DRAFT: Learning and Recall in Bayesian Decision-making.

David R. Just
University of California, Berkeley

An immense body of literature has developed to explain individual violations of expected utility in laboratory experiments and in the real world. Most of the explanations for these violations center on individual distortion of the value of the potential outcomes (utility) and the probability associated with each outcome (probability weighting). A separate literature has developed to describe personal bias when making probability assessments and forecasts. I suggest a model of belief updating based on the ndings of the judgement bias literature. This updating process has implications for decision-making under uncertainty which explain many of the expected utility violations. Some agricultural applications are suggested.

Acknowledgement 1 *I would like to thank John D. Hey for providing the data used in this paper. Also, I wish to thank David Zilberman, Brian Wright, Peter Berck, Ethan Ligon, and Rashi Glazer for guidance and comments.*

1 Introduction

Since Simon [42, 43, 44] introduced the notion of bounded rationality, there has been a drive in economics to focus not only on what people ought to do given full information, but to also describe the mechanisms that determine how they percieve the problems they face. Simon stressed the importance of modeling *procedural rationality* (or optimization given ones ability to understand and process information) rather than *substantive rationality* (global optimization). He hypothesized that including human factors in models of decision-making would substantially improve our ability to predict, understand and regulate behavior.

Psychologists have long made a study of the limitations humans face in computing, forecasting, and perceiving the world around them. The judgment bias literature (see [27] for an anthology) in particular focuses on human ability to understand, use and make probability statements. There are several common biases that occur when individuals are asked to state beliefs concerning probabilistic events. Some economists have tested for the effects of these biases in economic markets (both experimental [17, 16, 7], and real world [11]). However, little has been done to model economically when these probabilty biases will matter, or when we may expect them to be manifest.

A very separate literature has developed to explain how individuals make decisions under uncertainty. Most of this literature focuses on exposing common violations of von Neumann and Morgenstern s expected utility axioms [34] within the laboratory, and proposing and testing alternate models designed to explain these violations. This literature has grown voluminous, and the number of leading models now total well into the double digits. Few have recognized that there might be a connection between individual failure to maximize expected utility and judgment biases (one exception is Grether [17]). It seems only reasonable that if biases occur when processing probabilities, that this would hamper an individuals ability to consistently maximize expected utility. In this paper I propose a model that makes the relationship between judgment bias and decision under uncertainty explicit.

In the following section I present a short and incomplete review of the judgment bias literature (for a more complete presentation see Just [25]). Characterizing these results yeilds a set of stylized facts that any complete model of judgment bias ought to imply. In section 3 I present the limited learning model. This model is designed to describe behavior found in the judgment bias literature, and provide some explanation for the biases.

Section 4 Contains a description of the non-expected utility literature, listing many of the common violations or paradoxes found through experimental tests. I also show examples of how the limited learning model may explain each of these violations. In section 5 I review many of the common properties of the limited learning model, and argue the usefulness of a model of probability judgment in decision-making under uncertainty. Recently new objections have been raised concerning the expected utility representation of preferences [37]. In section 6 I discuss these issues and give mathematical evidence that any preference functional that has the ability to describe observed behavior must have a form similar to the limited learning model.

In section 7 I use experimental data to estimate one functional form for the limited learning model. Further, I use procedures that have become common in the non-expected utility theory literature to demonstrate the descriptive ability of the model. Finally in section 8 I make a case for why the limited learning model may be of use outside the laboratory. In particular I review some evidence that we already observe the predicted violations of expected utility theory in markets. Taking such violations and biases into account may provide a different perspective on agricultural policy measures.

2 Judgment Bias

The judgment bias literature seeks to measure the accuracy with which individuals make probability assessments. There are two concepts which jointly measure accuracy. The rst of these is calibration. An individual is well calibrated when events that he or she predicts with x % probability happen x % of the time. The second measure of accuracy is resolution. An individual displays good resolution (or knowledge) if their probability assessments are generally close to 1 or 0. There is some trade off between the two notions of accuracy. I can, of course, predict that the Dow Jones Industrial Average (DJIA) will close at 10,000 tomorrow with certainty. This prediction has very high resolution (I make a very narrow prediction), but terrible calibration (what I predict with certainty will happen with very small probability). On the other hand I could predict that the DJIA will close at some real number with certainty and always be correct. Individuals tend to be poorly callibrated especially when predicting events for which they have little experience [35]. Training has helped in some experiments, but the training does not appear to be transferable [1]. In this section I will characterize many of the biases that arise when individuals make probability statements. These biases appear to cause greater resolution at the expense of calibration.

