Ambiguity Aversion in Multi-armed Bandit Problems

Christopher M. Anderson*
Department of Environmental and Natural Resource Economics
University of Rhode Island
Kingston, RI 02881

September 18, 2000

Abstract

In multi-armed bandit problems, information acquired from experimentation is valuable because it tells the agent whether to select a particular option again in the future. This paper tests whether people undervalue this information because they are ambiguity averse, or have a distaste for the variance of the unknown mean of the payoff distribution. Using Kahn and Sarin's (1988) model of ambiguity aversion, I derive the prediction that ambiguity averse agents have lower than optimal Gittins indexes, appearing to undervalue information, but are willing to pay more than ambiguity neutral agents to learn the true mean of the payoff distribution, appearing to overvalue information. This prediction is tested with a laboratory experiment which elicits a Gittins index and a willingness to pay on six two-armed bandits. Consistent with the predictions of ambiguity aversion, the Gittins indexes are significantly lower than optimal and the willingness to pay are significantly higher than optimal. However, the magnitude of the errors is invariant to small changes in ambiguity.

*This draft is preliminary. Comments welcome at cma@uri.edu. I am grateful for insightful discussions with and encouragement from Jeff Banks, Colin Camerer and Paolo Ghirardato, an audiences at the 2000 Fall Meetings of the Economics Science Association and the University of Rhode Island for helpful comments, and the Russell Sage Foundation for financial support. The instruments for the experiments presented here are available at http://eeps6.caltech.edu/~cma/mab/ambig/OneAskMain.html. and http://eeps6.caltech.edu/~cma/mab/ambig/SellAskMain.html.

1 Introduction

In many economically significant environments, agents must repeatedly choose among uncertain alternatives about which they can learn only through experimentation. Examples include the situations of a shopper deciding whether to purchase his favorite brand of orange juice or experiment with a new one he has never tried and an oil company deciding whether to continue testing a tract of land or to move its equipment to another tract. If these agents do not experiment enough, they can lose considerable welfare: the shopper could miss out on a delicious new brand of juice he would purchase and enjoy in the future, and the oil company may engage in an expensive recovery operation based on too few good test results. On the other hand, if these agents experiment too much, they may lose welfare as they pursue inferior choices.

Despite the economic importance of this sort of experimentation, little is known about how agents approach such problems. Building on findings that agents do not experiment enough in search problems (Cox and Oaxaca, 1989; 1990; 1992; 1996; Pratt, Wise and Zeckhauser, 1979), Anderson (2000b) demonstrated that laboratory subjects do not experiment enough when choosing among uncertain alternatives. Similarly, when directly queried, laboratory subjects had lower than optimal Gittins indexes (Anderson, 2000a). However, the potential explanations of risk aversion, unobserved experimentation cost and hyperbolic discounting, did not explain the underexperimentation. Therefore, the extant research leaves us without an understanding of why agents' bandit strategies are suboptimal. This understanding could help economists develop carefully tailored public policies and corporate strategies to help improve the welfare of those facing bandit problems.

This paper considers considers the suboptimality could be attributable to ambiguity aversion. It develops Kahn and Sarin's (1988) model of ambiguity aversion into a model of behavior in bandits with Bernoulli payoffs and beta priors. The formal model generates a seemingly paradoxical prediction: agents' Gittins indexes will be lower than optimal, so it appears they do not value information enough, but they will pay more than optimal for information about the value of an uncertain arm, so it appears they value information too much. This prediction is tested directly in a laboratory experiment.

The next section of this paper explains some of the many applications of the bandit environment. Section 3 defines the a multi-armed bandit and presents the notation which will be used in subsequent sections. Section 4 distinguishes risk from ambiguity, and introduces the model of ambiguity aversion which will be adapted to bandits. Section 5 discusses the interpretation of ambiguity aversion, especially the crucial assumption that, although ambiguity aversion is modeled by a failure to reduced compound lotteries, agents use Bayes rule to update their prior beliefs to reflect observed payoffs. It then goes on to prove that the Gittins index exists and is unique for the ambiguity averse agent. Section 6 proves several properties of the Gittins index, and derives the predictions that ambiguity averse agents will have lower Gittins indexes than ambiguity neutral agents, but will also be willing to pay more to learn the true mean of an ambiguous arm than ambiguity neutral agents. The experimental design used to test this prediction is presented in Section 7, and the results of the experiment are in Section 8. Section 9 discusses the results and considers the implications ambiguity aversion for policymaking.

2 Applications of the Experimentation Environment

Conceptually, experimentation problems focus on the value the agent assigns to the information obtained from experimentation. This information value arises from the expected increase in future payoffs based on the information. A surprising array of practical and economically significant decisions can be explained in terms of experimentation and information value:

Brand Choice: As mentioned above, a consumer shopping for a product he frequently buys, like orange juice or window cleaner, faces an experimentation problem: he must decide whether to purchase the best brand he's tried so far, or to experiment with new brands. He knows how good his favorite brand is on average, and how much it varies in quality, but he can learn about the new brand only by trying it. Therefore, he must consider whether the value of the information obtained about the quality of the new brand is worth foregoing his favorite brand. If he learns the new brand is better, he can use this information to improve his future utility by buying the new brand again. On the other hand, if it is worse, he has missed out on his favorite brand once, but he can return to it on the next purchase. If he underestimates the impact better orange juice will have on his future utility, he may never try the new brand and deprive himself of a possible gain.

Exploration: The oil company also faces an experimentation problem. Any agent exploring for natural resources tests land parcels to decide whether to mine or drill them. In this case, both additional testing and moving to a new tract are experimentation. The company can improve its estimate of how much oil is in the current tract with additional testing, or it can conclude that additional testing is so unlikely to influence its recovery decision that its equipment would be better used exploring another tract. If the company undervalues the information it would gain from additional testing on the current tract, it might decide to drill based on too few good test results, embarking on an expensive recovery operation in an area with few resources, or it might decide to abandon the parcel based on too few bad test results, leaving valuable resources in the ground.

Research and Development: Researchers want to allocate their time among a number of projects in a way that will maximize their chance of making an important discovery. For instance, a pharmaceutical company might experiment with several different approaches to treating a disease. The information acquired from experimentation can be used to focus subsequent research on the most promising alternatives, reducing the costs that they would incur by pursuing unpromising ones. However, if the company undervalues the information additional research on a specific treatment would provide, they may abandon an effective and profitable treatment whose promise was not immediately apparent.

Job and Price Search: Search problems are a special case of experimentation problems. Searcher's choices are somewhat different than those just described. Rather than repeatedly choosing from among multiple alternatives, at least one of which gives an uncertain payoff, searchers must decide whether to exit the problem with a known payoff stream (i.e., accept an offer) or to experiment by waiting for another offer. The information value here represents not the value of information *per se*, but rather the expected increase in future payoffs arising from the chance that future offers will be better.

For example, a worker looking for a job must decide to accept a wage offer, and receive that wage forever, or to experiment by continuing to look for a better offer. For low offers, she can expect to receive a better offer in the future, and this possibility constitutes the information value. If she does not experiment with enough different prospective employers, she could end up underemployed.

Similarly, a consumer looking for the best price on a product must decide whether to buy from the closest store at that store's price, or to experiment by searching other stores for a better price. The information value in this problem arises from the possibility that other stores have lower prices, and so the consumer may gain from searching. If the consumers do not experiment with different stores, stores can charge high prices, knowing consumers will not seek lower prices elsewhere.

Given the range and economic significance of these naturally occurring bandits, it is important that economists understand why people do not make optimal decisions. Public policies and corporate strategies can be developed based on this understanding which will help agents facing bandit problems to improve their welfare.

3 Formalizing the Experimentation Environment

To conduct a careful study of behavior in experimentation problems, the experimentation environment must be formalized. This section builds the theoretical foundations necessary to understand the extensions of experimentation problems discussed here. First, it introduces the multi-armed bandit, a formal framework for studying experimentation. It then proceeds to explain how uncertain alternatives can be valued using a certain alternative: the expected payoff from a certain alternative which makes an agent indifferent between the certain and uncertain alternatives captures the discounted present value of present experimentation.

3.1 Multi-armed Bandits

The experimentation problems described earlier can all be formally modeled as multi-armed bandits. The term bandit is used because each alternative can be thought of as a different slot machine. Each alternative, or arm, has two levels of randomness. First, an arm's payoffs are are randomly distributed. Second, one or more of the parameters of the arm's payoff distribution are unknown, but are drawn from known distributions themselves. In the case of the shopper looking for orange juice, his favorite brand which he has tried many times is a "known" average payoff arm, because he knows how much quality varies, and has a very clear idea of how good it is on average. The new brand, on the other hand, has unknown average payoff. The shopper has beliefs about how good it is on average, and about how much it varies, but he does not know for sure; he can update his beliefs by experimenting with the new brand.

In addition to a collection of arms, a multi-armed bandit must also have a discount sequence which indicates the present value of payoffs received in each future period. This is usually idiosyncratic to the agent. The agent combines her beliefs about the likelihood of different average payoffs with her beliefs about the variance of payoffs around the average to formulate a strategy which maximizes the present discounted value of payoffs received.

3.1.1 Information Value

The key concept in bandit problems, and the one which will eventually be used to identify the causes of undervaluation, is information value. The information value is the present discounted value of the expected increase in future payoffs arising from information gained by present experimentation. The consumer seeking orange juice can select the new brand, assuming its uncertainty, but also expect to gain from it. If the new juice is bad, he can switch back to his favorite brand next time. But if the new juice is good, he will have found a better juice, which he will buy and enjoy every period in the future and which he would not have found if he had not experimented. The information value captures the expected contribution to future payoffs arising from the possibility the new juice is better; it reflects the possibility the new juice is bad only in the present period because the shopper can switch back to his favorite brand.

If agents underestimate the information value, they will not experiment enough and may lock onto an alternative which gave good payoffs early, but which is not necessarily the one with the best average payoff. On the other hand, if agents overestimate the information value, they will experiment too much and waste choices on alternatives with low average payoffs. This intuition provides the basis for the experiment described in Section 7. It asks subjects for the information value they perceive from a single unknown arm. Their reported information values can be used to test for optimalityg by comparing it to the optimal information value for an exponential discounter.

3.2 Bandit Notation

For simplicity, attention is restricted to two-armed bandits. Otherwise, the notation I use largely follows that of Berry and Fristedt (1985).

3.2.1 Arms

An arm consists of a distribution from which payoffs are drawn, a set of distributions from which the distribution of payoffs is selected and a prior over the set of distributions. Let $Q \in \mathcal{D}$ denote the distribution from which a payoff is drawn when the arm is chosen, where $\mathcal{D}$ is the set of possible payoff distributions. The agent's prior over the elements in $\mathcal{D}$ is denoted G. Although, except where specified, the theory given here works for general

Q, $\mathcal{D}$ and G, those who prefer concreteness may consider Q to be a normal distribution with known variance σ^2 and unknown mean μ, $\mathcal{D}$ the set of normal distributions with variance σ^2 and $\mu \in \Re$, and G a normal distribution from which μ is drawn with known mean ν and known variance τ^2. I will also consider the case where $\mathcal{Q}$ is binomial payoff distribution with an unknown mean θ, $\mathcal{D}$ is the set of possible binomial distributions, and $\mathcal{G}$ is a beta distribution with unknown parameters α and β.

When an arm is selected, a payoff X is drawn from Q. The agent uses Bayes' rule to update her beliefs that Q is a particular element in $\mathcal{D}$. Let F on $\mathcal{D}$ denote the updated set of beliefs. Further, let $(X)F$ on $\mathcal{D}$ denote that the beliefs F have been updated to reflect the payoff X.

The two-armed bandits I consider will have one arm F, and a second arm with a known Q. Since Q will have only one parameter, the mean of the normal distribution, this known arm will be denoted λ, where λ is the value of the mean of the known Q.

3.2.2 Discount Sequences

A bandit consists of two elements: a collection of arms following the description above, and a discount sequence giving the discounted present value of payoffs in future periods. A general discount sequence will be denoted $A = (\alpha_1, \alpha_2, \alpha_3, \ldots)$, where α_t denotes the relative value of payoffs received in period t. In this notation, an exponential discount sequence is $A = (1, \delta, \delta^2, \ldots)$. When it is convenient, $A^{(1)}$ will be used to denote the one-period-ahead continuation of A, $(\alpha_2, \alpha_3, \ldots)$.

Given these elements, the two-armed bandits on which this paper focuses can be written $(F, \lambda; A)$, where F is the unknown Q bandit, λ is the known Q bandit, and A is the discount sequence. Of particular interest will be the case where A is exponential, which will be denoted $(F, \lambda; \delta)$.

Berry and Fristedt (1985) characterize the set of discount sequences for which a bandit reduces to an optimal stopping problem.

Definition 1 *For any discount sequence $A = (\alpha_1, \alpha_2, \alpha_3, \ldots)$, let $\gamma_t = \sum_{\tau=t}^{\infty} \alpha_\tau$. Then A is* regular *if, for $t = 1, 2, \ldots$*

$$\frac{\gamma_{t+2}}{\gamma_{t+1}} \leq \frac{\gamma_{t+1}}{\gamma_t} \tag{1}$$

provided that $\gamma_{t+1} > 0$.

Knowing this is important because optimal stopping problems are much better understood, and much easier to compute solutions for, than the general bandit problem.

3.2.3 Strategies, Worths and Values

A strategy in a bandit is a series of history-dependent arm selections σ, designating an arm choice in each period for each possible F in that period. The worth of a strategy (what it is expected to pay) is given by

$$W(F, \lambda; A; \sigma) = E_\sigma[\sum_{\tau=1}^{\infty} \alpha_\tau X_\tau] \tag{2}$$

where X_τ is the payoff received at time τ from whichever arm is prescribed by σ given the F at time τ.

The value of the bandit is the expected payoff given that the agent plays the optimal strategy (assuming it exists),

$$V(F, \lambda; A) = \sup_\sigma W(F, \lambda; A; \sigma). \tag{3}$$

Two other expressions of value are of interest. Let $V^F(F, \lambda; A)$ be the value of selecting F in the current period and then continuing optimally and $V^\lambda(F, \lambda; A)$ be the value of selecting λ initially and then continuing optimally.

$$V^F(F, \lambda; A) = \alpha_1 E[X|F] + E[V((X)F, \lambda; A^{(1)})] \tag{4}$$

$$V^\lambda(F, \lambda; A) = \alpha_1 \lambda + V(F, \lambda; A^{(1)}) \tag{5}$$

These expressions will be useful in understanding the value function. The difference in these functions, $V^F(F, \lambda; A) - V^\lambda(F, \lambda; A)$ will be denoted $\Delta(F, \lambda; A)$.

