THE EFFECT OF OVERCONFIDENCE ON SEQUENTIAL DECISION MAKING: AN EXPERIMENTAL INVESTIGATION OF THE CLASSICAL SECRETARY PROBLEM

A THESIS

Presented to

The Faculty of the Department of Economics and Business

The Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Kate Hoffman

May 2014

THE EFFECT OF OVERCONFIDENCE ON SEQUENTIAL DECISION MAKING:
AN EXPERIMENTAL INVESTIGATION OF THE CLASSICAL SECRETARY PROBLEM

Kate Hoffman

May 2014

Economics

Abstract

This experimental economics thesis uses the Classical Secretary Problem (CSP) to simulate a sequential observation and selection problem in the context of employer hiring decisions. The CSP is paired with an overconfidence test to examine whether there is a relationship between a subject's level of overconfidence and his or her success in making optimal hiring decisions. No significant differences were found between subjects with varying overconfidence levels, but significant deviation was found between subjects' behaviors and the selections dictated by the optimal policy. Significant learning among all subjects was also discovered between the first and second half of the CSP, indicating a tendency among subjects to revise strategies and correct mistakes.

KEYWORDS: (Experimental economics, Overconfidence, Individual decision making, Secretary Problem, Sequential search)

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

Say you've just accepted a new job, and you now have to relocate to a new city. You don't know anyone who lives in this city, and you've never been there. Yet, despite your lack of knowledge about the city, you must find a new apartment to rent. You organize a number of apartment tours and showings, but due to your lack of knowledge about the real estate market or the city, all you know when you view an apartment is how it compares to the apartments you've already seen. Unfortunately, the real estate market is booming so you have to decide on the spot, during the showing, whether you'd like to accept or reject the current option. If you accept it, the realtor will draw up the papers and you'll sign the lease. If you reject it, however, you can't change your mind and recall it later – someone else will have already rented it. And, given that your new job is starting soon, if you reach the very end of your options without making a selection, you must accept whatever comes last – which could be anything from the very best apartment to the very worst.

Sequential observation and selection problems like the one outlined above exist in many real-world situations. Whether attempting to rent a new apartment (Zwick, Rapoport, Lo, & Muthukrishnan, 2003), searching for the cheapest gas station along the highway, picking a marriage partner or buying a used car, the decision maker is often only able to see his or her options one at a time. Although the goal at hand is, of course, to select the *best* of the available choices, it is often difficult to do so with limited information about the alternatives that have yet to be seen.

There is, however, a solution – an optimal policy that outlines a specific number of options that are to be skipped, regardless of how they rank relative to each other. Once this theoretical predetermined block is skipped over, the equilibrium solution is to select the next best option that comes along.

Prior empirical research, however, has shown that most people tend to stop searching too early, with respect to the optimal policy. Research has also found that overconfident individuals tend to overestimate their ability to make optimal decisions, as their confidence is not aligned with the reality of their knowledge. This applies to sequential observation and selection problems, as excessive self-assurance in regard to making decisions may cause overly confident individuals to be more likely to stop searching too early or make a rash decision, thus decreasing their likelihood of making an optimal decision.

In this thesis, an experiment will be used to investigate whether overconfidence affects subjects' behaviors and decisions in a sequential search decision task.

2. LITERATURE REVIEW

One of the most common examples of a sequential observation and selection decision problem is the "Secretary Problem" (SP). The Secretary Problem, which simulates an employer's hiring decisions, first appeared in February 1960, in a column written by Martin Gardner in *Scientific American* (Ferguson, 1989). Lindley (1961) followed soon after as the first to solve the SP. Gilbert and Mosteller (1966) contributed a more basic and generalized paper on the SP, which helped to fuel the explosion of ideas, generalizations and effort surrounding the SP since 1972 (Seale & Rapoport, 1997). There are many variations of the Secretary Problem, but its original, simplest form – referred to as the Classical Secretary Problem (CSP) – forms the basis of this study.

Many experiments have been conducted involving the CSP and its generalizations; it has been especially common to conduct research on the effects of relaxing one or more of its common assumptions. What has not been thoroughly researched, however, is the relationship between overconfidence and decision makers' performances on the CSP. This is important to investigate because "generally, overconfidence has a negative effect on decision quality" (Zacharakis & Shepherd, 2001, p. 313).

Overconfidence – also known as the overconfidence bias – refers to "the widespread prevalence of positive illusions and self-enhancement biases" among all people (Fast, Sivanathan, Mayer, & Galinsky, 2012, p. 250). "Overconfident people make probability judgments that are more extreme than they should, given the evidence and their knowledge" (Zacharakis & Shepherd, 2001, p. 311). Klayman, Soll, González-Vallejo, & Barlas (1999, p. 219) agree, as they add: "With regard to confidence, people's judgments about the quality of their information include some unsystematic error. Given an imperfect correlation between accuracy and confidence, it is inevitable that low accuracy is on average associated with not-so-low confidence, and so on." This applies directly to the CSP, as overconfident individuals may overestimate the true probability that a potential employee being interviewed is indeed the very best applicant; this bias may be further increased in scenarios when decision makers are especially incentivized to pick the very best option for a monetary reward.

Overconfidence is widespread and of great practical importance (Griffin & Varey, 1996), and can be seen in many common everyday decisions; consumers, for example, are often overly confident in their decisions about what to buy, and think they know more than they actually do (Alba & Hutchinson, 2000). People are constantly processing incomplete information – a consumer who is searching for new shoes, for example, may find a pair and be confident the shoes are being sold at the lowest price available. The same shoes may be discounted even more just a few stores over, yet the consumer is overly confident in his or her decision to buy and thus stops the search before reaching the best choice (Alba & Hutchinson, 2000). Indeed, overconfident decision makers may prematurely limit their information search, "commit resources without pausing to consider additional information" (Mahajan, 1992, p. 329), and lose perspective of the true limits of their metaknowledge (Zacharakis & Shepherd, 2001).