2.1 The Law of Small Numbers

The law of small numbers is a somewhat derisive name given to the phenomenon of ignoring sample size [46]. In other words individuals when combining data to make inferences often take sample properties of very small samples to be the population properties. By not adjusting the weight of evidence to suit the sample size, too much will be inferred from small samples, and too little inferred from large samples. Of most interest, is the fact that this property is so often displayed by scientists, even within journal articles [46, 9].

The lack of adjustment for sample size suggests that there must be some sort of cost involved in correctly weighting samples in belief updating. The effects of ignoring sample size in scienti c studies include: overstated power of statistical tests, overcon dence in trends and replicability, and attempts to nd causal explanations for outlying data. Individuals making decisions may be lead to believe items are positively correlated when they are independent, and beliefs may uctuate too quickly or too slowly given the information obtained by the individual. The law of small numbers is somewhat to blame for all of the following biases, causing individuals to update using incorrect inferrences.

2.2 Representativeness

Representativeness bias occurs when individuals ignore base rates (prior informaiton) [28]. The name representativeness is suggestive of individuals believing the process that is most represented by new information to be most likely, disreguarding any prior probabilities. This process is similar to the process of maximum likelihood estimation in that prior information is ignored, and the process with the highest likelihood is considered the most probable process. This bias is one of the most studied in economics. Grether [17] and Camerer [6] have found signi cant evidence of representativeness in experimental markets. DeBondt and Thaler [11] nd evidence from New York Stock Exchange data. Essentially, when base rates are ignored, individuals beliefs will drift wildly whenever new information comes in. Arrow

[4] suggests that representativeness typifies the movements of stock market and commodities market prices.

2.3 Conservatism

Conservatism is the opposite of representativeness. When individuals display conservatism they are learning too slowly for the information they recieve. Edwards [12] finds that in some cases it will take individuals 5 iterations to do the belief updating work Bayes theory does in one. Less has been done to find conservatism in commodities markets. Conservatism would occur if individuals failed to incorporate new price information. Prices may linger too long when true values of stocks or commodities had already changed. If some individuals learn too slowly in some instances, they may create a market bubble as individuals overvalue (or undervalue) stocks given the information present in the market.

2.4 Overconfidence

Representativeness and conservatism are opposing biases in belief updating. While some have speculated the conditions under which each may occur, little is known directly about the circumstances leading to one bias or the other. A clue may be gained from observing static beliefs of individuals. Alpert and Raiffa [1] and Oskamp [35] describe a phenomenon known as overconfidence. In each of their studies, individuals are asked to assess confidence intervals (or quantiles) for many varying physical events (e.g. the number of business school Ph. D. students at Harvard). Individuals overwhelmingly stated confidence intervals that were too narrow. Support has been found for this finding among several different elicitations in varying fields (see [9] for a review) . It appears that whatever process determines the weight of prior and likelihood in belief updating has as its effect that beliefs are less diffuse than optimal Bayesian updating would suggest. Individuals tend to ignore outlying events when assessing probability, perhaps due to the cost of accounting for low probability events. This means that risk is underassessed, and individuals are overconfident about their predictions. This leads me to believe that the opposing processes of representativeness and conservatism may be due to a cost of incorporating diffuse information.

2.5 Empirical Models and the Speed of Learning

Grether [17] and Edwards [12] both use a similar modi cation of Bayes rule, dubbed generalized Bayes rule, to estimate effects of representativeness (in the case of Grether) and conservatism (in the case of Edwards). The generalized Bayes rule has the following form

$$p_{t+1}(\theta) = \frac{p_t(\theta)^r \, l(x|\theta)^{1-r}}{\int p_t(\theta)^r \, l(x|\theta)^{1-r} \, d\theta}, \tag{1}$$

where x represents the data presented to the individual, θ is the parameter to be forecasted, p_t are beliefs in period t, l is the likelihood of observaing x given the parameter θ, and r is a constant in the unit interval. Edwards rst suggested using this form when he noticed that, for a given situation, Bayesian log-likelihood ratios of stated beliefs seemed to be a linear transformation of the beliefs implied by Bayes rule. This rule has also been used by Zellner [49] in a statistical inferrence setting. While both Grether and Edwards nd the generalized Bayes rule to t beliefs well in any given question, they nd that weights differ between questions. Edwards suggests modeling the dependence of r on environmental characteristics as a promising area of research.

