A TOURNAMENT PROBLEM

LESTER R. FORD, JR.* AND SELMER M. JOHNSON, The RAND Corporation

Introduction. In his book,† Steinhaus discusses the problem of ranking n objects according to some transitive characteristic, by means of successive pairwise comparisons. In this paper we shall adopt the terminology of a tennis tournament by n players. The problem may be briefly stated: "What is the smallest number of matches which will always suffice to rank all n players?"

Steinhaus proposes an inductive method whereby, the first k players having been ranked, the $(k+1)$-st player is matched against the median player in the first k, and by a "halving" process is finally ranked into this chain. Then the $(k+2)$-nd player is ranked into the new chain of $k+1$ players in the same manner.

Using this process, a player can be ranked into a chain of k others in $S(k) = 1+[\log_2 k]$ matches. Steinhaus thus shows that $M(n)$ matches always suffice for n players where

$$M(n) = 1 + nS(n) - 2^{S(n)}.$$

He then states, "It has not been proved that there is no shorter proceeding possible, but we rather think it to be true."

The purpose of this note is to present an improved procedure, compare it with Steinhaus' $M(n)$ as an upper bound and with a lower bound $L(n)$ derived from information theory, and to discuss the asymptotic behavior of these three functions for large n.

A lower bound, $L(n)$, is easily seen to be $L(n)=1+[\log_2 (n!)]$, since each pairing can do no more than divide the remaining possibilities into two complementary sets; the results of the comparison then selects one or the other of these. Observing $n!$ possibilities initially, with halving the best we can do at each stage, we are led directly to the above formula for $L(n)$.

The improved procedure. This may be explained inductively in three steps (illustrated for the case $n=19$ by Fig. 1). Suppose $n=2r$ or $2r+1$.

1. Pair off $2r$ of the players and let the pairs play in the first round leaving one man out if n is odd.

2. By a continuation of the present method applied to r players, give a complete ranking of these r first round winners.

3. The third step is best explained by a diagram.

At this point in the ranking we have a hierarchy of the form illustrated in Figure 1 for $n=19$. First round winners $J, I, \cdots, B$ are ranked in that order with J the best player. A is the first round loser to B, and other first round losers

* Now at General Analysis Corporation.

† H. Steinhaus, Mathematical Snapshots, New York, 1950, pp. 37–40.

are indicated directly below their respective victors. The odd man drawing a first round bye is considered a loser and put in the position at the extreme left of the diagram.