Accordingly, J. Edward Russo and Paul J. H. Schoemaker (1992) offer a "confidence quiz" that "measures something called *metaknowledge*: an appreciation of what we do know and what we do not know... Metaknowledge concerns a higher level of expertise: understanding the nature, scope, and limits of our basic, or *primary knowledge*" (p. 8). Russo and Schoemaker (1992) used this quiz in many different contexts, most notably between employees and managers. This is important in the realm of decision making, as "to size up and factor uncertainty into our judgments is crucial to successful decision making. Experimental evidence suggests that this is a serious weakness in human judgment" (Russo & Schoemaker, 1992, p. 9).

The overall trend of decision makers stopping earlier than the optimal point in a sequential search has been observed in many studies related to the CSP. The question becomes, then, to what extent overconfidence may influence a decision maker's choice to stop sooner than otherwise predicted by economic theory. After all, according to Klayman et al. (1999, p. 219), "people want to think they are intelligent and knowledgeable," and thus overconfident people may be much more self-assured in their decisions to accept applicants and choose to search less; people who are not overconfident, on the other hand, may be more likely to search for a longer period of time, as they do not have the same conviction as those who are overconfident.

3. THEORY

In the classical version of the secretary problem (CSP), a fixed and known number of applicants (*n*) are presented to the decision maker (DM) sequentially, and in random order. As each potential employee is interviewed (viewed), the DM must either accept or reject the applicant. If the DM accepts, the open position is filled and thus the trial is terminated. If the DM rejects, the next applicant in the random order is presented, and the DM must again decide whether to accept or reject. In the version of the CSP used in the present study, the subject is only paid if s/he correctly selects the very best of the *n* applicants available. Thus, given this 0 – 1 payoff function, the DM's objective is to maximize the probability that the chosen applicant is indeed the best possible choice.

The CSP is based on eight assumptions (Bearden, Rapoport, & Murphy, 2006). They are as follows:

1. There is only one position to be filled.
2. *n,* the number of applicants for the position, is known before the search starts.
3. The decision maker (DM) can rank the *n* applicants from best to worst without any ties.
4. The *n* applicants are interviewed sequentially, one at a time and in a random order. Every possible ordering of the *n* applicants is equally likely.

5. As each individual applicant is being interviewed, the DM must choose to accept or reject the applicant. Accepting an applicant will terminate the search, while rejecting an applicant will continue the search onto interviewing the next applicant, if there is one.
6. Each decision to either accept or reject the current applicant will be solely based on the relative ranks of all applicants interviewed so far.
7. Once rejected, an applicant cannot be recalled.
8. The DM's objective is to select the best applicant. In this scenario, only selecting the best applicant constitutes a win – anything less is a loss.

These assumptions are arguably very restrictive, especially compared to what would usually be experienced in a real decision-making scenario. Each of the eight assumptions above can, however, be relaxed to offer more realistic instances of individual choice behavior. The first assumption can be relaxed to assume that several positions are available, instead of only one (Gilbert & Mosteller, 1966). Assumption two can be changed so that the DM only knows the distribution of the value of n, instead of knowing the precise number of applicants (Gianini-Pettit, 1979; Presman & Sonin, 1973; Rasmussen & Robbins, 1975). Assumption seven has many generalizations, as it is considered possibly the most restrictive of the eight assumptions (Seale & Rapoport, 1997). Such generalizations include the possibility of recalling rejected applicants, with an associated probability that they are no longer available (Smith, 1975). Another generalization associated with the seventh assumption is that subjects are allowed to recall one of the last m applicants ($m < n$) with the stipulation that if the previously rejected applicant is no longer available, the subject must continue on with the interview process (Corbin, 1980; Yang, 1974). Lastly, the eighth assumption can also be generalized in many ways. In the present study, the DM is indeed only satisfied with the best. In other scenarios, however, the DM may receive alternate variations of utilities for selecting applicants other than the very best one. In other words, the DM may receive utility u_i if the applicant selected is the i^{th} best (Seale & Rapoport, 1997).

Following the dynamic programming work of Lindley (1961), a numerical procedure for determining the optimal policy for the CSP under investigation in the present study was used. In the CSP, the state of the decision process at each period is described by two integers (r, s), where r is the number of applicants presented so far and s is the relative rank of the rth or last presented applicant (Seale & Rapoport, 1997). When each new item is presented, the new state of the decision process is ($r + 1$, s'), where s' has equal likelihood of being any one of the integers 1, 2, . . . , $r + 1$ (Seale & Rapoport, 1997).

When $s = 1$, the applicant is referred to as a "candidate" for acceptance. The probability that it is indeed the best of all n applicants is r/n. *If* $s \neq 1$, on the other hand, the rth item cannot possibly be the best applicant, so there is no point in choosing to accept. The equation for the maximum probability of choosing the best applicant, with (r, s) as the state of the decision process and with dynamic programming (Lindley, 1961), can be defined by equation 3.1.

$$a_r = 1/r + 1/(r + 1) + . . . + 1/(n - 1) \qquad (3.1)$$

The optimal decision in state (r, 1) is to stop if $a_r < 1$ and continue if $a_r > 1$. The integer r^* represents r when $a_{r-1} \geq 1 > a_r$. Thus, the optimal policy is to reject the first $r^* - 1$ applicants and accept the next candidate ($s = 1$) that comes along. This is a *cutoff* policy, and rejecting the first $r^* - 1$ applicants allows the associated probability of winning to be defined by equation 3.2.

$$(r^* - 1)a_{r^*-1}/n \qquad (3.2)$$

As n approaches α, both r^*/n and the probability of choosing the best applicant approach $1/e = 0.368$. In this specific experiment, where $n = 30$, $a_{r-1} \geq 1 > a_r$ when $r = 12$. Thus, $r^* = 12$

and the optimal policy in this case is to skip the first eleven ($r^* – 1$) applicants and pick the next candidate (relative rank = 1) thereafter. The associated probability of selecting the optimal candidate when $r^* = 12$ is 0.3787. Additionally, $r^*/n = 0.4$.