In separate studies much evidence has been found that the speed of learning is dependent on the diffusion (or variance) of the simuli. In fact there is now a mountain of evidence that learning is slowed beyond what is suggested by Bayes rule when new information is diffuse [19]. This suggests that we might expect r to depend on variance of the distributions p_t and l. Hogarth and Einhorn [22] also nd a dependence upon how complicated prior and likelihood information are. These observations forms the basis for the limited learning model. In the next section I present this model and discuss some of its properties.

3 The Limited Learning Model

To construct a simple model of learning, consider an individual who uses the generalized Bayes rule to combine incoming information with previous beliefs. Suppose that the individual must exert some mental effort (or cost) in combining information, and that this cost is dependent upon the complexity of the two distributions to be combined and the degree to which the individual incorporates both. Let z be the choice variable, ϵ be the random parameter the individual is learning about, and let $U(z,\epsilon)$ be the individuals Bernoulli utility function. Suppose that in this instance the weights yielding the most accurate information are according to Bayes rule, and are hence equal. The individual s true expected utility is thus given by

$$\int_{-\infty}^{\infty} \frac{U(z,\epsilon)\, p(\epsilon)\, l(\epsilon)}{\int_{-\infty}^{\infty} p(\epsilon)\, l(\epsilon)\, d\epsilon} d\epsilon - c^1(z) - c^2(p,l,r), \tag{2}$$

where $c^1(z)$ is the cost associated with choice z, and $c^2(p,l,r)$ is the cost associated with combining prior p and likelihood l using weight r. The individual however does not observe this expected utility, but his anticipated expected utility

$$\int_{-\infty}^{\infty} \frac{U(z,\epsilon)\, p(\epsilon)^r\, l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r\, l(\epsilon)^{1-r}\, d\epsilon} d\epsilon - c^1(z) - c^2(p,l,r). \tag{3}$$

Thus the individual must maximize (3) with respect to z yeilding rst order conditions

$$\int_{-\infty}^{\infty} \frac{U_z(z,\epsilon)\, p(\epsilon)^r\, l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r\, l(\epsilon)^{1-r}\, d\epsilon} d\epsilon - c_z^1(z) = 0. \tag{4}$$

Thus, so long as the functions are well behaved, any solution must have the individual meeting this rst order condition. While the individual is not able to observe the true expected utility, it may be reasonable that an individual would know from previous experience what type of tradeoffs to expect from changing the weighting, r. Even if this is a new situation and the indivual does not have this knowledge, the individual may hazard some guess as to the tradeoff and hence perform an optimization similar to the following[1]:

$$\begin{array}{ll} \max_{z,r} & \int_{-\infty}^{\infty} \frac{U(z,\epsilon)p(\epsilon)l(\epsilon)}{\int_{-\infty}^{\infty} p(\epsilon)l(\epsilon)d\epsilon} d\epsilon - c^1(z) - c^2(p,l,r) \\ \text{subject to} & \int_{-\infty}^{\infty} \frac{U_z(z,\epsilon)p(\epsilon)^r l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r l(\epsilon)^{1-r} d\epsilon} d\epsilon - c_z^1(z) = 0. \end{array} \tag{5}$$

This is similar to the rst order approach used in contract theory and is subject to the same regularity conditions for sufficiency of rst order conditions

[1] I have dropped conditioning from the likelihood function to eliminate some clutter.

(see Macho-Stadler and Perez-Castrillo [32]). The rst order conditions for (5) are given by

$$\int_{-\infty}^{\infty} \frac{U_z(z,\epsilon)\,p(\epsilon)\,l(\epsilon)}{\int_{-\infty}^{\infty} p(\epsilon)\,l(\epsilon)\,d\epsilon} d\epsilon - c_z^1(z) + \lambda \left[c_{zz}^1(z) - \int_{-\infty}^{\infty} \frac{U_{zz}(z,\epsilon)\,p(\epsilon)^r\,l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r\,l(\epsilon)^{1-r}\,d\epsilon} d\epsilon \right] = 0 \quad (6)$$

$$- c_r^2(p,l,r) - \lambda \int_{-\infty}^{\infty} \frac{U_z(z,\epsilon)\,[\ln p(\epsilon) - \ln l(\epsilon)]\,p(\epsilon)^r\,l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r\,l(\epsilon)^{1-r}\,d\epsilon} d\epsilon$$
$$+ \lambda \int_{-\infty}^{\infty} \frac{[\ln p(\epsilon) - \ln l(\epsilon)]\,p(\epsilon)^r\,l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r\,l(\epsilon)^{1-r}\,d\epsilon} d\epsilon \int_{-\infty}^{\infty} U_z(z,\epsilon)\,p(\epsilon)^r\,l(\epsilon)^{1-r}\,d\epsilon = 0 \quad (7)$$

