete pairing are the same and after $n - 2$ comparisons we have 1 player (say n) that has never played and for the remaining 2^r we have a best player (say $n - 1 = 2^r$) and exactly r contenders for 2nd best. If player n does not play in the next comparison then two of the r contenders play and exactly one is eliminated in each such game. If all of these contenders but 1 is removed before player n plays then the remaining one plays against player n for the last comparison; this takes exactly r games after the pairings. If player n comes in earlier under R_E (actually for small r he plays first under R_E) then it still takes exactly r comparisons to complete the procedure. If player n wins then he continues to play each of the contenders for 2nd best until he loses. Since 1 contender is removed as a result of each comparison, it again takes exactly r comparisons after the pairings. Thus for $n = 2^r + 1$ and any procedure R_C that uses 'cycle pairing for $n - 1$' we have

$$(3.5) \qquad E\{T|R_C\} = 2^r + r - 1 = M\{T|R_C\}.$$

This result is analogous to (4.3) in [5]; it is proved to be M-optimal in (3.23) below, but the table above shows that it is not E-optimal.

We could also have gone through a cycle-pairing procedure for $n = 2^r$ and obtained the result $2^r + r - 2$ as in (4.3) of [5] but, as our table shows, this result can be improved upon.

Clearly any of the attained values for the $t = 2$ ordering problem in [5] can be used as an upper bound for the optimal procedure in the $t = 2$ selection problem.

Since we have a special interest in procedures that use cycle pairing for $n - 1$ we let R_C denote any such procedure and study its properties.

For $n = 2^r + 1$ the cycle-pairing and complete pairing are the same and after $n - 2$ comparisons we have 1 player (say n) that has never played and for the remaining 2^r we have a best player (say $n - 1 = 2^r$) and exactly r contenders for 2nd best. If player n does not play in the next comparison then two of the r contenders play and exactly one is eliminated in each such game. If all of these contenders but 1 is removed before player n plays then the remaining one plays against player n for the last comparison; this takes exactly r games after the pairings. If player n comes in earlier under R_C (actually for small r he plays first under R_C) then it still takes exactly r comparisons to complete the procedure. If player n wins then he continues to play each of the contenders for 2nd best until he loses. Since 1 contender is removed as a result of each comparison, it again takes exactly r comparisons after the pairings. Thus for $n = 2^r + 1$ and any procedure R_C that uses cycle pairing for $n - 1$ we have

$$E\{T|R_C\} = 2^r + r - 1 = \text{Max}\,\{T|R_C\}. \qquad (4.5)$$

This result is analogous to (4.3) in [1]; it is proved to be M-optimal in (4.2) below, but the table above shows that it is not E-optimal.

We could also have gone through a cycle-pairing procedure for $n = 2^r$ and obtained the result $2^r + r - 2$ as in (4.4) of [1] but, as our table shows, this result can be improved upon.

Clearly any of the attained values for the $t = 2$ ordering problem in [1] can be used as an upper bound for the optimal procedure in the $t = 1$ selection problem.

We define a comparison $C(a \text{ v } b)$ to be of level $j (j = 1, 2\ldots, [\log n])$ and denote it by C_j if the two players a and b each have exactly $2^{j-1} - 1$ inferiors, the two sets of inferiors are disjoint and neither a or b has any proven superiors. We want to prove a result about the reduction in entropy for any comparison C_j of level j.

Lemma: If two players a, b, with no proven superiors, are best in disjoint subsets S_a, S_b respectively, with common size I for each, then the expected reduction in entropy $E\{\Delta\}$ due to the comparison a v b is given by

$$E\{\Delta\} = \frac{2I(2n-1-3I)}{n(n-1)}, \tag{3.6}$$

regardless of any knowledge previously obtained that affects only the relative ordering of the remaining $n - 2I$ players.

Proof: Suppose that the two players a and b each have exactly $I - 1$ inferiors with no overlap (in our application $I = 2^{j-1}$) and, to begin with, there is no other previous information.

Those states of nature which can still be correct are partitioned into subsets so that each subset corresponds to one of the possible true decisions. The decisions $D(x, y)$ corresponding to these subsets will be grouped into a convenient table, and the number of cases given for each, before we play a v b. Let $S_\alpha (\alpha = 1, 2, \ldots, m)$ denote the seconds (or immediate inferiors) of a in the connected set of size I and let M_α denote the number of ways of linearizing this set of size I with S_α in second position (consistent with all known order relations); let $M = M_1 + M_2 + \ldots + M_m$. Then $S'_\beta (\beta = 1, 2, \ldots, m')$, M'_β and M' are defined similarly for player b. Let p (or p_i) denote any one of the $n - 2I$ players not in these two connected subsets.