When used in the context of the present study's experimental design, this optimal policy means that a subject should reject the first eleven applicants and select the next candidate that comes along. If no candidate comes along after the initial eleven applicants have been skipped, then the subject should continue to search until s/he reaches the last of the 30 applicants in the round.

In the present study, subjects' decisions during the CSP are compared to their results on an overconfidence quiz. This quiz, which was presented to subjects as a "general knowledge test" so as to prevent the intentions from skewing subjects' answers, measures metaknowledge, which is defined by Russo and Schoemaker (1992) as "an understanding of the limits of our knowledge" (p. 7). The quiz used in this study asks each subject to provide a low value and a high value for ten general knowledge questions (see Appendix I) such that s/he is 90% sure that the correct answer is contained within the given range. In other words, the subject should answer nine of the ten questions on the overconfidence quiz correctly. This is a difficult task given the average person does not precisely know the answer to any of these questions. Russo and Schoemaker (1992) claim, however, "whether you know a lot or a little about a subject, you are still responsible for knowing how much you don't know" (p. 9). Thus, this quiz can help to size up and factor uncertainty into our judgments to help with more successful future decision making. In this study, the overconfidence quiz allowed subjects to be sorted into separate classifications in order to determine if the behaviors and decisions made during the CSP vary significantly between groups of subjects with differing overconfidence levels.

This quiz measures overconfidence by testing for subjects' calibration, which is the degree to which a subject's confidence matches his or her accuracy. If a subject is well calibrated, s/he should answer nine of the ten questions on the overconfidence quiz correctly. If a subject answers less than nine correctly, s/he shows overconfidence bias in his or her degree of knowledge. The lower the amount of correct answers, the more severe the bias.

I hypothesized that subjects who tested as less overconfident would successfully choose optimal candidates more often than those who tested as extremely overconfident. Likewise, I hypothesized that very overconfident subjects would stop searching earlier than those who were less overconfident.

4. METHODS

Forty subjects (twenty female and twenty male) participated in this experiment. All participants were undergraduate Colorado College students recruited through visits to three undergraduate classes. Each recruiting visit included a brief description of the individual decision-making experiment designed by the researcher and the payment offered for both showing up and for optimal performance. Interested students were asked to sign up to participate during a specific time slot over a four-day period.

The experiment was conducted in a private office within the Colorado College Economics and Business Department. Subjects participated in the experiment individually and as each subject arrived, s/he was seated across from the researcher. Before beginning the experiment, each subject was asked to read and sign a consent form. Once the form was signed and handed to the researcher, the subject was given verbal and written instructions for the first part of the experiment: the overconfidence quiz. This overconfidence quiz was given to the subject under the guise of a "general knowledge test," so as to avoid alerting him/her to the specific nature of what was being measured.

The instructions for the overconfidence quiz asked each subject to write down a low and high value for ten specific questions (see Appendix I) such that s/he was 90% sure that each range contained the correct answer for the corresponding question (Russo & Schoemaker, 1992). The challenge was to be neither too narrow nor too wide; if the subject had no idea of the answer s/he could give a wide range, but if s/he was quite certain, s/he should give a narrow range. No outside help (computer, phones, tablets, etc.) was permitted.

Once the subject completed the overconfidence quiz, s/he was asked to hand the quiz to the researcher and listen as the researcher explained the instructions for the CSP. The instructions emphasized that 30 applicants (who could be ranked from 1 – 30 with no ties) would be presented in random order, and that only the relative rankings of applicants would be displayed. The instructions also placed special emphasis on how relative ranks changed with each new applicant.

After hearing the instructions, the subject was asked to complete two practice rounds to verify understanding of the sequential decision problem at hand. Two rounds of $n = 10$ applicants were given (absolute ranks: 4, 8, 1, 6, 5, 7, 2, 10, 3, 9 and 7, 4, 2, 5, 6, 3, 8, 9, 10, 1) and the updating of relative ranks throughout each round was illustrated. Once a subject completed both practice exercises, thereby demonstrating full understanding of the experimental instructions, the researcher answered any questions asked by the subject and reminded him/her once more that payoff was contingent on performance. The researcher then presented the experimental rounds contained in the binders.

In terms of payment, each subject received $5.00 for showing up to participate. Additional payment was dependent upon performance, as the subject earned an additional $1.00 each time s/he successfully picked the optimal candidate (relative rank = 1). Subjects could, therefore, earn up to $20.00 on top of the show up fee, for a maximum of $25.00 if s/he chose the optimal candidate in all twenty rounds.

Each subject faced twenty independent replications (trials) of the CSP with $n = 30$ applicants. All of the pages in the binders were compiled to display the updated rank information for each of the 30 applicants in all twenty trials (see Appendix II for examples). Each trial followed the same pattern. The relative rank of the first applicant was displayed (by definition, the relative rank of the first applicant is always "1"). The subject was allowed to accept (hire) or reject the applicant (thus choosing to interview the next applicant). If the subject decided to interview the next applicant, the relative rankings of both the past and the current applicants were displayed on the next page in the binder. After interviewing two applicants, for example, the relative rankings of the last and previous to last applicants were either (2, 1) or (1, 2). Similarly, after three applicants, the relative rankings were (1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2) or (3, 2, 1).

Each time the subject rejected an applicant, s/he was given another opportunity to either hire the current applicant or interview another. This process continued until the subject either made a selection (accepted an applicant), or interviewed all 30 applicants. If the subject reached the final applicant without making a selection, s/he was forced to accept whoever came last. If a selection occurred, however, the researcher immediately flipped to the last page for that specific trial within the binder to reveal the absolute ranking of the chosen applicant, as well as the absolute rankings of the other 29 applicants. The subject was then informed of the absolute ranking of the selected applicant.

If the applicant selected was indeed the optimal candidate (absolute rank = 1), the researcher informed the subject that s/he had made a correct selection, and added $1.00 to the subject's cumulative earnings. The updated cumulative earnings were continuously displayed in a small book that was placed in front of the subject, and this amount was updated with each additional correct candidate selection.