We can combine (6) and (4) to obtain

$$\int_{-\infty}^{\infty} \frac{U_z(z,\epsilon)\,p(\epsilon)\,l(\epsilon)}{\int_{-\infty}^{\infty} p(\epsilon)\,l(\epsilon)\,d\epsilon} d\epsilon - \int_{-\infty}^{\infty} \frac{U_z(z,\epsilon)\,p(\epsilon)^r\,l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r\,l(\epsilon)^{1-r}\,d\epsilon} d\epsilon$$
$$+ \lambda \left[c_{zz}^1(z) - \int_{-\infty}^{\infty} \frac{U_{zz}(z,\epsilon)\,p(\epsilon)^r\,l(\epsilon)^{1-r}}{\int_{-\infty}^{\infty} p(\epsilon)^r\,l(\epsilon)^{1-r}\,d\epsilon} d\epsilon \right] = 0. \quad (8)$$

Equation (7) is a very interesting equation. In physics, entropy is de ned as the expected log height of a density function. Information is then de ned as negative entropy. Equation (7) displays a tradeoff between anticipated marginal utility of information (on the left of the equation) and the marginal cost of processing information. Each of the two constraint terms involve the *anticipated* (note the weighting) expected difference in log height of prior and likelihood multiplied by the marginal bene t of z. Using the de nition of information from physics, this is simply the marginal bene t of information added through incorporation of the likelihood function. So, quite literally, the model implies one must pay for understanding. Equation (8) displays the tradeoff in true expected utility and cost of action. Theoretically we could solve this model for the optimal $r = R(l,p,y)$, where y is a variable representing the environmental factors of the decision situation. This weighting function would also be dependent on the type of problem the individual might be facing and his ability to understand the tradeoffs involved. Thus we can represent beliefs as

$$p_2(x) = \frac{p(x)^{R(l,p,y)}\, l(\theta|x)^{L(l,p,y)}}{\int\limits_{-\infty}^{\infty} p(x)^{R(l,p,y)}\, l(\theta|x)^{L(l,p,y)}\, dx}, \tag{9}$$

I will refer to this as the *Limited Learning Model*. Here R (for recall) is a function of the likelihood function, the prior function, and other factors represented by y such as passage of time, interceding data or environmental facotrs; L (learning) is also a function of these three objects. The weight R is a measure of how well the individual remembers and incorporates previous information. Thus it represents primacy effects from the psychology literature, or what Simon called information storage ability. The weight L is a measure of how well an individual incorporates new information in decisions and, hence, represents recency effects, or what Simon called information processing ability.

This model differs from that proposed by Grether or Edwards in that I allow weighting of the prior and likelihood function to be dependent on context. In accordance with Hogarth and Einhorn [22] these weightings should be a function of the complication involved in processing new information and retaining old information. By allowing the weights to vary, the limited learning model may better represent true human behavior.

Note that in the Bayesian model, a prior is necessary so that learning is not perfect and immediate. In the event that no prior information is available, the individual is assumed to create a prior based on the situation and other similar experiences. Fox and Irwin [14] cite evidence that individuals prior beliefs are incorporated into information presented in the laboratory when calculating probabilistic beliefs. While this approach may challenge the abilities of formal modeling in real world contexts, it is more realistic than perfect and instantaneous learning from the rst stimulus. In other words, this is more reasonable than using only a likelihood function to represent initial learning. The use of the entire distribution is in the spirit of the rank-dependent models proposed by Quiggin [36].

3.1 Explanation

3.1.1 The Prior

A prior function is used in decision theory to represent beliefs regarding the distribution of a random variable before observing new information. In the context of this paper, it represents the beliefs of an individual regarding the distribution of money payoffs. Even in the absence of previous experience, it is likely that individuals have opinions as to how much money they are likely (or unlikely) to win in a given setting (e.g. laboratory experiment, farming, etc.). Let us focus for now on an experimental setting.

Suppose the individual is told that they will have probability $l(x)$ of gaining x, and $l(y)$ of gaining y, where $l(x) + l(y) = 1$. Upon being told the numerical probabilities, the individual tries to convert the probabilities into understanding. Understanding within the individual need not be numerically based. However, I assume that we can represent the nal understanding numerically. Within this processing, individuals bias the information toward their prior beliefs [14, 17].

