Statistical Science
1989, Vol. 4, No. 3, 282–296

# Who Solved the Secretary Problem?

**Thomas S. Ferguson**

*Abstract.* In Martin Gardner's *Mathematical Games* column in the February 1960 issue of *Scientific American*, there appeared a simple problem that has come to be known today as the Secretary Problem, or the Marriage Problem. It has since been taken up and developed by many eminent probabilists and statisticians and has been extended and generalized in many different directions so that now one can say that it constitutes a "field" within mathematics-probability-optimization. The object of this article is partly historical (to give a fresh view of the origins of the problem, touching upon Cayley and Kepler), partly review of the field (listing the subfields of recent interest), partly serious (to answer the question posed in the title), and partly entertainment. The contents of this paper were first given as the Allen T. Craig lecture at the University of Iowa, 1988.

*Key words and phrases:* Secretary problem, marriage problem, search problem, relative ranks, stopping times, minimax rules.

## 1. INTRODUCTION

In the late 1950's and early 1960's there appeared a simple, partly recreational, problem known as the secretary problem, or the marriage problem, or the dowry problem, that made its way around the mathematical community. The problem has a certain appeal. It is easy to state and has a striking solution. It was immediately taken up and developed by certain eminent probabilists and statisticians, among them Lindley (1961), Dynkin (1963), Chow, Moriguti, Robbins and Samuels (1964), and Gilbert and Mosteller (1966). Since that time, the problem has been extended and generalized in many different directions so that now one can say that it constitutes a "field" of study within mathematics-probability-optimization. One can see from the review paper by Freeman (1983) how extensive and vast the field has become; moreover, the field has continued its exponential growth in the years since that paper appeared.

One main objective of the present article is historical, to review the history of the problem with the aim of determining who was the first to solve the secretary problem. The historical review may take us far, but I think you will find the journey interesting, and the conclusion surprising.

## 2. STATEMENT OF THE PROBLEM

The reader's first reaction to the title might well be to ask, "Which secretary problem?". After all, as I have just implied, there are many variations on the problem. The secretary problem *in its simplest form* has the following features.

1. There is one secretarial position available.
2. The number $n$ of applicants is known.
3. The applicants are interviewed sequentially in random order, each order being equally likely.
4. It is assumed that you can rank all the applicants from best to worst without ties. The decision to accept or reject an applicant must be based only on the relative ranks of those applicants interviewed so far.
5. An applicant once rejected cannot later be recalled.
6. You are very particular and will be satisfied with nothing but the very best. (That is, your payoff is 1 if you choose the best of the $n$ applicants and 0 otherwise.)

This basic problem has a remarkably simple solution. First, one shows that attention can be restricted to the class of rules that for some integer $r \geq 1$ rejects the first $r - 1$ applicants, and then chooses the next applicant who is best in the relative ranking of the observed applicants. For such a rule, the probability,

*Thomas S. Ferguson is Professor of Mathematics at University of California, Los Angeles. His mailing address is: Department of Mathematics, University of California, 405 Hilgard Avenue, Los Angeles, California 90024.*

$\phi_n(r)$, of selecting the best applicant is $1/n$ for $r = 1$, and, for $r > 1$,

$$(2.1)\qquad \begin{aligned}\phi_n(r) &= \sum_{j=r}^{n} P\begin{pmatrix} j\text{th applicant is best} \\ \text{and you select it}\end{pmatrix} \\ &= \sum_{j=r}^{n}\left(\frac{1}{n}\right)\left(\frac{r-1}{j-1}\right) = \left(\frac{r-1}{n}\right)\sum_{j=r}^{n}\frac{1}{j-1}.\end{aligned}$$

The optimal $r$ is the one that maximizes this probability. For small values of $n$, the optimal $r$ can easily be computed. Of interest are the approximate values of the optimal $r$ for large $n$. If we let $n$ tend to infinity and write $x$ as the limit of $r/n$, then using $t$ for $j/n$ and $dt$ for $1/n$, the sum becomes a Riemann approximation to an integral,

$$(2.2)\qquad \begin{aligned}\phi_n(r) &= \left(\frac{r-1}{n}\right)\sum_{j=r}^{n}\left(\frac{n}{j-1}\right)\left(\frac{1}{n}\right) \\ &\to x\int_x^1 \left(\frac{1}{t}\right) dt = -x\log(x).\end{aligned}$$

The value of $x$ that maximizes this quantity is easily found by setting the derivative with respect to $x$ equal to zero and then solving for $x$. When this is done we find that

$$(2.3)\qquad \begin{aligned}&\text{optimal } x = 1/e = .367879\cdots, \quad \text{and} \\ &\text{optimal probability} = 1/e.\end{aligned}$$

Thus for large $n$, it is approximately optimal to wait until about 37% of the applicants have been interviewed and then to select the next relatively best one. The probability of success is also about 37%.

This derivation may seem a little loose, but it gives the right answer. For a lucid presentation of these and related results, see the basic paper of Gilbert and Mosteller (1966).