3.2.4 Gittins indexes and Information Values

Two more quantities are useful for comparing the value of information among arms. The Gittins index, denoted $\Lambda(F, A)$, is the value of a known mean arm for which the agent is indifferent between selecting the unknown arm and the known mean arm in the current period. The information value, the present discounted expected value of additional payoffs attributable to the information gathered from experimentation, is the Gittins index minus the expected value of the arm.

4 Ambiguity Aversion

The concept of ambiguity aversion was most directly expressed by Ellsberg (1961). He posed a thought experiment where agents had to choose between an urn which contains 30 red balls and 60 balls in some unknown combination of black and yellow. Agents were asked to rank two pairs of bets: X gave a prize if a ball drawn from the urn was red, and Y if the ball drawn was black; X' gave a prize if the ball is either red or yellow, and Y' gives a prize if the ball is either black or yellow.

Most people prefer X to Y and Y' to X'. This is paradoxical because this pair of preferences is inconsistent with any fixed belief about the mixture of black and yellow balls in their urn: choosing X over Y implies a belief that there are fewer than 30 black balls, and therefore more than 30 yellow balls which implies X' should be preferred to Y'.

The Ellsberg paradox demonstrates that people prefer known probability bets to unknown probability bets, and provides an operationalization of ambiguity. However, no consensus has emerged of a precise definition of ambiguity. Partly, this stems from the variety of circumstances in which different notions of ambiguity seem suitable. For instance, the credibility of a source of information, or the degree of disagreement between multiple sources, such as expert witnesses, captures one notion of ambiguity (Einhorn and Hogarth, 1985). Ambiguity may also represent uncertainty about probability stemming from information which could be known, but is not (Frisch and Baron, 1988). True subjectivists may reject the notion of ambiguity altogether, since all subjective probability distributions are equally well known to ourselves (deFinetti, 1977).

The notion of ambiguity considered here is based on second order probabilities. A second order probability is the distribution of possible distributions. In the Ellsberg problem, the second order probability is the probability distribution over the number of black balls in the urn. Second order probability is also commonly encountered as statistical confidence. Consider, for instance, two coins, one which has been flipped twice yielding one head and one tail and another which has been flipped 1000 times, yielding 500 heads and 500 tails. Both coins have $P(\text{heads})=0.5$, but the coin with more flips has a lower-variance second order probability.

This approach is not without its drawbacks. There is some evidence that agents prefer known second order probabilities to unknown second order probabilities (Yates and Zukowski, 1976), suggesting third order probabilities may also affect ambiguity. Also, if people have difficulty understanding second order probabilities, then higher order probabilities are probably more difficult to consider.

In bandits, however, the second order probability is a natural interpretation of ambiguity because there is a unique, known second order probability, $\mathcal{G}$. In some applications, it may be difficult to argue that this probability is in fact known, but for the abstract version of the problem that can be tested in the laboratory, a sensible definition of ambiguity can be based on the variance of $\mathcal{G}$.

This representation of ambiguity draws a distinction between ambiguity and risk. Following Anscombe and Aumann (1963), the following formal notions of risk and ambiguity will be used.

Definition 2 *An arm F is **riskier than** an arm F' if the variance of $\mathcal{Q}$ is greater than the variance of $\mathcal{Q}'$.*

Therefore, agents who are risk aversion mind only variance in their payoff distribution. This variance is independent of the variance of the distribution of means of the payoff distributions.

Definition 3 *An arm F is **more ambiguous than** an arm F' if the variance of $\mathcal{G}$ is greater than the variance of $\mathcal{G}'$.*

Therefore, agents who are ambiguity averse do not like the variance in the distribution from which the means of the payoff distributions are drawn. Although this may not be intuitively distinct from variance in the payoff distribution (especially if, as a good economist, you reduce compound lotteries), ambiguity aversion and risk aversion are only weakly correlated within individuals (Hogarth and Einhorn, 1990).

4.1 Models of Ambiguity Aversion

Models of ambiguity aversion fall into three categories, those which leverage unique second order probabilities (and relax reduction of compound lotteries), those which allow multiple probabilities, and those which rely on nonadditive probabilities. Nonadditive probability models (e.g., Schmeidler, 1989) do not require that the subjective probabilities of an event in a set which will occur with objective probability one sum to one. Therefore, if an outcome will be either A or B, then $P(A) + P(B)$ does not have to equal one. Therefore, ambiguity aversion is represented because, when computing expected payoffs using the subjective probabilities, not all the probability weight will be represented. The remaining probability, $1 - P(A) - P(B)$, is a measure of faith in the the evidence on which agents' beliefs are based.

Another way to think about ambiguity is to consider independently the set of possible payoff distributions without reference to a second order probability which generates them. There are a variety of ways that agents may use these multiple probabilities to decide among actions. Many of these models suggest agents use a minimax decision rule (e.g., Gilboa and Schmeidler, 1989). Ambiguity aversion arises in these models as agents may have multiple probabilities and pessimistically evaluate each lottery using the probability which generates the lowest expected utility.

Given that bandits provide a known, unique second order probability, and the definition of ambiguity is based on a second order probability, the most natural set of models to apply are those using second order probabilities. These models relax the assumption of reduction of compound lotteries implied by subjective expected utility theories by applying a nonlinear weighting function to transform the second order probability distribution before computing an expectation. When the nonlinear weighting rule moves decision weight from high values to low values, agents will be ambiguity averse.

To formalize the notion of second order probabilities, let F be a distribution with a parameter θ, where $F(\theta)$ is the probability the agent will receive some prize X. Let θ have a second order distribution G, and let $\bar{\theta}$ be the expected value of θ.

Ambiguity attitudes are represented by a decision weighting function $\omega(X)$, which gives the expected utility from winning the prize X, adjusting for the ambiguity associated with selecting an ambiguous lottery. Using this, ambiguity aversion can be formally defined.

Definition 4 *An agent's decision weighting function $\omega(X)$ is ambiguity averse if $E[\omega(X)|F] \leq E[X|F]$.*

Formally, ambiguity aversion occurs when the value of a choice given its ambiguity is less than its expected value. Typically, $\omega(X)$ is decreasing in the variance of G.

Several forms for the function $\omega(X)$ have been proposed. The next two sections discuss two of them. The first discusses a model by Segal (1987), who proposes a broad class of functions which could represent ambiguity aversion. The second presents a specific function used by Kahn and Sarin (1988).

4.1.1 Segal

One model of ambiguity aversion was developed by Segal (1987), based on an anticipated utility model developed by Quiggan (1982) and others. If the state probabilities are ordered such that $\rho_1 < \rho_2 < \ldots$, then

$$W = u(x)(\theta_1) + u(x)\sum_{i=2}^{m}[f(\theta_i) - f(\theta_{i-1})]f(\sum_{j=i}^{m} p(\theta_j)). \tag{6}$$

Segal shows that if $f(0) = 0$, $f(1) = 1$ and $f(\cdot)$ satisfies a condition called nondecreasing elasticity (a slightly stronger form of convexity), then $E[\omega(X)|F] \leq E[X|F]$, or agents are ambiguity averse.

4.1.2 Kahn and Sarin

Kahn and Sarin (1988) propose a specific form for $\omega(X)$:

$$\omega(X) = u(x)\bar{\theta} + u(x)\int_{\theta=0}^{1}(\theta - \bar{\theta})e^{[-\xi(\theta-\bar{\theta})]/\sigma}dG(\theta)d\theta \tag{7}$$

where $\sigma = \sqrt{\int_{\theta=0}^{1}(\theta - \bar{\theta})^2 dG(\theta)d\theta}$ is that standard deviation of the second order probability distribution, and $\bar{\theta}$ is the expected value of θ.

If $\xi \neq 0$, then $(\theta - \bar{\theta})$ is the scale of the impact of the ambiguity. If $\xi > 0$, the second order probabilities are adjusted by underweighting the chance of higher than average θs and overweighting the probabilities of lower than average θs, leading to ambiguity aversion. A negative ξ does the opposite, representing ambiguity preference. Since it captures attitudes toward ambiguity, ξ is a characteristic of the individual, and therefore a primitive of the model.

The decision weighting function can be interpreted as an expectation, where $\bar{\theta}u(x)$ is the expected utility of the choice and the $u(x)$ times the integral is the psychological cost (or reward) associated with making an ambiguous choice. This interpretation requires that the cost be incurred when the choice is made, so it is subtracted from any outcome realized.

5 The Gittins Index of Ambiguity Averse Agents

Ambiguity aversion may play an intuitive role in bandit problems. Agents may dislike trying alternatives about which they have less information because they do not know the mean of the payoff distribution; the larger the variance of possible means, the more they dislike unknown alternatives. This attitude could lead to underexperimentation, as agents avoid ambiguous alternatives, and lower than optimal Gittins indexes.

This section formalizes this intuition and develops some theory on how ambiguity aversion affects behavior in bandits. It proves that a Gittins index exists, and that a two-armed bandit with one arm known is a stopping problem for the ambiguity averse agent. These results form the foundation for extending Kahn and Sarin's model of ambiguity aversion to Bernoulli arms with beta priors.

5.1 Bayes Rule and Reduction of Compound Lotteries

Extending ambiguity aversion to bandit problems requires some interpretation because these models have, in the past, applied only to choices among compound lotteries. Models which leverage second order probabilities assume that, in computing their expected payoffs, agents do not use Bayes rule to reduce compound lotteries; they use some other function which expresses their ambiguity aversion.

In bandits, however, ambiguity neutral agents apply Bayes rule not only to reduce compound lotteries to compute their expected payoff, but also to update their priors over the parameters of the payoff distribution and to compute the probabilities of continuations. How to best extend models of choice among compound lotteries to bandits depends on whether ambiguity aversion is the product of a fundamental problem in applying Bayes rule, or whether it is some other phenomenon which is well modeled by using some substitute for Bayes rule.

I am not aware of any evidence on either side of this question. However, for purposes of this paper, I will assume that agents understand and use Bayes rule to update their prior beliefs and to compute continuation probabilities. Therefore, it is only the act of computing an expected value, considering payoffs themselves, in the context of ambiguity which leads to non-Bayesian behavior.

To emphasize this distinction, I will use $E[\omega(X)|F]$ to indicate an expected payoff adjusted for ambiguity. Probabilities will be denoted $P(X = 1|F)$, and are not affected by the agents' ambiguity attitudes. To indicate that a given arm is being evaluated by an ambiguity averse agent, its payoff distribution will be written F_ω. However, because probabilities are unaffected by ambiguity attitudes this does not correspond to a different distribution. Further, because Bayes rule is applied properly in updating prior beliefs, $(X)F_\omega$ reflects that F has been updated to reflect X prior to being evaluated under the ambiguity attitude represented in $\omega(\cdot)$.

5.2 Existence of a Dynamic Allocation Index

Before developing a specific model of ambiguity aversion in bandits, I prove that a Gittins index exists for ambiguity averse agents.

Theorem 1 *For each nonincreasing discount sequence A with $A \neq 0$ and each distribution F on $\mathcal{D}$, there exists a unique function $\Lambda(F_\omega, A)$ such that the F arm is optimal initially in the $(F_\omega, \lambda; A)$ bandit if and only if $\lambda \leq \Lambda(F_\omega, A)$ and the λ arm is optimal initially if and only if $\lambda \geq \Lambda(F_\omega, A)$.*

This proof follows the proof of existence of a Gittins index for hyperbolic discounters. First, it is necessary to prove some preliminary results about $V(F_\omega, \lambda; A)$ and $\Delta(F_\omega, \lambda; A)$.

Lemma 1 *For all F and for all A, $V(F_\omega, \lambda; A)$ is continuous and nondecreasing in λ.*

An increase in λ can affect the value function in two ways: it increases the value of arm λ whenever it is chosen, and it expands the set of F over which the optimal strategy prescribes the λ arm to include those of higher expected value. Given this, an increase in λ could not result in a reduction of the value function because an increase in the value function never makes it more likely F will be chosen, and it strictly increases the value of any choice of the λ arm.
Proof: Suppose $\lambda^* > \lambda$ and σ is optimal in the $(F_\omega, \lambda; A)$ bandit. Suppose σ is followed in the $(F_\omega, \lambda^*; A)$ bandit. The only change compared with $(F_\omega, \lambda; A)$ is when arm 2 is selected and λ^* is received instead of λ.

Therefore,

$$\begin{aligned} V(F_\omega, \lambda; A) &= W(F_\omega, \lambda; A; \sigma) && (8) \\ &\leq W(F_\omega, \lambda^*; A; \sigma) = V(F_\omega, \lambda^*; A) && (9) \end{aligned}$$

Therefore, $V(F_\omega, \lambda; A)$ is nondecreasing in λ.

For continuity, let σ^* be optimal for the $(F_\omega, \lambda^*; A)$. Then the only difference between $W(F_\omega, \lambda; A; \sigma^*)$ and $W(F_\omega, \lambda^*; A; \sigma^*)$ is the result of arm 2 when it is chosen.

$$\begin{aligned} V(F_\omega, \lambda^*; A) &= W(F_\omega, \lambda^*; A) && (10) \\ &\leq W(F_\omega, \lambda; A; \sigma^*) + (\lambda^* - \lambda)\gamma_1 && (11) \\ &\leq V(F_\omega, \lambda; A) + (\lambda^* - \lambda)\gamma_1 && (12) \end{aligned}$$

Since Equation 9 proved $V(F_\omega, \lambda; A) \leq V(F_\omega, \lambda^*; A)$, this implies that this relationship must approach equality as $\lambda^* \to \lambda$. Therefore, $V(\cdot, \lambda; A)$ is continuous in λ. ♡

In order to show a dynamic allocation index exists, I also need a result showing how the size of the error made by choosing the the wrong arm varies with λ. Define the function $\Delta(F_\omega, \lambda; A)$ as the difference in the value functions from choosing the F arm first and then continuing optimally and choosing the λ arm first and then continuing optimally;

$$\Delta(F_\omega, \lambda; A) = V^F(F_\omega, \lambda; A) - V^\lambda(F_\omega, \lambda; A). \tag{13}$$

The absolute value of this quantity can be thought of as the cost of making an error by selecting the wrong arm initially. This quantity turns out to be very important, as the following lemma does most of the work in proving Theorem 1.

