Working Papers

Institute of Mathematical Economics

413

February 2009

The Best Choice Problem under Ambiguity

Tatjana Chudjakow and Frank Riedel

IMW · Bielefeld University
Postfach 100131
33501 Bielefeld · Germany

email: imw@wiwi.uni-bielefeld.de
http://www.wiwi.uni-bielefeld.de/~imw/Papers/showpaper.php?413
ISSN: 0931-6558

The Best Choice Problem under Ambiguity

Tatjana Chudjakow Frank Riedel

Institute of Mathematical Economics
Bielefeld University

February 2, 2009

Abstract

We model and solve Best Choice Problems in the multiple prior framework: An ambiguity averse decision maker aims to choose the best among a fixed number of applicants that appear sequentially in a random order. The decision faces ambiguity about the probability that a candidate – a relatively top applicant — is actually best among all applicants. We show that our model covers the classical secretary problem, but also other interesting classes of problems. We provide a closed form solution of the problem for time-consistent priors using minimax backward induction. As in the classical case the derived stopping strategy is simple. Ambiguity can lead to substantial differences to the classical threshold rule.

1 Introduction

The classical secretary problem is one of the most popular problems in the area of Applied Probability, Economics, and related fields. Since its appearance in 1960 a rich variety on extensions and additional features has been discussed in the scientific[1] and popular[2] literature.

In the classical secretary problem introduced by Gardner (1960) an employer sequentially observes N girls applying for a secretary job appearing in a random order. She can only observe the relative rank of the current girl compared to applicants already observed and has no additional information on their quality. The applicants can be strictly ordered. Immediately after the interview the employer has to accept the girl or to continue the observation. Rejected applicants do not come back. Based on this information the agent aims to maximize the probability of finding the best girl[3].

Most of the classical literature in this field assumes that the girls come in random order where all orderings are equally likely. The solution is surprisingly simple. It prescribes to reject a known fraction of girls, approximately $\frac{N}{e}$, and to accept afterwards the next *candidate*, i.e. a girl with relative rank 1. Such stopping rules are called simple. This strategy performs very well: Indeed, the chance of success is approximately $36,8\% \approx \frac{1}{e}$ for large N.

This nice and surprising result is based on the strong assumption that the girls arrive randomly, all possible orderings being equally likely. There are good reasons to care about the robustness of this assumption. From a subjective point of view, the decision maker might not feel secure about the distribution of arrivals — she might face "ambiguity". Even if we take a more objective point of view, we might want to perform a sensitivity analysis of the optimal rule. While there is certainly a degree of randomness in such choice situations, it is not obvious that the arrival probability would be independent of the girl's quality, e.g. It might well be that more skilled applicants find

[1] Freeman (1983) gives an overview of the development until the eighties. Ferguson (1989) contains historical anecdotes. A lot of material is covered in Berezovski and Gnedin (1984).

[2] Gardner's treatment is the first instance here. A most recent example is the treatment in a book about "love economics" by a German journalist Beck (2005). It plays also a role in psychological experiments, see Todd (2009).

[3] This is an extreme utility function, of course. On the other hand, the analysis based on this extreme assumption serves as a benchmark for more general utility functions. The results are usually similar, at least in the Bayesian setting, see Ferguson (2006), e.g.

open jobs earlier[4]. In this paper, we present a way of dealing with these questions by embedding the best choice problem into a multiple prior framework as introduced by Gilboa and Schmeidler (1989) for the static case, and extended to dynamic frameworks by Epstein and Schneider (2003b). The agent works here with a class of possible prior distributions instead of a single one, and uses the minimax principle to evaluate her payoffs. We use then the general theory for optimal stopping under ambiguity developed in Riedel (2009) to analyze the model.

Our main result is that the optimal stopping rule is still *simple*, or a cutoff rule. The agent rejects a certain number of girls before picking the next applicant that is better than all previous ones — in the literature on best choice problems, such applicants are usually called *candidates*. The optimal strategy thus consists in building up a "database" by checking a certain number of applicants, and to take the first candidate thereafter[5]. We are able to obtain an explicit formula for the optimal threshold that determines the size of the database.

In best choice problems, ambiguity can lead to earlier or later stopping compared to the Bayesian case, in contrast to the analysis in Riedel (2009) where ambiguity leads to earlier stopping. The reason for this is that the original payoff process in best choice problems is not adapted. Indeed, when the employer accepts a candidate, she does not know if that candidate is the best among all applicants. She would have to observe all of them to decide this question. She thus uses her current (most pessimistic) belief about the candidate indeed being the best applicant. Two effects work against each other then. On the one hand, *after* picking a candidate, the agent's pessimism leads her to believe that the probability of better candidates to come is very high — this effect makes her cautious to stop. On the other hand, *before* acceptance, she uses a very low probability for computing the chance of seeing a candidate. This effect makes her eager to exercise her option. We illustrate these effects with three classes of examples.

In general, the optimal threshold can be quite different from the classical case (and in this general sense, the 37 %–rule described above

[4]Another obvious way of introducing ambiguity concerns the number of applicants. This question is not pursued here; see Engelage (2009) for a treatment of best choice problems with an unknown and ambiguous number of applicants.