## 3. HISTORICAL BACKGROUND

There is some obscurity as to the origins of this problem. It seems to be generally agreed among workers in the field that the first statement of the problem to appear in print occurred in the February 1960 column of Martin Gardner in *Scientific American*, where it is attributed to Fox and Marnie. A solution to the problem is outlined in the March 1960 issue of *Scientific American* and attributed to Moser and Pounder. In 1963, it appeared as a problem in *The American Mathematical Monthly* contributed by Bissinger and Siegel (1963); in 1964, a solution appeared due to Bosch. But Mosteller learned of the problem in 1955 from Andrew Gleason, "who claimed to have heard it from another," (Gilbert and Mosteller, 1966). Herb Robbins recalls discussing the problem in 1953–54 at Columbia University, and Merrill Flood recalls presenting a version of the problem that he called the fiancé problem at a conference on mathematical problems in logistics held at George Washington University in January 1950 (personal communication, 1988). I personally remember working on extensions of the problem in the summer of 1959. In any case, many people knew of the problem by the time it appeared in print.

Lindley (1961) seems to be the first to solve the problem in a scientific journal. He extends the problem to an arbitrary utility based on the rank of the applicant selected, and considers in particular the problem of minimizing the expected rank of the applicant selected, rank 1 being best. However, the difficult problem of finding the asymptotic, for large $n$, optimal strategies and expected rank was left open, and finally solved neatly by Chow, Moriguti, Robbins and Samuels (1964). Dynkin (1963) considers the problem as an application of the theory of Markov stopping times, and shows that, properly interpreted, the problem is monotone so that the one-stage look-ahead rule is optimal.

Then, in 1966, came the basic paper of Gilbert and Mosteller, with elegant derivations and extensions in a number of important directions. In particular, they allow $r$ choices to obtain the best; they consider the problem of obtaining the best or the next best; they treat the "full-information" case, in which one is allowed to observe the actual values of the applicants (presumed to be chosen independently from a given distribution), for both the best-choice problem and for the minimum-rank problem; they analyze some game theoretic versions of the problem; etc. This paper, more than the others, foreshadowed the explosion of ideas, generalizations, and effort that would impact this area starting in 1972 and that continues strongly today. Some of these contributions will be mentioned in Section 5.

This, briefly, is the official early history of the secretary problem. Since we are to attempt to discover who first solved this problem, we shall, as is customary in historical papers, proceed backward in time, looking for the germ of the idea hidden in forgotten literature. The first possibility occurs in the work of Arthur Cayley.

## 4. CAYLEY'S PROBLEM

The distinguished English mathematician, Arthur Cayley (1821–1895), is perhaps best known for his seminal work in the theory of algebraic invariants. He was also one of the most prolific mathematicians the world has ever known. His collected works contain some 966 papers touching on many subjects in mathematics, theoretical dynamics and astronomy. Paper #705 contains some 50 pages of problems and solutions that Cayley submitted to the *Educational Times* from

1871 to 1894. One of these problems Cayley (1875), is as follows:

> 4528. (Proposed by Professor Cayley) A lottery is arranged as follows: There are $n$ tickets representing $a, b, c, \cdots$ pounds respectively. A person draws once; looks at his ticket; and if he pleases, draws again (out of the remaining $n - 1$ tickets); and so on, drawing in all not more than $k$ times; and he receives the value of the last ticket drawn. Supposing that he regulates his drawings in the manner most advantageous to him according to the theory of probabilities, what is the value of his expectation?

From the "Solution by the Proposer," we see that Cayley believes it to be understood that $k$, $n$, and the $a, b, c, \cdots$ are known numbers. (Note that here $k$, rather than $n$, represents the total number of drawings.) He solves the problem by what is now known as the method of backward induction of dynamic programming. As an example, he takes $n = 4$, and $a, b, c, d = 1, 2, 3, 4$, and finds for $k = 1, 2, 3, 4$, the values of the most advantageous expectation to be $^{10}\!/_4$, $^{38}\!/_{12}$, $^{85}\!/_{24}$, 4 resp.

Cayley's problem was resurrected from oblivion by Moser (1956), who also reformulated the problem in a neater guise which may be viewed as an approximation to the Cayley problem when $n$ is very large and $a, b, c, \cdots$ are $1, 2, \cdots, n$: You observe, sequentially, random variables $X_1, \cdots, X_k$ known to be iid from a uniform distribution on the interval (0, 1); if you stop after observing $X_j$, then you receive $X_j$ as your reward. The optimal rule, as found by Moser, is to stop when there are $m$ observations left to be observed if the value of the present observation is greater than $E_m$, where the $E_m$ are defined recursively by $E_0 = 0$ and $E_{m+1} = (1 - E_m^2)/2$. The corresponding equations are discussed in Guttman (1960) for the normal distribution, in Karlin (1962) for the exponential distribution and in Gilbert and Mosteller (1966) for the inverse power distribution.

