Economic Theory 10, 55–77 (1997)

Economic Theory

An experimental analysis of the bandit problem

Jeffrey Banks[1], Mark Olson[2], and David Porter[3]

[1] Department of Economics, University of Rochester, Rochester, NY 14627, USA
[2] CREED, University of Amsterdam, Amsterdam, THE NETHERLANDS
[3] Division of Humanities and Social Sciences, Caltech, Pasadena, CA 91125, USA

Received: October, 27, 1994; revised version February 27, 1996

Summary. We investigate, in an experimental setting, the behavior of single decision makers who at discrete time intervals over an "infinite" horizon may choose one action from a set of possible actions where this set is constant over time, i.e. a bandit problem. Two bandit environments are examined, one in which the predicted behavior should always be myopic (the two-armed bandit) and the other in which the predicted behavior should never be myopic (the one-armed bandit). We also investigate the comparative static predictions as the underlying parameters of the bandit environments are changed. The aggregate results show that the behavior in the two bandit environments are quantitatively different and in the direction of the theoretical predictions.

JEL Classification Numbers: C91, C73, D81, D83.

1 Introduction

Models of search and learning have become quite pervasive in the field of microeconomics in the last two decades. For example, in the area of labor economics the "matching" models of Jovanovic (1979), Wilde (1979), and Viscusi (1979) all have, as their basic scenario, a worker who periodically receives information about her current job's true characteristics (and hence the wages she can expect), and has the opportunity to remain with her current employer or switch to a new job where again information about future wages may accrue.[1] In industrial organization, Rothschild (1974) models a monopolist attempting to learn the true state of consumer demand for its product: by setting different prices, the monopolist can gain differential information about demand. Ishikida (1992) models the decentralized assignment of a digital pipe as a search process where users randomly send packets

[1] Cf. Mortensen (1985) for an in-depth survey of these and other search models in labor economics.

Correspondence to: D. Porter

ment of a digital pipe as a search process where users randomly send packets and can cause delays to other users of a communication network.

Many of these models share an underlying canonical form of individual decision making under uncertainty, namely that of a *bandit problem*.[2] The structure of a typical bandit problem has a single decision maker, who at discrete time intervals over an infinite horizon may choose one action from a set of possible actions (i.e. the arms of the bandit), where this set is constant over time. Each arm, if employed in a period, generates a reward to the decision maker according to some time-invariant distribution, where for each arm there is a set of possible reward distributions known as the *types* of the arm. The decision maker begins with some prior belief about an arm's true type, and any additional information would be useful in ascertaining the best arm to play. Further, it is commonly assumed that the true types of different arms are independent; hence knowledge of an arm's type can only be generated by employing the arm for at least a single trial. Finally, the decision maker is interested in maximizing the discounted sum of her expected rewards.

Given the temporal stationary of the above decision problem, there is an optimal strategy for the decision maker which depends only on her current belief (i.e. probability distribution over types) about the different arms. This stationary can equivalently be thought of as a particular type of path independence, in that any two paths of reward realizations leading to the same updated belief should generate the same behavior by an optimizing individual. Indeed, a result by Gittins and Jones (1974) for finite-armed bandits shows that an *index* strategy is actually optimal: for each arm, and for each possible belief about that arm's true type, one can assign a number depending only on the characteristics of that arm such that an optimal strategy simply prescribes selecting in any trial the arm with the highest number.[3]

The presence of such an index, while highlighting the stationary of the solution to the decision maker's problem, also renders certain types of comparative statics exercises more tractable. For instance, by changing one distribution associated with an arm, or more simply by changing the prior belief associated with an arm, one can deduce how this alters the index and consequently the index strategy. This stationary of the optimal strategy, along with certain comparative statics hypotheses, provide the motivation for our experimental investigation.

This paper describes experiments in which individuals were faced with one of two relatively simple bandit problems similar to the one detailed above. In the first, labeled Bandit Problem I below, there are two arms, one of which pays a *certain reward*, while the other generates either a high or low reward, and where the probability of a high reward can take on one of two

[2] Cf. Banks and Sundaram (1992b).

[3] This result has been extended to the case of a countable infinity of arms by Banks and Sundaram (1992b).

possible values: a "high" probability of generating a high payoff (which we refer to as the "good" type) and a "low" probability of generating a high payoff (the "bad" type). In Bandit Problem II there are again two arms, but now both arms generate uncertain payoffs. Each arm again produces either a high or low reward, and each can be either a "good" or "bad" type, where the true types are drawn independently. Therefore the only difference in the arms is the decision maker's current belief about each arm's type, i.e. the probability she assesses to each arm being "good".

In both of these problems the predictions implied by the theory alluded to above are straightforward: in Bandit Problem I, once an individual has begun playing the arm producing the certain payoff, she should remain there forever, since no new information is being generated and hence the indices of the arms remain unchanged. Alternatively, if she employs the uncertain arm and a high payoff results she should remain with that arm, since such a high payoff should lead her to increase the expected benefits from continual play of that arm; that is, the index on the uncertain arm increases. More generally, the theory predicts that an individual will have a *critical belief* about the uncertain arm, wherein if her current belief about the uncertain arm being "good" is above this level she should employ that arm in the current period, while if her belief is below this level she should play the certain arm. Further, this critical belief will be a function of the specific reward probabilities, as well as an individual's discount factor and attitudes towards risk.

In Bandit Problem II, the theoretical prediction is considerably more stark: the optimal strategy for an individual regardless of the reward probabilities, her discount factor or her attitudes towards risk, is *myopic;* that is, in any period the choice of arm consistent with maximizing the discounted sum of expected payoffs is simply the arm that generates the highest expected one-period reward; or equivalently the arm with the higher probability of being "good".[4] The intuition behind this result is readily apparent given that index strategies are optimal: when there are only two possible types, the index on an arm will be increasing in the probability the arm is the "good" type; and since in Bandit Problem II the arms are identical up to the current belief about type, the indices will be the same up to this belief as well. Therefore the arm with the higher probability of being "good" will have the higher index, and hence will constitute the optimal choice.

This predicted myopia in Bandit Problem II is in contrast to the behavior predicted in Bandit Problem I, where if an individual's belief is such that either the certain or the uncertain arm would generate the same expected one-period payoff, the optimal strategy always prescribes the uncertain arm. The reason for this is the "option value" or learning aspect inherent in the uncertain arm: even if playing this arm gives the same payoff as the certain arm today, it may, in addition, generate information about the arm which

[4] Cf. Banks and Sundaram (1992a)

will be useful tomorrow, information which (by definition) is not generated by playing the certain arm. In this sense, then, behavior in Bandit Problem I should *never* be myopic, whereas behavior in Bandit Problem II should *always* be myopic.