Lemma 2 *For any F on $\mathcal{D}$, $\Delta(F_\omega, \lambda; A)$ is decreasing in λ when A is non-increasing with $A \neq 0$.*

Proof: Fix $\lambda^* > \lambda$.

This proof proceeds in three parts. Part (i) derives an expression for $\Delta(F_\omega, \lambda^*; A) - \Delta(F_\omega, \lambda; A)$. Part (ii) performs a finite induction on the horizon to establish that $\Delta(F_\omega, \lambda^*; A) - \Delta(F_\omega, \lambda; A)$ is negative. Part (iii) extends the result of Part (ii) to infinite horizons.

(i) The value of choosing the F arm first and then proceeding optimally is given by

$$V^F(F_\omega, \lambda; A) = \alpha_1 E[\omega(X)|F] + E[V((X)F_\omega, \lambda; A^{(1)})]. \tag{14}$$

Similarly, the value of selecting the λ and then continuing optimally is given by

$$V^\lambda(F_\omega, \lambda; A) = \alpha_1 \lambda + V(F_\omega, \lambda; A^{(1)}). \tag{15}$$

Now define two more functions, which will prove to be of considerable algebraic convenience. $\Delta^+(F_\omega, \lambda; A) = \max[0, \Delta(F_\omega, \lambda; A)]$ and $\Delta^-(F_\omega, \lambda; A) = \max[0, -\Delta(F_\omega, \lambda; A)]$ so that

$$V^F(F_\omega, \lambda; A) = V(F_\omega, \lambda; A) - \Delta^-(F_\omega, \lambda; A) \tag{16}$$
$$V^\lambda(F_\omega, \lambda; A) = V(F_\omega, \lambda; A) - \Delta^+(F_\omega, \lambda; A). \tag{17}$$

Mnemonically, Δ^- is nonzero when $\Delta(F_\omega, \lambda; A)$ is negative, or when λ is the optimal arm.

Using these definitions, substitute for $V((X)F_\omega, \lambda; A)$ and $V(F_\omega, \lambda; A)$ in Equations 14 and 15 above. This gives

$$V^F(F_\omega, \lambda; A) = \alpha_1 E[\omega(X)|F] + E[V^\lambda((X)F_\omega, \lambda; A^{(1)}) + \Delta^+((X)F_\omega, \lambda; A^{(1)})] \quad (18)$$
$$V^\lambda(F_\omega, \lambda; A) = \alpha_1 \lambda + V^F(F_\omega, \lambda; A^{(1)}) + \Delta^-(F_\omega, \lambda; A^{(1)}). \quad (19)$$

These expressions can then be used to compute $\Delta(F_\omega, \lambda; A)$. The first two terms in Equation 18 represent the value of selecting arm F in the first period, arm λ in the second and then continuing optimally. Similarly, the first two terms in Equation 19 represent the value of selecting arm λ in the first period, arm F in the second and then continuing optimally. Given this interpretation, subtracting the first two terms in Equation 19 from those in Equation 18 gives $(\alpha_1 - \alpha_2)[E[\omega(X)|F] - \lambda]$. This gives

$$\Delta(F_\omega, \lambda; A) = (\alpha_1 - \alpha_2)[E[\omega(X)|F] - \lambda] + E[\Delta^+((X)F_\omega, \lambda; A^{(1)})] - \Delta^-(F_\omega, \lambda; A^{(1)}). \quad (20)$$

Using this expression to compute $\Delta(F_\omega, \lambda^*; A) - \Delta(F_\omega, \lambda; A)$ gives

$$\begin{aligned} \Delta(F_\omega, \lambda^*; A) - \Delta(F_\omega, \lambda; A) &= (\alpha_1 - \alpha_2)[\lambda - \lambda^*] + \\ E[\Delta^+((X)F_\omega, \lambda^*; A^{(1)}) &- \Delta^+((X)F_\omega, \lambda; A^{(1)})] + \\ \Delta^-(F_\omega, \lambda; A^{(1)}) &- \Delta^-(F_\omega, \lambda^*; A^{(1)}). \end{aligned} \quad (21)$$

(ii) Proving the lemma requires that the difference in Equation 21 be negative. This section performs induction on a finite horizon to demonstrate that this is true.

Let A_n be a nonincreasing discount sequence with finite horizon n, so elements after the n^{th} are zero.

First, suppose $n = 1$. Then, for all A_1, $\Delta(F_\omega, \lambda^*; A_1) - \Delta(F_\omega, \lambda; A_1)$ is negative implies

$$\begin{aligned} E[\omega(X)|F] - \lambda^* &< E[\omega(X)|F] - \lambda \\ \lambda^* &> \lambda \end{aligned} \quad (22)$$

which is true by assumption.

Now suppose that the horizon $n > 1$ and $\Delta(F_\omega, \lambda^*; A_n) < \Delta(F_\omega, \lambda; A_n)$ for any nonincreasing A_n. Now I will use this induction hypothesis to show that $\Delta(F_\omega, \lambda^*; A_{n+1}) < \Delta(F_\omega, \lambda; A_{n+1})$.

Equation 21 can be rewritten with the truncated discount sequence

$$\begin{aligned} \Delta(F_\omega, \lambda^*; A_{n+1}) - \Delta(F_\omega, \lambda; A_{n+1}) &= (\alpha_1 - \alpha_2)[\lambda - \lambda^*] + \\ E[\Delta^+((X)F_\omega, \lambda^*; A_{n+1}^{(1)}) &- \Delta^+((X)F_\omega, \lambda; A_{n+1}^{(1)})] + \\ \Delta^-(F_\omega, \lambda; A_{n+1}^{(1)}) &- \Delta^-(F_\omega, \lambda^*; A_{n+1}^{(1)}). \end{aligned} \quad (23)$$

The first term on the righthand side of Equation 23 is nonpositive because $\lambda^* > \lambda$ by assumption and $\alpha_1 \geq \alpha_2$ by hypothesis.

The remaining terms are negative for similar reasons. Consider the second term. Since $A^{(1)}_{n+1}$ is nonincreasing and has horizon n, we have

$$\begin{aligned} &E[\Delta^+((X)F_\omega, \lambda^*; A^{(1)}_{n+1}) - \Delta^+((X)F_\omega, \lambda; A^{(1)}_{n+1})] \\ &= E[\max[0, \Delta((X)F_\omega, \lambda^*; A_n)] - \max[0, \Delta((X)F_\omega, \lambda; A_n)]]. \end{aligned} \quad (24)$$

The induction hypothesis gives that $\Delta(F_\omega, \lambda^*; A_n) < \Delta(F_\omega, \lambda; A_n)$ for all F, in particular σF. Therefore, the second term in Equation 24 is always larger than the first, implying that the second term in Equation 23 is negative.

Consider the third term. Since $A^{(1)}_{n+1}$ is nonincreasing and has horizon n, we have

$$\begin{aligned} &\Delta^-(F_\omega, \lambda; A^{(1)}_{n+1}) - \Delta^-(F_\omega, \lambda^*; A^{(1)}_{n+1}) \\ &= \max[0, -\Delta(F_\omega, \lambda; A_n)] - \max[0, -\Delta(F_\omega, \lambda^*; A_n)]. \end{aligned} \quad (25)$$

The induction hypothesis gives that $\Delta(F_\omega, \lambda^*; A_n) \leq \Delta(F_\omega, \lambda; A_n)$. Therefore, the second term in Equation 25 is always larger than the first, implying that the third term in Equation 23 is negative.

Since the first term in Equation 23 is nonpositive and the other two are strictly negative, the difference Equation 21 is negative for every finite horizon. Now we let the horizon go to infinity to show it is negative for infinite horizons.

(iii) Suppose $n = \infty$. Let A_T denote the truncation of A_∞ at finite T, so A_T coincides with A_∞ up to time T and has zeros afterwards. Letting $T \to \infty$ in the result from Part (ii) gives

$$\Delta(F_\omega, \lambda^*; A_\infty) < \Delta(F_\omega, \lambda; A_\infty). \quad (26)$$

Since $A = A_\infty$, we have $\Delta(F_\omega, \lambda^*; A) \leq \Delta(F_\omega, \lambda; A)$ for all horizons. This is sufficient to prove the lemma. ♡

Given this result, the existence of a dynamic allocation index is easy to prove.

Proof of Theorem 1: This proof begins by defining $\Lambda(F_\omega, A) = \inf\{\lambda \in \mathcal{D} :$ the λ arm is optimal for the $(F_\omega, \lambda; A)$ bandit $\}$. Then I show that this definition implies that F is uniquely optimal if $\lambda < \Lambda(F_\omega, A)$. Then I use Lemma 2 to show that λ is uniquely optimal if $\lambda > \Lambda(F_\omega, A)$. Indifference at $\lambda = \Lambda(F_\omega, A)$ then follows from the continuity of V.

For $\lambda < \Lambda(F_\omega, A)$, we have

$$V^F(F_\omega, \lambda; A) > V^\lambda(F_\omega, \lambda; A) \quad (27)$$

from the definition of $\Lambda(F_\omega, A)$. Because $\Lambda(F_\omega, A)$ is the infimum value of λ for which λ is optimal, it must be that F is uniquely optimal.

The case where $\lambda > \Lambda(F_\omega, A)$ is a little harder because there may be values of λ above $\Lambda(F_\omega, A)$ where F is optimal. However, the fact that $\Delta(F_\omega, \lambda; A)$ is strictly decreasing in λ, as shown in Lemma 2, proves that this cannot be. Therefore

$$V^F(F_\omega, \lambda; A) < V^\lambda(F_\omega, \lambda; A) \tag{28}$$

for all $\lambda > \Lambda(F_\omega, A)$and λ is uniquely optimal.

Finally, if $\lambda = \Lambda(F_\omega, A)$, we have that neither F nor λ is uniquely optimal. Extending the continuity of $V(F_\omega, \lambda; A)$ to $V^\lambda(F_\omega, \lambda; A)$ and $V^F(F_\omega, \lambda; A)$, the previous cases sandwich possible values of $V^\lambda(F_\omega, \lambda; A)$ and $V^F(F_\omega, \lambda; A)$ to give

$$V^F(F_\omega, \Lambda(F_\omega, A); A) = V^\lambda(F_\omega, \Lambda(F_\omega, A); A), \tag{29}$$

which is equivalent to both arms being optimal initially for the $(F_\omega, \Lambda(F_\omega, A); A)$ bandit.♡

5.3 The Stopping Property

The index will be much easier to compute and discuss theoretically if ambiguity averse agents' strategies satisfy the stopping property.

Theorem 2 *If A is nonincreasing, then for every $(F_\omega, \lambda; A)$ bandit, there is an optimal strategy for which every selection of λ is followed by another selection of λ.*

First, I need a result which indicates when λ will never be optimal.

Lemma 3 *If $E[\omega(X)|F] \geq \lambda$, then the F arm is optimal for any A.*

Proof: Suppose that σ is an optimal strategy for the $(F_\omega, \lambda; A^{(1)})$ bandit. Let σ^* be a strategy which indicates the F arm initially and then follows σ, ignoring the initial realization from F. Then

$$\begin{aligned} W(F_\omega, \lambda; A; \sigma^*) &= \alpha_1 E[\omega(X)|F] + V(F_\omega, \lambda; A^{(1)}) && (30) \\ &\geq \alpha_1 \lambda + V(F_\omega, \lambda; A^{(1)}) = W(F_\omega, \lambda; A; \sigma_2) && (31) \end{aligned}$$

where σ_2 is a strategy which chooses λ initially and then proceeds optimally. Since there is a strategy which starts with the F arm and is least as good as the optimal strategy which starts with the λ arm, the F arm is optimal. $\heartsuit$

Proof of Theorem 2: Let A_n denote a nonincreasing discount sequence with horizon n. The proof is by induction on the horizon.

If $n = 1$, then the proposition is trivially true since there is no further selection of either arm.

Suppose $n \geq 2$ and for every $(F_\omega, \lambda; A_{n-1})$ bandit, there is an optimal strategy for which every selection of λ is followed by another selection of λ. Then $A_n^{(1)} \in A_{n-1}$.

Assume it is optimal to select F initially. Then the inductive hypothesis shows that there exists a continuation which never switches back to F after its first selection of λ.

On the other hand, if it is optimal to select λ initially, the inductive hypothesis applies trivially unless there is an optimal strategy σ^* which indicates a selection of λ, then selects F at stages $2 \ldots N$, and then λ thereafter. This might be the case if I were indifferent at time 1 and returned to indifference after several selection of F. I will now show such a strategy cannot be better than a strategy which never switches back to F.

Since the value of λ is known, we can assume that σ^* does not depend on the initial observation from λ. So for each m, $\{N > m\}$ is measurable with respect to the σ-field generated on the outcomes $(X_2, \ldots, X_m)$.

Lemma 3 implies that if the sequence of outcomes through time m contains s successes and $f = m - s - 1$ failures while following σ^*, then

$$N = m \Rightarrow E[X|\sigma^s \phi^f F_\omega] < \lambda \tag{32}$$

This condition says that if λ is going to be selected in the current period, then the sequence of successes and failures must be such that it is no longer optimal to select F. This bound comes from Lemma 3.

Now I will show that there is a strategy σ which starts with F and is at least as good as σ^*. Let σ select F initially and then imitate σ^* by selecting the arm prescribed by σ^* one period earlier.

$$W(F_\omega, \lambda; A; \sigma^*) = E_{\sigma^*}\left[\alpha_1 \lambda + \sum_{m=2}^{N} \omega(X_m)\alpha_m + \lambda \sum_{m=N+1}^{\infty} \alpha_m | F\right] \tag{33}$$

which must best at least $\gamma_1 \lambda$ since σ^* is optimal. Therefore,

$$\sum_{m=2}^{\infty} E_{\sigma^*}\left[(\omega(X_m)-\lambda)I\{N \geq m\}|F\right]\alpha_m = E_{\sigma^*}\left[\sum_{m=2}^{N}(\omega(X_m)-\lambda)\alpha_m|F\right] \geq 0 \tag{34}$$

The value under σ is given by

$$W(F_\omega, \lambda; A; \sigma) = E_\sigma\left[\sum_{m=2}^{N}\omega(X_m)\alpha_{m-1} + \lambda\sum_{m=N+1}^{\infty}\alpha_{m-1}|F\right] \tag{35}$$

The stopping strategy σ is better if the difference in these worths is positive.

$$W(F_\omega, \lambda; A; \sigma) - W(F_\omega, \lambda; A; \sigma^*) = \sum_{m=2}^{\infty} E_{\sigma^*}\left[(\omega(X_m)-\lambda)I\{N \geq m\}|F\right](\alpha_{m-1}-\alpha_m) \tag{36}$$

The term $(\alpha_{m-1} - \alpha_m)$ is weakly positive because A_n is nonincreasing. The expectation weakly is positive because $E[\omega(X_m)|F] \leq \lambda$ by Lemma 3.