Although there are strong points of similarity between Cayley's problem and the secretary problem, there is one important difference. The payoff is not one or zero depending on whether you select the best or not; it is a numerical quantity depending on the intrinsic value of the object selected. This difference plays a big role in the feeling of the problem and has led to a class of problems, called variously the house hunting problem, the problem of selling an asset, or the search problem, with a literature as large as for the secretary problems. The basic problem is the Cayley–Moser problem with an infinite horizon, with the payoff modified to make certain that one will wish to stop in a finite time. For example, Sakaguchi (1961) and Chow and Robbins (1963): Random variables $X_1, X_2, \cdots$ are observed sequentially at a cost of $c > 0$ per observation; if you stop after observing $X_n$, then you receive $X_n - nc$. If recall of past observations is allowed, the payoff for stopping after observing $X_n$ is $\max(X_1, \cdots, X_n) - nc$; such problems were treated by MacQueen and Miller (1960), Derman and Sacks (1960) and Chow and Robbins (1961). In Karlin (1962), the problem is solved with a discount rather than a cost. See DeGroot (1970) for a fuller account of these developments.

This class of problems forms a rather distinct set of problems that is still being confused with the secretary problems. Since there are so many variations of the basic secretary problem (each of the 6 conditions listed in Section 2 has been modified by at least one author), I think it is worthwhile to try to define what a secretary problem is. My definition is: *A secretary problem is a sequential observation and selection problem in which the payoff depends on the observations only through their relative ranks and not otherwise on their actual values.*

With this definition then, Cayley's problem is not even a secretary problem. We must look elsewhere to see who solved the secretary problem. Proceeding farther back in time, we come to the first practical application I could find of these sequential observation and selection techniques: the selection of a wife by Johannes Kepler.

## 5. KEPLER'S PROBLEM

When the celebrated German astronomer, Johannes Kepler (1571–1630), lost his first wife to cholera in 1611, he set about finding a new wife using the same methodical thoroughness and careful consideration of the data that he used in finding the orbit of Mars to be an ellipse. His first, not altogether happy, marriage had been arranged for him, and this time he was determined to make his own decision. In a long letter to a Baron Strahlendorf on October 23, 1613, written after he had made his selection, he describes in great detail the problems he faced and the reasons behind each of the decisions he made. He arranged to interview and to choose from among no fewer than eleven candidates for his hand. The process consumed much of his attention and energy for nearly 2 years, what with the investigations into the virtues and drawbacks of each candidate, her dowry, negotiations with her parents, natural hesitations, the advice of friends, etc. The book of Arthur Koestler (1960) contains an entertaining and insightful exposition of the process. The book of Carola Baumgardt (1951) contains much supplementary information.

Suffice it to say that of the eleven candidates interviewed, Kepler eventually decided on the fifth. It may be noted that when $n = 11$, the function $\phi_n(r)$ of (2.1) takes on its maximum value when $r = 5$. Perhaps, if Kepler had been aware of the theory of the secretary problem, he could have saved himself a lot of time and trouble.

Of course, in all practical applications of theoretical results, the assumptions are never exactly satisfied, and in the present instance this is especially true as we can see from Kepler's letter. For example, after interviewing candidate number 5 and being strongly attracted to her, Kepler listened to the advice of friends who were concerned with her lack of high rank, wealth, parentage and dowry (she was an orphan), and who persuaded him to propose to number 4. Thus, clearly Kepler thought he could recall past applicants in violation of assumption 5 of Section 2. Certainly, Kepler would have been interested in the papers of Yang (1974), Petruccelli (1981, 1984), Rose (1984), Ferenstein and Enns (1988) and others, who allow backward solicitation with a cost or with a probability $q$ of being accepted. The probability $q$ is not 1 in Kepler's case since candidate number 4 turned him down. He had waited too long.

That Kepler went on after his failure with number 4 shows that he was not just interested in getting the best (assumption 6 of Section 2). Perhaps he was minimizing the expected rank or some other utility function, in which case the papers of Chow, Moriguti, Robbins and Samuels (1964), Mucci (1973), Lorenzen (1979), Frank and Samuels (1980), etc., would have interested him. Since he had been married before, it is unrealistic to assume that he knew nothing about women (assumption 4 of Section 2); he would have enjoyed the full-information problem of Gilbert and Mosteller (1966) or Tamaki (1986). But it is also unreasonable to assume that he knew everything about women (who does?), so the models with partial information and learning of Stewart (1978), Samuels (1981), Campbell and Samuels (1981) or Campbell (1982) are more to the point.

On the other hand, he actually interviewed all 11 candidates and could have gone on. Perhaps he was expecting a random number of available candidates (violating condition 2 of Section 2), in which case he would have enjoyed reading the papers of Presman and Sonin (1972), Gianini-Pettitt (1979), Abdel-Hamid, Bather and Trustrum (1982), and Bruss and Samuels (1987). Or perhaps there was a cost of observation as in Bartoszynski and Govindarajulu (1978), Lorenzen (1981) or Samuels (1985) or a discount factor as in Rasmussen and Pliska (1976). He would certainly be interested in the random arrival models of Sakaguchi (1976, 1986), Cowan and Zabczyk (1978) and Bruss (1987), in the game theoretic models of Presman and Sonin (1975), Fushimi (1981) and Enns and Ferenstein (1987), and in the multiple criteria formulation of Stadje (1980), Gnedin (1983), Berezovskiy, Baryshnikov and Gnedin (1986) or Samuels and Chotlos (1986). Possibly, Kepler was concerned with the actual value of his bride and not just with her ranking among the other candidates. This would make it not a secretary problem at all, but a stopping-rule problem more like Cayley's problem or the search problem discussed in the previous section.