Previous experimental studies of individual choice under uncertainty have focused on Bayesian updating or search models. The experiments of Cox and Oaxaca (1989, 1991) and Harrison and Morgan (1990)) examine search models where an agent's information is unaffected by an action {search, stop searching} unless the agents belief of the initial prior distribution of rewards is misspecified, (e.g. an agent has a prior with support (0,1) but observes an outcome of 2). Grether (1992) studied Bayesian updating and other heuristics that agents may use in making decisions under uncertainty. His results suggest that "in making judgments under uncertainty individuals use different decision rules in different decision situations" and what we want to discover as economists are the variables or factors in terms of which decision strategies are stable.[5] The focus of our experimental design is on the stationarity of strategies and the comparative static properties of the variables that theory predicts will influence behavior for relatively simple bandit problems.

The next section provides a more formal description of our Bandit problems along with the predictions and notation we will use in analyzing data from our experiments. We then present the experimental design and results.

2 Description of the bandit problems

A decision maker selects an option or arm $a \in \{A, B\}$, available at time $t = 1, 2, 3, \ldots$ After choosing an arm a reward is obtained which is a realization from a distribution with a possibly unknown parameter.

In Bandit Problem I arm A pays a constant amount of .50 each time it is selected, whereas option B's reward is uncertain. Specifically, B can be one of two possible types, good (G) or not (N), where B's type is the same for each time t but is unknown ex ante by the decision maker; let p be the prior belief that B is good. Type G generates a reward of 1 with probability g and 0 with probability $1 - g$, whereas a bad type generates 1 and 0 with probabilities n and $1 - n$, respectively, where $g > n$. Assume the decision maker has a per period utility function, updates her beliefs in a Bayesian fashion, discounts rewards by the factor $\delta \in (0, 1)$, and seeks to maximize discounted expected utility. An optimal strategy exists for this problem; indeed the following simple characterization describes the **stationary** optimal strategy:

[5] Recently there has been an emerging literature concerning Bayesian learning in games (see for example "Symposium on Bounded Rationality and Learning", Economic Theory, vol. 4,6,1994). However, this literature focuses on learning how to play a game against others.

> There exists a $p^* \in [0,1]$ such that at time t if the updated prior p_t concerning arm B is greater than p^* arm B is selected; otherwise arm A should be selected.

Bandit Problem II is the same as I, except that now A is structurally similar to B, *viz*. A can be one of two types, G and N (where these are the same possible types as for B), where the true types of the two arms are drawn independently. Let p_0^j be the prior belief $j \in \{A, B\}$ is a good type and p_t^j is the current belief about j being a good type. Then the optimal strategy is characterized as follows:

A (resp. B) is an optimal selection at time t if and only if $p_t^A \geq p_t^B$ (resp.$\leq$).

In both Bandit Problem I and Bandit Problem II the optimal decision rule depends on the decision maker's updated beliefs concerning whether the uncertain arm is good or bad, which given the setup here is a function of the number of 1's and 0's an arm has generated. To make this updating relatively transparent, in the experiments below, we assume that $g = 1 - n$; under this assumption, Bayesian updated beliefs will depend only on the *difference* between the 1's and 0's generated. With this in mind, define $C(B)$ for Bandit Problem I inductively as follows:

$$C(B_0) = 0$$

$$C(B_t) = \begin{cases} C(B_{t-1}) + 1 & \text{if } B \text{ selected at } t-1 \text{ and high payoff observed} \\ C(B_{t-1}) - 1 & \text{if } B \text{ selected at } t-1 \text{ and low payoff observed} \\ C(B_{t-1}) & \text{if } A \text{ selected at } t-1 \end{cases}$$

For Bandit Problem II

$$C(B_t) = \begin{cases} C(B_{t-1}) - 1 & \text{if } A \text{ selected at } t-1 \text{ and high payoff observed } \textit{or} \\ & B \text{ selected at } t-1 \text{ and low payoff observed} \\ C(B_{t-1}) + 1 & \text{if } A \text{ selected at } t-1 \text{ and low outcome } \textit{or} \\ & B \text{ selected at } t-1 \text{ and high outcome} \end{cases}$$

Using $C(B_t)$ we can restate the implications provided in the previous paragraph. Let $\sigma_t : C \to \{A, B\}$ denote the strategy based on these counts; then

(i) For Bandit Problem I there exists a **critical belief cut point**

$$c \in \{\ldots, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2 \ldots\}$$

$$\sigma_t = \begin{cases} B & \text{if } C(B_{t-1}) > c \\ A & \text{if } C(B_{t-1}) < c \\ A \text{ or } B & \text{if } C(B_{t-1}) = c \end{cases}$$

(ii) For Bandit Problem II the optimal strategy is

$$\sigma_t = \begin{cases} B & \text{if } C(B_{t-1}) > 0 \\ A & \text{if } C(B_{t-1}) < 0 \\ A \text{ or } B & \text{if } C(B_{t-1}) = 0 \end{cases}$$

The comparative statics for this problem can be computed to find that the cut point c will decrease as an individual's discount rate or level of risk aversion increases and c decreases as the probability of obtaining a high reward increases.

In our experiments, we will examine:

1. *Stationary strategies*, that is, strategies based on counts
2. *Myopic* behavior predicted for Bandit Problem II and nonmyopic behavior for Bandit Problem I, and
3. The comparative statics properties of cut points with respect to discount rates, reward probabilities, and risk aversion.

3 Experimental procedures and design

3.1 Risk attitudes and infinite horizons

From the discussion above, we note that the theory implies that an individual's specific cut point in Bandit Problem I depends on risk attitudes, discounting, priors, payoffs, etc. In Bandit Problem II none of these parameters are predicted to have an effect on individual cut points. Parameters such as the probabilities and arm payoffs can be induced and/or controlled; controlling discounting and risk is more problematic.