Following the completion of trial twenty, subjects were paid their cumulative earnings and dismissed from the office. Subjects completed the experiment on average in approximately 25 minutes. The mean payoff per subject was \$10.73 (min = \$7.00, max = \$14.00), which included the \$5.00 guaranteed for showing up to participate.

5. RESULTS

The 40 subjects completed 800 trials in total, selected candidates in 539 trials (67.38%), hired applicants who were not candidates in 148 trials (18.50%) and interviewed all 30 applicants without making a selection in 113 trials (14.13%). Overall, the subjects made correct selections in 229 trials (28.63%). The subjects selected candidates in an average of 13.48 of their twenty total trials (s = 1.92, min = 9, max = 17), selected non-candidates in an average of 3.70 trials (s = 2.83, min = 0, max = 10) and made no selection in an average of 2.83 trials (s = 1.89, min = 0, max = 7). Subjects correctly selected the very best candidate in an average of 5.73 trials (s = 1.84), with a minimum of two correct choices and a maximum of nine. Accordingly, mean earnings per subject were $5.73. Table 5.1 displays a synopsis of the experimental results for all participants.

Table 5.1

Summary of Experimental Results for All Subjects

Condition	Total	Mean ($\bar{x}$)	Median	Mode	Std dev (s)	Min	Max
Candidates selected	539	13.48	14	14	1.92	9	17
Non-candidates selected	148	3.70	4	1	2.83	0	10
No selections	113	2.83	2	1	1.89	0	7
Correct selections	229	5.73	6	7	1.84	2	9
Completed trials	800						

Note: Std dev = standard deviation; Min = minimum; Max = maximum.

Across all subjects and all twenty trials, the modal stopping point was $r = 30$. The mode is the number that appears most often, so this indicates that subjects selected the 30^{th} and final applicant more than any other within the 30 total applicants in each round. Adhering to the optimal policy, the theoretical equilibrium also results in a modal stopping point of $r = 30$.

The median stopping point among all subjects was found to be $r = 17$, which is earlier than the median stopping point of $r = 20$ that was found by using the optimal policy. This value of $r = 20$ was found by skipping the first eleven applicants in each round (as outlined in the optimal policy, as $r^* = 12$), and then choosing the next candidate (relative rank = 1) that came along, if there was one. Once the optimal policy's stopping points for all twenty rounds were determined, the middle-most number was selected. Thus, the median stopping point among subjects was earlier than that of the optimal policy, which suggests that subjects tended to stop their searching too early.

Subjects also appear to stop too early when evaluated by the mean stopping point, as they stopped on average at $r = 17.85$, which is earlier than the optimal policy's average stopping point of $r = 21.70$. This value of $r = 21.70$ was calculated in the same fashion as the median stopping point for the optimal policy, with the difference being that instead of taking the middle-most number, the stopping points for all rounds were averaged. When a t-test was used to compare all subjects' average stopping points against the stopping points dictated by the optimal policy, the difference was significant ($p = 0.025$). Additionally, when all subjects' stopping points for individual trials (not averaged) were compiled and compared to those of the optimal policy, the difference was also found to be significant ($p = 0.017$). These comparisons indicate that the behaviors of subjects differed significantly from the behaviors that result from adhering to the optimal policy.

In searching for evidence of learning, mean stopping times for the first half of the session (block 1: trials 1 – 10) and the second half (block 2: trials 11 – 20) were computed. A paired sample t-test was used to test the null hypothesis of no difference between mean stopping times during the two blocks. Mean stopping time for all subjects in the first block ($r = 17.15$) was earlier than in the second block ($r = 18.56$). This difference between the mean stopping time in block 1 versus block 2 is highly significant ($p < 0.001$), which suggests that the propensity of the subjects to select applicants too early decreased with experience in playing the CSP, indicating that learning did in fact take place between block 1 and block 2.

Table 5.2

Proportion of Correct Selections for All Subjects

	All trials	Block 1	Block 2
All subjects	0.29	0.22	0.36
Extremely overconfident	0.32	0.26	0.37
Mildly overconfident	0.28	0.18	0.38
Optimal policy	0.35	0.20	0.50

Note: Block 1 = trials 1 – 10; Block 2 = trials 11 – 20. Extremely confident subjects are subjects who correctly answered 0 – 2 of the ten questions on the overconfidence quiz, and mildly overconfident subjects are those who answered 5 – 10 questions correctly. The proportions of correct selections for all subjects, extremely overconfident subjects and mildly overconfident subjects refer to the average proportion of correct selections ($\bar{x}$) for all 40 subjects. The proportion of correct selections for the optimal policy refers to the specific set of selections that results from adhering to this strategy.

Table 5.2 shows the mean proportion of correct selections among all 40 subjects for all trials, compared to the proportion of correct selections found when adhering to the optimal policy. Among all trials and all subjects, the mean proportion of correct selections was 0.29, which is less than the optimal policy's proportion of 0.35 correct selections. The mean proportions of correct selections found in block 1 and block 2 among all subjects can also be seen in Table 5.2, in order to once again examine the results for evidence of learning. The mean proportion of correct selections among all subjects in block 1 ($\bar{x}$ = 0.22) was less than in block 2 ($\bar{x}$ = 0.36). When compared to the optimal policy's proportion of 0.20 correct selections in block 1 and 0.50 correct selections in block 2, the mean proportion of correct selections for subjects was slightly higher in block 1, but much lower in block 2. A paired-sample t-test was used to test the null hypothesis of no difference between the proportion of correct selections in the two blocks, and the result was highly significant ($p < 0.001$). Thus, subjects learned to stop selecting applicants too early – as seen when examining the mean stopping point – and were also significantly more successful at picking the correct candidate in the second half of the CSP.