It is clear that much more research needs to be done to clarify which of these problems Kepler was actually solving. Whichever one it was, there can be no doubt that the outcome was favorable for him. His new wife, whose education, as he says in his letter, must take the place of a dowry, bore him seven children, ran his household efficiently, and seems to have provided the necessary tranquil homelife for his erratic genius to flourish.

## 6. THE GAME OF GOOGOL

Let us return to the question: Of which of the many different versions of the secretary problem am I trying to find the solver? As historians, we should take as *the* secretary problem, the problem as it first appeared in print, in Martin Gardner's February 1960 column in *Scientific American*, where it was called the game of googol and described as follows.

> Ask someone to take as many slips of paper as he pleases, and on each slip write a different positive number. The numbers may range from small fractions of 1 to a number the size of a googol (1 followed by a hundred 0's) or even larger. These slips are turned face down and shuffled over the top of a table. One at a time you turn the slips face up. The aim is to stop turning when you come to the number that you guess to be the largest of the series. You cannot go back and pick up a previously turned slip. If you turn over all slips, then of course you must pick the last one turned.

The astute reader may notice that this is not the simple form of the secretary problem described in Section 2. The actual values of the numbers are revealed to the decision maker in violation of condition 4. Also, there is this "someone" who chooses the numbers, presumably to make your problem of selecting the largest as difficult as possible. The game of googol is really a two-person game.

This raises two questions. First, can you guarantee a higher probability of selecting the largest number if you allow your decision rule to depend on the actual values of the numbers? In other words, does the stated solution give the lower value of the game? Second, if you are told how this "someone" is choosing the numbers to place on the slips, can you now guarantee a higher probability of selecting the largest number? In other words, does the value of the game exist, and can we find optimal or $\varepsilon$-optimal strategies for the sequence chooser? These questions were not addressed in the solution presented in *Scientific American* in March 1960.

The statement of the problem as it appeared in *The American Mathematical Monthly* in 1963 is much the same, but somewhat more nebulous since you are not told where the numbers come from. Perhaps it is a game against nature, so the second question above does not arise. In any case, the solution presented for the problem in 1964 contains the same oversight.

Those distinguished statisticians who worked on the secretary problem in the 1960's were more careful in their statements of the problem in specifying what information could be used in the decision rule, but none of them attacked the above problem. Therefore, to see who first solved the problem, we must proceed into the 1970's and beyond.

Suppose you were this "someone" who must choose the numbers to write on the slips. How would you go about choosing the numbers to make it as difficult as possible for me to obtain the largest? You could not just choose the numbers $1, 2, \cdots, n$, because then I could wait until the number $n$ appeared and thus obtain the largest number with probability 1. You could not just choose them iid from some fixed distribution, because this would lead to the full-information case solved by Gilbert and Mosteller (1966), who showed that I can obtain the largest number with probability at least $0.58016 \cdots$, which is the limiting value for large $n$. So, you must choose the numbers in some dependent fashion. But you might as well choose them to be an exchangeable process since the numbers are put in random order anyway before being shown to me. Thus, you are drawn to the partial information models of Stewart (1978), Petruccelli (1980) and Samuels (1981).

## 7. PARTIAL INFORMATION MODELS

Let $X_1, \cdots, X_n$ denote the values of the numbers on the slips. These may be considered as the parameters of our statistical problem. We want to find a prior distribution for $X_1, \cdots, X_n$, with respect to which the Bayes rule is the usual optimal rule based on the relative ranks of the $X_j$.

In the paper of Stewart (1978), the $X_j$ are chosen iid from a uniform distribution on the interval $(\alpha, \beta)$, denoted by $U(\alpha, \beta)$, and $(\alpha, \beta)$ is chosen from the three-parameter Pareto distribution. Specifically,

$$\begin{aligned}&(\alpha, \beta) \quad \text{is } Pa(k, l_0, u_0), \quad \text{and}\\ &X_1, \cdots, X_n, \quad \text{given } (\alpha, \beta), \quad \text{are iid } U(\alpha, \beta),\end{aligned} \tag{7.1}$$

where the three-parameter Pareto distribution, $Pa(k, l_0, u_0)$ with $k > 0$ and $l_0 < u_0$, is the distribution with density

$$\begin{aligned}&g(\alpha, \beta \mid k, l_0, u_0)\\ &\quad = \frac{k(k+1)(u_0 - l_0)^k}{(\beta - \alpha)^{k+2}} I(\alpha < l_0, u_0 < \beta)\end{aligned} \tag{7.2}$$

where $I$ represents the indicator function. This is a conjugate family of distributions for the uniform distributions, and the posterior distribution of $(\alpha, \beta)$ given $X_1, \cdots, X_j$ is also Pareto,

$$\begin{aligned}&(\alpha, \beta), \quad \text{given } X_1, \cdots, X_j,\\ &\text{is } Pa(k + j, l_j, u_j), \quad \text{where}\\ &l_j = \min\{l_0, X_1, \cdots, X_j\} \quad \text{and}\\ &u_j = \max\{u_0, X_1, \cdots, X_j\}.\end{aligned} \tag{7.3}$$

Thus, one can interpret the prior information as being equivalent to a sample of size $k$ from a uniform distribution with a minimum of $l_0$ and a maximum of $u_0$.