We define a comparison $G(a \vee b)$ to be of level j ($j = 1, 2, \ldots, [\log n]$) and denote it by G_j if the two players a and b each have exactly $2^{j-1} - 1$ inferiors, the two sets of inferiors are disjoint and neither a or b has any proven superiors. We want to prove a result about the reduction in entropy for any comparison G_j of level j.

Lemma: If two players a, b, of the n given players are best in disjoint subsets S_a, S_b, respectively, with common size I for each, then the expected reduction in entropy $H(I)$ due to the comparison $a \vee b$ is given by

$$H(I) = \frac{n(n-I)}{2I(2n-1-2I)} . \qquad (3.6)$$

regardless of any knowledge previously obtained that affects only the relative ordering of the remaining $n - 2I$ players.

Proof: Suppose that the two players a and b each have exactly $I - 1$ inferiors with no overlap (in our application $I = 2^{j-1}$) and, to begin with, there is no other previous information.

The states of nature which we can still be correct are partitioned into subsets so that each subset corresponds to one of the possible true decisions. The decisions $D(\pi, \pi')$ corresponding to these subsets will be grouped into a convenient table, and the number of cases given for each, before we play $a \vee b$. Let S_α^a ($\alpha = 1, 2, \ldots, m$) denote the records (or immediate inferiors) of a in the connected set of size I and let M_α denote the number of ways of linearizing this set of size I with S_α^a in second position (consistent with all known order relations); let $M = M_1 + M_2 + \ldots + M_m$. Then S_β^b ($\beta = 1, 2, \ldots, m'$), M_β' and M' are defined similarly for player b. Let p (or p_i) denote any one of the $n - 2I$ players not in these two connected subsets.

	Type of Decision	Number of Each Type	Number of Cases per Decision
1.	$D(a, b)$	1	$2MM' \frac{(n-2)!}{(I-1)!(I-1)!}$
2.	$D(a, s_\alpha)$	$(\alpha = 1, 2, \ldots, m)$	$M_\alpha M' \frac{(n-2)!}{I!(I-2)!}$
3.	$D(a, p)$	$n - 2I$	$2MM' \frac{(n-2)!}{I!(I-1)!}$
(3.7) 4.	$D(b, s_\beta)$	$(\beta = 1, 2, \ldots, m')$	$MM'_\beta \frac{(n-2)!}{I!(I-2)!}$
5.	$D(b, p)$	$n - 2I$	$2MM' \frac{(n-2)!}{I!(I-1)!}$
6.	$D(p_1, p_2)$	$\binom{n - 2I}{2}$	$2MM' \frac{(n-2)!}{I!\ I!}$

The uncertainty before playing $a \vee b$ can now be obtained directly from the above table. The total number of cases is easily checked to be $n!\ MM'/(I!)^2$ and with this as denominator and the number of cases above as numerator, we have the probability for each decision type.

We now play $a \vee b$ and suppose $a > b$. Then types 2 and 3 in the table remain in their entirety and we also have half the number of cases in type 1 and half for each decision of type 6. Since types 3 and 5 have the same number of cases and the sum over α for type 2 equals the sum over β for type 4, it follows that the comparison $a \vee b$ partitions the entire set of cases exactly in half. Hence we use a simple average for finding the expected uncertainty after playing $a \vee b$.

The uncertainty U_1 after finding that $a > b$ is given by

$$(3.8) \qquad U_1 = \frac{2I(I-1)}{n(n-1)} \sum_{\alpha=1}^{m} \frac{M_\alpha}{M} \log \frac{M\, n(n-1)}{2M_\alpha I(I-1)} + (n - 2I) \frac{4I}{n(n-1)} \log \frac{n(n-1)}{4I}$$
$$+ \frac{2I^2}{n(n-1)} \log \frac{n(n-1)}{2I^2} + \binom{n-2I}{2} \frac{2}{n(n-1)} \log \frac{n(n-1)}{2}$$

and we omit the corresponding U_2 after finding that $b > a$. The orginal uncertainty U_o from the above table is

Type of Decision	Number of Each Type	Number of Cases per Decision
1. $D(a, b)$	1	$NM \frac{(n-2)!}{(N-1)!(M-1)!}$
2. $D(a, g_i)$	$(i = 1, \dots, m)$	$N_i M \frac{(n-2)!}{(N-1)!\,M!}$
3. $D(a, b)$	$n - 2$	$2NM \frac{(n-3)!}{(N-1)!\,M!}$
(3.4) 4. $D(b, g_i)$	$(i = 1, \dots, m)$	$N_i M \frac{(n-2)!}{(N-1)!\,M!}$
5. $D(b, c)$	$n - 2$	[illegible]
6. $D(g_i, g_j)$	[illegible]	[illegible]

The uncertainty before playing a v b can now be obtained directly from the above table. The total number of cases is easily found to be $n!\,NM/(N!)$, and with this as denominator and the number of cases above as numerator, we have the probability for each decision type.