The lottery procedure of Berg, Daley, Dickhaut, and O'Brian (1986) (hereafter BDDO) is sometimes used in experiments to induce specific risk attitudes on subjects. BDDO is a generalization of a procedure proposed by Roth and Malouf (1979). The BDDO procedure induces any prespecified risk preference. It is a two phase decision process. First, subjects choose actions that yield "points" which are stochastically related. Second, the number of points determine the probability of winning some dollar amount in a lottery. Walker, Smith and Cox (1986) concluded that the "lottery payoff" procedure does not seem to work for the first price auction. Results by Rietz (1993) suggest that the procedure may work if carefully applied, however, Cox and Oaxaca (1993) demonstrate that Rietz's results are inconclusive at best. Other authors such as Cooper et al. (1989) have indicated that BDDO does not appear to change behavior.

For our experiments we use the elicitation technique of Becker, DeGroot, and Marschak (1964) (hereafter BDM) to obtain certainty equivalents to see if there is a correlation between responses from the BDM procedure and decisions made in the Bandit Problems part of our experiments.

To induce discounted infinite horizons we employ a probabilistic end rule used in previous experimental studies (see for example Camerer and Weigelt (1993)). That is, after each decision, there is a fixed and known probability that the period will end. This procedure has some problems that cause concern. The inability of a subject to understand the probabilistic nature of the end rule (or probability at all) is possible. Second, the (small)

chance that an experiment will continue for more than several hours is not credible. Nonetheless, the ability to understand probability is an attribute of the population, as such we realize that only comparative statics may be valid with this procedure. Concerning the second issue, the beliefs that subjects may have on the actual ending rule, we simply announced a specific ending rule (a maximum time for the experiment or a maximum number of periods whichever happens first).

3.2 Environment parameters (treatments)

For all experiments there was a choice between one of two options (arms) called A and B. In Bandit Problem I there is one certain (A) and one uncertain (B) arm and in Bandit Problem II both arms are uncertain. If the arm is uncertain then it can be one of two types called good and bad, with the prior probability of an arm being good fixed at .50. The payoffs of the uncertain arm (high and low) is fixed for all experiments (high $= 100$ tokens low $= 0$ tokens). The payoff for the certain arm is 50 tokens. Ten tokens were equivalent to 5 cents.

We induce discounting in our "infinite" horizon models with the probabilistic end rules with either of two **probability of continuation** (δ) values 0.80 and 0.90. Each subject was advised of this fact along with the average period length before end. These values are provided in Table 1.

These "discount rates" were selected because δ less than 0.8 will have an expected length less than 5, and we will not be able to observe long runs. Any δ greater than 0.9 has a high probability of having a large number of rounds and thus there will likely be a number of periods that will end from the time limit rule.[6]

In addition to the "discount rate", we also vary g, the probability that the good type has a high payoff (recall that we always set $n = 1 - g$). We use $g = .7$ and .9. Given the values of g and δ we can calculate the optimal critical belief cut point c^* for a risk neutral player. Table 2 lists the treatments, the associated optimal risk neutral cut point (c^*) and posterior probability prediction. Both c and p^* describe the stationary optimal

Table 1. Probability of period end and expected number of rounds

Chance period will last	$\delta = .8$	$\delta = .9$
5 more rounds	32 out of 100	59 out of 100
10 more rounds	10 out of 100	34 out of 100
5 more rounds	3 out of 100	20 out of 100
30 more rounds	1 out of 100	4 out of 100
Expected number of rounds per period	5	10

[6] For $\delta = 0.95$, there is an 8% chance that the number of trials is greater than 50, and there is almost a 2% chance that the number of rounds is greater than 100.

Table 2. Parametric treatment conditions

Treatment	c	p^*
$\delta = .8,\ g = .7$	-0.5	0.32
$\delta = .8,\ g = .9$	-0.5	0.20
$\delta = .9,\ g = .7$	-1.0	0.23
$\delta = .9,\ g = .9$	-0.5	0.12

strategy for Bandit Problem I as described in Section 2. For Bandit Problem II the optimal critical belief cut point (c) is always equals to zero. In Bandit Problem I the optimal probability cut point (p^*) describes the subjects stationary optimal strategy in terms of the probability that arm B is good; that is, if the probability that arm B is good is greater than (p^*), then the optimal strategy is to choose arm B. The optimal critical belief cut point (c) describes the subject's stationary optimal strategy in terms of the realization of low and high payoffs; that is, if the count of high and low payoffs $C(B)$ is greater than c then the optimal strategy is to choose arm B. Since we have set $n = 1 - g$ the stationary optimal strategy based on c is equivalent to the stationary optimal strategy based on (p^*). In the following analysis we will focus attention on the optimal critical belief cut point strategy based on c.

3.3 Experimental procedures

Subjects were recruited from the student population at the University of Amsterdam and all experiments were conducted at the Center for Experimental Economics and Political Decisionmaking (CREED). An experimental session was constructed as follows (instructions were computerized; an abbreviated set of instructions is supplied in Appendix A):

(a) Each subject was seated at an individual personal computer, the experimental program is started with a predetermined set of parameters and instructions. Each subject proceeds through the instructions for the BDM process at their own pace, and can practice as much as they like without any time constraints. Each subject's experimental session is independent of any other subject's session.
(b) The subject is then asked to answer 4 different BDM questions in which they are paid based on their response/outcome. The specific BDM lottery questions are provided in the instrutions in Appendix A.
(c) After the BDM procedure the subject proceeds through the instructions for one of the bandit problems.
(d) Each experimental session consists of periods and rounds in a period. At the beginning of each period the state of each uncertain arm is drawn according to the probabilities that are given to the

subject at the beginning of the experiment. Information concerning payoffs and probabilities was provided to subjects.

(e) A sequence of rounds is then run in each period. At the beginning of a round the subject is asked to choose an arm. Given the subjects choice and the fixed and known probabilities, the subject's payoff is drawn. A random number is then drawn to determine if the period is to continue. We incorporate a computerized roulette wheel for the random draw of the stopping rule, this feature was used to reduce subject boredom and automatic responses. If the period continues a new round is run. If the period ends then a new period begins and new state variables are drawn (i.e. the types of the arm good or bad) the subject is informed that this is being done. The subject screen is provided below:

(f) Each subject plays one of the bandit problems for a maximum amount of time (60 minutes) or a maximum number of periods (5) whichever comes first.

(g) Each subject repeats each period with the same parameters given to them at the beginning of the experimental session.

3.4 Experimental design summary

Table 3 provides a summary of the experiments we conducted.

4 Experimental results

We focus on cut points to describe individual decision-making in our experiments. In the subsections that follow we calculate the *best cut point c* for each subject. The cut point c is determine by finding the value that results in the smallest number of observed deviations. For example, Figure 2 supplies decisions made by a subject in our experiments.