[5]Optimal stopping rules need not be simple. For example, in the situation with incomplete information about the number of objects a Bayesian approach does not lead to simple stopping rules in general, see Presman and Sonin (1975).

is not robust). When the highest probability of finding a candidate decays sufficiently fast, the threshold– number of applicants–ratio can be very close to zero; indeed, it is independent of the number of applicants for large N. In such a situation, one has rather an absolute than a relative threshold. Instead of looking at the first 37 % of applicants, one studies a fixed number of them before choosing the first relatively top applicant.

On the other hand, if the probability of finding a candidate at time n is in the interval $[\frac{\gamma}{n}, \frac{1}{\gamma n}]$ for some parameter $\gamma \in (0,1)$, the threshold– number of applicants–ratio can converge to any positive number between 0 and 1. For $\gamma \to 1$, we obtain again the 37 % – rule. In this sense, the classical secretary problem is robust.

Last not least, we give an example where the ambiguity about applicants being candidates remains constant over time. This example can be viewed as the outcome of independent coin tosses with identical ambiguity as described in Epstein and Schneider (2003a)[6]. The aim is to pick the last 1 in this series of zeros and ones. We show that the agent optimally skips all but a finite number of applicants. In this case, we the ratio converges to 1 for large N. Different parametrizations of this example show that ambiguity can lead to earlier stopping (when the probability of finding a candidate are known to be small) as well as later stopping compared to the Bayesian case.

On the modeling side, our approach succeeds in finding a model that allows to introduce ambiguity into best choice problems. Note that one has to be careful when introducing ambiguity into dynamic models because one can easily destroy the dynamic consistency of the model[7]. To do so, we reformulate the classical secretary problem in the following way. The agent observes a sequence of ones and zeros, where 1 stands for "the current applicant is the best among the candidates seen so far". The agent gets the payoff of 1 if she stops at a 1 and there is no further 1 coming afterwards. In the secretary problem, the probability of seeing a 1 at time n is $1/n$ as all orderings are equally likely. We then allow for ambiguity by introducing an interval $[a_n, b_n]$ for this probability; finally, we construct a time–consistent model by pasting all these marginal probabilities together as explained in Epstein and Schneider (2003b).

The analysis of the stopping problem proceeds in three steps. In

[6]see also the examples of this type discussed in Riedel (2009).

[7]See Epstein and Schneider (2003b) for a general discussion of time–consistency in multiple prior models, and Riedel (2009) for the discussion in an optimal stopping framework.

a first step, we have to derive an equivalent model with adapted payoffs — note that the payoff function is not adapted here because the agent's payoff depends on the events that occur after stopping. We pass to adapted payoffs by taking conditional expectations prior by prior; it is not at all clear that this leads to the same ex ante payoff, though. Time–consistency and the corresponding law of iterated expectations for multiple priors[8] ensure this property. In the second step, we compute explicitly the relevant minimal conditional expectations. After having stopped, the agent uses the *maximal* probability for seeing a 1 afterwards. Intuitively, the agent's pessimism induces him to suppose that the best candidate is probable to come later after having committed herself to an applicant. After this, we have arrived at an optimal stopping problem that can be solved with the methods developed in Riedel (2009). Indeed, the problem at hand turns out to be a *monotone* problem: the worst–case measure can thus be identified as the measure under which the probabilities of seeing a candidate are minimal (until the time of stopping, of course). It then remains to solve a classical Bayesian stopping problem, and we are done.

The paper is organized as follows: Section 2 introduces the model and provides the stepwise solution as well as the main theorem. Section 3 contains three classes of examples that allow to discuss in more detail the effects of ambiguity in best choice problems.

2 Best Choice Problems under Ambiguity

Let us start with formalizing the classical best choice problem in a way that allows a natural generalization to ambiguity. In the classical secretary problem, the agent observes sequentially the relative ranks of applicants, say $R_1 = 1$ for the first one, $R_2 \in \{1, 2\}$ for the second, $R_3 \in \{1, 2, 3\}$ for the third and so on. The random ordering implies that the random variables (R_n) are independent[9]. As we are only interested in finding the best girl, we can discard all applicants with a relative rank higher than 1, and call candidates those girls that are relatively top at the moment. Let us introduce the $0-1$–valued sequence $Y_n = 1$ if $R_n = 1$ and $Y_n = 0$ else. The random variables

[8] See, e.g., Riedel (2009), Lemma 11.

[9] See Ferguson (2006) or Chow, Robbins, and Siegmund (1971) for the technical details.

(Y_n) are also independent, of course, and we have

$$P[Y_n = 1] = \frac{1}{n}$$

because all permutations are equally likely.

A *simple* stopping rule first rejects $r-1$ applicants and accepts the next candidate, if it exists, i.e.

$$\tau(r) = \inf\{k \geq r | Y_n = 1\}$$

with $\tau(r) = N$ if no further candidate appears after applicant $r-1$. One uses independence of the (Y_n) and monotonicity of the value function to show that optimal stopping rules must be simple, see Section 2.1.4 below for the argument in our context. It then remains to compare the expected success of the different simple rules. The event that girl n is a candidate and also the best of all girls means that no further girl has relative rank 1. In terms of our variables (Y_n), this means that $Y_n = 1$ and $Y_k = 0$ for all $k > n$. The success of a simple strategy is then

$$\begin{aligned}
\phi(r) &:= P[\tau(r) \text{ picks the best girl}] = \sum_{n=r}^{N} P[\tau(r) = n, \text{girl } n \text{ is best}] \\
&= \sum_{n=r}^{N} P[Y_r = 0, \ldots, Y_{n-1} = 0, Y_n = 1, Y_k = 0, k > n] \\
&= \sum_{n=r}^{N} \prod_{j=r}^{N} P[Y_j = 0] \, \frac{P[Y_n = 1]}{P[Y_n = 0]} \\
&= \prod_{j=r}^{N} \frac{j-1}{j} \sum_{n=r}^{N} \frac{1/n}{1-1/n} = \frac{r-1}{N} \sum_{n=r}^{N} \frac{1}{n-1}\,.
\end{aligned}$$

The sum approximates the integral of $1/x$, so the value is approximately

$$\phi(r) = \frac{r-1}{N} \log \frac{N}{r-1}\,.$$

The maximum of the function $-x \log x$ is in $1/e$, so we conclude that the optimal r is approximately $[N/e] + 1$.