We now play a v b and suppose a > b. Then types 4 and 5 in the table remain in their entirety and we also have half the number of cases in type 1 and half for each decision of type 6. Since types 2 and 3 have the same number of cases and the sum over all of type 2 equals the same over type 3, it follows that the comparison a v b partitions the entire set of cases exactly in half. Hence we use a simple average for finding the expected uncertainty after playing a v b.

The uncertainty U_1 after finding that a > b is given by

$$U_1 = \frac{n(n-1)}{2N(N-1)} \sum_{i=2}^{M} \frac{N_i}{N} \log \frac{2N(N-1)}{N_i\, n(n-1)} + (n-2)\frac{n(n-1)}{2N} \log \frac{2N}{n(n-1)} \quad (3.5)$$

$$+ \frac{n(n-1)}{2N^2} \log \frac{N^2}{n(n-1)} + (N-2n)\frac{n(n-1)}{N^2} \log \frac{2N}{n(n-1)} +$$

and we omit the corresponding U_1' after finding that b > a. The original uncertainty U_0 from the above table is

$$(3.9) \qquad U_o = \frac{I(I-1)}{n(n-1)} \sum_{\alpha=1}^{m} \frac{M_\alpha}{M} \log \frac{M\, n(n-1)}{M_\alpha I(I-1)} + \frac{I(I-1)}{n(n-1)} \sum_{\beta=1}^{m'} \frac{M'_\alpha}{M'} \log \frac{M' n(n-1)}{M'_\beta\, I(I-1)}$$

$$+ \; 2(n-2I)\, \frac{2I}{n(n-1)} \log \frac{n(n-1)}{2I} + \frac{2I^2}{n(n-1)} \log \frac{n(n-1)}{2I^2}$$

$$+ \; \binom{n-2I}{2} \frac{2}{n(n-1)} \log \frac{n(n-1)}{2} \; .$$

To obtain the expected reduction in uncertainty we first remove $\log \frac{1}{2} = -1$ from each term in U_1 and U_2 and we then easily obtain from (3.8) and (3.9)

$$(3.10) \qquad U_o - \left(\frac{U_1+U_2}{2}\right) = 1 - \frac{2I^2}{n(n-1)} - \frac{(n-2I)(n-1-2I)}{n(n-1)}$$

$$= \frac{2I(2n-1-3I)}{n(n-1)} \; .$$

If there is present previous knowledge about some of the remaining $n - 2^j$ players (say, among r of them) then we first consider each possible fixed relative ordering among these r players separately. Each of the numbers in the third column above is affected (we actually divide by $r!$) but this constant $r!$ also divedes the total. Hence the probabilities and the uncertainty is unchanged. Since the previous knowledge about these r players can be written as a union of such fixed relative orders among them, we now average our result over all the relative orders of these r players consistent with our previous knowledge and obtain the same result; this proves the lemma.

For the application of this lemma that is needed, we have the

<u>Corollary</u>: The expected reduction in entropy $E\{\Delta_j\}$ due to any j^{th} level comparison $C_j(a \vee b)$ for $1 \leq j \leq [\log n]$ is given by

$$(3.11) \qquad E\{\Delta_j\} = \frac{2^j(2n-1-3\cdot 2^{j-1})}{n(n-1)} \; ,$$

regardless of any knowledge previously obtained that affects only the relative ordering of the remaining $n - 2^j$ players.

We now use the lemma above to obtain lower bounds for all procedures that start with any given pairing scheme. We consider only the 'cycle-pairing for $n-1$' schemes since they give the better results, but the method can also be applied to complete-pairing schemes, cycle-pairing schemes (for n), ordinary-pairing schemes, etc. For $n = 2^r + c$ $(0 < c \leq 2^r)$ any cycle-pairing procedure R_c has at least 2^{r-i} comparisons of level $i (i = 1, 2, \ldots, r-1)$ and we assume it has exactly that many among the first $2^r - 1$ comparisons. Then the (expected) reduction in entropy due to these comparisons is, by (3.6),

$$(3.12) \qquad Q = \sum_{j=1}^{r} \frac{2^{r-j}2^j(2n-1-3\cdot 2^{j-1})}{n(n-1)} = \frac{2^r\{(2n-1)r-3(2^r-1)\}}{n(n-1)} .$$

Let the total number of comparisons T be partitioned into T_1 and T_2 where $T_1 = 2^r - 1$ are the pairings and $T_2 = T - T_1$ are the remaining comparisons under procedure R_C. Since the total uncertainty at the outset is $\log \binom{n}{2}$ and 1 is an upper bound for the reduction in entropy in any one step, it follows that

$$(3.13) \qquad Q + 1\cdot E\{T_2\} \geq \log \binom{n}{2}.$$

Hence, using (3.13), we obtain the desired lower bound for the expectation under any cycle-pairing procedure R_C with $n > 2^r$

$$(3.14) \qquad E\{T|R_C\} = 2^r - 1 + E\{T_2\} \geq 2^r - 1 + \log \binom{n}{2} - \frac{2^r\{(2n-1)r-3(2^r-1)\}}{n(n-1)} .$$