Stewart obtained a rather striking result for this prior distribution of the $X_j$, as follows. First, the payoff is changed so that you win only if you stop on the largest $X_j$ and if that $X_j$ is greater than $u_0$. Then, one shows that

$$\begin{aligned}&P(X_j = u_n \mid X_1, \cdots, X_j)\\ &\quad = \frac{k + j + 1}{k + n + 1} I(X_j = u_j).\end{aligned} \tag{7.4}$$

This implies that attention can be restricted to rules that depend only on the relative ranks of the observations including $u_0$. In fact, the problem becomes equivalent to the secretary problem of Section 2 with $n + k + 1$ applicants, in which you start with $k + 1$ applicants already rejected, the largest having value $u_0$. In particular, the Bayes rule among rules that use all the information is the rule that rejects the first $r' - 1$ applicants, and selects the next applicant who is relatively best (and better than $u_0$), where $r' = \max(1, r - k - 1)$ and $r$ is the optimal value of $r$ for the secretary problem of Section 2 with $n + k + 1$ applicants. In addition, for all values of $k$, the probability of win under an optimal rule tends to $1/e$ as $n \to \infty$! Since for sufficiently large $n$, the maximum $X_j$ will be greater than $u_0$ with probability close to 1, this means that given any $\varepsilon > 0$, there is an $N$ and a distribution of the form (7.1) such that for $n > N$ if $(\alpha, \beta)$ is chosen from this distribution and then the $X_j$ are chosen as from a uniform distribution on $(\alpha, \beta)$, the probability of win in the game of googol is less than $1/e + \varepsilon$.

Thus Stewart has solved googol asymptotically for $n$ large. Unfortunately, for fixed finite $n$, one cannot find, for all $\varepsilon > 0$, $\varepsilon$-optimal distributions for choosing the $X_j$ among the distributions he suggests. We must look further. The paper of Petruccelli (1980) considers some other distributions for the $X_j$. If the $X_j$ are uniform on the interval $(\theta - 0.5, \theta + 0.5)$, then the best invariant stopping rule gives a probability of win asymptotic to $0.43517\cdots$ as $n \to \infty$. If the distributions are normal with mean $\mu$ and variance 1, the situation is even worse (from our point of view). The best invariant rule gives a probability of win asymptotic to the full-information case, namely, $0.58016\cdots$. Asymptotically, you might as well tell your opponent what $\mu$ you are using.

Finally, we come to the paper of Samuels (1981), who extends the results of Stewart. Samuels shows that in the model where the $X_j$ are chosen from the uniform distribution on $(\alpha, \beta)$, the usual rule (the optimal rule based on relative ranks) is minimax for each $n$. Since the usual rule is an equalizer rule (it gives the same probability of success no matter how the $X_j$ are chosen), we see that it is minimax against general, nonparametric alternatives as well. In other words, the usual solution achieves the lower value of the game. Thus, I believe that Steve Samuels deserves credit for having solved (the more difficult half of) the secretary problem as it appeared in *The American Mathematical Monthly.*

However, this exhausts my search through the relevant literature. What can be said about the game of googol? I can finally give you my answer to the question in the title of this article. Who solved *the* secretary problem? Nobody.

## 8. A RESOLUTION

Let me hasten to apologize for this anticlimax, and to venture the opinion that the reason no one has solved this problem is that possibly no one was interested in googol as a game, or perhaps realized there was a problem yet to be solved. To remedy the situation, let us try to find the solution now. With the hint given by the paper of Stewart, it turns out not to be hard.

In fact, since we are considering only the best choice problem, we may consider a simpler class of distributions than that of Stewart—the one-parameter uniform and the two-parameter Pareto distributions. Thus, we take

$$\text{(8.1)} \quad \begin{aligned} &\theta \quad \text{is } Pa(\alpha, 1), \quad \text{and} \\ &X_1, \cdots, X_n, \quad \text{given } \theta, \quad \text{are iid } U(0, \theta), \end{aligned}$$

where the two-parameter Pareto distribution, $Pa(\alpha, m_0)$, is the distribution with density

$$\text{(8.2)} \qquad g(\theta \mid \alpha, m_0) = \alpha m_0^{\alpha}/\theta^{\alpha+1} I(\theta > m_0),$$

where $\alpha > 0$ and $m_0 > 0$. For simplicity, we take $m_0 = 1$. This class of Pareto distributions forms a conjugate prior for $U(0, \theta)$, and contains the posterior distribution of $\theta$ given $X_1, \cdots, X_j$:

$$\text{(8.3)} \quad \begin{aligned} &\theta, \quad \text{given } X_1, \cdots, X_j, \quad \text{is } Pa(\alpha + j, m_j), \\ &\text{where} \quad m_j = \max(m_0, X_1, \cdots, X_j). \end{aligned}$$

Let us pretend that we are playing a game of googol, I choosing the $X_j$ and you choosing the stopping rule. For a given $\varepsilon > 0$, I will choose the $X_j$ according to (8.1), and find an $\alpha$ (sufficiently close to zero) so that you will win with a probability less than $\phi_n + \varepsilon$, where $\phi_n$ is the maximum probability you can guarantee using strategies that depend only on the relative ranks, $\phi_n = \max_r \phi_n(r)$.