Therefore, there is a strategy which satisfies the stopping property and which is at least as good as the optimal strategy.♡

6 Bandits with Kahn-Sarin Ambiguity Averse Agents

This section establishes properties of the behavior of ambiguity averse agents in bandit problems. The analysis here is restricted to the case of a Bernoulli arm whose parameter has a known beta(α, β) distribution.

The primary concern of this section is that adding the dynamic element of probability updating to the ambiguous choice problem does not affect the way the model represents ambiguity aversion. It is not obvious from Kahn and Sarin's paper that, when an agent acquires more information about an ambiguous alternative, $\xi > 0$ will still lead to an $\omega(\cdot)$ which is ambiguity averse. As it turns out, we do not need to be concerned that we will observe some sequence of successes and failures for which the Kahn-Sarin transformation will result in behavior other than ambiguity aversion.

Theorem 3 *Let F be the distribution of the parameter of a Bernoulli arm with a beta(α, β) prior. If ξ is positive, then the Kahn-Sarin agent will be ambiguity averse for all informative priors ($\alpha > 0$ and $\beta > 0$).*

The proof of this theorem reduces the problem to a difference between two Kummer's functions, or confluent hypergeometric functions of the first kind. (Spanier and Oldham, 1987; Abramowitz and Stegun, 1972) This function is

$$ {}_1F_1(a,b,z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{zt} t^{a-1} (1-t)^{b-a-1} dt. \tag{37}$$

Kummer's function has many uses in theoretical physics, and it provides one class of solutions to Kummer's confluent hypergeometric differential equation, $zy'' + (b-z)y' - ay = 0$ with initial conditions ${}_1F_1(a,b,0) = 1$ and $\frac{\partial}{\partial z}\, {}_1F_1(a,b,z)|_{z=0} = a/b$. There is no obvious relationship between its physical interpretation and its appearance in a model of ambiguity aversion; it is used here because existing results about Kummer's function simplify proof of Theorem 3.

Before proving Theorem 3, it is necessary to prove a property of Kummer's function.

Lemma 4 *Kummer's confluent hypergeometric function has the property that* ${}_1F_1(a+1,b+1,z) -_1 F_1(a,b,z)$ *is strictly negative for all* $b > a > 0$ *and* $z < 0$.

Proof: Kummer's first theorem allows $z < 0$ to be transformed into $z > 0$. Then Kummer's function can be related to an Euler beta function, which provides a closed-form solution to the definite integral in ${}_1F_1(\cdot,\cdot,z)$. The closed-form solution can be manipulated to demonstrate that each term in the summation in the Euler beta function is strictly negative.

To keep track of the sign of z, let z_- denote $z < 0$, and $z_+ = -z_-$ denote $z > 0$. Kummer's first theorem (Slater, 1960, p. 6) holds that ${}_1F_1(a,b,z) = e^z\, {}_1F_1(b-a,b,-z)$ for all z. Therefore,

$$ {}_1F_1(a,b,z_-) = e^{z_-}\, {}_1F_1(b-a,b,z_+) \tag{38}$$

$$ {}_1F_1(a+1,b+1,z_-) = e^{z_-}\, {}_1F_1(b-a,b+1,z_+). \tag{39}$$

The difference then reduces to

$$ \frac{\Gamma(b)e^{z_-}}{\Gamma(b-a)\Gamma(a)} \left[\frac{b}{a} \int_0^1 e^{tz_+} t^{b-a-1} (1-t)^a dt - \int_0^1 e^{tz_+} t^{b-a-1} (1-t)^{a-1} dt \right]. \tag{40}$$

The exponential term in the integral can be broken into its representation as an infinite sum (Slater, 1960, p. 34), yielding

$$ \frac{\Gamma(b)e^{z_-}}{\Gamma(b-a)\Gamma(a)} \left[\frac{b}{a} \int_0^1 \sum_{n=0}^{\infty} \frac{z_+^n}{n!} t^{b-a-1+n} (1-t)^a dt \right.$$

$$-\int_0^1 \sum_{n=0}^{\infty} \frac{z_+^n}{n!} t^{b-a-1+n}(1-t)^{a-1} dt \Bigg]. \tag{41}$$

The terms which depend on z_+ can be moved out of the integral, and the difference can be written

$$\frac{\Gamma(b)e^{z_-}}{\Gamma(b-a)\Gamma(a)} \sum_{n=0}^{\infty} \frac{z_+^n}{n!} \int_0^1 \left(\frac{b}{a}(1-t) - 1\right) t^{b-a-1+n}(1-t)^{a-1} dt. \tag{42}$$

The integral is a (difference of) Euler beta functions, and its solution can therefore be represented in terms of gamma functions. The integral equals

$$= \Gamma(a) \frac{(b-a)\Gamma(b+1+n)\Gamma(b-a+n) - b\Gamma(b+n)\Gamma(b-a+1+n)}{a\Gamma(b+n)\Gamma(b+1+n)} \tag{43}$$

$$= \frac{\Gamma(a)\Gamma(b-a+n)}{a(b+n)\Gamma(b+n)} [(b-a)(b+n) - b(b-a+n)] \tag{44}$$

$$= -an \frac{\Gamma(a)\Gamma(b-a+n)}{a(b+n)\Gamma(b+n)} \tag{45}$$

The entire difference can be represented as

$$-\frac{\Gamma(b)e^{z_-}}{\Gamma(b-a)\Gamma(a)} \sum_{n=0}^{\infty} \frac{z_+^n}{n!} \frac{n\Gamma(a)\Gamma(b-a+n)}{(b+n)\Gamma(b+n)}. \tag{46}$$

Since $b > 0$, $a > 0$, $b - a > 0$ and $n > 0$, every gamma function in this expression is strictly positive. Similarly, $z_+ > 0$, so every power of z_+ is strictly positive also. Therefore, every term in the infinite sum is strictly positive. Additionally, $e^{z_-} > 0$ is strictly positive. Therefore, the entire difference is negative for any $b > a > 0$. ♡

The proof that $z > 0$ leads to a positive difference (and ultimately ambiguity-seeking behavior) follows along similar lines. However, the application of Kummer's First Theorem is not necessary because all powers of z are positive, without needing to change its sign.

I can now prove the main result.

Proof of Theorem 3: This proof manipulates the Kahn-Sarin model of ambiguity aversion with a beta distribution (the conjugate prior for the Bernoulli distribution). The ambiguity term in the decision weight expression is shown to be proportional to the difference between two confluent hypergeometric functions of the first kind. Lemma 4 proves this difference is strictly negative, implying ambiguity aversion.

The Kahn-Sarin model holds that agents make choices based on a decision weight equal to

$$E[\omega(x)|\alpha,\beta] = \bar{\theta} + \int_0^1 (\theta - \bar{\theta}) e^{-\xi(\theta-\bar{\theta})/\sigma} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \quad (47)$$

where $\bar{\theta} = \frac{\alpha}{\alpha+\beta}$ is the expected value of x and $\sigma = \sqrt{\frac{\alpha+\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$ is the standard deviation of x.

Ambiguity aversion holds when $E[\omega(x)|\alpha,\beta] < E[x|\alpha,\beta]$. This is equivalent to showing that the integral is strictly negative. Evaluating it gives

$$\begin{aligned} &\frac{1}{\Gamma(\alpha)\Gamma(\beta)} \Big[\Gamma(\alpha+1)\Gamma(\beta) e^{\xi\bar{\theta}/\sigma} \; {}_1F_1(\alpha+1, \alpha+\beta+1, -\xi/\sigma) \\ &\qquad - \Gamma(\alpha)\Gamma(\beta)\bar{\theta} e^{\xi\bar{\theta}/\sigma} \; {}_1F_1(\alpha, \alpha+\beta, -\xi/\sigma) \Big] \\ &= \frac{e^{\xi\bar{\theta}/\sigma}}{\Gamma(\alpha)} \left[\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)} \; {}_1F_1(\alpha+1, \alpha+\beta+1, -\xi/\sigma) \right. \\ &\qquad \left. - \frac{\bar{\theta}\Gamma(\alpha)}{\Gamma(\alpha+\beta)} \; {}_1F_1(\alpha, \alpha+\beta, -\xi/\sigma) \right]. \end{aligned} \quad (48)$$

Applying the fact that $\Gamma(k+1) = k\Gamma(k)$ on the numerator and denominator of the left-hand term gives

$$\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha\Gamma(\alpha)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \bar{\theta}\frac{\Gamma(\alpha)}{\Gamma(\alpha+\beta)}. \quad (49)$$

The integral therefore evaluates to

$$\frac{\bar{\theta} e^{\xi\bar{\theta}/\sigma}}{\Gamma(\alpha+\beta)} \left[{}_1F_1(\alpha+1, \alpha+\beta+1, -\xi/\sigma) - {}_1F_1(\alpha, \alpha+\beta, -\xi/\sigma) \right]. \quad (50)$$

Since the coefficient is positive under the assumptions of the theorem, the sign of the integral is determined by the sign of the difference between the Kummer's functions. Lemma 4 proves that this difference is strictly negative. Therefore, the value of the integral is strictly negative, and $E[\omega(x)|\alpha,\beta] < E[x|\alpha,\beta]$, which proves the claim. ♡

Corollary 1 *Kahn-Sarin ambiguity aversion is preserved under all possible sequences of successes and failures.*

Bayes rule prescribes that the posterior of a beta distribution with Bernoulli observations is given by beta$(\alpha+s, \beta+f)$ after observing s successes and f failures. Since $s \geq 0$ and $f \geq 0$, $\alpha' = \alpha + s > 0$ and $\beta' = \beta + f > 0$. Therefore, Theorem 3 applies with beta(α', β'). ♡

6.1 The Gittins Index of an Ambiguity Averse Agent

For ambiguity aversion to be a descriptive model of behavior, it should predict the lower-than-optimal Gittins indexes observed.

Theorem 4 *Suppose A is regular with $\alpha_1 > 0$ and F_ω is the ambiguity-aversion adjusted version of F. Then $\Lambda(F, A) \geq \Lambda(F_\omega, A)$.*

Before proving this directly, we need a result about the impact of ambiguity aversion on the value function.

Lemma 5 *Suppose F is a Bernoulli arm and the agent is ambiguity averse. Then*

$$V(F, \lambda; A) \geq V(F_\omega, \lambda; A) \tag{51}$$

for all A.

Proof: This proof proceeds by induction on the horizon.

Suppose the horizon of A is zero. Then

$$V(F, \lambda; A_0) = V(F_\omega, \lambda; A_0) = 0 \tag{52}$$

because neither arm is ever selected and no reward is received.

For any $n \geq 1$, assume that $V(F, \lambda; A_n) \geq V(F_\omega, \lambda; A_n)$ for any F satisfying the conditions of the proposition and for any A with horizon less than n.

Suppose the horizon of A is n, and τ is an optimal strategy in the $(F_\omega, \lambda; A_n)$ bandit. Let τ' have the same first selection as τ and then proceed optimally in the $(F, \lambda; A_n)$ bandit.

It is sufficient to show that

$$W(F, \lambda; A_n; \tau') \geq V(F_\omega, \lambda; A_n). \tag{53}$$

Without loss of generality, assume that the first move under both τ and τ' is arm 1. Then

$$\begin{aligned} W(F, \lambda; A_n; \tau') &= \alpha_1 E[X|F] + P(X = 1|F) V(\sigma F, \lambda; A_n^{(1)}) \\ &\quad + P(X = 0|F) V(\phi F, \lambda; A_n^{(1)}) \\ V(F_\omega, \lambda; A_n) &= \alpha_1 E[\omega(X)|F] + P(X = 1|F) V(\sigma F_\omega, \lambda; A_n^{(1)}) \\ &\quad + P(X = 0|F) V(\phi F_\omega, \lambda; A_n^{(1)}). \end{aligned} \tag{54}$$

Note that the probabilities $E[X|F]$ in the second equation are not transformed by the ambiguity function; the agent understands the probabilities, but does not like the uncertainty they imply for her payoffs.

$$\begin{aligned} & W(F,\lambda;A_n;\tau') - V(F_\omega,\lambda;A_n) \\ &= \alpha_1\left[E[X|F] - E[\omega(X)|F_\omega]\right] \\ &\quad + P(X=1|F)\left[V(\sigma F,\lambda;A_n^{(1)}) - V(\sigma F_\omega,\lambda;A_n^{(1)})\right] \\ &\quad + P(X=0|F)\left[V(\phi F,\lambda;A_n^{(1)}) - V(\phi F_\omega,\lambda;A_n^{(1)})\right] \end{aligned}$$

The first term is nonnegative because $\omega()$ represents ambiguity aversion.

The second term is nonnegative because the induction hypothesis applies to any F, in particular σF. The third term is nonnegative because the induction hypothesis applies to any F, in particular ϕF.

Therefore, the difference is nonnegative, so the claim is proven. ♡

Proof of Theorem 4: This proof is by contradiction.

Suppose $\Lambda(F,A) < \Lambda(F_\omega,A)$. Then arm 1 is optimal initially in the $(F,\Lambda(F,A);A)$ and $(F_\omega,\Lambda(F,A);A)$ bandits. Then

$$\begin{aligned} & V(F,\Lambda(F,A);A) - V(F_\omega,\Lambda(F,A);A) \\ &= \alpha_1\left[E[X|F] - E[\omega(X)|F]\right] \\ &\quad + P(X=1|F)\left[V(\sigma F,\Lambda(F,A);A^{(1)}) - V(\sigma F_\omega,\Lambda(F,A);A^{(1)})\right] \\ &\quad + P(X=0|F)\left[V(\phi F,\Lambda(F,A);A^{(1)}) - V(\phi F_\omega,\Lambda(F,A);A^{(1)})\right] \end{aligned} \tag{55}$$

The first term is nonnegative by the definition of ambiguity aversion. The second and third terms are each positive by the proposition above.