We now use the lemma above to obtain lower bounds for all procedures that start with any given pairing scheme. We consider only the cycle-pairing for $n-1$ schemes since they give the better results, but the method can also be applied to complete-pairing schemes, cycle-pairing schemes (for n), ordinary-pairing schemes, etc. For $n = 2^r + c$ $(0 < c \leq 2^r)$ any cycle-pairing procedure P_C has at least 2^{r-k} comparisons of level k $(k = 1, 2, \ldots, n-1)$ and we assume it [illegible] that many among the first $2^r - 1$ comparisons. Thus the (expected) reduction in entropy due to these comparisons is, by (3.6),

$$J = \sum_{k=1}^{r} \frac{2^{r-k} [illegible]}{[illegible]} = \frac{[illegible]}{n(n-1)} . \qquad (3.14)$$

Let the total number of comparisons T be partitioned into T_1 and T_2, where $T_1 = 2^r - 1$ are the pairings and $T_2 = T - T_1$ are the remaining comparisons under procedure P_C. Since the total uncertainty at the outset is $\log \binom{n}{2}$ and 1 is an upper bound for the reduction in entropy in any one step, it follows that

$$J + 1 \cdot T_2 \geq \log \binom{n}{2} . \qquad (3.15)$$

Hence, using (3.15), we obtain the desired lower bound for the expectation under any cycle-pairing procedure P_C with $n > 2$:

$$E\{T \mid P_C\} = 2^r - 1 + E\{T_2\} \geq 2^r - 1 + \log \binom{n}{2} - \frac{[illegible]}{n(n-1)} . \qquad (3.16)$$

For $n = 2^r + 1$ we obtain an improvement over (3.14) by using (3.3). For $n = 2^r$ we can use the same result (3.14) provided we replace r by $r - 1$. The values of (3.14) are given as the LBE in the table in Section 2. It follows from the above construction that the LBE is strictly increasing in n. Asymptotically $(n \to \infty)$, the value of (3.14) is between $(n/2) + \log n$ (the limit for $n = 2^r$ as $r \to \infty$) and $n + (\log n)/n$ (the limit for $n = 2^r + 1$ as $r \to \infty$).

We now derive an exact formula for the expectation under procedure R_A. Let 2^s denote the number of players involved in the initial cycle-pairing. If $n = 2^r + c$ (with $0 \leq c < 2^r$) then s will usually be $r - 1$ or r; the value of n where it changes from $r - 1$ to r is close to $3 \cdot 2^{r-1}$.

The probability that the q^{th} newcomer is the first one to beat the original x_1 is easily shown to be

$$(3.15) \qquad \frac{1}{2^s+q} \prod_{\alpha=0}^{q-2} \left(\frac{2^s+\alpha}{2^s+\alpha+1}\right) = \frac{2^s}{(2^s+q-1)(2^s+q)} \quad (q = 1, 2, \ldots, n-2^s)$$

and the sum of these is $1 - 2^s/n$. It follows that the probability that x_1 is beaten by the last newcomer or by no one is $2^s/(n-1)$. Conditional on the event that the q^{th} newcomer is the first to beat x_1, we utilize one extra comparison (beyond the basic one needed for each newcomer) under endings E_1 and E_2 for earlier newcomers n_α $(1 \leq \alpha \leq q-1)$ with probability $(2^{s-1} + \alpha)^{-1}$ and for later newcomers n_β, not including the last two, (i.e., for $q + 1 \leq \beta \leq n - 2 - 2^s$) with probability $2(2^s + \beta)^{-1}$. In addition the q^{th} newcomer also uses 1 extra comparison to beat x_1. For $\beta = n - 1 - 2^s$ (and $n \geq 2^s + 3$) we add 2 extra comparisons only if n_0 beats x_2, which then has probability $3/n$. Given ending E_3, which, as stated above, has probability $2^s/(n-1)$, we have to play $s - 1$ extra games after n_0 plays x_2 and each of the previous newcomers n_α $(1 \leq \alpha \leq n-1-2^s)$

For $n = 2^r + 1$ we obtain an improvement over (1.13) by using (1.4). For $n = 2^r$ we can use the same result (1.14) provided we replace r by $r - 1$. The values of (1.14) are given as the LBN in the table in Section 2. It follows from the above construction that the LBN is strictly increasing in n. Asymptotically $(n \to \infty)$, the value of (1.14) is between $(n/2) + \log n$ (the limit for $n = 2^r$ as $r \to \infty$) and $n + (\log n)/n$ (the limit for $n = 2^r + 1$ as $r \to \infty$).