Overconfidence quiz results and classifications. The CSP used in the present study was preceded by an overconfidence quiz. This quiz examined how overconfident subjects were in their judgments[1]. Table 5.3 shows the distribution of subjects' scores on the overconfidence quiz. Thirteen subjects provided correct ranges for 0 – 2 of the ten questions; seventeen subjects provided correct ranges for 3 – 4 questions; nine subjects provided correct ranges for 5 – 6 questions; and one person provided correct ranges for 7 – 10 questions.

Table 5.3

Distribution of All Subjects' Results on the Overconfidence Quiz

	Overconfidence quiz score				
	0 – 2 correct	3 – 4 correct	5 – 6 correct	7 – 8 correct	9 – 10 correct
Number of subjects	13	17	9	1	0

All 40 subjects in the present study answered less than nine questions correctly and, as a result, tested as overconfident. These results are not altogether surprising, however, as Russo and

[1] The overconfidence quiz asked subjects to provide a low value and a high value for ten general knowledge questions such that they were 90% confident the right answer was contained. Thus, if done correctly, nine out of the ten ranges should have contained the correct answers.

Schoemaker (1992) noted: “Of the 2,000-plus individuals to whom we have given a ten-question quiz using 90 percent confidence intervals, fewer than 1 percent were not overconfident” (p. 9).

All subjects tested as overconfident, so for the purposes of this study I have chosen to examine whether there is a difference in the decisions made by the most overconfident subjects (those who answered 0 – 2 questions correctly) and subjects who were less overconfident (those who answered 5 – 10 correctly). The thirteen subjects who answered 0 – 2 questions correctly on the overconfidence quiz are referred to as *extremely overconfident*, and the ten subjects who answered 5 – 10 questions correctly are referred to as *mildly overconfident*. I chose to omit the results of the seventeen subjects who answered 3 – 4 questions correctly because, given that all subjects tested as overconfident, I will focus on examining whether differences exist between the two extremes of the available spectrum of scores.

Table 5.4

Summary of Experimental Results within Classifications of Subjects

Condition	Extremely overconfident	Mildly overconfident	Optimal policy
Candidates selected	14.15 (70.75%)	12.80 (64.00%)	13.00 (65.00%)
Non-candidates selected	3.62 (18.10%)	3.90 (19.50%)	0.00 (00.00%)
No selections	2.23 (11.15%)	3.30 (16.50%)	7.00 (35.00%)
Correct selections	6.31 (31.55%)	5.50 (27.50%)	7.00 (35.00%)

Note: Numbers without parentheses denote the mean result for each condition, within each classification, out of the twenty experimental trials. Percentages, which are denoted by parentheses, were calculated by dividing each mean by twenty, which was the total number of trials presented to each individual subject.

Extremely overconfident subjects. Table 5.4 displays a synopsis of the results for subjects who answered 0 – 2 questions correctly on the confidence quiz, alongside those who answered 5 – 10 correctly and the results of the optimal policy. These thirteen extremely overconfident subjects completed 260 trials, selected candidates in 184 trials (70.75%), interviewed all 30 applicants without making a selection in 29 trials (11.15%) and hired applicants who were not candidates in 47 trials (18.10%). They made correct selections in 82 trials (31.55%), and mean earnings per subject in this classification were $6.31, which is higher than the overall mean earnings of $5.73 found among all subjects. Although the mean number of correct selections among the extremely overconfident subjects ($\bar{x}$ = 6.31) was less than the seven correct selections that result from the optimal policy, it was more than both the mean success rate found among all subjects ($\bar{x}$ = 5.73), and among those who were classified as mildly overconfident ($\bar{x}$ = 5.50). Yet when separate t-tests were run, neither difference was found to be significant (p = 0.17 and p = 0.15, respectively).

I computed the mean stopping time (period) for the extremely overconfident subjects across all twenty trials. The mean stopping time for these subjects was r = 17.57, which is earlier than the mean stopping time that results from the optimal decision rule (r = 21.70). In searching for evidence of learning, I separately computed the mean stopping times for blocks 1 and 2 within this classification. Mean stopping time in the first block (r = 17.21) was earlier than in the second block (r = 17.94). A paired sample t-test was used to test the null hypothesis of no difference between mean stopping times during the two blocks. These numbers suggest that the propensity of the subjects to select applicants too early decreased with experience in playing the CSP, yet the difference between block 1 and block 2 is not significant (p = 0.15).

Once again searching for evidence of learning, I examined the mean proportion of correct selections during block 1 (trials 1 – 10) and block 2 (trials 11 – 20) for the thirteen subjects who tested as extremely overconfident. These calculations can be seen in Table 5.2. When the mean proportions of correct selections for block 1 ($\bar{x}$ = 0.26) and block 2 ($\bar{x}$ = 0.37) are compared, it is apparent that subjects in this classification were more successful in block 2. When a paired sample t-test was used to test the null hypothesis of no difference between mean proportions of correct selections during the two blocks, it did indeed become clear that the null hypothesis could be rejected, as subjects in this classification were significantly more successful in the second half of the trials (p = 0.021). Thus, learning is once again evident between block 1 and block 2.

Mildly overconfident subjects. A synopsis of the results for subjects who answered 5 – 10 questions correctly on the confidence quiz can be seen in Table 5.4. These ten mildly overconfident subjects completed 200 trials, selected candidates in 128 trails (64.00%), hired applicants who were not candidates in 39 trials (19.50%) and interviewed all 30 candidates without making a selection in 33 trials (16.50%). Overall, the subjects in this classification made correct selections in 55 trials (27.50%), and each earned an average of $5.50. The mean number of correct candidate selections among the mildly overconfident subjects ($\bar{x}$ = 5.50) was less than the optimal policy's seven correct candidate selections. It was also less than the number of correct selections found among all subjects ($\bar{x}$ = 5.73) and that of those who were classified as extremely overconfident ($\bar{x}$ = 6.31). When separate t-tests were run, however, neither difference was found to be significant (p = 0.36 and p = 0.15, respectively).