However, we also know that the strategy using arm 2 at every stage is optimal for the $(F,\Lambda(F,A);A)$ bandit. Therefore

$$\begin{aligned} V(F,\Lambda(F,A);A) - V(F_\omega,\Lambda(F,A);A) &= \gamma_1\Lambda(F,A) - V(F_\omega,\Lambda(F,A);A) \\ &\leq 0 \end{aligned} \tag{56}$$

The inequality follows because $\Lambda(F,A) < \Lambda(F_\omega,A)$ implies that arm 2 cannot be the optimal choice in the first period of the $(F_\omega,\Lambda(F,A);A)$ bandit. Therefore, the value of the $(F_\omega,\Lambda(F,A);A)$ bandit must be strictly greater than $\gamma_1\Lambda(F,A)$ (arrived at through some strategy which selects arm 1 initially). However, Equation 56 being negative contradicts Equation 55 being positive. Therefore, it cannot be that $\Lambda(F,A) < \Lambda(F_\omega,A)$. ♡

6.2 Willingness to Pay for Information About the True Average Payoff of an Ambiguous Arm

Theorem 5 *If A is regular, an ambiguity averse agent will pay more to learn the value of the parameter θ than an optimal agent.*

Proof: This proof calculates the value of learning the value of θ and demonstrates that it is increasing in λ. The result follows from Theorem 4 which demonstrates that $\Lambda(F, A) \geq \Lambda(F_\omega, A)$.

Because A is regular, the value of playing an uncertain bandit is given by $\gamma_1 \Lambda(F, A)$.

If agents knew the value of θ, they would select the arm yielding $\Lambda(F, A)$ if $\theta < \Lambda(F, A)$, and the arm yielding θ otherwise. This yields a payoff of

$$\gamma_1 \left[\int_0^{\Lambda(F,A)} \Lambda(F, A) dF(\theta) + \int_{\Lambda(F,A)}^1 \theta dF(\theta) \right]. \tag{57}$$

The value of knowing θ, or the amount the agent would be willing to pay to learn θ, is given by

$$\begin{aligned} &\gamma_1 \left[\int_0^{\Lambda(F,A)} \Lambda(F, A) dF(\theta) + \int_{\Lambda(F,A)}^1 \theta dF(\theta) \right] - \gamma_1 \Lambda(F, A) \\ = \;& \gamma_1 \left[\int_0^{\Lambda(F,A)} \Lambda(F, A) dF(\theta) + \int_{\Lambda(F,A)}^1 \theta dF(\theta) \right] \\ & - \gamma_1 \left[\int_0^{\Lambda(F,A)} \Lambda(F, A) dF(\theta) + \int_{\Lambda(F,A)}^1 \Lambda(F, A) dF(\theta) \right] \\ = \;& \gamma_1 \int_{\Lambda(F,A)}^1 (\theta - \Lambda(F, A)) dF(\theta). \end{aligned}$$

This expression is positive since $(\theta - \Lambda(F, A)) \geq 0$ at every point in the range of integration. It is zero only if $\Lambda(F, A) = 1$.

It remains to show that this value is decreasing in λ. Now suppose $\lambda^* > \lambda$. This will show that changing from λ^* to λ results in an increase in the value of learning θ.

At λ^*, the value of learning θ is

$$\gamma_1 \int_{\lambda^*}^1 (\theta - \lambda^*) dF(\theta). \tag{58}$$

A smaller index has two effects on this quantity. First, it extends the range of the integral to the range of $\lambda \leq \theta < \lambda^*$. Second, it increases the integrand by increasing the difference. The value of learning θ under λ can be written

$$\gamma_1 \left[\int_{\lambda}^{\lambda^*} (\theta - \lambda) dF(\theta) + \int_{\lambda^*}^{1} [(\theta - \lambda^*) + (\lambda^* - \lambda)] \, dF(\theta) \right]. \qquad (59)$$

Both of the new terms are positive, indicating that the value is decreasing in λ.

Since $\Lambda(F, A) \geq \Lambda(F_\omega, A)$, the ambiguity averse agent will be willing to pay more for information about the value of θ. ♡

The result of this theorem provides a surprising contrast to Theorem 4. Because agents are willing to pay more than optimal to learn about the value of θ, it appears that they overvalue information. On the other hand, because their Gittins index is lower than optimal, ambiguity averse agents appear to undervalue information.

This paradox does not arise in the other models studied. Ambiguity averse agents value the counterfactual universe where they know θ without ambiguity aversion; agents who are hyperbolic discounters, or who cannot properly solve the dynamic programming problem, will bring these sub-optimalities to the counterfactual calculation. Therefore, this surprising prediction of ambiguity aversion is excellent grounds for testing the model.

A slightly different result is needed to test this prediction in an experiment. Because I do not know the agents Gittins indexes, I cannot set the value of the second arm at the value that makes the agent indifferent. However, Theorem 5 can be generalized.

Theorem 6 *An ambiguity averse agent will pay more than an ambiguity neutral agent for information about the true value of θ in the $(F, \lambda; A)$ for any λ. If $\lambda < \Lambda(F, A)$, the ambiguity averse agent will pay strictly more.*

Proof: If the agent knows θ, then her expected payoff will be $\gamma_1 \max[\theta, \lambda]$ because the agent will simply select whichever arm gives the higher payoff.

If the agent does not know θ, she computes a Gittins index $\Lambda(F., A)$ (where $F.$ is either F or F_ω). Then she expects to receive $V(F., \lambda; A)$ if $\lambda < \Lambda(F., A)$ and $\gamma_1 \lambda$ if $\lambda \geq \Lambda(F., A)$.

Given that $\max[\theta, \lambda]$ can be written as in Equation 57, the agents should be willing to pay

$$\gamma_1 \left[\int_{0}^{\lambda} \lambda dF(\theta) + \int_{\lambda}^{1} \theta dF(\theta) - \lambda \right] \qquad (60)$$

if $\lambda > \Lambda(\cdot, A)$, and

$$\gamma_1 \left[\int_0^\lambda \lambda dF(\theta) + \int_\lambda^1 \theta dF(\theta) \right] - V(F, \lambda; A) \tag{61}$$

if $\lambda < \Lambda(\cdot, A)$.

There are three cases to consider.

(i) If $\lambda \leq \Lambda(F_\omega, A) < \Lambda(F, A)$, then both an ambiguity averse and an ambiguity neutral agent will select F in the first period, and Equation 61 applies. Lemma 5 shows $V(F_\omega, \lambda; A) < V(F, \lambda; A)$, so Equation 61 will be larger for the ambiguity averse agent.

(ii) If $\Lambda(F_\omega, A) < \Lambda(F, A) \leq \lambda$, then it is optimal for both agents to select λ in the first period (and indefinitely), so they will both determine the value of learning θ using Equation 60. This value does not depend on attitude toward ambiguity, so both agents will pay the same amount.

(iii) If $\Lambda(F_\omega, A) \leq \lambda < \Lambda(F, A)$, then the ambiguity averse agent will select λ in the first period and determine the value of learning θ using Equation 60, but the ambiguity neutral agent will select F in the first period and determine the value of learning θ using Equation 61.

For any F and any value of λ, Equation 60 is larger than Equation 61. Subtracting Equation 61 from Equation 60 yields

$$V(V, \lambda; A) - \gamma_1 \lambda > 0. \tag{62}$$

Positivity follows because the fact that an agent is using Equation 61 implies F is the optimal first period choice, and for this to be the case, it must allow the agent to do better than $\gamma_1 \lambda$. Therefore, Equation 60 is larger than Equation 61, and the ambiguity averse agent will pay more to learn θ.

Note that the ambiguity averse agent will pay strictly more in cases (i) and (iii), and exactly the same in case (ii).

7 Experimental Design

Testing the prediction that agents with suboptimal indexes will pay more than optimal for initial information about an ambiguous arm requires two treatments: one to determine a Gittins index, and the second to determine the subject's willingness to pay for information about the value of an ambiguous arm.

In each treatment, subjects will play six bandits in a random order. Each bandit will have a known, fixed horizon and no induced discounting. Each bandit will have two arms, one with an unknown (ambiguous) average payoff, and one with a known (unambiguous) average payoff.

	Prior (α, β)	Prior Mean	Prior Std	Periods	Gittins Index	Optimal WTP
A	(1,1)	0.50	0.29	4	0.62	0.25
B	(2.5,2.5)	0.50	0.20	4	0.56	0.21
C	(1.1,3.9)	0.22	0.17	8	0.31	0.06
D	(3.9,1.1)	0.78	0.17	5	0.83	0.04
E	(2,3)	0.40	0.20	5	0.47	0.21
F	(3,2)	0.60	0.20	8	0.69	0.24

Table 1: Properties of the six unknown arms used in the experiment

7.1 Ambiguous Arms

In each treatment, subjects may choose to receive payoffs from a Bernoulli arm with an unknown (ambiguous) probability of paying off. The probability of payoff (the parameter of the Bernoulli distribution) is distributed beta(α, β), where α and β are known parameters. This probability is represented to subjects in terms of balls and urns. They are told that they are choosing between urns (arms) which contain 100 balls in some combination of red and white. When they choose an urn, one ball is drawn at random from the urn, and they earn $1 if the ball is red and nothing if it is white. The different priors are related using tables which give the chance that there are exactly (PDF) and less than (CDF) each possible number of red balls in the unknown urn.

In separate treatments, subjects play each of six bandits with the characteristics represented in Table 1. These bandits were chosen to allow direct testing for three possible effects: a mean effect, a variance effect and a length effect. Bandits A and B have the same mean and same length but different variances, so their data can be compared to discover the impact of a change in variance. Bandits C and D (and E and F) have the same variance, but different means and different numbers of periods. A mean effect can be tested by comparing data from bandits C and E (which have means below 0.5) with that from bandits D and F (which have means above 0.5). On average, these pooled data will have the same number of periods, and the same variance. A length effect can be tested by comparing the data from bandits C and F (which have length 8) and D and E (which have length 5). On average, these pooled data will have the same means and the same variance. If more than one of these effects is present, standard statistical techniques can be used to control for confounding effects.

The instructions used in each treatment are in Appendix A.

7.2 Gittins Index Treatment

In the Gittins index treatment, the subject is given the table which describes the prior over the arm mean and told the number of periods in the bandit. She is then asked the minimum true mean of the known mean arm for which she would choose the known mean arm in the first period. The mean of the known mean arm is then announced, and the subject's first period arm choice is made for the subject based on her reported index: the known mean arm is chosen for her if her reported index is lower than the known mean, and the unknown mean arm is chosen otherwise. In subsequent periods, the subject can choose either the known or unknown mean arm.

The minimum mean of a known mean arm for which she would choose a known mean arm in the first period is elicited using a simple titration mechanism where the subject can respond Yes or No to a question like, "Would you choose the known mean arm this period if it paid X% of the time?" This question was repeated, with successive values of X given by a bisection algorithm, until the subject's indifference point was narrowed to the nearest percent. This value is the subject's Gittins index.

That this mechanism is incentive compatible for eliciting a Gittins index is proven in Anderson (2000a).

7.3 Willingness to Pay Treatment

In the willingness to pay treatment, the subject must choose between the unknown mean arm and an arm which pays with probability one half. Before the first period, each subject is given the opportunity to pay the experimenter to tell her the actual mean of the ambiguous arm. If she buys this information, she knows the true means of both arms, and will choose the one with the higher mean in each period. If she does not, she must choose between a known and an unknown mean arm in each period.

The amount the subject is willing to pay is elicited using a simple titration mechanism like that used to elicit the Gittins index. The subject can respond Yes or No to a question like, "Would you be willing to pay $X.XX to learn the average of the unknown mean arm?" This question was repeated, with successive values of $X.XX (between $0.00 and $1.00) given by a bisection algorithm, until the subject's indifference point was narrowed to the nearest penny. This value is the subject's willingness to pay.

Once the subject's willingness to pay is established, it is compared to a randomly determined selling price. If the subject's willingness to pay is higher than the random price, then the subject is told the true mean of the unknown mean arm and the selling price is deducted from her total payoff; if her willingness to pay is lower than the price, she is not charged, and is not told the true mean of the unknown mean arm.

Subjects' willingness to pay was assessed at the same six priors as Gittins indexes are elicited. The value of learning θ, however, also depends on λ. Theorem 6 indicates that for any fixed $\lambda < \Lambda(F, A)$, the ambiguity averse agent will pay strictly more than an ambiguity neutral agent to learn the value of θ; if $\lambda \geq \Lambda(F, A)$, then both agents would pay the same. This prediction of ambiguity aversion can be tested by holding λ fixed and changing the prior from which the value of θ is drawn. If ambiguity aversion contributes to suboptimal experimentation, agents should pay much more than optimal when high values of θ are likely, but exactly optimal when low values of θ are likely. By using $\lambda = 0.5$, the different priors will progress through the range where ambiguity averse agents will pay more for information into that where they will pay just as much as ambiguity neutral agents.

7.4 Comments on Experimental Design

One challenge of this design is that risk aversion cannot be perfectly controlled. In the experiments with normal payoff distributions, risk can easily be controlled by having the variance of the payoff distribution be the same for the known mean and unknown mean arms. With Bernoulli arms, however, the variance is a function of the probability of payoff. Therefore, it is impossible to have both arms be equally risky.

Although perfect control is not possible, some measure of control is attained by having the known arm be risky as well. As an uncertain payoff, it deprives agents of a guaranteed reward, which may put them in a different frame of mind than when they have access to a certain payoff.[1]

A slightly higher degree of control is possible in the willingness to pay treatment, where the known mean arm pays with probability one half. This maximizes the variance of the Bernoulli payoff distribution, thereby maximizing risk. Therefore, risk averse subjects will never prefer the ambiguous arm to the unambiguous arm due to risk preference. Even if risk aversion is a significant factor in behavior, it will draw agents toward the almost surely less risky arm, the ambiguous arm. This should reduce the subject's willingness to pay for information about the true mean of the ambiguous

[1] There is some evidence that choice over events which occur with probability one is not the limit of choice over risky lotteries. Therefore introducing a little uncertainty changes the way subjects process the information and may attenuate the effects of risk aversion relative to a choice where one alternative is certain.

arm. Therefore, if subjects pay more than is optimal, it is attributable to ambiguity aversion, and *in spite of* any possible risk aversion.

A second complication of this design is that it compares the value of information based on an elicited Gittins indexes with the value of information based on a willingness-to-pay procedure. One might argue that any possible difference observed is attributable to a framing effect due to the differing procedures. There is no natural control for eliciting willingness-to-pay in the Gittins context, or vice-versa, so these effects are difficult to control for. It should be noted, however, that the phenomenon of underexperimentation is robust to procedural variances, as it appears in both the direct choice environment and the Gittins elicitation procedures. Therefore, we might expect it to be invariant to a willingness-to-pay procedure as well.