We now derive an exact formula for the expectation under procedure T_1. Let 2^s denote the number of players involved in the initial cycle-pairing. If $n = 2^r + c$ (with $0 \le c < 2^r$) then s will usually be $r - 1$ or r; the value of n where it changes from $r - 1$ to r is close to $3 \cdot 2^{r-1}$.

The probability that the q^{th} newcomer is the first one to beat the original n_1 is easily shown to be

$$\frac{1}{2^s+q}\prod_{m=1}^{q-2}\left(\frac{2^s+m}{2^s+m+1}\right) = \frac{2^s}{(2^s+q-1)(2^s+q)} \quad (q = 1, 2, \ldots, n-2^s) \qquad (1.15)$$

and the sum of these is $1 - 2^s/n$. It follows that the probability that n_1 is beaten by the last newcomer or by nobody is $2^s/(n-1)$. Conditional on the event that the q^{th} newcomer is the first to beat n_1, we utilize one extra comparison (beyond the basic one needed for each newcomer) under endings E_1 and E_2 for earlier newcomers n_i $(1 \le i \le q-1)$ with probability $(2^{s-1} + i)^{-1}$ and for later newcomers n_i not including the last two, (i.e., for $q + 1 \le i \le n - 2 - 2^s$) with probability $2(2^s + i)^{-1}$. In addition the q^{th} newcomer also uses 1 extra comparison to beat n_1. For $i = n - 1 - 2^s$ (and $q \le i + 1$) we add 2 extra comparisons only if n_1 beats n_2, which then has probability $2/n$. Given ending E_1 which, as stated above, has probability $2^s/(n-1)$, we have to play $s - 1$ extra games after n_1 plays n_2 and each of the previous newcomers n_i $(1 \le i \le n-1-2^s)$

plays an extra game with probability $(2^{s-1}+\alpha)^{-1}$. If we add these extra games to the set of $n - 2^s$ basic games, one for each newcomer, and the 2^s-1 comparisons used in the initial pairing, then we obtain

$$(3.16)\qquad E(T|R_A) = (2^s-1) + (n-2^s) + \frac{2^s}{n-1}\{s-1 + \sum_{\alpha=1}^{n-1-2^s} \frac{1}{2^{s-1}+\alpha}\}$$

$$+ \sum_{q=1}^{n-1-2^s} (1 + \sum_{\alpha=1}^{q-1} \frac{1}{2^{s-1}+\alpha} + \sum_{\beta=q+1}^{n-2-2^s} \frac{2}{2^s+\beta}) \frac{2^s}{(2^s+q)(2^s+q-1)}$$

$$+ 2(\tfrac{3}{n})(\tfrac{2}{n-1}) \sum_{q=1}^{n-2-2^s} \frac{2^s}{(2^s+q)(2^s+q-1)} .$$

In (3.16) we make extended use of the elementary identity

$$(3.17)\qquad \sum_{i=a+1}^{n-b} \frac{1}{(c+i+1)(c+i)} = (\frac{1}{c+a} - \frac{1}{c+n-b})h_n(a+b+1)$$

where $h_n(x) = 1$ for $n \geq x$ and $= 0$ otherwise; in particular each of the double sums can be summed or simplified by (3.17). After some straightforward algebra and simplification we obtain

$$(3.18)\qquad E\{T|R_A\} = n - 1 + 2^s(\frac{s-1}{n-1}) + 2 \sum_{\alpha=1}^{n-2-2^s} \frac{1}{\alpha+2^{s-1}}$$

$$+ h_n(2+2^s)\{\frac{n2^s}{(n-1)(n-2)} + \frac{12(n-2-2^s)}{n(n-1)(n-2)} + \frac{2^s}{(n-1)(n-1-2^{s-1})} - 1\}.$$

Under 'cycle-pairing for n-1' with $n = 2^r + 1$ we set $s = r$ and we easily obtain $n + r - 2$ in agreement with (3.5).

Asymptotically $(n \to \infty)$ we obtain the maximum and minimum of (3.18) by searching for the value of n for which $g(r-1, n) = g(r, n)$ where $g(s, n)$ is the right side of (3.18). In fact we obtain from (3.18) the quadratic equation

$$(3.19)\qquad n^2 \ln 2 - n\, 2^{s-2}(s + 2 \ln 2) + (s + 1)2^{2s-3} \approx 0$$

plays an extra game with probability $(2^{s-1}+i)^{-1}$. If we add these extra games to the set of $n - 2^s$ basic games one for each nonwinner, and the 2^s-1 comparisons used in the initial pairing, then we obtain