In period 1 of the figure we find that any cut point less than 0 describes the choice pattern. For period 2 the best cut point is any number between -1 and -2, while for period 3 it is -1. We will use the midpoint of the period best cut point intervals for each subject to construct a cut point distribution to investigate "stationary."

Using the best cut points, we investigate the major comparative predictions of the model of optimal choice. Specifically, we want to know whether behavior is more myopic behavior in Bandit Problem II than Bandit Problem I, whether the comparative static predictions hold for the discount rate (δ) and probability of a high value draw (g), and if the strategies selected are stationary.

4.1 Aggregate cut point behavior

Consider the decisions made by the subject graphed in Figure 2. In period 1 of the experiment (Bandit Problem I with low discount and high reward

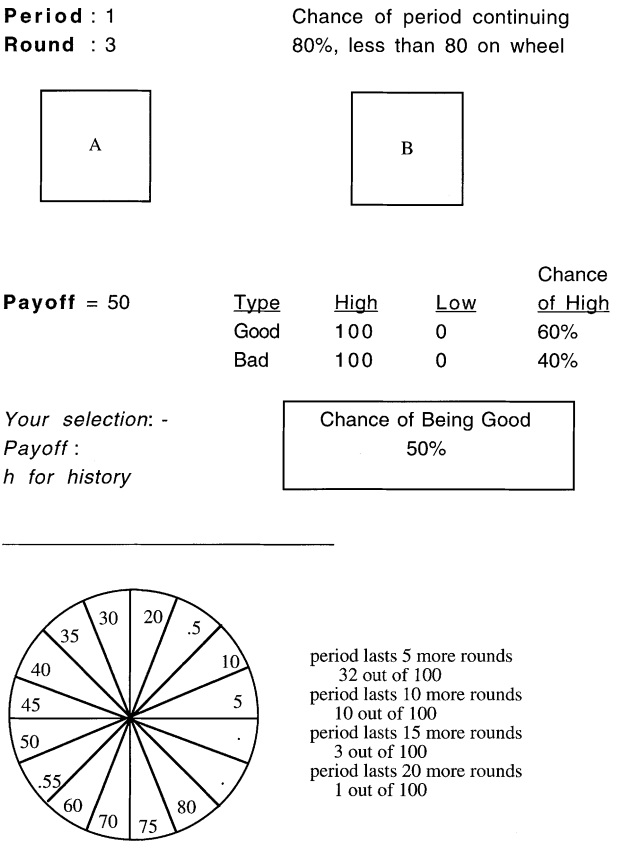


Figure 1. Subect screen layout

Table 3. Experimental design summary

Treatment	Number of Experiments (Subjects)
Bandit Problem I: $\delta = .8,\ g = .7$	20
Bandit Problem I: $\delta = .8,\ g = .9$	19
Bandit Problem I: $\delta = .9,\ g = .7$	18
Bandit Problem I: $\delta = .9,\ g = .9$	19
Bandit Problem II: $\delta = .8,\ g = .7$	14
Bandit Problem II: $\delta = .8,\ g = .9$	12
Bandit Problem II: $\delta = .9,\ g = .7$	11
Bandit Problem II: $\delta = .9,\ g = .9$	13

probability) the subject selected Arm *B* each time and obtained the high payoff each time. Thus, the best cut point for period for this subject is any number less than 0. For period 2 the best cut point is any number between -1 and -2. In period 3 the best cut point is -1. Using the midpoint of the best cut point interval for our estimate of c we obtain the following result.

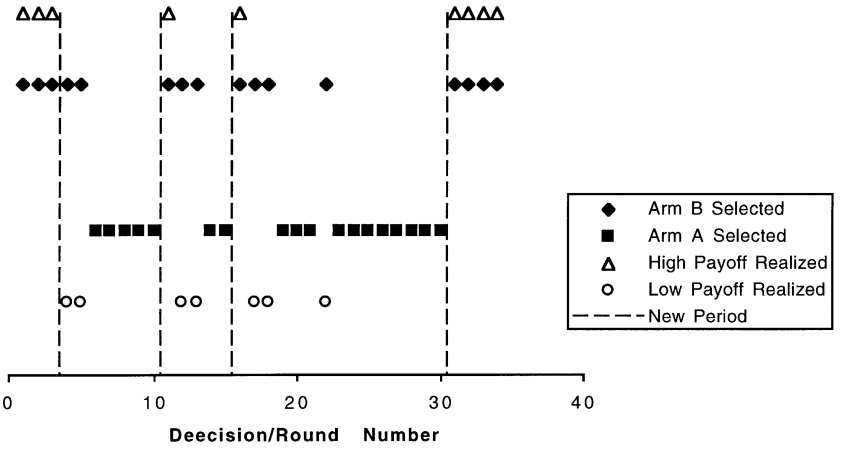


Figure 2. Subject 3 decisions: Bandit Problem I ($\delta = 0.8, g = 0.9$)

Observation 1. The distribution of best cut points are significantly different between the two Bandit Problems. Specifically, the cut point distribution is consistent with myopic behavior with Bandit Problem II but not for Bandit Problem I.

Support. Figures 3 and 4 show that the distribution of cut points and Table 4 shows the descriptive statistics for these distributions for Bandit Problems I and II. Table 4a provides the mean, standard deviation and median of cut points taken across all periods for each Bandit Problem. In addition, it provides the degrees of freedom for the F-statistic and the probability level for the Hotelling T^2 tests.[7]. The test for the equality of means across the Bandit Problems shows that there is a significant difference in cut points for the two Bandit Problems. We can also reject the hypothesis that behavior is myopic for Bandit Problem I; we cannot reject the hypothesis of myopic behavior for Bandit Problem II. The Hotelling T^2 test treats each subject's 5 period cut points as an individual observation. Table 4b provides the mean cut points per period for each Bandit Problem.