The most commonly used model of information combining is that of Bayesian updating. In this case, if the individual s prior were to be known to be of the same quality as the likelihood information told to the individual, the optimal updating rule would imply the following posterior distribution:

$$p_2(z) = \begin{cases} \frac{p(x)l(x)}{p(x)l(x)+p(y)l(y)} & \text{for } z = x \\ \frac{p(y)l(y)}{p(x)l(x)+p(y)l(y)} & \text{for } z = y \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The posterior probability of any outcome in the case of two possible outcomes always lies in the interval whose endpoints are the prior and likelihood values of that outcome.

It is not likely that experimenters would give out bad information. If it could be known that experimenters were telling the truth, then prior information would be of lesser quality. In fact if it was known that the likelihood function represented truth, a weight of $R = 0$ would be given to all prior information under optimal updating. This yields

$$p_2(z) = \begin{cases} l(x) & \text{for } z = x \\ l(y) & \text{for } z = y \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

In other words, an optimal updating of beliefs, given truth of the information observed, implies beliefs are identical to the truth. Alternatively, knowing that the prior is truth leads to the likelihood receiving a weight of zero:

$$p_2(z) = \begin{cases} p(x) & \text{for } z = x \\ p(y) & \text{for } z = y \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

In this way Bayesian updating can represent the biasing of beliefs toward some prior beliefs. It is important that Bayesian updating alters probabilities based on the outcome associated with a probability. Two outcomes with the same probability will not necessarily be distorted in the same way. The distortion of each will depend on the context of other possible outcomes and the weights given them.

3.1.2 Non-Optimal Weighting

While an econometrician can weight based on the size of data sets involved in the prior and likelihood functions, individuals may not know the quality of information they are told. Beyond not knowing the quality, there are also limitations in one s ability to process information, which were described previously. For these reasons it is unreasonable to assume that individuals will always behave optimally when combining prior beliefs and incoming information.

These limitations mean that low quality priors may be too prevalent in posterior beliefs, or low quality information may prevail against high quality prior beliefs. Hogarth and Einhorn [22] suggest that the principle determinant of whether recency or primacy effects will be observed is the relative degree of difficulty in maintaining a prior and processing incoming information.

This means that instead of basing the weighting of prior and likelihood solely on quality as in the statistical model, or having a static weighting as Zellner suggests, the weighting appears to be a function of how complicated the information they represent are to understand. In this way beliefs will be closer to the prior distribution of payoffs when information to be learned is in some way complicated to process. When the likelihood functions of two lotteries are similar, distortions arising from inefficient use of information may cause seeming violations of expected utility theory. This is an argument similar to Buschena and Zilberman [5] and Rubenstein [40].

Other factors may also affect the weighting of information, such as reliability of the source, familiarity with similar information, experience with similar situations, or something as simple as typeface used to write the information. Schultz [41] arguments about education in decision-making suggest that educational factors may be appropriate to include in any weighting function.

3.1.3 Measurement of Complication

In order to identify which distributions are more heavily weighted in information processing, we need a measure of how complicated a distribution is to understand. In the absence of direct observation of complication, some instrument will be necessary. This measure will form the basis of the weighting functions.

One obvious choice is the variance of the distribution. A distribution with widely varying outcomes often appears complicated and will likely be harder to interpret than very tightly dispersed distributions. In all of the analysis of this article I will use this measure, as it is widely understood, and an intuitive measure. Also, Haruvy et. al [19] cites substantial evidence that variance tends to slow learning in experimental and eld studies.

Variance has weaknesses as a measure of complication, however. In particular small probabilities of large monetary outcomes are likely viewed as complicated and outside the realm of updating experience (see the example in the next section). Yet gambles with this description would have low variance. A better measure would be *entropy*. Entropy was introduced through the eld of physics as a representation of the amount of disorganization in a distribution. Literally, negative entropy is interpreted as information. The formula for the entropy of a density $f(\cdot)$ is

$$H(f(x)) = -\int \log\left(\frac{f(x)}{m(x)}\right) f(x)\, dx, \tag{13}$$

where $m(\cdot)$ is a weighting function. If $m(x) = 1$, then this is just negative the expected log height of the density. The higher the expected log height, the more concentrated the distribution is. By incorporating a weighting function, entropy can give more importance to certain ranges of payoffs in determining how complicated a distribution. In this way distributions with low concentration around extreme payoffs may also be included as complicated distributions. Entropy is also preferable as it measures the degree of concentration anywhere in the distribution, and not just the concentration about the mean.