7.5 Subjects

Subjects consisted of 33 Caltech undergraduates who did not necessarily have any training in economics, though many had participated in unrelated economics experiments. Experimental sessions lasted about an hour and a half, and payments averaged $18, ranging from $3 to $24.

8 Results

Figure 1 is a box-and-whiskers plot of the data for each bandit. Gittins indexes and willingness to pay are represented as differences from optimal, with a positive difference corresponding to a higher than optimal Gittins index or willingness to pay. Each bar indicates the distribution of the data. The thin black horizontal line is at the median response, the grey box covers the middle 50%, and the long vertical lines cover 90% of the data. Note that this is the distribution of the data, and does not naturally correspond to confidence intervals of the central tendency of the data.

Based on this graph, it appears the data are consistent with ambiguity aversion. In every bandit, the median Gittins index is too low and the median willingness to pay is too high. In five of the six bandits, about 90% of the willingness to pay errors are higher than the median Gittins index error.

It is also noteworthy that the level of distribution of each error remains fairly constant from one bandit to the next. There is no dramatic effect of changes in prior standard deviation (ranging from 0.169 in (1.1,3.9) and (3.9,1.1) to 0.289 in (1,1)), prior mean (ranging from (1.1,3.9) to (3.9,1.1)) or horizon length.

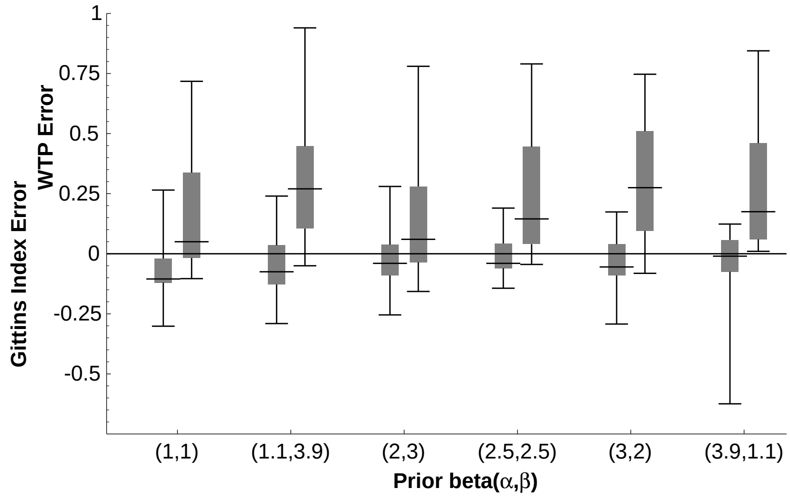

Figure 1: Box-and-whiskers plot of the difference between subjects' responses and the optimal Gittins index (on the left) and the optimal willingness to pay (on the right)

The rest of this section develops formal statistical tests for the patterns which appear in this graph.

Result 1 *Most subjects had higher than optimal willingness to pay and lower than optimal Gittins indexes.*

Ambiguity aversion predicts that subjects will have WTPs which are too high and Gittins indexes which are too low. The frequency with which each subject made these errors is reported in Table 2. The columns represent the number of the six bandits in which the subject reported a higher than optimal willingness to pay, and the rows represent the number of the six bandits in which the subject reported a lower than optimal Gittins index. The number in each cell is the number of subjects who made that combination of errors. For instance, three subjects made had higher than optimal willingness to pay and lower than optimal Gittins indexes in each of the six bandits in which each measurement was elicited.

An overwhelming majority of subjects fall in the lower righthand corner of this table, where they frequently have higher than optimal willingness to pay and lower than optimal Gittins indexes. Of the 33 subjects, 25 of them, have higher than optimal willingness to pay in at least five of the six bandits; 17 of the 33 have lower than optimal Gittins indexes. This pattern

# gi ≤	# wtp ≥ optimal						
optimal	0	1	2	3	4	5	6
0				1	2		
1						1	1
2							3
3						3	2
4							3
5	1					1	4
6				2	2	4	3

Table 2: Number of subjects who reported a higher than optimal willingness to pay in X bandits and lower than optimal Gittins indexes in Y bandits

Prior	(1,1)	(2.5,2.5)	(1.1,3.9)	(3.9,1.1)	(2,3)	(3,2)	Pooled
Willingness to Pay							
N	33	33	33	32	33	33	197
Median overpay	0.05	0.15	0.29	0.19	0.08	0.29	0.17
# overpay	24.00	29.00	30.00	29.00	23.00	31.00	166.00
p-value	0.0045	7E-06	1E-06	2E-06	0.012	2E-07	0
Gittins Index							
N	33	33	33	33	33	33	198
Median undervalue	0.10	0.04	0.07	0.01	0.03	0.04	0.04
# too low	27.00	19.00	24.00	19.00	20.00	22.00	131.00
p-value	0.0001	0.192	0.0045	0.192	0.112	0.0278	3E-06

Table 3: One-tailed p-values that WTPs and Gittins indexes are optimal for each arm and for the pooled data, based on the median response

of response is consistent with ambiguity aversion.

Result 2 *The median willingness to pay is significantly higher than optimal in each bandit, and the median Gittins index is significantly lower than optimal in three of the six bandits. In the pooled data, the median willingness to pay is significantly higher than optimal and the median Gittins index is significantly lower than optimal.*

Table 3 presents the median overpayment and median undervaluation for each arm, as well as for the pooled data. Because I am particularly interested in overpayment and undervaluation, these errors will both be defined as positive; a negative overpayment corresponds to underpayment, and a negative undervaluation corresponds to a higher than optimal Gittins index.

For every arm, the median overpayment is positive, meaning the median willingness to pay is higher than optimal. The third row indicates the number of overpayments which are positive. This number can be used to calculate a p-value for the hypothesis that the true median overpayment is equal to zero using Mosteller and Rourke's (1973) technique for calculating a nonparametric confidence interval for the median. Mosteller and Rourke establish a confidence interval by computing the probability that the true median is between the i^{th} and $(N-i+1)^{st}$ largest observations by computing the chance that between i and $(N-i+1)$ observations fall to the left of the median.[2] This idea can be extended to this circumstance by using the cumulative binomial to calculate the probability that, if the true median is zero, only $n \leq N$ observations are negative.

The last row presents this p-value, which is significant at all conventional levels for each arm, and extremely significant for the pooled data. Therefore, the median willingness to pay is significantly higher than optimal, consistent with the prediction of ambiguity aversion.

The second section of the table presents the same information for undervaluation. The median undervaluation is significantly positive at conventional levels for three of the six bandits. However, the median of the pooled data is highly significantly positive, consistent with the prediction of ambiguity aversion.

8.1 Tests of Simple Effects

The data in Table 3 suggest that ambiguity aversion is a significant factor in bandit problems, leading subjects to paradoxically undervalue information by having lower than optimal Gittins indexes and overvalue information by having higher than optimal willingness to pay. However, ambiguity aversion also predicts that as ambiguity increases, the undervaluation and overpayment should be more severe. However, a direct test of this prediction does not support ambiguity aversion.

Result 3 *When the mean and bandit horizon are constant, an increase in*

[2]This probability is represented by the binomial distribution and is given by

$$2\sum_{j=0}^{i}\left(\begin{array}{c}N\\j\end{array}\right)\left(\frac{1}{2}\right)^{N}. \tag{63}$$

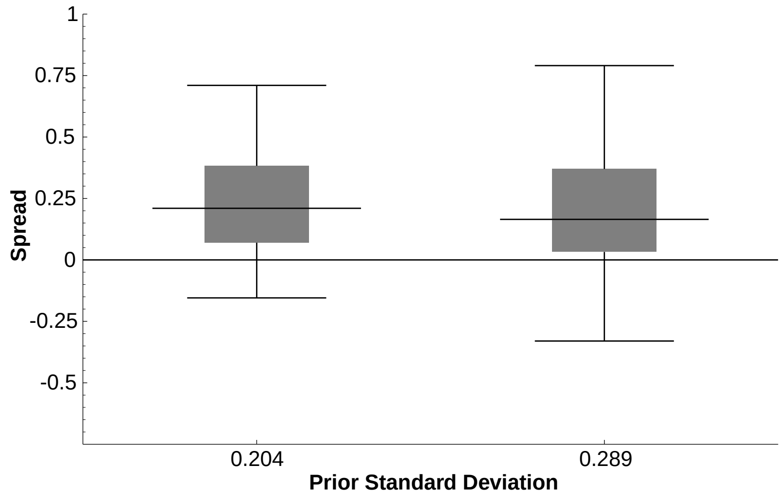

Figure 2: Box-and-whiskers plot of the spread for two bandits which differed only in prior variance

	(1,1)	(2.5,2.5)	p-value
Spread	0.17	0.26	1.000
Overpayment	0.05	0.15	0.317
Undervaluation	0.10	0.04	0.140

Table 4: Median spread, overpayment and undervaluation for the (1,1) bandit (with prior standard deviation 0.289) and the (2.5,2.5) bandit (with prior standard deviation 0.204), and p-values for the hypothesis that the two medians are the same.

variance does not result in a significant change in subjects' Gittins indexes or willingness to pay.

Whether variance affects the overpayment and undervaluation in the way predicted by ambiguity aversion can be tested directly by comparing the overpayment and undervaluation of the (1,1) and (2.5,2.5) bandits. Both have four periods and a prior mean of 0.5, and therefore differ only in variance. The prior standard deviation of the (1,1) bandit is 0.289 and of the (2.5,2.5) bandit is 0.204, so ambiguity aversion predicts the overpayments and undervaluations to be larger for the (1,1) bandit.

This can be tested directly by looking at the spread, the sum of the undervaluation and overpayment. Figure 2 is a box-and-whiskers plot of the spread for the (1,1) and (2.5,2.5) bandits. The (1,1) bandit has a slightly

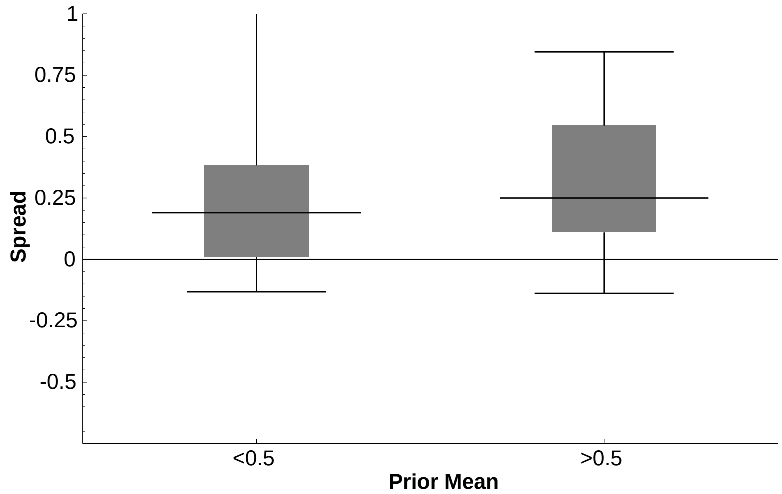

Figure 3: Box-and-whiskers plot of the spread for bandits which with Gittins indexes above and below the known mean arm

smaller median spread than the (2.5,2.5) bandit, the opposite of what ambiguity aversion would predict.

Table 4 shows the median spread for each bandit, and the p-value for the test that the medians of the two samples are the same. The continuity-corrected test statistic for the two-sample median test (Siegel and Castellan, 1988, Section 6.3) is exactly zero, leading to a p-value of one. Therefore, although the median of the higher variance bandit is lower, the difference is not significant.

The other two rows of the table, which show how overpayment and undervaluation change with ambiguity, illustrates that, although undervaluation increases with variance, overpayment seems to decrease (though not significantly).

Result 4 *When the variance and bandit horizon are constant, an increase in mean results in no change in overpayment and a borderline significant decrease in undervaluation.*

Ambiguity aversion makes a subtle prediction about how the mean will affect the spread. Although undervaluation should not be affected, the subjects' willingness to pay should be exactly the same as that of an ambiguity neutral agents when the mean is below 0.5. The reason for this is that, when the Gittins index is below 0.5, both the ambiguity averse and ambiguity neutral subject expects to pick the 50/50 arm in the first period (and in

	Prior Mean $\leq .5$	Prior Mean $\geq .5$	p-value
Spread	0.20	0.26	0.336
Overpayment	0.18	0.26	0.336
Undervaluation	0.06	0.02	0.056

Table 5: Median spread, overpayment and underpayment for the bandits with a prior mean below one half and those with a prior mean above one half, and p-values for the hypothesis that the two medians are the same.

every period thereafter) if he does not learn the mean of the unknown arm. Therefore, there is no ambiguity in the bandit, even when there is no information. Since the two bandits with means less than 0.5 also have optimal Gittins indexes below 0.5, the ambiguity averse agents must have subjective Gittins indexes below 0.5. Therefore, when they are calculating their willingness to pay, the ambiguity averse subjects consider the two-armed bandit with the ambiguous arm to be an unambiguous problem, because they will never encounter the ambiguity in optimal play.

This suggests that the level of overpayment should increase when the mean increases from below 0.5 to above 0.5. Figure 3 is a box-and-whiskers plot of the spread of bandits (1.1,3.9) and (2,3) on the left and (3,2) and (3.9,1.1) on the right. There is a slight increase in the median, and in the middle two quartiles, when the mean increases.

Table 5 presents the median spread, overpayment and undervaluation for the bandits with a prior mean below one half and those with a prior mean above one half. Although the median spread for the means above 0.5 is higher than that for the means below 0.5, the difference is not significant; nor is the difference in median overpayment. Interestingly, undervaluation is slightly lower than for priors with higher means, and this difference is borderline significant.

It is possible that this lack of increase in the spread is attributable to ambiguity aversion, if subjects' Gittins indexes for the arms with means above 0.5 are, due to ambiguity aversion, also below 0.5. The reported Gittins indexes, though not infrequently below the prior mean of the ambiguous arm, are rarely that much lower. Therefore, the lack of difference is probably not attributable to ambiguity aversion.