In fact, I will give you an additional slight advantage and still keep your probability of success below $\phi_n + \varepsilon$. I will say that you win if you stop at the largest $X_j$, or if all the $X_j$ are no greater than 1. This will allow you to restrict attention to stopping rules that stop only at a relatively largest $X_j$ that is greater than 1. The probability you win is

$$\text{(8.4)} \qquad P(\text{win}) = P(\text{all } X_j \le 1) + P(\text{win*}),$$

where win* represents the event {max $X_j > 1$ and you choose it}. The first term does not depend on your strategy and is easily computed:

$$\text{(8.5)} \quad \begin{aligned} P(\text{all } X_j \le 1) &= \int_1^\infty P(\text{all } X_j < 1 \mid \theta) g(\theta \mid \alpha, 1)\, d\theta \\ &= \alpha \int_1^\infty (1/\theta)^n (1/\theta)^{\alpha+1}\, d\theta \\ &= \alpha/(n+\alpha). \end{aligned}$$

Your problem is to maximize the second term of (8.4).

First, we find the probability, conditional at stage $j$, that $m_j > 1$ is already as large as it will get.

$$\text{(8.6)} \quad \begin{aligned} &P(m_j = m_n \mid X_1, \cdots, X_j) \\ &\quad = E(P(m_j = m_n \mid \theta, X_1, \cdots, X_j) \mid X_1, \cdots, X_j) \\ &\quad = \int_{m_j}^\infty (m_j/\theta)^{n-j} g(\theta \mid \alpha + j, m_j)\, d\theta \\ &\quad = (\alpha + j)/(\alpha + n) \end{aligned}$$

independent of $X_1, \cdots, X_j$. This is an analog of Stewart's result (7.4). If you have a new candidate at stage $j$, that is if $X_j = m_j$, it is optimal to select it if and only if

$$\frac{\alpha + j}{\alpha + n} \ge P(\text{win* with best strategy from stage } j + 1 \text{ on}).$$

The right side of this inequality is a nonincreasing function of $j$, since any strategy available at stage $j + 2$ is also available at stage $j + 1$. Since the left side of the inequality is an increasing function of $j$, an optimal rule may be found among rules of the form for some $r \geq 1$: reject the first $r - 1$ applicants and accept the next applicant for which $X_j = m_j$, if any.

Using such a strategy, the probability of a win* may be computed as

$$P(\text{win}^*) = \sum_{j=r}^{n} P(\text{select } j)P(j \text{ is best} \mid \text{select } j).$$

The probability that $j$ is best given you select it is just (8.6). The probability that you select $j$ can be found using (8.6):

$$\begin{aligned} P(\text{select } j) &= P(m_{r-1} = m_{j-1} \text{ and } m_j > m_{j-1}) \\ &= \left(\frac{\alpha + r - 1}{\alpha + j - 1}\right)\left(1 - \frac{\alpha + j - 1}{\alpha + j}\right) \\ &= \frac{\alpha + r - 1}{(\alpha + j - 1)(\alpha + j)}. \end{aligned}$$

Combining these into (8.4) and letting $\phi_n(r, \alpha)$ denote the probability of a win, we find

$$\phi_n(r, \alpha) = \left(\frac{\alpha}{\alpha + n}\right) + \left(\frac{\alpha + r - 1}{\alpha + n}\right) \sum_{j=r}^{n} \frac{1}{(\alpha + j - 1)}.$$

Now, note that $\phi_n(r, \alpha)$ is continuous in $\alpha$ for $\alpha > 0$, and that as $\alpha \to 0$, $\phi_n(r, \alpha) \to \phi_n(r)$ for all $r = 1, \cdots, n$, where $\phi_n(r)$ is given by (2.1). Hence, as $\alpha \to 0$,

$$\max_r \phi_n(r, \alpha) \to \max_r \phi_n(r) = \phi_n.$$

*Therefore, an $\varepsilon$-optimal method of choosing the $X_j$ is given by* (8.1), *where $\alpha = \alpha(n, \varepsilon)$ is chosen so that*

$$|\max_r \phi_n(r, \alpha) - \phi_n| < \varepsilon.$$

This derivation may be considered an alternate proof of the minimaxity result of Samuels mentioned in Section 7. It is interesting to note that this result cannot be obtained using the distributions of (7.1). The optimal rule based on relative ranks is exactly Stewart's rule with $k = -1$; but $k$ must be positive for (7.1) to be a distribution; hence, we cannot approximate the case $k = -1$ with distributions. (In addition, the term corresponding to (8.5) does not go to zero.) If it seems strange that the distributions (7.1) were considered before the simpler distributions (8.1), the reason is that Stewart and Samuels treated in their papers problems with more general payoffs, not just the best choice problem, and needed a broader class of distributions. Unfortunately, (7.1) does not contain (8.1). If there is a moral to this, maybe it is that the simpler cases should always be examined first.