Thus, not only do we see that behavior is quite different between each of the Bandit Problems, they are consistent with the predicted direction for the behavior of the cut points. Recall that for Bandit Problem II the cut point should always be 0 and is not affected by risk aversion or any of the treatments we used. On the other hand, Bandit Problem I has clear comparative static predictions; $\partial c/\partial \delta < 0$, $\partial c/\partial g < 0$ and the more risk

[7] The Hotelling T^2 test is the multivariate analog of Student's t-test (see for example Giri (1995)). We use this test since for each subject we have multiple observations (estimates for each of 5 periods). We cannot assume that these observations are independent. The Hotelling T^2 test does not require independence since each subject's period cut points are treated as a single 5-dimensional multivariate observation from a 5-dimensional normal distribution with no restrictions on the covariance matrix

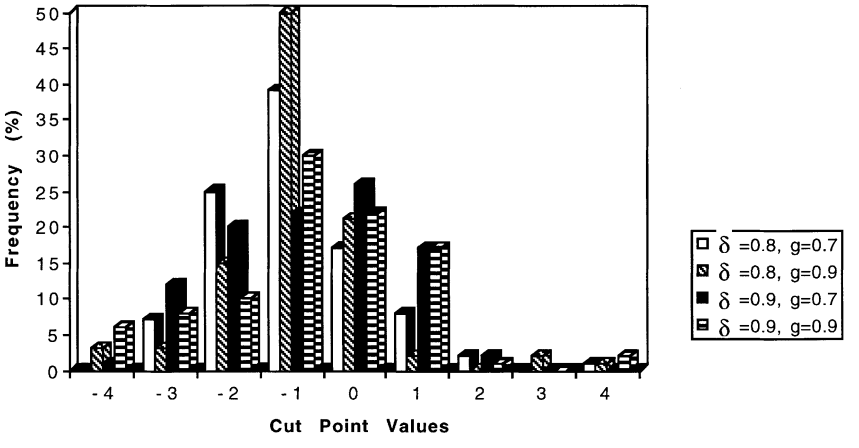


Figure 3. Distribution of cut points (Bandit Problem I)

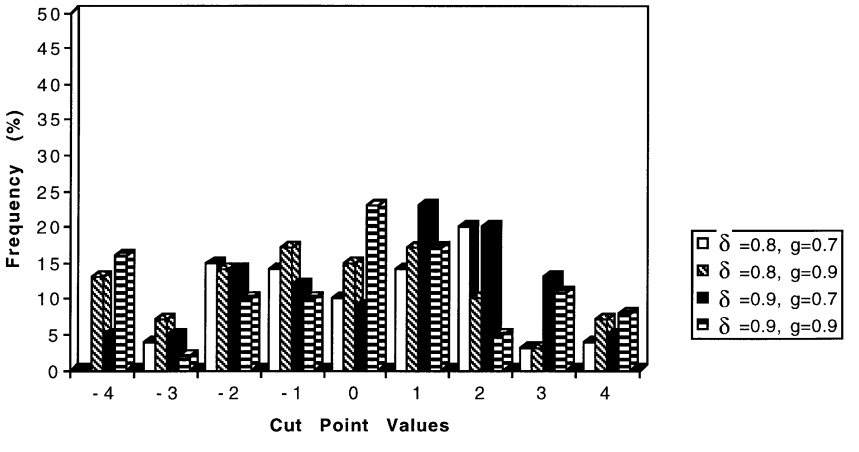


Figure 4. Distribution of cut points (Bandit Problem II)

averse the subject is, the lower his cut point. To investigate the comparative static properties of the Bandit Problems we estimate the following statistical model:

$$c = \alpha + \beta^* \Gamma + \gamma^* \Delta + \theta^* \Gamma\Delta + \lambda^* x + \nu^* r + \varepsilon$$

where

$$\Gamma = \begin{cases} 1 & \text{if } \delta = .9 \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta = \begin{cases} 1 & \text{if } g = .9 \\ 0 & \text{otherwise} \end{cases}$$

Table 4a. Best cut point estimates

Problem	Mean	Standard Deviation	Median	Degrees of Freedom	T^2F statistic	p-value
Bandit I	-0.80	1.38	-0.50	(5,71)	27.75	0.00
Bandit II	-0.03	1.44	0.00	(5,45)	1.84	0.12
Bandit I = Bandit II	NA	NA	NA	(5,120)	8.92	0.00

Table 4b. Mean cut points by period

Problem	1	2	3	4	5
Bandit I	-0.85	-0.80	-1.06	-0.72	-0.59
Bandit II	0.45	-0.30	0.18	-0.35	-0.14

$$\Gamma\Delta = \text{interaction term}$$

$$x = \begin{cases} 1 & \text{if subject is experienced} \\ 0 & \text{otherwise} \end{cases}$$

The r variable above is our proxy for individual's risk attitudes which we derived from subject responses to the BDM procedure used in our experiments. We construct a nonparametric estimate by calculating the number of times in which subject i's selling price for the BDM lottery $j(S_{ij})$ is greater than or less than the risk neutral selling price for lottery $j(S_{rj})$. Specifically, for subject i we calculate the value:

$$r_i = \#(S_{ij} > S_{rj}) - \#(S_{ij} < S_{rj})$$

where $\#(x)$ is the number of instances in which x is true.

Thus $r_i \epsilon [-4, 4]$, $r_i = 0 \Rightarrow$ risk neutrality, and a larger r_i shows increasing risk aversion.[8]

The theory presented in section 2 predicts that the treatment effects should all be negative and the risk preference effect should be positive for Bandit Problem I. For Bandit Problem II all of the parameters should be zero.

Observation 2. In Bandit Problem II none of the treatments or risk preferences has an effect on subjects cut point strategies

Support. In the table below, the estimates of the statistical model with their associated t-statistics and p-values are provided. Each of the estimates are not significantly different from zero. Thus, we cannot reject the hypothesis that the parameters have no effect on individual cut point strategies.

[8] We calculate two other measures based on the BDM responses which we supply in Appendix B. In all cases the measures are consistent with subject risk aversion. Furthermore, none of the measures affect the conclusions listed in observations 2 and 3.

Model estimates for Bandit Problem II

Variable Name	Estimate	*t*-statistic	*p*-value
Constant(α)	0.009	0.061891	0.951
Discount Rate (β)	−0.146	−0.710164	−0.481
High Value Prob.(γ)	0.075	0.351988	0.727
Interaction (θ)	0.110	0.368470	0.714
Experience (λ)	0.027	0.168079	0.867
Risk Preference (ν)	−0.001	−0.032690	−0.974
# of observations	50		

Observation 3. In Bandit Problem I the model estimates are consistent with the comparative static predictions, i.e. the sign of the coefficients are as predicted. However, none of the treatments or individual risk preference estimates are significant.

Support. The Bandit Problem I estimates and statistics are provided in the table below. Thus, while cut points are nonmyopic (we reject the hypothesis that $\alpha = 0$) the strategies are not significantly affected by discount rates, probability of good draws or risk aversion.