The key to the lack of increase may lie in the the fact that the premise of the prediction does not hold: agents do not have optimal willingness to pay in the arms with means below 0.5. Table 3 shows that bandits (1.1,3.9) and (2,3) (and (1,1) and (2.5,2.5)) have significantly higher than optimal willingness to pay. Considering why this is so may provide insight into why

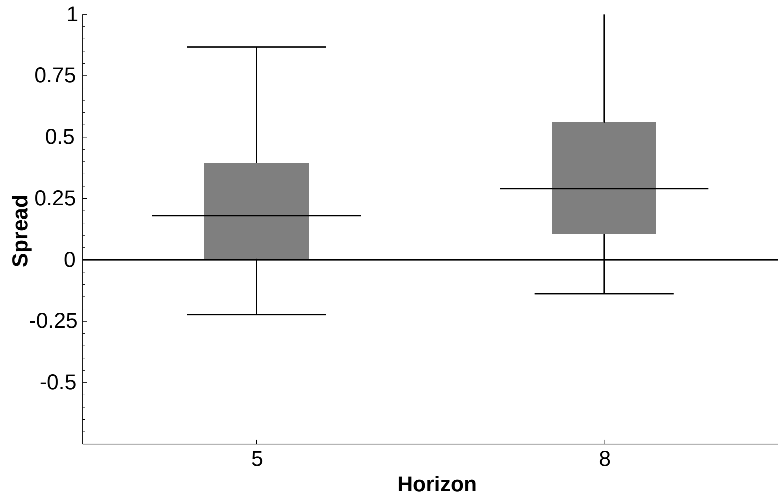

Figure 4: Box-and-whiskers plot of the spread for bandits with different horizons

	5 Periods	8 Periods	p-value
Spread	0.18	0.30	0.011
Overpayment	0.14	0.25	0.067
Undervaluation	0.02	0.07	0.056

Table 6: Median spread, overpayment and underpayment for the bandits with a five period horizon and those with an eight period horizon, and p-values for the hypothesis that the two medians are the same.

there is no mean effect.

Result 5 *When the variance and mean are constant, an increase in horizon results in an increase in both overpayment and undervaluation.*

Ambiguity aversion does not hold any role for the horizon, so any change in spread is not attributable to ambiguity aversion by may provide some insight into subjects' decision process.

Figure 4 is a box-and-whiskers plot of the spread of the (3.9,1.1) and (2,3) bandits on the left and the (1.1,3.9) and (3,2) bandits on the right. There is a slight increase in the median, and in the middle two quartiles, with the longer horizon.

Table 6 presents the median spread, overpayment and undervaluation for the bandits with a five period horizon and those with an eight period horizon.

There is a statistically significant increase in spread from the five period to the eight period horizon, generated by borderline significant increases in both overpayment and undervaluation.

Both the optimal Gittins index and the optimal willingness to pay are increasing in the horizon. These results suggest the subjects are too sensitive to the increase in willingness to pay, but not sensitive enough to the increase in the Gittins index.

This sort of relationship might arise from a simplification of Equation 61.[3] In this experiment γ_1 is simply the number of periods in the bandit. Subjects may try to approximate $V()$ by some linear function of the number of periods, rather than solve the dynamic programming problem. If this approximation yielded an average per-period value which was lower than the true per-period average value (which would be consistent with observing lower than optimal Gittins indexes), subjects would have an average per-period willingness to pay which was higher than optimal. Furthermore, their simple model would be linearly increasing in the horizon, so the amount of the overpayment would be increasing in the horizon.

8.2 Testing for Multiple Effects

The previous section used the design of the experiment to test simple hypotheses about the impact of changes in prior mean and variance and the horizon. However, this approach throws out some of the data, as not all six bandits are used in any of these tests. This section uses a simple linear model to test for these effects in the whole sample.

Result 6 *When tested on the entire sample, only the horizon effect is significant.*

The simple test for a change in the spread related to a change in the variance only used one third of the sample, and only two variances. Figure 5 shows the distribution of the data in the bandits with every variance used. As with the simple test, there is no dramatic trend toward larger spreads with higher standard deviations. However, this larger sample does not carefully control for mean and horizon effects.

The regressions in Table 7 use the whole sample and control for linear mean and horizon effects, and come to the same conclusion: the spread does

[3] After one session, a subject described to me exactly Equation 61 (these are Caltech students) and then asked, "but how do you compute the value of not having the information?"

	Constant	Prior Mean	Prior Std	Log(Horizon)
Spread	-0.03	0.12	-0.28	0.19
	(-0.14)	(0.75)	(-0.54)	(2.59)
Overpayment	0.07	0.07	-0.52	0.16
	(0.31)	(0.64)	(-1.31)	(2.08)
Undervaluation	0.12	0.09	0.28	0.03
	(-0.85)	(0.57)	(0.84)	(1.25)

Table 7: Results of regressions of spread, overpayment and undervaluation on prior mean and variance and the log of horizon (t-statistics are in parentheses)

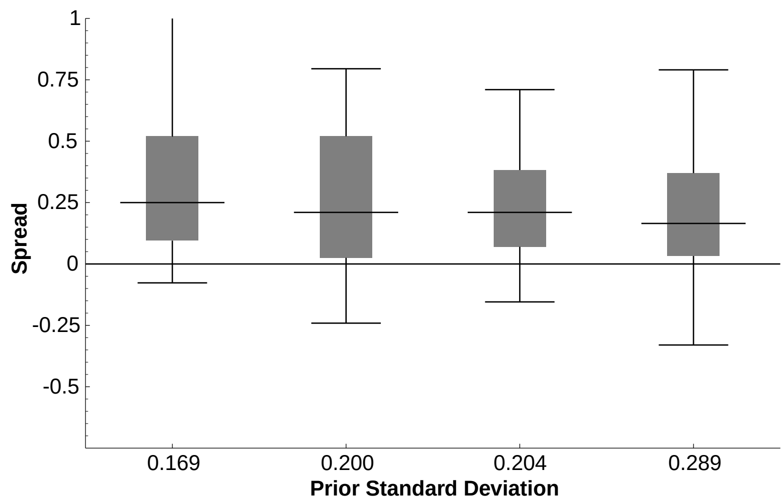

Figure 5: Box-and-whiskers plot of the spread for all bandits ordered by prior standard deviation

not vary significantly with changes in the variance. This is an important prediction of ambiguity aversion, and without support for it, it is difficult to say conclusively that the data in Table 3 are actually attributable to it. However, it may be that the small range of standard deviations used here is not large enough to generate a detectable difference.

The mean effect is also still not significant, but the horizon effect remains significant. Given that the horizon effect is consistent with a particular type of systematic error in the computation of the value function and the Gittins index, it seems possible that overpayment is attributable to the same cause as undervaluation.

9 Discussion

This paper showed how a distaste for ambiguity, the variance of the prior distribution of the mean, might lead to the lower than optimal Gittins indexes implied by results in search problems, as well as in Anderson's experiments. Ambiguity averse agents dislike receiving payoffs from ambiguous process, and therefore are willing to accept less valuable unambiguous alternatives than would be an ambiguity neutral agent. This means their Gittins indexes are lower than optimal.

Lower than optimal Gittins indexes make it seem as though ambiguity averse agents do not place enough value on the information they gain from experimentation. It is paradoxical, then, that if the problem is reframed to explore how much they would be willing to pay for information about an ambiguous alternative, they are willing to pay more than an ambiguity neutral agent, appearing to overvalue information.

This surprising prediction of ambiguity aversion is unique among theories which assume that agents correctly formulate and solve the dynamic programming problem. Other theories of systematic deviations, such as hyperbolic discounting and risk aversion, affect the calculation of the value function in both the ambiguous and unambiguous cases, so the difference in willingness to pay will not be generated. Of course, if agents do not correctly formulate and solve dynamic programming problems, almost any effect can be predicted.

The experiment presented here tested this surprising prediction, using one treatment which elicited subjects' Gittins index on six bandits, and another treatment which elicited subjects' willingness to pay on the same six bandits. The basic results are exactly as ambiguity aversion predicts: subjects have significantly lower than optimal Gittins indexes, and are willing to pay significantly more than optimal to learn the true mean of the ambiguous arm.

However, ambiguity aversion also predicts that undervaluation and overpayment should increase with the ambiguity of the alternative. Similarly, there should be an increase in overpayment as the Gittins index moves from below 0.5 to above 0.5. However, neither of these predicted effects is significant. Instead, the only significant change in overpayment and undervaluation arises from increases in the number of periods in the bandit.

The increase in the spread as the horizon increases is mostly attributable to an increase in overpayment. Why would agents be willing to pay more to remove ambiguity from a longer bandit? They may be willing to pay more to avoid having to solve the longer problem explicitly, giving up some money to save mental computation cost. However, this does not explain why Gittins indexes are lower than optimal.

Rather, because willingness to pay is higher than optimal when means are below 0.5, it appears likely that at least some of the suboptimality observed is attributable to a difficulty in solving the dynamic programming problem. If this difficulty results from a simplification of Equation 61, overpayment could increase with the horizon, without there necessarily being any corresponding impact on undervaluation.

That difficulty solving the dynamic programming problems is causing undervaluation and overpayment can be tested in a couple ways. A simple to solve bandit, such as one in which the unknown arm either always paid or never paid (with known probabilities) could test subject's intuitive dynamic programming ability. In retrospect, it would have been nice to include such a bandit among those used in this experiment. Alternately, subjects could be brought in for a thorough training session on dynamic programming which could elaborate on the direction and magnitude of the effects discussed in the strategy section of the instructions.

Like risk aversion, ambiguity aversion is a characteristic of individual agents. Therefore, rather than try to eliminate it, policies should focus on reducing ambiguity or on insuring against it. Insuring against ambiguity is more difficult than insuring against risk because it is difficult to insure against bad second order probability outcomes. This is because the outcome itself, the mean of the payoff distribution on the arm, is not observable. In the job search context, ambiguity may be represented by the current state of the job market or level of unemployment. Unlike the individual, however, the government policymaker has access to aggregate job market data, and may be able to use other people's success rates to estimate the true mean of the wage offer distribution. This information can be used to offer workers unemployment insurance against bad second order probability outcomes, and to encourage ambiguity averse agents to search more.

In other applications, public policy is unnecessary because other agents in the economy may have an interest in helping agents overcome their tendency to underexperiment. For instance, companies introducing new brands and stores with low prices would both like consumers to experiment with them. If these agents know consumers do not experiment enough, they can take steps to encourage experimentation. A company with a new brand might offer free samples at the supermarket or through the mail, or generous coupons. Stores with low prices may aggressively advertise, or even, as some new dot-coms are doing, offer first purchases for free. These measures all provide information which reduce ambiguity, and will benefit the consumer in the long run.

Finally, this paper accomplished a broader goal of demonstrating how simple behavioral models can be integrated into larger bodies of theory to produce testable predictions based on sound psychology. Although there are a number of different models of ambiguity aversion, the Kahn and Sarin model leverages the second order probability available to bandit agents. With some interpretation, the behavioral model was integrated into the sophisticated mathematics of bandits to derive the surprising and testable prediction that agents will have lower than optimal Gittins indexes, but be willing to pay more than optimal to eliminate ambiguity. Hopefully the exercise of developing the theory and testing it will inspire others to integrate formal behavioral models into their theories.

Citations

Citations

Abramowitz, M. and C. Stegun. (Eds.). Confluent Hypergeometric Functions. Ch. 13 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. New York: Dover, 1972. 503-515.

Anderson, C. A Laboratory Test of Present Bias in Bandit Problems. Mimeo. 2000a.

Anderson, C. Behavioral Models of Strategies in Multi-Armed Bandit Problems. California Institute of Technology Dissertation (working). 2000b.

Anscombe, F. and R. Aumann. A Definition of Subjective Probability. *Annals of Mathematical Statistics* 34, 1963. 199-205.

Banks, J., M. Olson and D. Porter. An Experimental Analysis of the Bandit Problem. *Economic Theory*, 10:55-77, 1997.

Berry, D. and B. Fristedt. *Bandit Problems.* New YorK: Chapman and Hall, 1985.

Cox, J. and R. Oaxaca. Testing Job Search Models: The Laboratory Approach. In *Research in Labor Economics* vol 15. Greenwich, CT: JAI Press, 1996. 171-207.

——. Direct Tests of the Reservation Wage Property. *The Economic Journal*, 102:1423-1432, 1992.

——. Unemployment Insurance and Job Search. In *Research in Labor Economics* vol 11. Greenwich, CT: JAI Press, 1990. 223-240.

——. Laboratory Experiments with a Finite-Horizon Job-Search Model. *Journal of Risk and Uncertainty*, 2:301-329, 1989.

de Finetti, B. Probabilities of Probabilities. In *New Directions in the Application of Bayesian Methods.* Ed. A. Aykac and C. Brumat. Amsterdam: North Holland, 1977. 1-10.

Einhorn, H. and R. Hogarth. Ambiguity and Uncertainty in Probabilistic Inference. *Psychology Review* 92:433-461, 1985.

Ellsberg, D. Risk, Ambiguity and the Savage Axioms. *Quarterly Journal of Economics* 75:643-669, 1961.

Frisch, D. and J. Baron. Ambiguity and Rationality. *Journal of Behavioral Decision Making* 1:149-157, 1988.

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

Gittins, J. *Multi-Arm Bandit Allocation Indicies.* New York: John Wiley and Sons, 1989.

Gittins, J. and D. Jones. A Dynamic Allocation Index for the Sequential Design of Experiments. In *Progress in Statistics*, ed. J. Gani et al. Amsterdam: North Holland, 1974. 241-66.

Hogarth, R. and H. Einhorn. Venture Theory: A Model of Decision Weights. *Management Science* 36:780-803, 1990.

Kahn, B. and R. Sarin. Modelling Ambiguity in Decisions Under Uncertainty. *Journal of Consumer Research* 15:265-272, 1988.

Kahneman, D. and A. Tversky. On Prediction and Judgement. *ORI Research Monograph 12.* 1972.

Mosteller, F. and R. Rourke. *Sturdy Statistics.* Reading, MA: Addison-Wesley Publishing, 1973.

Pratt, J., D. Wise and R. Zeckhauser. Price Differences in Almost Competitive Markets.

Quiggan, J. A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization* 3:323-343, 1982.

Schmeidler, D. Subjective Probability and Expected Utility without Additivity. *Econometrica* 57:571-587, 1989.

Segal, U. The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach. *International Economic Review* 28:175-202, 1987.

Siegel, S. and N. Castellan. *Nonparametric Statistics for the Behavioral Sciences*, 2 ed. New York, NY: McGraw-Hill Book Company, 1988.

Slater, L. *Confluent Hypergeometric Functions.* Cambridge: Cambridge University Press, 1960.