$$(3.16) \qquad E(T|K_A) = (2^s-1) + (n-2^s) + \frac{2^s}{n-1}\Big(s-1 + \sum_{i=1}^{n-1-2^s} \frac{1}{2^{s-1}+i}\Big)$$

$$+ \sum_{q=1}^{n-1-2^s}\Big(1 + \sum_{i=1}^{q-1}\frac{1}{2^{s-1}+i} + \sum_{i=q+1}^{n-2-2^s}\frac{2}{2^s+i}\Big)\frac{2^s}{(2^s+q)(2^s+q-1)}$$

$$+ 2\Big(\frac{n}{\cdot}\Big)\Big(\frac{2}{n-1}\Big)\sum_{q=1}^{n-2-2^s}\frac{2^s}{(2^s+q)(2^s+q-1)} .$$

In (3.16) we make repeated use of the elementary identity

$$(3.17) \qquad \sum_{i=a+1}^{n-b} \frac{1}{(c+i+1)(c+i)} = \Big(\frac{1}{c+a} - \frac{1}{c+n-b}\Big)h(n-a-b-1)$$

where $h(x) = 1$ for $x \geq 0$ and $= 0$ otherwise; in particular each of the double sums can be summed or simplified by (3.17). After some straightforward algebra and simplification we obtain

$$(3.18) \qquad E[T|K_A] = n - 1 + 2^s\Big(\frac{s+1}{n-1}\Big) + 2\sum_{i=1}^{n-2-2^s}\frac{1}{2^{s-1}+i}$$

$$+ h_n(2^{s-1})\Big[\frac{n2^s}{(n-1)(n-2)} + \frac{12(n-2-2^{s-1})}{n(n-1)(n-2)} + \frac{2^s}{(n-1)(n-1-2^{s-1})} - 1\Big].$$

Under cycle-pairing for $n-1$, with $n = 2^r + 1$ we set $s = r$ and we easily obtain $n + r - 2$ in agreement with (3.[illegible]).

Asymptotically $(n \to \infty)$ we obtain the minimum value of (3.18) by searching for the value of n for which $g(r-1, n) = g(r, n)$, where $g(s, n)$ is the right side of (3.18). In fact we obtain from (3.18) the quadratic equation

$$(3.19) \qquad n^2 \ln 2 - n\,2^{s-2}(s + 2\ln 2) + (s+1)2^{2s-[illegible]} = 0$$

where $\ln x$ is the natural logarithm of x. Since we are looking for a final change for s as n increases, we take the largest root for n in (3.19) and obtain

$$(3.20) \qquad n \approx 2^{s-2} \left(\frac{s}{\ln 2} + \frac{\ln 2}{s}\right).$$

The limiting value of (3.18) for this value of n (which is the same for both $s = r$ and $s = r - 1$ and hence is the desired limiting value) is by (3.20)

$$(3.21) \qquad E\{T|R_A\} = n + 2 \ln s + \mathcal{O}(1) = n + 2 \ln \ln n + \mathcal{O}(1).$$

Based on the form of (3.21) we conjecture that a lower bound for all procedures for $n \geq 3$ is given by

$$(3.22) \qquad CLB = n - 2 + (2 \ln 2) \log \log n.$$

A lower bound for the minimax problem can be obtained using a modification of a method due to Slupecki [4]. Our result is stated in two parts; the first gives some easily derived inequalities that any lower bound has to satisfy and the second shows that what we got in the first part is an attainable lower bound for the maximum length of any branch of the procedure.

Lemma: For $n \geq 3$ players a lower bound (LBM) for the maximum branch length of any procedure for the selection $(t = 2)$ problem must satisfy the inequalities

$$(3.23) \qquad LBM \leq \begin{cases} n - 2 + [\log (n-1)] & \text{if } n \text{ is of the form } 2^r + 1 \\ n - 1 + [\log (n-1)] & \text{otherwise.} \end{cases}$$

Proof: The second part of (3.23) is the same as the Schreier [3] and Slupecki [4] result for the $t = 2$ ordering problem. Since the selection problem cannot require more than the ordering problem, this part is clear. For the first part we need only put one unit aside until the very last comparison (see procedure R_M above) and use the second part again for $n - 1$ units.

where $\ln x$ is the natural logarithm of x. Since we are looking for a final change for s as n increases, we take the largest root for n in (3.14) and obtain

$$(3.23) \qquad n \sim 2^{s-2} \left(\frac{s}{\ln 2} + \frac{\ln 2}{s}\right).$$

The limiting value of (3.11) for this value of n (which is the same for both $s = t$ and $s = t - 1$ and hence is the desired limiting value) is by (3.23)

$$(3.24) \qquad E[T_N^s] = n + 2 \ln s + O(1) = n + 2 \ln \ln n + O(1).$$

Based on the form of (3.24) we conjecture that a lower bound for all procedures for $n \geq 3$ is given by

$$(3.25) \qquad CLD = n + 2 + (2 + \epsilon) \log \log n.$$

A lower bound for the minimax problem can be obtained using a modification of a method due to Slupecki [7]. The result is stated in two parts; the first gives some easily derived inequalities that any lower bound has to satisfy and the second shows that what we got in the first part is an attainable lower bound for the maximum length of any branch of the procedure.