I computed the mean stopping time (period) for each mildly overconfident subject across all twenty trials. The mean stopping time for these subjects was r = 18.32, which is earlier than the expected average stopping time under the optimal policy (r = 21.70). It is also later than the overall average stopping time for all subjects (r = 17.85), yet not significantly so (p = 0.27). In searching for evidence of learning, the mean stopping times for blocks 1 and 2 were calculated for this classification as well. A paired sample t-test was used to test the null hypothesis of no difference between mean stopping times during the two blocks. Mean stopping time in the first block (r = 17.73) was earlier than in the second block (r = 18.90). This result suggests that the propensity of the mildly confident subjects to select applicants too early decreased with experience in participating in the CSP, yet this difference between block 1 and block 2 is not significant ($p = 0.10$).

I again separately computed the proportions of correct selections for block 1 (trials 1 – 10) and block 2 (trials 11 – 20) among the ten subjects who tested as mildly overconfident (see Table 5.2). When the mean proportion of correct selections for block 1 ($\bar{x} = 0.18$) and block 2 ($\bar{x} = 0.38$) are compared, it is clear that subjects in this classification were more successful in block 2. When a paired sample t-test was used to test the null hypothesis of no difference between the proportions of correct selections during the two blocks, it did indeed become clear that once again the null hypothesis could be rejected, as subjects in this classification were significantly more successful in the second half of the trials ($p = 0.003$).

Cross-classification analysis. The results then beg the question of whether there are significant differences between the behaviors of mildly and extremely overconfident subjects. First, an independent t-test was run to test the null hypothesis of no difference between mean overall stopping times for the two groups. Mean stopping time during all trials among extremely overconfident subjects ($r = 17.57$) was earlier than that of mildly overconfident subjects ($r = 18.32$), yet this difference was not found to be significant ($p = 0.22$). The difference between mean stopping time in block 1 for extremely overconfident subjects ($r = 17.21$) and mildly overconfident subjects ($r = 17.73$) was also found to be insignificant ($p = 0.34$), as was the block 2 stopping time difference between extremely overconfident subjects ($r = 17.94$) and mildly overconfident subjects ($r = 18.90$) ($p = 0.14$).

I ran t-tests comparing extremely and mildly overconfident subjects' overall proportions of correct selections ($p = 0.19$) as well as in block 1 ($p = 0.06$) and block 2 ($p = 0.42$), but these three tests also failed to yield results compelling enough to be able to reject the null hypotheses of no difference between the two groups of subjects. Likewise, a t-test comparing the number of correct selections for extremely overconfident and mildly overconfident subjects in all twenty trials yielded insignificant results ($p = 0.19$), as did a t-test of the number of non-candidates selected by either classification ($p = 0.41$). In fact, all t-tests comparing extremely overconfident subjects and mildly overconfident subjects resulted in $p > 0.05$, which, for the purposes of this study, indicates that there are no significant differences between the two classifications. Additionally, no significant differences were found between the behaviors of the extremely overconfident subjects or the mildly overconfident subjects versus the behaviors of all 40 subjects as a whole.

6. DISCUSSION

I was unable to reject the null hypothesis that there is no difference between the extremely overconfident and mildly overconfident subjects. This may be caused by problems within the experiment itself; more diversity among subjects' overconfidence levels, more trials and more applicants per trial may have yielded significant differences. This may also be explained by large variation in decision-making processes within the separate classifications themselves. Subjects with similar scores on the overconfidence quiz do not automatically behave identically; extremely overconfident subjects, for example, may have very diverse approaches when faced with the CSP. Another possible reason for the lack of significant difference between subjects of differing overconfidence levels is the variation of educational and experiential background. When faced with this experiment, for example, a student who is classics major may make decisions very differently than a student who is an economics major, independent of overconfidence levels. Thus, this and other confounding variables may have affected the results.

Whatever the reason, no significant difference was found between the behaviors of the extremely overconfident subjects and the mildly overconfident subjects. Data such as the proportion of correct selections and the average stopping point were compared over all twenty trials as a whole, as well as between block 1 (trials 1 – 10) and block 2 (trials 11 – 20). Yet no compelling evidence was found to support either of the original hypotheses – the first hypothesis being that less overconfident subjects would successfully choose optimal candidates more often, and the second being that very overconfident subjects would stop searching earlier than those who were less overconfident.

Although neither hypothesis was supported, data collected from this experiment supported previous findings of significant learning between block 1 and block 2. This was observed between all subjects as a whole, as well as within overconfidence classifications. In all cases, the subjects' mean proportion of correct selections were larger in block 2 than in block 1. Evidence of learning is important in this context, as it shows how subjects of all overconfidence levels adjusted their strategies for solving the situation simulated by the CSP. This may also help to explain the lack of significant difference between overconfidence classifications, as all subjects were observed correcting mistakes such as stopping too early, making rash decisions, accepting non-candidates and rejecting candidates that should have otherwise been selected.

The results of this experiment have also illuminated that all subjects deviated significantly from the behaviors dictated by the optimal policy. When adhering to the optimal policy in the present experiment, the first eleven applicants in each trial are skipped and the next candidate (relative rank = 1) to follow is selected. If no such candidate appears after the eleven skipped applicants, the subject continues rejecting applicants until s/he reaches the very last one. The actual behaviors of subjects in this experiment, however, did not align with this optimal strategy. Some subjects selected non-candidates (relative rank $\neq$ 1), some selected candidates too early and some rejected candidates that should have been selected. This was found among subjects of all overconfidence levels, and may be caused by the fact that subjects do not know how best to approach the given scenario at hand. It is, after all, difficult to know when to stop searching.

7. CONCLUSION

This thesis used experimental means to investigate whether overconfidence has an effect on individual decision making, specifically in the context of employer hiring decisions. During the experiment, subjects participated in both an overconfidence quiz and a simulation of the Classical Secretary Problem (CSP). Previous studies have examined various generalizations of the CSP, but none have examined its relationship to subjects' overconfidence levels.