2.1 Best Choice under Ambiguity

2.1.1 Formulation of the Problem

We generalize now the above model to ambiguity by allowing that the probabilities

$$P[Y_n = 1 | Y_1, \ldots, Y_{n-1}] \in [a_n, b_n]$$

for all histories $Y_1, \ldots, Y_{n-1}$ come from an interval instead of being a known number. Throughout the paper, we assume that $0 < a_n \leq b_n < 1$.

Modeling ambiguity in dynamic settings requires some care if one wants to avoid traps and inconsistencies. We view the random variables (Y_n) as outcomes of independent, but ambiguous experiments where in the nth experiment the distribution of Y_n, i.e. the number $P[Y_n = 1]$ is only known to come from an interval $[a_n, b_n]$. From these marginal distributions, the agent has to construct all possible joint distributions for the sequence (Y_n). She does so by choosing any number $p_n \in [a_n, b_n]$ after having observed $Y_1, \ldots, Y_{n-1}$. A possible prior then takes the form

$$P[Y_1 = 1] = 1 \tag{1}$$

because the first applicant is always a candidate, and

$$P[Y_n = 1 | Y_1, \ldots, Y_{n-1}] = p_n \in [a_n, b_n] \tag{2}$$

for a predictable sequence of one–step–ahead probabilities p_n. Note that we allow p_n to depend on the past realizations of $(Y_1, \ldots, Y_{n-1})$. For a time–consistent worst–case analysis this is important because different one–step–ahead probabilities might describe the worst case after different histories. From now on, we work with class $\mathcal{P}$ of all probability measures that satisfy (1) and (2) for a given sequence $0 < a_n \leq b_n < 1$, $n = 1, \ldots, N$. For more on the foundations of dynamic decisions under ambiguity, we refer the reader to Epstein and Schneider (2003b) and Epstein and Schneider (2003a), see also **?**.

The astute reader might now wonder why we speak about independent realizations if the conditional probabilities are allowed to depend on past observations. Independence in a multiple prior setting is to be understood in the sense that the interval $[a_n, b_n]$ is independent of past observations, just as it means that the conditional probability of the event $\{Y_n = 1\}$ given the past observations is independent of these observations in classical probability. In this sense, the agent does not learn from past observations about the degree of ambiguity of the nth experiment.

We are now ready to formulate our optimization problem. Based on the available information the agent chooses a stopping rule τ that maximizes the expected payoff which is 1 if she happens to find the best girl, and 0 else. A way to describe this in our model is as follows: applicant n is the best if she is a candidate (she has to be relatively best among the first n, of course), and if she is not topped by any subsequent applicant: in other words, we have $Y_n = 1$ and there is no further candidate afterwards, or $Y_k = 0$ for $k > n$. Let us define

$$Z_n = \begin{cases} 1 & \text{if } Y_n = 1, Y_k = 0, k > n \\ 0 & \text{else} \end{cases} .$$

The agent aims to choose a stopping rule τ that maximizes

$$\inf_{P \in \mathcal{P}} \mathbb{E}^P[Z_\tau]. \tag{3}$$

2.1.2 Reformulation in Adapted Payoffs

The next problem that we face is that the sequence (Z_n) is not adapted to the filtration generated by the sequence (Y_n) because we do not know at the time when we pick an applicant if she is best or not. As in the classical case, we therefore take first conditional expectations of the rewards (Z_n) before we can apply the machinery of optimal stopping theory. In the multiple prior framework we thus consider

$$X_n = \operatorname*{ess\,inf}_{P \in \mathcal{P}} \mathbb{E}[Z_n | \mathcal{F}_n] .$$

where $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. In the Bayesian framework, it is relatively easy to show that the expected payoffs $\mathbb{E}[Z_\tau] = \mathbb{E}[X_\tau]$ are the same for all stopping times τ. In the multiple prior framework, this is less obvious. Indeed, the identity

$$\inf_{P \in \mathcal{P}} \mathbb{E}^P[Z_\tau] = \inf_{P \in \mathcal{P}} \mathbb{E}^P[X_\tau]$$

does require a condition on the set of priors which has become known as *rectangularity* or *m–stability*. In our model, this condition is satisfied[10], and we have

[10] Compare Epstein and Schneider (2003b) or Riedel (2009), Section 4.1.

Lemma 1 *For all stopping times τ we have*

$$\inf_{P\in\mathcal{P}} \mathbb{E}^P[Z_\tau] = \inf_{P\in\mathcal{P}} \mathbb{E}^P[X_\tau].$$

We can thus reformulate our problem as

$$\text{maximize } \inf_{P\in\mathcal{P}} \mathbb{E}^P[X_\tau] \tag{4}$$

over all stopping times $\tau \leq N$.