Model Estimates for Bandit Problem I

Variable Name	Estimate	*t*-statistic	*p*-value
Constant (α)	−0.492	−4.229269	0.000
Discount Rate (β)	−0.221	−1.370419	0.175
High Value Probability (γ)	−0.082	−0.503916	0.616
Interaction (θ)	0.338	1.463447	0.148
Experience (λ)	−0.212	−1.494695	0.139
Risk Preference (ν)	0.008	−0.287399	0.775
# of observations	76		

Notice that in both cases the risk preferences parameter has the least affect on the best cut point, while the probability of a high draw has the biggest impact. This outcome could be an indication that the BDM procedure may not be accurate in eliciting true risk preferences.

All of the conclusions we have presented to this point utilizes a best cut point estimate for each period. However, one of the conclusions of the model is that the cut point strategies are stationary, i.e., the cut points should not change from period to period. We now check to see if this is the case.

Observation 4. The best cut point estimates are not stationary.

Support. Using the Hotelling T^2 test we examine the hypothesis that the cut points are equal across periods. The F-statistics for the test are provided for each treatment along with their corresponding p-values. We see that in seven

out of eight of the treatments we can safely reject the hypothesis of stationary cut points.

Bandit Problem I

Treatment	Degrees of freedom	*F*-statistic	*p*-value
$\delta = .8\ g = .7$	(4,16)	9.83	0.00
$\delta = .8\ g = .9$	(4,15)	3.16	0.05
$\delta = .9\ g = .7$	(4,14)	9.88	0.00
$\delta = .9\ g = .9$	(4,15)	3.84	0.02

Bandit Problem II

Treatment	Degrees of freedom	*F*-statistic	*p*-value
$\delta = .8\ g = .7$	(4,10)	3.94	0.04
$\delta = .8\ g = .9$	(4,8)	5.85	0.02
$\delta = .9\ g = .7$	(4,7)	14.27	0.00
$\delta = .9\ g = .9$	(4,9)	1.59	0.26

While the aggregate cut points are not stationary, these cut points select a large number of the individual choices. Using the best cut point estimates for a subject, we can determine how many choices would be consistent with the cut point and how many would not be consistent. In the table below we pool over all subject and periods, the number of consistent choices from the best cut points as a percentage of all choices made in the experiments. Notice that the best cut point predicts over 90% of the choices in most cases and no less than 80% for any one treatment.

Percent of choices consistent with best cut point estimates

	Bandit Problem I			Bandit Problem II		
Treatment	% Consistent	Standard Deviation	# of choices	% Consistent	Standard Deviation	# of choices
$\delta = .8\ g = .7$	95	11	100	93	14	70
$\delta = .8\ g = .9$	99	5	95	92	14	60
$\delta = .9\ g = .7$	90	15	90	80	18	55
$\delta = .9\ g = .9$	94	12	95	90	15	65

5 Conclusions

Our research program presented above was relatively straightforward: design experiments to investigate individual decision-making behavior in two simple but separate simple bandit problems. Then, determine the pattern of behavior in the two environments relative to the following theoretical predictions:

1. Cut point behavior should be consistent with stationary cut point strategies for both environments.

2. In one of the environments behavior should be consistent with myopic behavior while the other should exhibit non-myopic behavior.
3. For various parameter choices (discount rates and Probabilities of good outcomes) the comparative static predictions should be neutral in the myopic case important in the non-myopic environment.

The experimental results suggest that there are individual violations of stationarity in cut point behavior but that the most likely collection of decision rules that explains subject behavior are stationary strategies.[9] For all the measures we used there is a clear distinct pattern in the two bandit environments: behavior is more myopic in the environment in which that form of behavior was predicted. The comparative static results show that none of the treatments have a significant effect on choices made in the myopic case. While the treatment effects are not significant for the nonmyopic environment, all of the effects have the predicted direction of influence.

Appendix A: Instructions

Below is an abbreviated set of the computerized instructions we used.

Screen 1

Instructions. You are about to participate in an individual decision making experiment. The decisions you will make during the experiment will result in Dutch guilder profits that will be yours to keep. Thus, you should follow the instructions carefully to understand how you can make profits. In this experiment all values will be stated in terms of tokens. Each token you earn can be redeemed into guilders at a rate of ___ tokens per Dutch guilder. The experiment will be broken-up into two different parts in which you make decisions and earn profits.

⟨Press any key to continue⟩

Screen 2

Instructions for Part 1. In this portion of the experiment you will be asked a series of questions. Given the answer to the questions a spin of a computerized roulette wheel will be made. Given your answer and the outcome of the roulette spin you will earn profits and proceed to part 2 of the experiment.

We will now take you through an example of how decisions made during this part translates into profits. In questions 1 and 2 a roulette wheel containing the twenty numbers $5, 10, 15, \ldots, 100$ will be spun (all numbers are equally likely to be selected). If the wheel selects a certain set of numbers

[9] In Banks et al. (1994), we use the likelihood search procedure of El-Gamal and Grether (1995) over cutpoint strategies to further support this result.

you will receive nothing, if the wheel selects any number not in that set you will receive a fixed amount of tokens. You will then be asked how much you would be willing to sell this game for. The roulette wheel will then be spun and if the number the wheel selects is equal or greater than the price you asked, you will be paid that amount. If the number the wheel selects is less than the price you ask, you will play the game.

We will now go through several sample questions of this type so that you can see how it works. The outcomes of these sample questions will not count toward your profits; this is only for practice.

⟨press B to go back or any other key to continue⟩

Screen 3

Practice round. In this game we will spin the roulette wheel with the twenty numbers $5, 10, 15, \ldots, 100$. If a number less than or equal to 50 is selected you win nothing. If the number is greater than 50 you will receive 100 tokens

How much are you willing to accept instead of playing this game?

Please enter a number between 0 and 100.

Given that you asking ___, the roulette wheel will be spun and if it selects a number that is greater than the price you are asking you will be paid that amount. Otherwise you will play the game.

Random spin of the Roulette Wheel

The wheel landed on the number ___ which less than your asking price of ___ Thus, you must play the game.

Random spin of the Roulette Wheel

The wheel landed on the number ___ which is less than or equal to ___ so you win 0 tokens.