Lemma: For $n \geq 3$ players a lower bound (LBM) for the maximum branch length of any procedure for the selection $(t = 2)$ problem must satisfy the inequalities

$$(3.26) \qquad LBM \geq \begin{cases} n - 2 + [\log(n-1)] & \text{if } n \text{ is of the form } 2^k + 1 \\ n - 1 + [\log(n-1)] & \text{otherwise.} \end{cases}$$

Proof: The second part of (3.26) is the same as the Schreier [] and Kislitsyn [4] result for the $t = 2$ ordering problem. Since the selection problem cannot require more than the ordering problem, this part is clear. For the first part we need only put one more catch at the very last comparison (see procedure P_N above) and use the second part again for $n - 1$ inputs.

Since this actually gives us an ordering of the $t = 2$ best in $n - 1$, it follows that we need only one more comparison (the unit set aside versus the second best of $n - 1$). This gives a total of

$$(3.24) \qquad n - 2 + [\log (n - 2)] + 1 = n - 2 + [\log (n - 1)],$$

which proves the lemma.

For the second part we follow the method of Slupecki and consider a class of procedures (or a system) S that is characterized by the fact that defeated players do not enter into the first $n - 1$ comparisons. We show below (3.25) that the maximum branch length of any procedure in S is equal to or greater than the right side of (3.23). Since the early inclusion of defeated players can only lengthen our procedure, it follows that our results holds a fortiori for all procedures. The two results taken together then prove the

Theorem: For $n \geq 3$ players the best lower bound (LB_0) for the maximum branch length in any procedure is given by the right side of (3.23) or for all $n \geq 3$ by

$$(3.25) \qquad LB_0 = n - 1 + [\log (n - 2)].$$

Proof: It is easily checked that the theorem holds for $n = 3$ and 4; we use these as starting values for an induction proof. Let k be any integer > 4 and suppose the theorem holds for every integer i such that $3 \leq i < k$. Let s (approximately equal to $k/2$) denote the number of games in the first round in which the k players take part. In the games that follow the first round we use the induction hypothesis on $k - s$ players (the winners of the first round plus any players that did not play in the first round). The first round gives us no information for the M-problem for selecting (or ordering) these $k - s$ players. After getting the two best of these $k - s$

Since this actually gives us an ordering of the $t = 2$ best in $n - 1$, it follows that we need only one more comparison (the unit set aside versus the second best of $n - 1$). This gives a total of

$$(4.24) \qquad n - 2 + \lceil \log (n - 2) \rceil + 1 = n - 2 + \lceil \log (n - 1) \rceil,$$

which proves the lemma.

For the second part we follow the method of Slupecki and consider a class of procedures (or a system) $\mathcal{S}$ that is characterized by the fact that defeated players do not enter into the first $n - 1$ comparisons. We wish to show that the minimum branch length of any procedure in $\mathcal{S}$ is equal to or greater than the right side of (4.23). Since the early inclusion of defeated players can only lengthen our procedure, it follows that our results holds a fortiori for all procedures. The two results taken together then prove the

<u>Theorem</u>: For $n \geq 3$ players the best lower bound (LB_0) for the minimum branch length in any procedure is given by the right side of (4.23) or for all $n \geq 3$ by

$$(4.23) \qquad LB_0 = n - 1 + \lceil \log (n - 2) \rceil.$$

<u>Proof</u>: It is easily checked that the theorem holds for $n = 3$ and 4; we use these as starting values for an induction proof. Let k be any integer > 4 and suppose the theorem holds for every integer i such that $3 \leq i < k$. Let s (approximately equal to $k/2$) denote the number of games in the first round in which the k players take part. In the games that follow the first round we use the induction hypothesis on $k - s$ players (the winners of the first round plus any players that did not play in the first round). The first round gives us no information for the M-problem for selecting (or ordering) these $k - s$ players. After getting the two best of these $k - s$

players, we then take into account the two people (at most) that they defeated in the first round.

We first note that the two expressions on the right side of (3.23) are equivalent to the single expression on the right side of (3.25). To see this we need only consider the cases $n = 2^r + 1$ and $n \neq 2^r + 1$ separately. Hence we can use either of these expressions; we prefer to use (3.23) for the induction hypothesis on $k - s$ players.