2.1.3 Reduction to a Monotone Problem

We are now in the position to apply the general theory of optimal stopping with multiple priors as developed in Riedel (2009). To this end, let us first have a closer look at the payoffs (X_n). It is clear that $X_n = 0$ if we do not have a candidate at n, i.e. $Y_n = 0$, so we need only to focus on the case $Y_n = 1$. We are then interested in the minimal (conditional) probability that all subsequent applicants are no candidates. It is quite plausible (but requires a proof, of course) that the probability is minimal under the measure $\bar{P}$ where the probabilities for being a candidate are maximal, $\bar{P}[Y_n = 1|Y_1, \ldots, Y_{n-1}] = b_n$. Under this measure, the (Y_n) are independent (because the conditional probabilities for $Y_n = 1$ are independent of past observations, but see the proof of Lemma 2 for the details), and we thus have

Lemma 2 *The payoffs (X_n) satisfy*

$$\begin{aligned} X_n =& Y_n \min_{P\in\mathcal{P}} P[Y_{n+1} = 0, \ldots, Y_N = 0] \\ =& Y_n \prod_{k=n+1}^{N} (1 - b_k) \\ =:& Y_n \cdot B_n \end{aligned} \tag{5}$$

The agent faces now a sequence of adapted payoffs that is monotone in Y_n (indeed, linear). The random variables (Y_n) are independent under the measure Q where the conditional probabilities for a candidate are

$$Q[Y_n = 1|Y_1, \ldots, Y_{n-1}] = a_n \,.$$

Moreover, the probabilities of finding a candidate are smallest under this measure in the whole class $\mathcal{P}$ in the sense of first–order stochastic dominance. We are thus in a situation that is called a monotone problem in Riedel (2009). The general theory there tells us that the optimal stopping rule with multiple priors coincides with the optimal stopping rule under the measure Q – the worst-case measure.

Theorem 1 *The optimal stopping rule* τ^* *for* (4) *is the same as the optimal stopping rule for the Bayesian problem*

$$\text{maximize } \mathbb{E}^Q[X_\tau]. \tag{6}$$

2.1.4 Optimal Stopping under the Worst–Case Measure Q

We are now back to a classical optimal stopping problem under the measure Q. A standard argument shows that optimal stopping rules must be simple. It works as follows. From classical optimal stopping we know that it is optimal to stop when the current payoff X_n is equal to the current value of the problem

$$v_n := \sup_{\tau \geq n} \mathbb{E}^Q \left[X_\tau | X_1, \ldots, X_n \right] .$$

The independence of the (X_n) under Q implies that the value of the problem after having rejected $n-1$ applicants is independent of the past observations, i.e.

$$v_n = \sup_{\tau \geq n} \mathbb{E}^Q[X_\tau]. \tag{7}$$

The sequence (v_n) is decreasing as we maximize over a smaller set of stopping times. On the other hand, the numbers

$$B_n := \prod_{k=n+1}^{N} (1 - b_k)$$

are increasing in n. Now suppose that it is optimal to take a candidate n. We have then $B_n = v_n$; therefore, we get

$$B_{n+1} \geq B_n = v_n \geq v_{n+1} ,$$

and it is also optimal to stop when a candidate appears at time $n+1$. We conclude that optimal stopping rules are simple.

Lemma 3 *The optimal stopping rule* τ^* *is simple, i.e. there exists a number* $1 \leq r^* \leq N$*, s.t.*

$$\tau^* = \tau(r^*) = \inf\{n \geq r^* | Y_n = 1\}.$$

In the next step we compute the optimal threshold r^* maximizing (7) over all simple strategies. Let us denote by

$$\phi(r) := \mathbb{E}^Q[X_{\tau(r)}]$$

the payoff from starting to search at applicant r. We then have

$$\phi(N) := \mathbb{E}^Q(X_{\tau(N)}) = a_N \tag{8}$$

and

$$\phi(r) = a_r \cdot B_r + (1 - a_r) \cdot \phi(r + 1) \tag{9}$$

for $r < N$.

While our recursive formula for $\phi(r)$ is useful for numerical computations, we record also the explicit solution of this linear difference equation. To simplify the interpretation of this expression, we introduce two concepts.

Definition 1 *For each $n \le N$ we call*

$$\alpha_n = \frac{1 - a_n}{1 - b_n} = 1 + \frac{b_n - a_n}{1 - b_n}$$

the degree of ambiguity and

$$\beta_n = \frac{a_n}{1 - b_n}$$

the ambiguous odds of applicant n.

The first ratio α_n measures the ambiguity persisting at the time n. The term tends to 1 as length of the interval $[a_n, b_n]$ decreases. In case of $a_n = b_n$ the node n is completely unambiguous and the decision maker faces only risk at n. Similarly, one can think of the product $\prod_{k=n}^{N} \alpha_k$ as the cumulated ambiguity persisting between n and N. The model is unambiguous if and only if $\prod_{k=1}^{N} \alpha_k = 1$. Note, that we call the ratio $p/(1-p)$ the odds for a zero–one bet. In a similar way, the ration β_n measures the odds of seeing a candidate at time n where we now use the (nonlinear) probability induced by our ambiguity model.