4 Non-Expected Utility

In this section I demonstrate the usefulness of the limited learning model of decision-making and compare it with several known expected utility violations. In order to be a preferred tool for modeling at least two requirements must be met. First, the model must explain available data. Second it must offer a simpler or more plausible explanation. In this problem, the latter may correspond to fewer departures from conventional theory of decision-making under uncertainty if individuals begin to approximate true probabilities in decisions, given experience.

To demonstrate the ability of this model to describe the data consider the following functions:

utility:

$$u(x) = 1 - e^{-.0001x} \tag{14}$$

prior:

$$p(\Delta x) = \begin{cases} \frac{1}{7000} - \frac{\Delta x}{(7000)^2} & \text{if } x \geq 0 \\ \frac{1}{7000} + \frac{\Delta x}{(7000)^2} & \text{if } x \square 0 \end{cases} \tag{15}$$

weighting:

$$R\left(\sigma_l^2, \sigma_p^2\right) = \begin{cases} 9.0909\frac{\sigma_l^2}{\sigma_p^2} & \text{if } \frac{\sigma_l^2}{\sigma_p^2} < .0022 \\ .02 + 17.3077\left(\frac{\sigma_l^2}{\sigma_p^2} - .0022\right) & \text{if } .0022 \square \frac{\sigma_l^2}{\sigma_p^2} < .0074 \\ .11 + 9.2457\left(\frac{\sigma_l^2}{\sigma_p^2} - .0074\right) & \text{if } .0074 \square \frac{\sigma_l^2}{\sigma_p^2} < .0306 \\ .3245 + .5\left(\frac{\sigma_l^2}{\sigma_p^2} - .0306\right) & \text{if } .0306 \square \frac{\sigma_l^2}{\sigma_p^2} \end{cases} \tag{16}$$

$$L\left(\sigma_l^2, \sigma_p^2\right) = 1 - R\left(\sigma_l^2, \sigma_p^2\right).$$

where u is utility, x is wealth, σ_p^2 and σ_l^2 are the variances of the prior and likelihood function, respectively. Figure 1, 2 and 3 display these graphs over the relevant intervals.