For $k = 4m$ we take $s = k/2$ and by the induction hypothesis the $k - s$ players require $(k - s - 1) + [\log (k - s - 1)] = k - s - 2 + [\log (k-2)]$ games. By the Schreier-Slupecki result we can assume that not only have we selected the two best of $k - s - 1$ but also that they are ordered. Hence we need only 1 extra game for completion (the contender for second best against the one defeated in the first round by the best player). This gives the total number of binary comparisons

$$(3.26) \qquad s + (k - s - 2) + [\log (k - 2)] + 1 = k - 1 + [\log (k - 2)],$$

which is the desired result for $k = 4m$.

For $k = 2(4m - 1)$ we take $s = k/2$ and the steps are similar to those above giving (3.26).

For $k = 2(4m + 1)$ we take $s = (k/2) - 1$ and again we obtain the right side of (3.26).

For $k = 4m - 1$ we take $s = (k - 1)/2$, so that $k - s = 2m$ is even and the steps are again the same.

For $k = 4m + 1$ we take $s = (k + 1)/2$, so that $k - s = 2m$ is even, and we obtain for the total

$$(3.27) \qquad s + (k - s - 1) + [\log (\frac{k-3}{2})] + 1 = k - 1 + [\log (k - 2)],$$

which completes the induction.

players. we then take into account the two people (at most) that they defeated in the first round.

We first note that the two expressions on the right side of (3.23) are equivalent to the single expression on the right side of (3.2). To see this we need only consider the cases $n = 2^c + 1$ and $n = 2^c + 1$ separately. Hence we can use either of these expressions; we prefer to use (3.23) for the induction hypothesis on $k - s$ players.

For $k = 4m$ we take $s = k/2$ and by the induction hypothesis the $k - s$ players require $(k - s - 1) + [\log (k - s - 1)] = k - s - 2 + [\log (k-2)]$ games. By the Schreier-Slupecki result we can assume that not only have we selected the two best of $k - s - 1$ but also that they are ordered. Hence we need only 1 extra game for completion (the contender for second best against the one defeated in the first round by the best player). This gives the total number of binary comparisons

$$s + (k - s - 2) + [\log (k - 2)] + 1 = k - 1 + [\log (k - 2)], \quad (3.26)$$

which is the desired result for $k = 4m$.

For $k = 2(4m - 1)$ we take $s = k/2$ and the steps are similar to those above giving (3.26).

For $k = 2(4m + 1)$ we take $s = (k/2) - 1$ and again we obtain the right side of (3.26).

For $k = 4m - 1$ we take $s = (k - 1)/2$ so that $k - s = 2m$ is even and the steps are again the same.

For $k = 4m + 1$ we take $s = (k + 1)/2$ so that $k - s = 2m$ is even, and we obtain for the total

$$s + (k - s - 1) + [\log (\tfrac{k-1}{2})] + 1 = k - 1 + [\log (k - 1)], \quad (3.27)$$

which completes the induction.

Since the procedure R_M has maximum length equal to the right side of (3.25) it follows that R_M is M-optimal for the $t = 2$ selection problem.

4. Acknowledgement.

The author wishes to thank Dr. Abdollah Hadian for many interesting conversations on this research. Thanks are also due to the referee who suggested that the proof of the lemma in section 3 could be generalized from the original form, which is now written as a corollary.

References

[1] Dubail, F. (1967). Algorithmes de questionnaires réalisables optimaux au sens de différents critères. Thesis à la Faculte des Sciences de l' Université de Lyons.

[2] Picard, C. (1965). Theorie des Questionnaires. Gauthiers-Villars, Paris.

[3] Schreier,J. (1932). On tournament elimination systems (in Polish). Mathesis Polska 7 154-160.

[4] Slupecki, J. (1949-51). On the system S of tournaments. Colloq. Math. II 286-290.

[5] Sobel, M. (1968). On ordering the t = 2 best of n items using binary comparisons. Submitted for publication.

[6] Sobel, M. and Groll, P. A. (1959). Group testing to eliminate efficiently all defectives in a binomial sample. Bell System Tech. J. 38 1179-1252.

References

[1] Dubail, F. (1961). Algorithmes de questionnaires réalisables optimaux au sens de différents critères. Thesis à la Faculté des Sciences de l'Université de Lyon.

[2] Picard, C. (1965). *Théorie des questionnaires.* Gauthiers-Villars, Paris.

[3] Schreier, J. (1932). On tournament elimination systems (in Polish). *Mathesis Polska* 7 154-160.

[4] Slupecki, J. (1950-51). On the system S of tournaments. *Colloq. Math.* II 286-290.

[5] Sobel, M. (1968). On an optimal search for the t best using only binary errorless comparisons: I. Submitted to Colloquium Publications (CNRS) on Questionnaire Theory, edited by C. Picard.

[6] Sobel, M. and Groll, P. A. (1959). Group testing to eliminate efficiently all defectives in a binomial sample. Bell System Tech. J. 38 1179-1252.