The solution of the linear difference equation (9) with boundary condition (8) is given by

$$\begin{aligned}\phi(r) =& B_{r-1}\left(\beta_n + \alpha_n\beta_{n+1} + \cdots \prod_{k=r}^{n-1}\alpha_k\beta_N\right) \\ =& B_{r-1}\cdot\left(\sum_{n=r}^{N}\beta_n\prod_{k=r}^{n-1}\alpha_l\right) .\end{aligned} \tag{10}$$

Let us check now that ϕ has a unique maximizer. From our recursion (9), we get that $\phi(r) - \phi(r+1) \geq 0$ is equivalent to $w_r \leq 1$ for

$$\begin{aligned}w_r :=& \frac{\phi(r+1)}{B_r} \\ =& \sum_{n=r}^{N}\beta_n\prod_{k=r}^{n-1}\alpha_k .\end{aligned} \tag{11}$$

As $\alpha_k > 1$ and $\beta_n > 0$, the sequence (w_r) is strictly decreasing. Thus, $(\phi(r))$ is increasing as long as $w_r > 1$ and decreasing afterwards, which shows that it has a unique maximum.

The maximizer is determined by

$$r^* = \inf\{r \geq 1 | w_r \leq 1\} \tag{12}$$

The optimal threshold r^* is determined by the weighted average of ambiguous odds weighted with the ambiguity persisting between r and n. Equation (11) and Equation (12) completely characterize the solution.

We summarize our findings in the following theorem.

Theorem 2

1. *The optimal stopping rule for* (3) *is simple, i.e. the agent first observes r^* candidates and takes then the first candidate that appears;*
2. *The optimal threshold r^* for the cutoff is given via* (11) *and* (12).

3 Comparative Statics

In this section we use the sequence (w_r) and the variables (α_n) and (β_n) defined above to analyze the effects of ambiguity on stopping and

the structure of the stopping strategy τ^*.
As it was shown in Riedel (2009), an ambiguity averse decision maker behaves like a Bayesian decision maker under a special worst-case probability measure constructed via backward induction. We have seen in the preceding section how to construct this measure, and that the optimal stopping rule is still simple. The central question by analyzing the effect of ambiguity is now the threshold r^*. In case of monotone problems where the payoff is known at the time of decision such as *House Selling Problem* or *Parking Problem* discussed in Riedel (2009) ambiguity leads to earlier stopping. The use of the worst-case measure lowers the value of the Snell envelope and forces the agent to stop earlier. The situation differs here because the agent faces actually two kinds of uncertainty. On the one hand, there is payoff uncertainty in the adapted version of the problem because the probability distribution of Y_n is not known. This effect leads to earlier stopping because it reduces the expected value from waiting. On the other hand, ambiguity also affects the chances that a better applicant is going to come after the current candidate. This ambiguity induces the agent to wait longer because she believes *after* stopping that candidates are going to appear with high probability. The two effects work against each other, and we thus proceed to study more detailed models in which we can disentangle them[11]. In addition, we compute the value of the threshold r^* and show that asymptotically, the relative fraction of applicants that the agent lets go by can assume any value between 0 and 1.

3.1 Ambiguous Secretary Problem

Our first example is the multiple prior version of the classical secretary problem. The decision maker is uncertain about the real distribution of the orderings for reasons explained in the introduction but has no additional information on the quality of the applicants. Doubting her strategy she aims to know what happens if she changes the measure slightly. Instead of $P[Y_n = 1] = \frac{1}{n}$, the ambiguity averse decision maker assumes that the probability lies in an interval near by $\frac{1}{n}$, i.e.

$$P[Y_n = 1|\mathcal{F}_{n-1}] \in \left[\min\{\frac{\gamma}{n}, 0,9999\}, \min\{\frac{1}{\gamma n}, 0,9999\}\right]$$

[11] A similar point has been made in a completely different model in Nishimura and Ozaki (2007) when there is uncertainty about the timing and uncertainty about the value from stopping.

N / γ	1	0.9	0.8	0.7	0.6	0.4	0.3	0.2	0.1
5	3	3	3	3	3	3	5	5	5
10	4	4	5	5	5	5	6	7	10
50	19	19	19	20	20	22	24	27	34
100	38	38	38	38	39	43	46	53	65
500	185	185	186	189	193	210	227	257	316
1000	369	369	372	376	385	419	453	513	630

Table 1: Absolute values of the threshold r^* for different values of N and levels of ambiguity γ. The threshold is increasing with ambiguity. The agent waits longer before accepting a candidate when ambiguity increases.

N / γ	1	0.9	0.8	0.7	0.6	0.4	0.3	0.2	0.1
5	60%	60%	60%	60%	60%	60%	100%	100%	100%
10	40%	40%	50%	50%	50%	50%	60%	70%	100%
50	38%	38%	38%	40%	40%	44%	48%	54%	68%
100	38%	38%	38%	38%	39%	43%	46%	53%	65%
500	37%	37%	37%	38%	39%	42%	45%	51%	63%
1000	37%	37%	37%	38%	39%	42%	45%	51%	63%

Table 2: Relative values of the threshold r^* for different values of N and levels of ambiguity γ. Also the relative threshold is increasing with ambiguity.

for some $\gamma < 1$, $2 \leq n \leq N$. We can use the analysis of the preceding section to compute the thresholds r^* that depends on γ and N, of course. Typical values are tabulated in Table 1 and 2 for the absolute and relative values of the threshold, resp. It is interesting to see that one waits longer as ambiguity increases. The effect of missing a potentially better applicant outweighs the lower expectation from ambiguity. We get here a potentially testable implication: the more uncertain the agent is, the longer she should wait before taking a decisive action in a best choice problem.