⟨Press any key to continue⟩

Bandit instructions

Bandit Problem 1

Screen 1

Instructions for Part 2. In part 2, the experiment will be broken up into ___ periods. Each period in turn will be divided into rounds in which you will make decisions and earn profits. At the beginning of a round, you will make a choice between two alternatives called A and B. The A alternative will pay you ___ tokens if you select it. The B alternative will be one of two possible types which we will call good and bad. If B is good, you will receive ___ tokens with a specified chance and ___ tokens with a specified chance. If B is bad, then your chance of obtaining ___ tokens will be lower and the chance of obtaining ___ tokens will be higher. After you select either alternative A or B you will be

informed of your paroff for the round and the roulette wheel will be spun. If the wheel selects a certain set of numbers the period will go to the next round, otherwise the period will end and we will proceed to a new period.

Before you go on to the experiment you will go through some practice periods. The outcomes of your decisions in these practice periods will not count toward your profits; they are only for practice.

Screen 2

In round 1, of each period of this experiment, the A alternative pays ___ tokens. The B alternative could be a good type that pays ___ tokens with a ___ percent chance and ___ tokens with a ___ percent chance. This means, if B is good, then over many rounds you could expect to earn, on average ___ tokens a round. On the other hand, B could be a bad type which pays ___ tokens with a ___ percent chance and ___ tokens with a ___ percent chance., This means if B is bad, then over many rounds you could expect to earn, on average, ___ tokens a round. You will be given information concerning the chance that B is a good or a bad type. Lastly, the chance that a period will continue after the current round is ___ percent. That means that, on average, in any round in a period,, you could expect the period to last ___ more rounds. Further information will be handed out to you.

⟨Press B to go back or any other key to continue⟩

Screen 3

We now summarize the specific features of this experiment:

1. Alternative A pays ___ tokens.,
2. At the beginning of a period, the B alternative will be selected as either a good type or a bad type with a fixed chance. It will remain that type for the entire period.
3. The chance that B is good is ___ percent.
4. If B is good it will pay ___ tokens with a ___ percent chance and ___, tokens with a ___ percent chance inthe first iteration of each, period.
5. If B is bad it will pay ___ tokens with a ___ percent chance and ___, tokens with a ___ percent chance.
6. The chance that the period continues at the end of the round is ___ percent.,
7. Your conversion rate is 1 Guilder for ___ tokens.

⟨Press B to go back or any other key to continue⟩

Screen 4

Before you begin the practice period, there are several features of the program that may be helpful. At any time you can press 'H' to see the history of your

choices and the outcomes. The screen always shows the payoff chances under a good type and the payoff chances for a bad type.

If you understand the process and want to practice press Enter, otherwise raise your hand and an experimenter dude will answer your questions.

There will be 2 practice periods.

⟨Press B to go back or any other key to practice⟩

Bandit Problem 2

Screen 1

Instructions for part 2. In part 2, the experiment will be broken up into periods. Each period in turn will be divided into rounds in which you will make decisions and earn profits. At the beginning of a round, you will make a choice between two alternatives called A and B. Alternatives A and B will be one of two possible types which we will call good and bad. A and B may be the same or different types. If your choice (A or B) is good, you will receive ___ tokens with a specified chance, and ___ tokens with a specified chance. If your choice (A or B), is bad, then your chance of obtaining ___ tokens will be lower and, the chance of obtaining ___ tokens will be higher. After you select either alternative A or B you will be informed of your payoff for the round and the roulette wheel will be spun. If the wheel selects a certain set of numbers the period will go to the next round, otherwise the period will end and we will proceed to a new period. We will now take you through an example of how decisions made during this part translates into profits. The outcomes of your decisions in these instructions will not count toward your profits. This is only for practice. Press any key to continue

Screen 2

Suppose we are in period 1, round 1. The alternatives A and B could be a good type that pays ___ tokens with a ___ percent chance, and ___ tokens with a ___ percent chance. This means that if A or B, is good, then over many rounds you could expect to earn, on average ___ tokens a round if you always choose that alternative. On the other hand, A or B could be a bad type which pays ___ tokens with, a ___ percent chance and ___ tokens with a ___ percent chance. This means if A or B is bad, then over many rounds you could expect to earn, average, ___ tokens a round. Remember that you do not always have to make the same choice, at each round you may either choose A or B. You will be given information concerning the chance that A or B is a good or a bad type. Lastly, the chance that a period will continue after the current round is ___ percent. That means that, on average, in any round in a period, you could expect the period to last ___ more rounds. Further information will be handed out to you.

⟨Press B to go back or any other key to continue⟩

Screen 3

We now summarize the specific features of this practice period:

1. At the beginning of a period, the A alternative will be selected as either a good type or a bad type. It will remain that type for the entire period.
2. At the beginning of a period, the B alternative will be selected as either a good type or a bad type. It will remain that type for the entire period.
3. The chance that alternative A is good is ___ percent.
4. The chance that alternative B is good is ___ percent.
5. If A or B is good it will pay ___ tokens with a ___ percent chance and ___ tokens with a ___ percent chance.
6. The chance that the period continues at the end of the round is ___ percent.

A sheet with this information will be given to you.

⟨Press B to go back or any other key to continue⟩

Screen 4

Before you begin the practice period, there are several features of the program that may be helpful. At any time you can press h to see the history of your choices and the outcomes. The screen always shows the payoff chances under a good type and the payoff chances for a bad type. If you understand the process and want to practice press enter, otherwise raise your hand and a monitor will answer your questions.

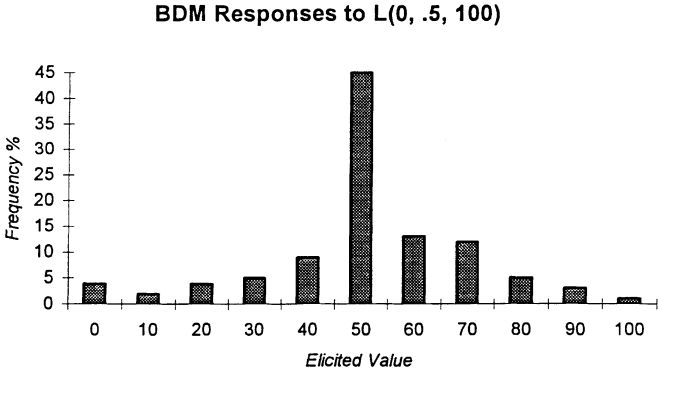
⟨Press B to go back or any other key to practice⟩

Appendix B

Three alternative measures of risk tolerance using data from the BDM procedure in our experiments are described below.