## *Note Added in Proof*

Steve Samuels has sent me a copy of the book, *Problems of Best Choice* (in Russian) by B. A. Berezovskiy and A. V. Gnedin, 1984, Akademia Nauk, USSR, Moscow. This book is devoted solely to the secretary problem and its variations. It contains not only a review of the field but also a careful exposition and new information as well. In their discussion of the partial information model of Stewart, they use the prior distribution (8.1) and derive (8.6). However, this is only used to prove the asymptotic minimaxity result of Stewart.

## REFERENCES

ABDEL-HAMID, A. R., BATHER, J. A. and TRUSTRUM, G. B. (1982). The secretary problem with an unknown number of candidates. *J. Appl. Probab.* **19** 619–630.

BARTOSZYNSKI, R. and GOVINDARAJULU, Z. (1978). The secretary problem with interview cost. *Sankhyā Ser. B* **40** 11–28.

BAUMGARDT, CAROLA (1951). *Johannes Kepler: Life and Letters.* Philosophical Library, Inc., New York.

BEREZOVSKIY, B. A., BARYSHNIKOV, YU. M. and GNEDIN, A. V. (1986). On a class of best choice problems. *Inform. Sci.* **39** 111–127.

BISSENGER, B. H. and SIEGEL, C. (1963). Advanced Problem 5086. *Amer. Math. Monthly* **70** 336. (Solution by A. J. Bosch. **71** (1964) 329–330.)

BRUSS, F. T. (1987). On an optimal selection problem of Cowan and Zabczyk. *J. Appl. Probab.* **24** 918–928.

BRUSS, F. T. and SAMUELS, S. M. (1987). A unified approach to a class of optimal selection problems with a random number of options. *Ann. Probab.* **15** 824–830.

CAMPBELL, G. (1982). The maximum of a sequence with prior information. *Sequential Anal.* **1** 177–191.

CAMPBELL, G. and SAMUELS, S. (1981). Choosing the best of the current crop. *Adv. in Appl. Probab.* **13** 510–532.

CAYLEY, A. (1875). Mathematical questions with their solutions. *The Educational Times* **23** 18–19. See *The Collected Mathematical Papers of Arthur Cayley* **10** 587–588 (1896). Cambridge Univ. Press, Cambridge.

CHOW, Y. S., MORIGUTI, S., ROBBINS, H. and SAMUELS, S. M. (1964). Optimal selection based on relative rank (The "secretary" problem). *Israel J. Math.* **2** 81–90.

CHOW, Y. S. and ROBBINS, H. (1961). A martingale system theorem and applications. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 93–104. Univ. California Press.

CHOW, Y. S. and ROBBINS, H. (1963). On optimal stopping rules. *Z. Wahrsch. verw. Gebiete* **2** 33–49.

COWAN, R. and ZABCZYK, J. (1978). An optimal selection problem associated with the Poisson process. *Theory Probab. Appl.* **23** 584–592.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

DERMAN, C. and SACKS, J. (1960). Replacement of periodically inspected equipment. *Naval Res. Logist. Quart.* **7** 597–607.

DYNKIN, E. B. (1963). The optimum choice of the instant for stopping a Markov process. *Soviet Math. Dokl.* **4** 627–629.

ENNS, E. G. and FERENSTEIN, E. Z. (1987). On a multiperson time-sequential game with priorities. *Sequential Anal.* **6** 239–256.

FERENSTEIN, E. Z. and ENNS, E. G. (1988). Optimal sequential selection from a known distribution with holding costs. *J. Amer. Statist. Assoc.* **83** 382–386.

FRANK, A. Q. and SAMUELS, S. M. (1980). On an optimal stopping problem of Gusein-Zade. *Stochastic Process. Appl.* **10** 299–311.

FREEMAN, P. R. (1983). The secretary problem and its extensions—A review. *Internat. Statist. Rev.* **51** 189–206.

FUSHIMI, M. (1981). The secretary problem in a competitive situation. *J. Oper. Res. Soc. Japan* **24** 350–359.

GARDNER, M. (1960). *Scientific American* Feb., Mar. (See also *Martin Gardner's New Mathematical Diversions* (1966) 35. Simon and Schuster, New York.)

GIANINI-PETTITT, J. (1979). Optimal selection based on relative ranks with a random number of individuals. *Adv. in Appl. Probab.* **11** 720–736.

GILBERT, J. and MOSTELLER, F. (1966). Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* **61** 35–73.

GNEDIN, A. V. (1983). Effective stopping on a Pareto-optimal option. *Avtomat. i Telemekh.* 44 (3) 166–170. (In Russian).

GUTTMAN, I. (1960). On a problem of L. Moser. *Canad. Math. Bull.* **3** 35–39.