The following result gives exact boundaries for the optimal threshold depending upon γ and N.

Theorem 3 *For given γ and N, the optimal threshold $r^*(\gamma, N)$ satisfies*

$$e^{-\frac{1}{\gamma}} \leq \frac{r^*(\gamma, N)}{N} \leq e^{-\frac{2\gamma}{1+\gamma}} + \frac{3}{N}. \tag{13}$$

In particular, the Secretary Problem is robust in the sense that

$$\lim_{N\to\infty,\gamma\uparrow 1} \frac{r^*(\gamma, N)}{N} = \lim_{N\to\infty} \frac{r^*(0)}{N} = e^{-1}\,. \tag{14}$$

3.2 Independent Coins with Identical Ambiguity

Our example corresponds to the Independent Indistinguishably Distributed case introduced in [ES1]. Here, the probability to meet a candidate remains constant over time. More generally, this is the case, where the decision maker does not know if the experiment changes over time. At the same time she has no reason to distinguish between periods. To express the uncertainty about the coin the agent uses a class of measures in each period.

We consider following bet: We observe an ambiguous coin being tossed N times and we win if we stop at the last time $\{head\}$ appears in the sequence. With this setup we are in the situation of the ambiguous best choice problem where the probabilities for $\{head\}$ remain constant over time:

$$P(n\text{-th toss is a head}|\mathcal{F}_{n-1}) \in [p-\varepsilon, p+\varepsilon]$$

for $\varepsilon \geq 0, \varepsilon \leq p$.

To get a feeling for the problem, let us start with the pure risk case, $\varepsilon = 0$. In this case, we get

$$w_r = \beta(N-r) = \frac{p}{1-p}(N-r)$$

and the optimal threshold is the first r such that

$$N - r \leq \frac{1-p}{p}\,.$$

In this problem, it is optimal to focus solely on the last $\left[\frac{1-p}{p}\right] + 1$ applicants, irrespective of the total number of applicants.

Let us now come to the ambiguous case. From Equation (11), we obtain for the degree of ambiguity $\alpha = \frac{1-p+\varepsilon}{1-p-\varepsilon} > 1$ and ambiguous odds $\beta = \frac{p-\varepsilon}{1-p-\varepsilon}$

$$w_r = \sum_{k=r}^{N} \beta \prod_{l=r}^{k-1} \alpha = \beta\, \frac{\alpha^{N-r+1} - 1}{\alpha - 1}\,.$$

The threshold r^* is given by the first r such that

$$\alpha^{N-r} \leq 1 + \frac{\alpha - 1}{\beta} = \frac{p+\varepsilon}{p-\varepsilon}\,.$$

We learn from this that the agent focuses only on the last

$$k(p,\varepsilon) \simeq \frac{\log \frac{p+\varepsilon}{p-\varepsilon}}{\log \frac{1-p+\varepsilon}{1-p-\varepsilon}}$$

applicants. This quantity is independent of N.

In this case we observe *memoryless stopping*: The decision about stopping does not depend on the number of the options already observed. Only the number of options left matters. Consequently, we obtain

$$\lim_{N\to\infty} \frac{r^*(N)}{N} = 1\,.$$

This example also allows us to demonstrate that ambiguity can lead both to earlier as well as to later stopping. For $p < \frac{1}{2}$, the quantity $k(p,\varepsilon)$ is increasing; consequently, the agent stops earlier when ambiguity increases. For $p = 1/2$, $k(p,\varepsilon)$ is independent of ε and ambiguity does not influence the stopping behavior. For $p > 1/2$, the agent stops later, in general.

3.3 Finite Stopping

In our last example we consider the case where the probability to meet a candidate falls very fast. Here, the value of waiting decreases very fast and becomes zero at some point. In this situation the future becomes worthless and interviewing additional candidates does not improve the expected payoff. Even if the pool of applicants is infinite the decision will be made in finite time. Here, we can compute the maximal amount of applicants that need to be interviewed in order to decide optimally.

To see how it works we first consider the value of waiting for a fixed number of candidates N and a given one-step-ahead probabilities $[a_n, b_n]$. Now we add an applicant with $P[Y_{N+1} = 1] \in [a_{N+1}, b_{N+1}]$. Clearly, adding applicants does not decrease the value of the problem. As we vary the number of applicants now, let us write w_r^N for the crucial sequence that determines the threshold $r^*(N)$. Clearly, w_r^N is increasing in N and the value of the threshold $r^*(N+1) \geq r^*(N)$. Now assume that $w_r^\infty := \lim_{N\to\infty} w_r^N$ exists. Then we can find $R \in \mathbb{N}$ s.t. $w_R^\infty < 1$ and therefore $w_R^N < 1$ for all N sufficiently large. Therefore, the value of the threshold $r^*(N)$ cannot exceed R. As $r^*(N)$ is an increasing, but bounded sequence of integers, it has to be constant from some point on, $r^*(N) = R$ for N sufficiently large.

In other words, the number of applicants does not matter here for large pools of applicants. The agent first studies a fixed number of applicants before taking the next candidate.