Let $L_j(l_j, p_j, h_j)$ denote the lottery j where you receive the amount l_j with probability $(l - p_j)$; you receive h_j with probability p_j.

The distribution of the selling price of the lottery $L(0, .5, 100)$ we used in the experiment is provided below (risk neutral response = 50):



Measures

1. Non-parametric statistic:
   For each subject i we calculate the statistic

$$r_i = \#(S_{ij} > S_{rj}) - \#(S_{ij} < S_{rj})$$

where

S_{ij} = subject i's selling price for lottery j;
S_{rj} = risk neutral selling price of lottery j;
$\#(\cdot)$ = number of instances in which the condition $\cdot$ is true

$r_i = 0$: risk neutral behavior
$r_i > 0$: "risk loving" behavior
$r_i < 0$: "risk averse" behavior

2. Median ratio log statistic:

r_i = median $_j[\ln(U_{vi})/\ln(U_{si})]$ – median over the j lotteries for subject $i\}$

where

$U_{vj} = (S_{rj} - l_i)/(h_j - l_i)$ – {normalized utility of risk neutral value of lottery $j\}$
$U_{sj} = (S_j - l_i)/(h_j - l_i)$ – {normalized utility of i's selling price for lottery $j\}$

$r_i = l$: risk neutral behavior
$r_i > l$: risk loving behavior
$r_i < l$: risk averse behavior

3. Median of selling prices:

$$r_i = \text{median of } S_j \text{ selling rices for subject } i$$

The median risk neutral selling price 52.5. Thus $r_i < 52.5$ implies risk averse behavior and $r_i > 52.5$ shows more risk loving behavior.

The table below supplies the descriptive statistics for each of the measures described above.

Measure	mean	Standard deviation	minimum	maximum	median	n
Non-parametric	−.38	2.05	−4	4	−1	126
Median ratio log	1.99	6.51	.13	59.9	.87	126
Median of selling prices	49.9	17.92	0	100	50	126

Finally, in the table below we provide correlations between the three measures of risk attitude and the selling price of the $L(0, .5, 100)$ lottery.

non-parametric	1.00	0.37	0.84	0.77
median ratio log	0.37	1.00	0.45	0.35
median of selling prices	0.84	0.45	1.00	0.90
$L(0, .5, 100)$	0.45	0.35	0.90	1.00

References

Akaike, H.: A new look at the statistical identification model IEEE Transactions on Automatic Control **19**, 716–723 (1974)

Banks, J. S., Sundaram, R.K.: A class of bandit problems yielding myopic optimal strategies. Journal of Applied Probability, **29**, 625–632 (1992a)

Banks, J. S., Sundaram, R. K.: Denumerable-armed bandits. Econometrica **60**, 1071–1096 (1992b)

Banks, J. S., Olson, M., Porter, D.: An experimental analysis of the two-armed bandit problem. Caltech Social Science Working Paper 892 (1994)

Becker, G.M., DeGroot, M.H., Marschak, J.: Measuring utility by a single-response sequential method. Behavioral Science **9**, 226–232 (1964)

Berg, J. E., Daley, L. A., Dickhaut, J. W., O'Brian, J. R.: Controlling preferences for lotteries on units of experimental exchange. Quarterly Journal of Economics **101**, 281–306 (1986)

Camerer, C.E., Weigelt, K.: A test of a probabilistic mechanism for inducing stochastic horizons in experiments. In: Research in Experimental Econimics, V 6, Isaac (ed.) R. M. (1993)

Cooper, R.W., DeJong, D. V., Forsythe, R., Ross, T. W.: Communication in Battle-of-the-Sexes games: some experimental results. Rand Journal of Economics **20**, 568–587 (1989)

Cox, J., Oaxaca, R.: Laboratory experiments with a finite horizon job-search model. Journal of Risk and Uncertainty **2**, 301–329 (1986)

Cox, J., Oaxaca, R.: Finite horizon search behavior with and without recall. Presented at the Economic Science Meetings, Tucson, AZ (1991)

Cox, J. C., Oaxaca, R.L.: Introducing risk neutral preferences: further analysis of the data. Presented at the Economic Science October Meetings (1993)

El-Gamal, M., Grether, D. M.: Are people bayesian ? Uncovering behavioral strategies. Journal of the American Statistical Association **90**, 1137–1145 (1995)

Giri, N.: Multivariate statistical inference. Academic Press, New York (1995)

Gittins, J., Jones, D.: A dynamic allocation index for the sequential allocation of experiments. In: Progress in Statistics J. Gani et al.(eds.) Amsterdam: North Holland, pp 241–266 (1974)

Grether, D. M.: Testing Bayes rule and the representativeness heuristic: Some experimental evidence. Journal of Economic Behavior and Organization **17**, 31–57 (1992)

Harrison, G.W., Morgan, P.B.: Search intensity in experiments. Economic Journal **100**, 478–486 (1990)

Ishikida, T.: Informational aspects of decentralized resource allocation. Technical Report, UC Berkeley, Interdisciplinary Group on Coordination Theory (1992)

Jovanovic, B.: Job-search and the theory of turnover. Journal of Political Economy **87**, 972–990 (1979)

Mellers, B., Ordonez, L., Birnbaum, M.: A change-of-process theory for contextual effects and preference reversals in risky decision making. Organizational Behavior and Human Decision Processes **52**, 331–369 (1992)

Mortensen, D.: Job-search and labor market analysis. In: Ashenfelter O., Layard R.(eds). Handbook of Labor Economics, Vol. II Amsterdam, North Holland, pp 849–919 (1985)

Rietz, T.A.: Implementing and testing risk-preference-induction mechanisms in experimental sealed-bid auctions. Journal of Risk and Uncertainty, **7**, 199–213 (1993)

Roth, A.E., Malouf, M.W.K.: Game-theortic models and the role of information in bargaining. Psychological Review, **86**, 574–594 (1979)

Rothschild, M.: A two-armed bandit theory of market pricing, Journal of Economic Theory **9**, 185–202 (1974)

Viscusi, W.: Job-hazards and worker quit rates: an analysis of adaptive worker behavior. International Economic Review **20**, 29–58 (1979)

Walker, J. M., Smith, L., Cox, J.C.: Inducing risk neutral preferences: an examination in a controlled market environment. Journal of Risk and Uncertainty **3**, 5–24 (1990)

Wilde, L.: An information-theoretic approach to job quits. In: S. Lippman, J. McCall, (eds.) Studies in the Economics of Search. New York: North Holland, pp 35–52 (1993)