Previous work in the realm of overconfidence suggests that overconfidence negatively affects accurate decision making, as overconfident individuals are less aware of what they don't know, and thus may be more prone to making rash decisions. The first hypothesis that shaped this experiment was that subjects who tested as more extremely overconfident would be less successful at picking optimal candidates in the CSP; the second hypothesis was that very overconfident subjects would stop searching earlier than subjects who were less overconfident.

These hypotheses, however, were not supported by the results of this experiment. In fact, no significant difference whatsoever was found between subjects of differing levels of overconfidence. When t-tests were used to compare the behaviors of extremely overconfident subjects with those of mildly overconfident subjects during all twenty trials, during block 1 (trials 1 – 10) and during block 2 (trials 11 – 20), all resulted in p-values greater than 0.05, which, in the context of the present study, indicated insignificance.

A significant difference was found, however, between the behavior of the forty subjects in this experiment and the choices dictated by the optimal policy. Significance was also found between the decisions made by all subjects in the first half of the rounds (trials 1 – 10) versus the second half (trials 11 – 20). The success of subjects in selecting the optimal candidate in the first half of the twenty trials (block 1) vs. the second half (block 2), for example, significantly different. This indicates that learning occurred as subjects progressed throughout the experiment, and this was true for the subjects as a whole, as well as within the extremely confident and mildly confident classifications. This is relevant because it shows how people of all confidence levels revise their decision-making strategies and learn from their mistakes, which range from stopping too early to making brash decisions.

Whether the decision in question involves choosing an apartment, a gas station along the highway or a new employee, the way to strategically secure the best possible results involves skipping the first third or so of the options, and picking the very best option that follows. When faced with a sequential search decision-making scenario of this type, some people tend to search too long, while others search too little. With a little learning, however, the chances of making a better decision can be improved, regardless of overconfidence.

APPENDIX I

The Overconfidence Quiz

General knowledge test

For each of the following ten questions, please provide a **low value** and a **high value** such that you are **90% confident** that the correct answer falls between the two.
Your challenge is to be neither too narrow nor too wide. If you have no idea, give a wide range; if you are quite certain, give a narrow range. Even if you have no idea what the correct answer is, you should still be able to enter a low and high value such that you are **90% certain** that the answer lies within your numerical range of responses. No outside help is permitted. Good luck!

Name: ______________________________ Grade: ________________
Date: ___________ Time: ___________ Gender: _____

---

1. **What was Martin Luther King Jr.'s age at the time of his death?**
   Low value: _______________ High value: _______________

2. **What is the length of the Nile (in miles)?**
   Low value: _______________ High value: _______________

3. **How many countries are members of the Organization of Petroleum Exporting Countries (OPEC)?**
   Low value: _______________ High value: _______________

4. **How many books are there in the Old Testament (or Jewish Scripture)?**
   Low value: _______________ High value: _______________

5. **What is the diameter of the moon (in miles)?**
   Low value: _______________ High value: _______________

6. **What is the weight of an empty Boeing 747 (in pounds)?**
   Low value: _______________ High value: _______________

7. **What year was Wolfgang Amadeus Mozart born?**
   Low value: _______________ High value: _______________

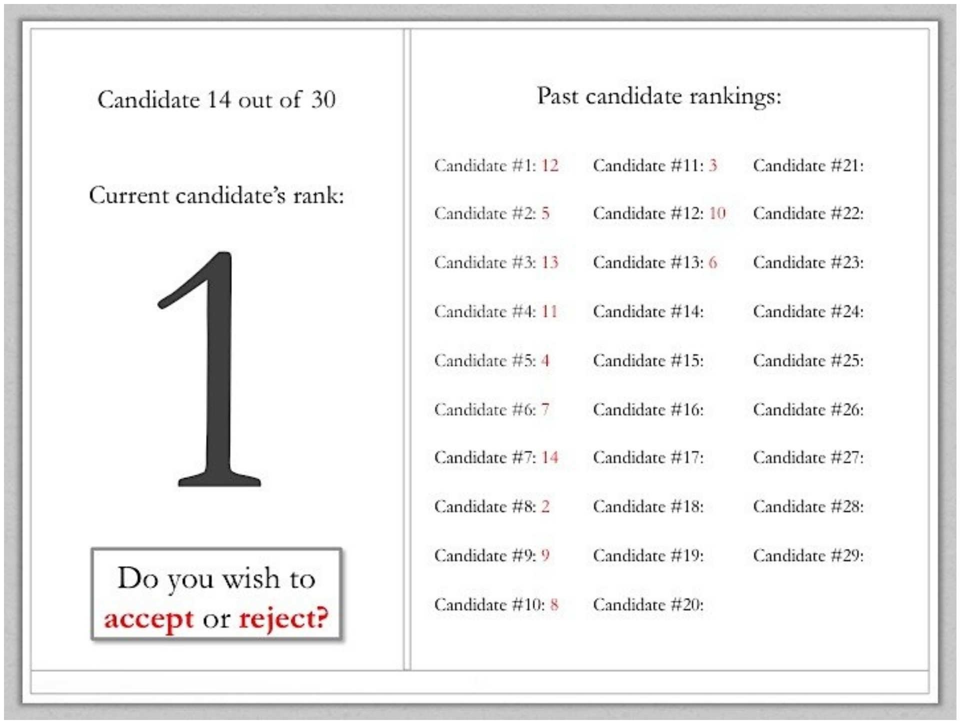
8. **What is the gestation period of an Asian elephant (in days)?**
   Low value: _______________ High value: _______________