Lemma 4 *If*

$$w^\infty := \lim_{N\to\infty} w_1^N \tag{15}$$

exists, then

1. *The value of the threshold $r^*(N)$ is bounded by a constant $R \in \mathbb{N}$ and for sufficiently large $N \in \mathbb{N}$, we have $r^*(N) = R$,*
2. *The fraction of rejected candidates converges to zero, i.e.*

$$\lim_{N\to\infty} \frac{r^*(N)}{N} = 0\,.$$

Let us reflect a moment under what condition the series $w^\infty = \sum_{k=1}^\infty \beta_k \prod_{l=r}^{k-1} \alpha_l$ is finite. By d'Alembert's ratio test, this is the case if we have

$$\limsup_{n\to\infty} \frac{1-a_n}{a_n} \frac{a_{n+1}}{1-b_{n+1}} < 1\,.$$

This condition holds true, e.g., when both (a_n) and (b_n) converge fast, say exponentially, to zero.

In this section we analyzed the observation period for different sets of measures. Depending on the structure of the set $\mathcal{P}$ the observation period converges to a constant $c \in (0,1)$ as in the case of the ambiguous secretary problem. Or it can converge to zero making the future worthless as in the finite stopping case. In the opposite case of memoryless stopping the observation period tends to 1, assigning zero value to the past.

4 Conclusion

We provide a closed form solution for the best choice problem in the multiple prior framework. An appropriate version of backward induction leads to the solution if the set of priors is time-consistent. Due to time-consistency most of classical arguments remain valid, the stopping rule is simple. The closed form solution allows to analyze the impact of ambiguity on the stopping behavior. Additionally, we show the robustness of the classical secretary problem in the multiple prior framework. A natural next step is to generalize the utility function. Additionally, one might extend the model to infinite settings.

A Appendix

A.1 Proof of Lemma 1

Proof: Fix a stopping time τ with values in $\{1, \ldots, N\}$. Our set of priors is compact, hence we can choose $Q^k \in \mathcal{P}$ that minimize $\mathbb{E}^P X_k 1_{\{\tau=k\}}$. Time–consistency of the set of priors implies that there exists a measure $Q \in \mathcal{P}$ such that

$$\sum_{k=1}^{N} \mathbb{E}^{Q^k} X_k 1_{\{\tau=k\}} = \mathbb{E}^Q \sum_{k=1}^{N} X_k 1_{\{\tau=k\}} = \mathbb{E}^Q X_\tau \,,$$

see, e.g., Lemma 8 in Riedel (2009). It follows that we have

$$\inf_{P \in \mathcal{P}} \mathbb{E}^P X_\tau = \inf_{P \in \mathcal{P}} \mathbb{E}^P \sum_{k=1}^{N} X_k 1_{\{\tau=k\}} = \sum_{k=1}^{N} \inf_{P \in \mathcal{P}} \mathbb{E}^P X_k 1_{\{\tau=k\}} \,.$$

By applying the law of iterated expectations for time–consistent multiple priors, this quantity is

$$= \sum_{k=1}^{N} \inf_{P \in \mathcal{P}} \mathbb{E}^P Z_k 1_{\{\tau=k\}}$$

and by applying time–consistency again, we get

$$= \inf_{P \in \mathcal{P}} \mathbb{E}^P Z_\tau \,.$$

□

A.2 Proof of Lemma 2

Proof: To show the independence we have to show that

$$\operatorname*{ess\,inf}_{P \in \mathcal{P}} P(Y_1 = y_1, \ldots, Y_N = y_N) = \prod_{i=1}^{N} \hat{P}(Y_i = y_i)$$

for a $\hat{P} \in \mathcal{P}$, $y_i \in \{0, 1\}$ for $1 \leq i \leq N$
Because of definition of $\mathcal{P}$ all events of above kind have positive probability under every $P \in \mathcal{P}$, i.e. $P(Y_1 = y_1, \ldots, Y_N = y_N) > 0$ for all sequences (y_i) with $y_i \in \{0, 1\}$ and all $P \in \mathcal{P}$. Therefore, using Bayes'

rule and the fact that one-step-ahead probabilities $[a_n, b_n]$ depend only on time we get

$$
\begin{aligned}
\min_{P\in\mathcal{P}} P[Y_1 = y_1, \ldots, Y_N = y_N] &= \min_{P\in\mathcal{P}} P[Y_N = y_N|\mathcal{F}_{N-1}]P[Y_i = y_i, i < N] \\
&= \min_{P\in\mathcal{P}} \prod_{n=1}^{N} P[Y_n = y_n|\mathcal{F}_{n-1}] \\
&= \prod_{n=1}^{N} \min_{x_n\in[\alpha_n,\beta_n]} P_{x_n}[Y_n = y_n|\mathcal{F}_{n-1}] \\
&= \prod_{n=1}^{N} \min_{x_n\in[\alpha_n,\beta_n]} P_{x_n}[Y_n = y_n]
\end{aligned}
$$

where P_{x_n} denotes the measure defined via $P_{x_n}[Y_n = y_n|\mathcal{F}_{n-1}] = x_n$.