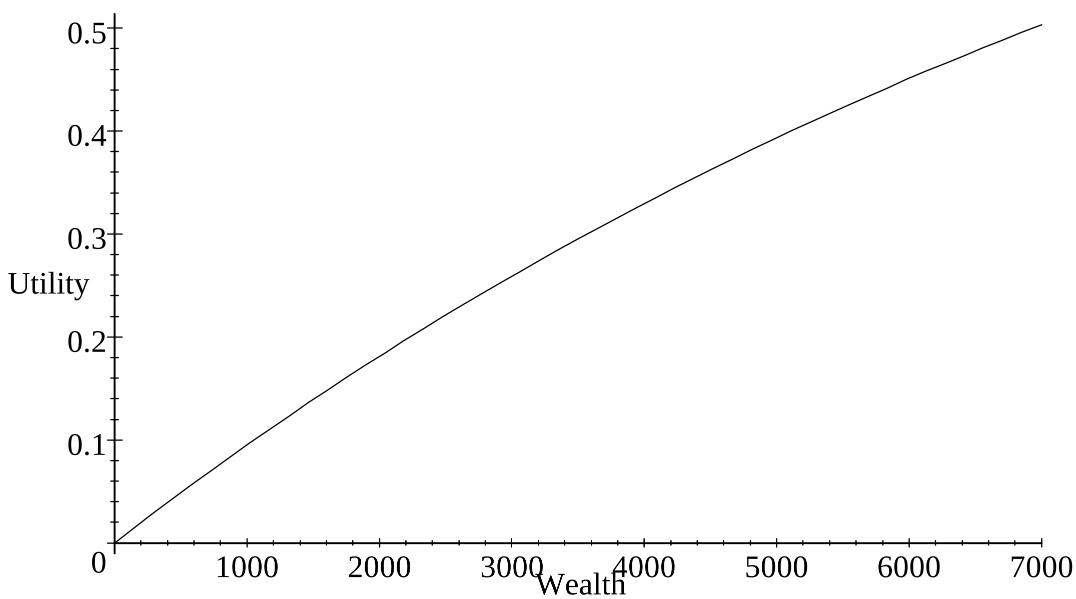

Figure 1: Utility as a Function of Wealth

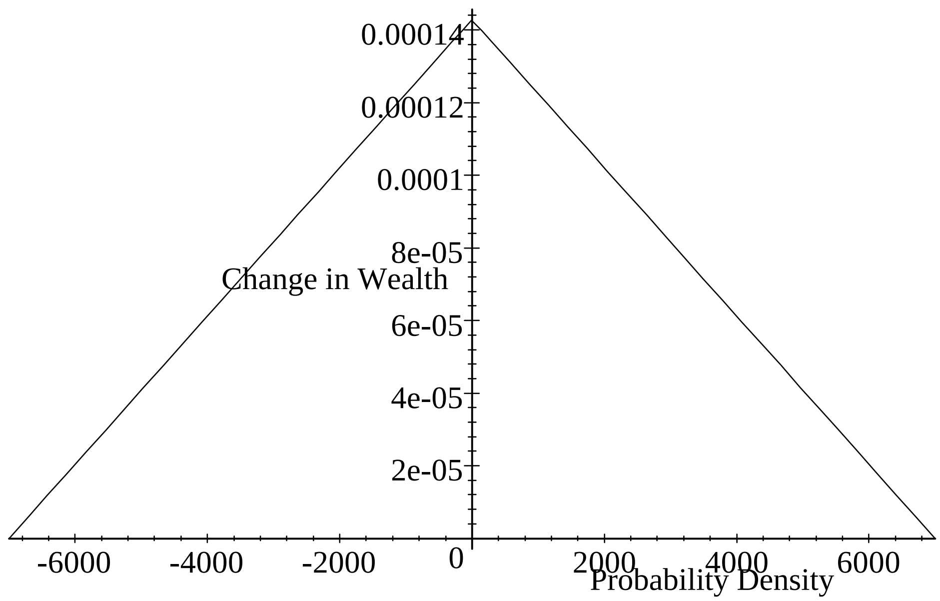

Figure 2: Prior Density Function

The utility function has constant absolute risk aversion of .0001.[2] It is important that the individual may display risk aversion (even over losses) in this framework. One of the major theses of non-expected utility theory is that individuals are risk loving over losses. Hence, the use of value functions that are convex below current wealth. This value function explains the what is called the reflection effect. My model suggests, however, that the utility function may be smooth, always concave, and still display the reflection effect. The use of a symmetric prior distribution creates the desired result.

I have used a tent-shaped distribution in this numerical example for simplicity and plausibility. Individuals may enter the experiment with some belief that their current wealth is not likely to change much (hence 0 is the modal value of the prior), and that they are about as likely to lose some amount x as they are to win x. Although I have not shown it in general, symmetry of the prior is all that is required to create the reflection effect in all examples discussed here.

I have used four connected line segments to describe a weighting function for the prior distribution. I derived this function by exploring the values that could explain the modal observations of the problems described in this section. While other functions (in particular smooth functions) could also describe the data, I consider this to be the simplest representation. It appears in Figure 3 that for a small standard deviation of wealth, individuals behave more rationally, relying mostly on the perfect information they receive. Very close to 0, the function R is convex.

When the standard deviation of monetary outcomes is larger, individuals depend more on their prior knowledge as Hogarth and Einhorn [22] would suggest. The function is concave in variance of wealth for larger values. Of particular interest is how the function changes slope around a standard deviation of $500. While any amount of uncertainty has a dramatic effect on the ability to process information relative to certainty, the rate of departure from super-rationality (or unconstrained rationality) decreases for larger variance of wealth.

This weighting function describes why individuals would give special

[2] This may seem like a very small degree of risk aversion. My reason for using such a small absolute level of risk aversion is purely to avoid problems with numerical accuracy. Higher absolute risk aversion creates a function that has slope zero over part of the relevant range with the accuracy of Matlab. Arrow [3] suggests that individuals should have relative risk aversion near 1, which would occur with this utility function if individuals have wealth of $10,000. All analysis is conducted with initial wealth of $10,000.

salience to certainty when making decisions, as 0 is the only standard deviation that appears to elicit a completely rational response (even if R were larger at certainty the individual would not consider the prior in weighting because a likelihood function that is only positive at one point will result in all posterior weight being placed on that point).

I will now proceed to demonstrate how these parameters potentially describe the responses observed by Kahneman and Tversky. I demonstrate their problems 1, 2 and 10, the Ellsberg paradox, a violation due to Loomes, and then list the expected utilities from all numerical problems found in Kahneman and Tversky [26]. For all examples assume a current wealth of $10,000. It is important to remember that I am not attempting to nd true parameters and functional forms. This is only an example to show that individuals may try to maximize expected utility under learning-recall limitations.