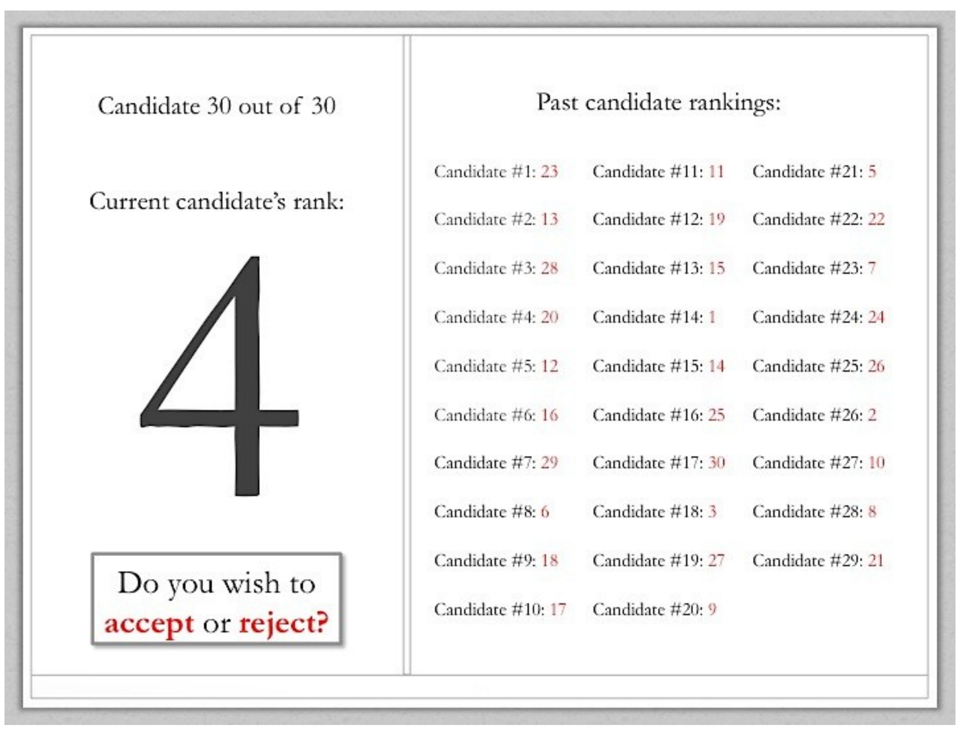
9. **What is the air distance from London to Tokyo (in miles)?**
   Low value: _______________ High value: _______________

10. **How deep is the deepest known point in the ocean (in feet)?**
    Low value: _______________ High value: ______________

APPENDIX II

Example Slides Used in Secretary Problem Binders (both are from trial 10)





REFERENCES

Alba, J. W., & Hutchinson, J.W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research*, 27(2), 123-156. doi: 10.1086/314317

Bearden, J. N., Rapoport, A., & Murphy, R. O. (2006). Sequential observation and selection with rank-dependent payoffs: An experimental study. *Management Science*, 52(9), 1437-1449. Retrieved from http://0-www.jstor.org.tiger.coloradocollege.edu/stable/20110615

Corbin, R. M. (1980). The secretary problem as a model of choice. *Journal of Mathematical Psychology, 21*(1), 1-29. doi:10.1016/0022-2496(80)90025-5

Fast, N. J., Sivanathan, N., Mayer, N. D., & Galinsky, A. D. (2012). Power and overconfident decision-making. *Organizational Behavior and Human Decision Processes*, 117(2), 249-260. doi:10.1016/j.obhdp.2011.11.009

Ferguson, T. S. (1989). Who solved the secretary problem? *Statistical Science, 4*(3), 282-289. doi:10.1214/ss/1177012493

Gianini-Pettitt, J. (1979). Optimal selection based on relative ranks with a random number of individuals. *Advances in Applied Probability*, 11(4), 720-736. doi:10.2307/1426856

Gilbert, J. P., & Mosteller, F. (1966). Recognizing the maximum of a sequence. *Journal of the American Statistical Association, 61*(313), 35-73. doi:10.1080/01621459.1966.10502008

Griffin, D.W., and Varey, C.A. (1996). Towards a consensus on overconfidence. *Organizational Behavior and Human Decision Processes,* 65(3), 227-231. doi:10.1006/obhd.1996.0023

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*(3), 216-247. doi:10.1006/obhd.1999.2847

Lindley, D. V. (1961). Dynamic programming and decision theory. *Journal of the Royal Statistical Society*. Series C (Applied Statistics), 10(1), 39-51. doi:10.2307/2985407

Mahajan, J. (1992). The overconfidence effect in marketing management predictions. *Journal of Marketing Research*, 29, 329-342. Retrieved from http://0-search.ebscohost.com.tiger.coloradocollege.edu/login.aspx?direct=true&db=ufh&AN=9602160656

Presman, E., & Sonin, I. (1973). The best choice problem for a random number of objects. *Theory of Probability & its Applications, 17*(4), 657-668. doi:10.1137/1117078

Rasmussen, W. T., & Robbins, H. (1975). The candidate problem with unknown population size. *Journal of Applied Probability*, 12(4), 692-701. doi:10.2307/3212720

Russo, J. E., & Schoemaker, P. J. (1992). Managing overconfidence. *Sloan Management Review*, 33(2), 7-17. doi:10.1137/1117078

Seale, D. A., & Rapoport, A. (1997). Sequential decision making with relative ranks: An experimental investigation of the "secretary problem". *Organizational Behavior and Human Decision Processes, 69*(3), 221-236. doi:10.1006/obhd.1997.2683

Smith, M. H. (1975). A secretary problem with uncertain employment. *Journal of Applied Probability*, 12(3), 620-624. doi:10.2307/3212880

Yang, M. C. K. (1974). Recognizing the maximum of a random sequence based on relative rank with backward solicitation. *Journal of Applied Probability*, 11(3), 504-512. doi:10.2307/3212694

Zacharakis, A. L., & Shepherd, D. A. (2001). The nature of information and overconfidence on venture capitalists' decision making. *Journal of Business Venturing*, 16(4), 311-332. doi:10.1016/S0883-9026(99)00052-X

Zwick, R, Rapoport, A., Lo, A. K. C., and Muthukrishnan, A. V. (2003). Consumer sequential search: Not enough or too much? *Marketing Science*, 22, 503-519. Retrieved from http://0-www.jstor.org.tiger.coloradocollege.edu/stable/4129735