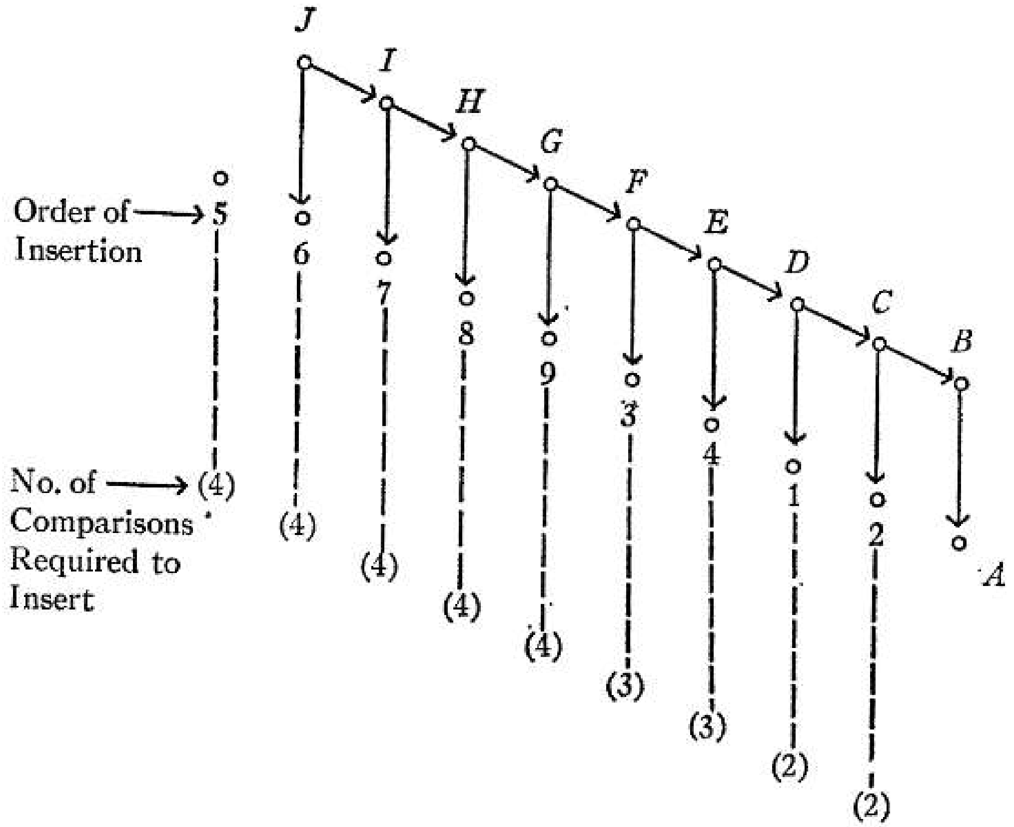

FIG. 1

The phrase "main chain" initially will refer to the chain $JIH \cdots CBA$, and the procedure will be to insert the numbered points in the main chain in the order indicated. The procedure is based on the fact that the insertion of a single point into a chain by the Steinhaus method is most efficient if the number of points on the chain is of the form $2^K - 1$.

Hence we start by inserting point 1 in the chain ABC. After this has been done, the "main chain" under point 2 consists of AB and possibly point 1; this insertion can also be performed with two comparisons.

We now turn to chains of length $2^3 - 1 = 7$, and observe that point 3 is as high as we can go, the main chain under 3 being composed of $ABCDE$ and 1 and 2, and so forth.

This represents a ranking technique, requiring $U(n)$ comparisons, where $U(n)$ is given recursively as follows.

$$U(1) = 0, \qquad U(2) = 1,$$

$$U(2k) = k + U(k) + \sum_{i=2}^{k} T(i),^*$$

$$U(2k+1) = k + U(k) + \sum_{i=2}^{k+1} T(i),$$

* The $T(i)$ are exactly the parenthetical numbers along the bottom of Figure 1.

where

$$\begin{aligned} T(i) &= 2 \quad \text{for} \quad 1 < i \leqq 3, \\ T(i) &= 3 \quad \text{for} \quad 3 < i \leqq 5, \\ T(i) &= 4 \quad \text{for} \quad 5 < i \leqq 11, \\ &\cdots\cdots\cdots\cdots \\ T(i) &= j \quad \text{for} \quad t_{j-1} < i \leqq t_j, \end{aligned}$$

and $t_j = \{2^{j+1} + (-1)^j\}/3$.

We shall now give a table of values of $M(n)$, $U(n)$, and $L(n)$ for some selected values of n, thereafter making some empirical observations.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
$M(n)$	0	1	3	5	8	11	14	17	21	25	29	33	37
$U(n)$	0	1	3	5	7	10	13	16	19	22	26	30	34
$L(n)$	0	1*	3	5	7	10	13	16	19	22	26	29	33

FIG. 2

It may be observed that $U(n) = L(n)$ for $n = 20$ and 21 also; thus the proposed procedure is optimal for those values of n in addition to the values $n \leqq 11$. We conjecture that the value $L(12) = 29$ can not be achieved, but it seems to be difficult of proof. We further conjecture that $U(n)$ is best possible, for all n, but have no mathematical grounds on which to attack a similar conjecture regarding $L(n)$.

Asymptotic formulae. We state the following formulae for the asymptotic behavior of $U(n)$, $L(n)$, and $M(n)$. The proofs are sufficiently cumbersome to be omitted. (The formulae for $U(n)$ and $M(n)$ are based on the subsequence of n which appear to give local minima for $U(n) - L(n)$, *i.e.*, $n = [2^k/3]$ for some k. These being the "best" values, the comparison may be somewhat invidious.) The decimal values which appear as coefficients for n are, of course, approximations.

$$\begin{aligned} M(n) &\sim n \log_2 n - .915n + O(\log_2 n), \\ U(n) &\sim n \log_2 n - 1.415n + O(\log_2 n), \\ L(n) &\sim n \log_2 n - 1.443n + O(\log_2 n). \end{aligned}$$

On the other hand, for the subsequence of $n = 2^k$ for some k, which are the "best" values for $M(n)$, the coefficients of n in the expressions for $M(n)$ and $U(n)$ are -1 and -1.333, respectively.

* A more correct statement for $L(n)$ should be "the least integer $K \geqq \log_2 n!$." Our formula is equivalent except for $n = 2$.