Spanier, J. and K. Oldham. The Kummer Function $M(a, c; x)$. Ch. 47 in *An Atlas of Functions.* Washington, DC: Hemisphere, 1987. 459-469.

Wolfram Research. Confluent Hypergeometric Function of the First Kind. In *Eric Weisstein's World of Mathematics.* http://mathworld.wolfram.com/ConfluentHypergeometricFunctionoftheFirstKind.html. Accessed 5/31/00.

Yates, F. and L. Zukowski. Characterization of Ambiguity in Decision Making. *Behavioral Science* 21:19-25, 1976.

A Instructions

A.1 Gittins Index Treatment

You are about to participate in an experiment designed to provide insight into decision processes. The amount of money you make will depend partly on decisions you make and partly on chance. If you follow the instructions carefully and make good decisions, you might earn a considerable amount of money.

A.1.1 How You Make Money

You will earn one dollar each time a red ball is drawn from an urn you choose. In each period, you may choose one of two urns: one with 100 balls in some known mixture of red and white, and one with 100 balls in some unknown mixture of red and white. When you choose an urn, one ball will be randomly drawn from that urn and it determines your payoff: one dollar if it is red, nothing if it is white.

You will be paid in cash for all of your earnings in excess of $28.

A.1.2 Order of the Experiment

This experiment will proceed as a number of rounds. Each round will have several periods. The number of periods in each round will be announced at the beginning of that round.

At the beginning of each round, the computer will randomly determine the number of red balls in the known mixture urn and in the unknown mixture urn. Before you learn the number of red balls in the known mixture urn, you will be asked the smallest number of red balls which would lead you to choose the known mixture urn in the first period. Your first period choice will be made for you by the computer based on your response. The computer will choose the known mixture urn for you if it contains more red balls than your smallest number, and the unknown mixture urn for you if the known mixture urn contains fewer red balls than your smallest number.

After the first period, you may select either urn. In each period, you will have to trade off selecting the known mixture urn with learning more about the number of red balls in the unknown mixture urn.

A.1.3 Urns

At the beginning of each round, you will be told the number of red balls in the known mixture urn. It may contain any number between zero and 100 red balls.

You will not know the mixture of red and white balls in the unknown urn. It may contain any number between zero and 100 red balls. However, not all mixtures are equally likely. The likelihood of different proportions of red balls is represented on the urn tables.

The number of red balls in both urns will remain constant from one period to the next, but will change at the beginning of each round.

A.1.4 Urn Tables

The chance that the unknown mixture urn contains a given number of red balls is represented in tables like the Practice Urn Table you have been given. A new table will be distributed at the beginning of each round.

The first column of the table shows a possible number of red balls. The second column indicates the chance that there are exactly X number of red balls in the unknown mixture urn. This is also illustrated in the graph below the table with a solid black line (read on the left axis). For instance, there is a 1% chance there are exactly 25 red balls (and 75 white balls) in the unknown mixture urn.

The third column of the table shows the chance that there are *at least* X red balls in the unknown mixture urn. This is illustrated in the graph below the table with a dashed lack line (read on the right axis). For instance, there is a 49.08% chance that there are fewer than 50 red balls in the unknown mixture urn.

A.1.5 Known Mixture Urn Cutoff

In the first period, the computer will ask you "Would you choose the known mixture urn in this period if it contained [Number] red balls?" If you would, click the "Yes" button, if not, click the "No" button. You will be asked a series of these questions, with a different [Number] each time, until the cutoff point at which you would just prefer the known mixture urn has been narrowed down to the nearest ball.

You should answer these questions carefully because your first period urn choice will be made for you based on your answers. The computer assumes you will choose the known mixture urn for all numbers of red balls larger than the cutoff, and the unknown mixture urn otherwise. Therefore, it will automatically choose the known mixture urn if it has more red balls than your cutoff, and the unknown mixture urn if the known mixture urn has fewer red balls than your cutoff.

A.1.6 Using the Computer

There are four panels on the computer screen. You may click in these panels with your mouse, but please do not attempt to use any other applications, look at the source code for this experiment or visit any other web sites during the experiment.

The History Panel

The long vertical panel on the left will contain your playing history. Please look at that panel now. For each period, it will show your choice of urn, your payoff and the minimum value for which you would choose the known mixture urn; recent periods will be added to the top of the list, though earlier periods will still be accessible by scrolling down.

The Information Panel

The top of the three panels on the right side provides you with information on the current period, the total number of periods and your total payoff. It also provides a "Best Guess" at the number of red balls in the unknown mixture urn. In the first period, it shows the average number of red balls that would be an urn based on the urn table. Once the unknown mixture urn is chosen, the "Best Guess" uses a law of probability called Bayes' rule. Bayes' rule uses the chance the observed combination of red and white balls arose from each possible mixture, and the chance of each mixture from the urn table, to determine the most likely average number of red balls in the unknown mixture urn, given the available information.

The Urn Choice Panel

Please look at the middle of the three righthand panels (which now has a "Begin" button). This is where you indicate your choice of urn each period. To indicate your choice of an urn, click once with the mouse in the circle in front of the name of the urn you wish to choose; a black dot will appear within the white circle. Then click the Submit button at the bottom of the panel one time with the mouse. Clicking the Submit button causes the computer to select a ball and calculate your payoff for the period.

The Instructions Panel

The bottom of the three right panels will contain these instructions. You may scroll through them and examine them at any point during the experiment.

A.1.7 Summary

1. At the beginning of the round, the Experimenter will distribute an urn table and announce the number of periods in the round.
2. The computer will randomly select the number of red balls in the known mixture urn, and in the unknown mixture urn.
3. You will be asked a series of questions to determine your known mixture urn cutoff, the minimum number of red balls in the known mixture urn for which you would choose it that period.
4. The computer will automatically choose the known mixture urn for you if its actual number of red balls is higher than your cutoff, and the unknown mixture urn otherwise.
5. The computer will randomly draw a ball from your chosen urn and announce your payoff: one dollar if the ball is red and nothing if it is white.
6. Fill in the record section of the urn table.
7. Wait for the experimenter to announce the beginning of the next period.
8. Choose between the known and unknown mixture urns, and return to Step 5.

A.1.8 Strategy

You have a chance to receive a payoff when you select either urn, but when you select the unknown mixture urn, you also gain some information about the number of red balls it contains, which may help you in future periods. At any point, your best guess may be higher or lower than the actual number of red balls in the urn. By trying the unknown mixture urn, you may learn it has more red balls than the known mixture urn, information you can use to improve your chance of getting a red ball in future periods; if the unknown mixture urn does not have more red balls, you can choose the known mixture urn in future periods.

This possibility of learning the unknown mixture urn is better than your initial best guess means it is sometimes advantageous to select the unknown mixture urn even when the known mixture urn contains more red balls than your best guess at the number in the unknown mixture urn. Whether it is worth experimenting with the unknown mixture urn depends on the difference between your best guess and the number of red balls in the known mixture urn, the number and color of the balls you have observed from the unknown mixture urn, the chance that the actual number of red balls in the unknown mixture urn is each value higher than your best guess (based on the urn table) and number of periods you have left to benefit from learning the unknown mixture urn has more red balls than your initial best guess.

A.2 Willingness to Pay Treatment

You are about to participate in an experiment designed to provide insight into certain features of decision processes. The amount of money you make will depend partly on decisions you make and partly on chance. If you follow the instructions carefully and make good decisions, you might earn a considerable amount of money.

A.2.1 How You Make Money

You will earn one dollar each time a red ball is drawn from an urn you choose. In each period, you may choose one of two urns: one with 50 red balls and 50 white balls, and one with 100 balls in some unknown mixture of red and white. When you choose an urn, one ball will be randomly drawn from that urn and it determines your payoff: one dollar if it is red, nothing if it is white.

You will be paid in cash for all of your earnings in excess of $28.

A.2.2 Order of the Experiment

This experiment will proceed as a number of rounds. Each round will have several periods. The number of periods in each round will be announced at the beginning of that round.

At the beginning of each round, the computer will randomly determine the number of red balls in the unknown mixture urn. You will then be asked how much you would be willing to pay to learn the number of red balls in the unknown mixture urn. If you are willing to pay more than the computer's randomly determined selling price, the computer will tell you then number of red balls in the unknown mixture urn and deduct the selling price from your total payoff.

If you are not willing to pay more than the computer's randomly determined selling price, you will not be charged, but you will only be able to learn about the the number of red balls in the unknown mixture urn by choosing it. In each period, you will have to trade off choosing the 50-50 urn with learning more about the number of red balls in the unknown mixture urn.

A.2.3 Urns

In each period you will be choosing between two urns. One urn will always be a 50-50 urn, containing 50 red balls and 50 white balls. On average, this urn will pay one dollar half the time.

You will not know the mixture of red and white balls in the unknown urn. It may contain any number between zero and 100 red balls. However, not all mixtures are equally likely. The likelihood of different proportions of red balls is represented on the urn tables.

The number of red balls in the unknown mixture urn will remain constant from one period to the next, but will change at the beginning of each round.

A.2.4 Urn Tables

The chance that the unknown mixture urn contains a given number of red balls is represented in tables like the Practice Urn Table you have been given. A new table will be distributed at the beginning of each round.

The first column of the table shows a possible number of red balls. The second column indicates the chance that there are exactly X number of red balls in the unknown mixture urn. This is also illustrated in the graph below the table with a solid black line (read on the left axis). For instance, there is a 1% chance there are exactly 25 red balls (and 75 white balls) in the unknown mixture urn.

The third column of the table shows the chance that there are **at least** X red balls in the unknown mixture urn. This is illustrated in the graph below the table with a dashed lack line (read on the right axis). For instance, there is a 49.08% chance that there are fewer than 50 red balls in the unknown mixture urn.

A.2.5 Willingness to Pay (WTP) for Information

At the beginning of each round, you will have the opportunity to buy from the experimenter the number of red balls in the unknown mixture urn. You may want to pay for this information because it tells you which urn has more red balls, and therefore is more likely to pay one dollar. If you do not have this information, you can learn whether there are many red balls in the unknown mixture urn only by choosing it and observing your payoffs. On the other hand, you do not want to pay more for this information than you can gain by having it.

A.2.6 Using the Computer to Purchase Information

After you have been shown the urn table and learned the number of periods in the round, you will be asked "Would you be willing to pay $X.XX to learn the number of red balls in the unknown mixture urn?" If you would be willing to pay that amount, click "Yes," if not, click "No." The computer will ask a series of these questions, with different values, until it has narrowed the amount you are willing to pay to the nearest cent.

Once it has determined how much you are willing to pay, the computer will compare your value to the randomly determined price at which it will sell the information. If your WTP is higher than the computer's price, the computer will tell you the number of red balls in the unknown mixture urn and deduct its price (not your WTP) from your total payoff. If your WTP is lower than the computer's price, you will not be charged, but you will not be told the number of red balls in the unknown mixture urn.

Be careful in selecting your WTP. If you enter a value which is higher than you are really willing to pay, you may have to pay more for the information than you want; if you enter a value which is lower than you are really willing to pay, you may not receive the information when the computer would be willing to tell you for a price you would be willing to pay.

A.2.7 Using the Computer to Choose an Urn

There are four panels on the computer screen. You may click in these panels with your mouse, but please do not attempt to use any other applications, look at the source code for this experiment or visit any other web sites during the experiment.

The History Panel

The long vertical panel on the left will contain your playing history. For each period, it will show your choice and the payoff you received; recent periods will be added to the top of the list, though later periods will still be accessible by scrolling down.

The Information Panel

The top of the three panels on the right side provides you with information on the current period, the total number of periods and your total payoff. It also provides a "Best Guess" at the number of red balls in the unknown mixture urn. In the first period, it shows the average number of red balls that would be in an urn based on the urn table. Once the unknown mixture urn is chosen, the "Best Guess" uses a law of probability called Bayes' rule. Bayes' rule uses the chance the observed combination of red and white balls arose from each possible mixture, and the chance of each mixture from the urn table, to determine the most likely average number of red balls in the unknown mixture urn, given the available information.

The Urn Choice Panel

The middle of the the three righthand panels is where you indicate your choice of urn in each period. To indicate your choice of an urn, click once with the mouse in the circle in front of the name of the urn you wish to choose; a black dot will appear within the white circle. Then click the Submit button at the bottom of the panel one time with the mouse. Clicking the Submit button causes the computer to generate a Random Value and calculate your payoff for the period.

The Instructions Panel

The bottom of the three right panels will contain these instructions. You may scroll through them and examine them at any point during the experiment.

A.2.8 Summary

1. At the beginning of the round, the Experimenter will distribute an urn table and announce the number of periods in the round.

2. The computer will randomly select the number of red balls in the

unknown mixture urn using the urn table.

3. The computer will ask you a series of questions to determine the maximum price you are willing to pay to learn the number of red balls in the unknown mixture urn.

4. The computer will compare your WTP to a random selling price.

   (a) If your WTP is higher than the selling price, the computer will tell you the number of red balls in the unknown mixture urn and deduct the selling price (not your WTP) from your total payoff.

   (b) If your WTP is lower than the selling price, the computer will not tell you the number of red balls in the unknown mixture urn.

5. The experimenter will instruct you to choose an urn.

6. The computer will randomly draw a ball from your chosen urn and announce your payoff: one dollar if the ball is red and nothing if it is white.

7. Fill in the record section of the urn table.

8. Wait for the experimenter to announce the beginning of the next period.

9. Choose between the known and unknown mixture urns, and return to Step 6.

A.2.9 Strategy

You have a chance to receive a payoff when you select either urn, but when you select the unknown mixture urn, you also gain some information about the number of red balls it contains, which may help you in future periods. At any point, your best guess may be higher or lower than the actual number of red balls in the urn. By trying the unknown mixture urn, you may learn it has more than 50 red balls, information you can use to improve your chance of getting a red ball in future periods; if the unknown mixture urn has fewer than 50 red balls, you can choose the 50-50 urn in future periods.

This possibility of learning the unknown mixture urn is better than your initial best guess means it is sometimes advantageous to select the unknown mixture urn even when your best guess at the number in the unknown mixture urn is less than 50. Whether it is worth experimenting with the unknown mixture urn depends on the difference between your best guess and

50, the number and color of the balls you have observed from the unknown mixture urn, the chance that the actual number of red balls in the unknown mixture urn is each value higher than your best guess (based on the urn table) and number of periods you have left to benefit from learning the unknown mixture urn has more red balls than your initial best guess.

Feel free to earn as much money as you can. Are there any questions?