KARLIN, S. (1962). Stochastic models and optimal policy for selling an asset. In *Studies in Applied Probability and Management Science* (K. J. Arrow, S. Karlin and H. Scarf, eds.) 148–158. Stanford Univ. Press, Stanford, Calif.

KOESTLER, A. (1960). *The Watershed. A Biography of Johannes Kepler.* Anchor Books, Garden City, NY.

LINDLEY, D. V. (1961). Dynamic programming and decision theory. *Appl. Statist.* **10** 39–51.

LORENZEN, T. J. (1979). Generalizing the secretary problem. *Adv. in Appl. Probab.* **11** 384–396.

LORENZEN, T. J. (1981). Optimal stopping with sampling cost. *Ann. Probab.* **9** 167–172.

MACQUEEN, J. and MILLER, R. P., JR. (1960). Optimal persistence policies. *Oper. Res.* **8** 362–380.

MOSER, L. (1956). On a problem of Cayley. *Scripta Math.* **22** 289–292.

MUCCI, A. G. (1973). On a class of secretary problems. *Ann. Probab.* **1** 417–427.

PETRUCCELLI, J. D. (1980). On a best choice problem with partial information. *Ann. Statist.* **8** 1171–1174.

PETRUCCELLI, J. D. (1981). Best-choice problems involving uncertainty of selection and recall of observation. *J. Appl. Probab.* **18** 415–425.

PETRUCCELLI, J. D. (1984). Best-choice problems involving recall with uncertainty of selection when the number of observations is random. *Adv. in Appl. Probab.* **16** 111–130.

PRESMAN, E. L. and SONIN, I. M. (1972). The best-choice problem for a random number of objects. *Theory Probab. Appl.* **17** 657–668.

PRESMAN, E. L. and SONIN, I. M. (1975). Equilibrium points in a game related to the best choice problem. *Theory Probab. Appl.* **20** 770–781.

RASMUSSEN, W. T. and PLISKA, S. R. (1976). Choosing the maximum from a sequence with a discount function. *Appl. Math. Optim.* **2** 279–289.

ROSE, J. (1984). Optimal sequential selection based on relative ranks with renewable call options. *J. Amer. Statist. Assoc.* **79** 430–435.

SAKAGUCHI, M. (1961). Dynamic programming of some sequential sampling design. *J. Math. Anal. Appl.* **2** 446–466.

SAKAGUCHI, M. (1976). Optimal stopping problems for randomly arriving offers. *Math. Japon.* **21** 201–217.

SAKAGUCHI, M. (1986). Best choice problems for randomly arriving offers during a random lifetime. *Math. Japon.* **31** 107–117.

SAMUELS, S. M. (1981). Minimax stopping rules when the underlying distribution is uniform. *J. Amer. Statist. Assoc.* **76** 188–197.

SAMUELS, S. M. (1985). A best-choice problem with linear travel cost. *J. Amer. Statist. Assoc.* **80** 461–464.

SAMUELS, S. M. and CHOTLOS, B. (1986). A multiple criteria optimal selection problem. In *Adaptive Statistical Procedures and Related Topics* (J. Van Ryzin, ed.) 62–78. IMS, Hayward, Calif.

STADJE, W. (1980). Efficient stopping of a random series of partially ordered points. *Multiple Criteria Decision Making. Theory and Application Lecture Notes in Econ. Math. Syst.* **177** 430–447. Springer, New York.

STEWART, T. J. (1978). Optimal selection from a random sequence with learning of the underlying distribution. *J. Amer. Statist. Assoc.* **73** 775–780.

TAMAKI, M. (1986). A full-information best-choice problem with finite memory. *J. Appl. Probab.* **23** 718–735.

YANG, M. C. K. (1974). Recognizing the maximum of a sequence with backward solicitation. *J. Appl. Probab.* **11** 504–512.

# Comment: Who Will Solve the Secretary Problem?

**Stephen M. Samuels**

Just like Johannes Kepler, who threw a new curve at the solar system, Tom Ferguson has given a different slant to the Secretary Problem. To its many practitioners who ritually begin by saying "all that we can observe are the relative ranks," Ferguson (citing historical precedent), in effect, responds "let's not take that assumption for granted." The heart of his paper, as I see it, is the following *Ferguson Secretary Problem*:

*Stephen M. Samuels is Professor of Statistics and Mathematics at Purdue University. His mailing address is: Department of Statistics, Mathematical Sciences Building, Purdue University, West Lafayette, Indiana 47907.*

Given $n$, either find an *exchangeable* sequence of continuous random variables, $X_1, X_2, \cdots, X_n$, for which, among all stopping rules, $\tau$, based on the $X$'s,

$$\sup_{\tau} P\{X_\tau = \max(X_1, X_2, \cdots, X_n)\}$$

is achieved by a rule based only on the *relative ranks* of the $X$'s—or prove that no such sequence exists.

Ferguson has come within *epsilon* of solving this problem. He has exhibited exchangeable sequences, for each $n$ and $\varepsilon > 0$, such that the best rule based only on relative ranks has success probability within $\varepsilon$ of the supremum. But he has left open the question of whether this supremum can actually be attained.

For $n = 2$, the answer is easy; there is no such sequence. The following elementary argument, which