□

A.3 Proof of Theorem 3

Proof: We denote by $w(\gamma)_n$ the sequence corresponding to the problem with ambiguity γ. Straightforward calculations show that

$$w_n(\gamma) = \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} \prod_{l=n}^{N} \left(1 + \frac{1-\gamma^2}{l\gamma - 1}\right) \tag{16}$$

To prove robustness we first show

$$e^{-\frac{1}{\gamma}} \leq \frac{r^*_\gamma}{N} \leq e^{-\frac{2\gamma}{1+\gamma}} + \frac{3}{N} \tag{17}$$

For the left-hand side of 17:

$$
\begin{aligned}
w_n^\gamma = \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} \prod_{l=n}^{N} \left(1 + \frac{1-\gamma^2}{l\gamma - 1}\right) &\geq \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} && (18) \\
&\geq \sum_{k=n}^{N} \frac{\gamma}{k} && (19) \\
&\geq \int_n^N \frac{\gamma}{k} dk && (20) \\
&= \gamma \log\left(\frac{N}{n}\right) && (21)
\end{aligned}
$$

For the threshold r_γ we obtain

$$1 = w_n^\gamma \geq \gamma \log\left(\frac{N}{n}\right) \tag{22}$$

$$\Leftrightarrow \tag{23}$$

$$\frac{r_\gamma^*}{N} \leq e^{-\frac{1}{\gamma}} \tag{24}$$

For the second inequality:

$$\begin{aligned}
w_n^\gamma &= \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} \prod_{l=n}^{N} \left(1 + \frac{1-\gamma^2}{l\gamma - 1}\right) \\
&= \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} \exp\left(\sum_{l=n}^{k-1} \ln\left(1 + \frac{1-\gamma^2}{l\gamma - 1}\right)\right) \\
&\leq \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} \exp\left(\sum_{l=n}^{k-1} \frac{1-\gamma^2}{l\gamma - 1}\right) \\
&\leq \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} \exp\left(\int_{l=n-1}^{k} \frac{1-\gamma^2}{l\gamma - 1} dl\right) \\
&\leq \sum_{k=n}^{N} \frac{\gamma^2}{k\gamma - 1} \left(\frac{k\gamma - 1}{(n-1)\gamma - 1}\right)^{\frac{1-\gamma^2}{\gamma}}
\end{aligned}$$

Using $\alpha := \frac{1-\gamma^2}{\gamma} - 1$ we obtain for $\gamma \geq 0.5$

$$\begin{aligned}
w_n^\gamma &\leq \int_{n-1}^{N} \frac{\gamma^2}{((n-1)\gamma - 1)^{\alpha+1}} (k\gamma - 1)^\alpha \, dk \\
&\leq \frac{\gamma^2}{1-\gamma^2} \left[\left(\frac{N}{n-3}\right)^{\alpha+1} - 1\right]
\end{aligned}$$

By setting $w_n^\gamma = 1$ we get

$$1 \leq \frac{\gamma^2}{1-\gamma^2}\left[\left(\frac{N}{n-3}\right)^{\alpha+1} - 1\right]$$

$$\Leftrightarrow$$

$$\begin{aligned}\frac{n-3}{N} &\leq \gamma^{\frac{2\gamma}{1-\gamma^2}} \\ &\leq \exp\left(\ln(\gamma)\frac{2\gamma}{1-\gamma^2}\right) \\ &\leq \exp\left(\frac{(\gamma-1)2\gamma}{1-\gamma^2}\right) \\ &\leq e^{-\frac{2\gamma}{1+\gamma}}\end{aligned}$$

□

A.4 Proof of Lemma 4

PROOF: Because of boundedness of w^∞ there exists a $R \in \mathbb{N}$, s.t.

$$\sum_{k=n}^{\infty} \beta_k \prod_{l=1}^{k-1} \alpha_l \leq 1 \text{ for all } n \geq R$$

and it follows that

$$r^*(N) \leq R \text{ for all } N$$

and

$$\frac{r^*(N)}{N} \leq \frac{R}{N} \to 0 \text{ for } N \to \infty$$

□

References

BECK, H. (2005): *Der Liebeskonom – Nutzen und Kosten einer Himmelsmacht*. Frankfurter Allgemeine Buch.

BEREZOVSKI, B., AND A. V. GNEDIN (1984): *Problems of Best Choice (in Russian)*. Akademia Nauk, Moscow.

Chow, Y., H. Robbins, and D. Siegmund (1971): *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin Comp., Boston.

Engelage, D. (2009): "Essays on Multiple Priors," Ph.D. thesis, Bonn Graduate School of Economics, Bonn, Germany.

Epstein, L., and M. Schneider (2003a): "IID: Independently and Indistinguishably Distributed," *Journal of Economic Theory*, 113(1), 32–50.

——— (2003b): "Recursive Multiple Priors," *Journal of Economic Theory*, 113, 1–31.

Ferguson, T. (2006): "Optimal Stopping and Applications," Electronic Text, http://www.math.ucla.edu/~tom/Stopping/Contents.html, University of California, Los Angeles.

Ferguson, T. S. (1989): "Who Solved the Secretary Problem?," *Statistical Science*, 4(3), 282–289.

Freeman, P. (1983): "The secretary Problem and its Extensions — a Review," *International Statistical Review*, 51, 189–206.

Gardner, M. (1960): *Scientific American*.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non–Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Nishimura, K. G., and H. Ozaki (2007): "Irreversible Investment and Knightian Uncertainty," *Journal of Economic Theory*, 136, 668–694.

Presman, E., and I. Sonin (1975): "Equilibrium Points in a Game Related to the Best Choice Problem," *Theory of probability and its Applications*, 20, 770–781.

Riedel, F. (2009): "Optimal Stopping with Multiple Priors," *Econometrica*.

Todd, P. (2009): "Searching for the Next Best Mate," Working Paper, Center for Adaptive Behavior and Cognition, Munich University.