4.0.4 Problem 1 and 2:

Kahneman and Tversky [26] induced the Allais paradox by asking individuals to choose between pairs of lotteries as follow:

Problem 1: Choose between

	$2500 with probability .33		
A:	$2400 with probability .66	B:	$2400 with certainty.
	$0 with probability .01		

Problem 2: Choose between

C:	$2500 with probability .33 $0 with probability .67	D:	$2400 with probability .34 $0 with probability .66.

They found that 82% of 72 respondents chose B in Problem 1, and 83% chose C in Problem 2. A simple calculation shows that these answers violate expected utility theory. Choosing B implies that $.33u(2500) + .01u(0) < .34u(2400)$, while choosing C implies that $.33u(2500) + .01u(0) > .34u(2400)$ an obvious contradiction.

While the prior I have assumed is positive over an interval, the likelihood function is zero at all values other than $2500, $2400, and $0. These values are , $p(2500) = 9.1837 \times 10^{-5}$, $p(2400) = 9.3878 \times 10^{-5}$, and $p(0) = 1.4286 \times 10^{-4}$. The likelihood function is as represented in Problem 1 above. The variance of the likelihood function and prior density are $\sigma_l^2 = 60819$ and $\sigma_p^2 = 8.1667 \times 10^6$ respectively. Thus the values of the weighting functions are $R\left(\sigma_l^2, \sigma_p^2\right) = .1104$ and $L\left(\sigma_l^2, \sigma_p^2\right) = .8896$, and the posterior density is

$$p_2(x) = \begin{cases} \frac{p(2500)^R l(2500)^{1-R}}{p(2500)^R l(2500)^{1-R} + p(2400)^R l(2400)^{1-R} + p(0)^R l(0)^{1-R}} & \text{if} \quad x = 2500 \\ \frac{p(2400)^R l(2400)^{1-R}}{p(2500)^R l(2500)^{1-R} + p(2400)^R l(2400)^{1-R} + p(0)^R l(0)^{1-R}} & \text{if} \quad x = 2400 \\ \frac{p(0)^R l(0)^{1-R}}{p(2500)^R l(2500)^{1-R} + p(2400)^R l(2400)^{1-R} + p(0)^R l(0)^{1-R}} & \text{if} \quad x = 0 \\ 0 & \text{otherwise} \end{cases}$$
$$= \begin{cases} .3444 & \text{if} \quad x = 2500 \\ .6395 & \text{if} \quad x = 2400 \\ .0161 & \text{if} \quad x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

The value of the utility function at the relevant values are $u(2500 + 10000) = .7135$, $u(2400 + 10000) = .7106$, and $u(0 + 10000) = .6321$. The expected utility of option A is $EU = .3444 \times .7135 + .6395 \times .7106 + .0161 \times .6321 = .7103$. The expected utility of option B is just $u(2400 + 10000)$, which is .7106. Thus, the individual who behaves as my assumed model suggests would choose option B as did most of the respondents.

The variance of the likelihood function in choice C of problem 2 is $\sigma_l^2 = 1381875$. The weights suggested by this example are $R\left(\sigma_l^2, \sigma_p^2\right) = .3938$ and $L\left(\sigma_l^2, \sigma_p^2\right) = .6062$. The posterior distribution is

$$p_2(x) = \begin{cases} .3536 & \text{if} \quad x = 2500 \\ .6464 & \text{if} \quad x = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

The expected utility from choice C is $EU = .3536 \times .7135 + .6464 \times 6321 = .6609$. The variance for choice D is $\sigma_l^2 = 1292544$. The weights are $R\left(\sigma_l^2, \sigma_p^2\right) = .3883$ and $L\left(\sigma_l^2, \sigma_p^2\right) = .6117$. The posterior distribution for D is

$$p_2(x) = \begin{cases} .3615 & \text{if} \quad x = 2400 \\ .6385 & \text{if} \quad x = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

and expected utility is $EU = .3615 \times .7106 + .6385 \times .6321 = .6605$. Thus, this individual would choose option C, as did the majority of the respondents.

The changes in probabilities are summarized in gures 4 through 6. Note in the tables, and in later discussions, that I have scaled the prior density values of possible outcomes to sum to one. Without using a model of learning based on prior experience, the two choices made here would contradict traditional theory. The increased probabilities for a \$2400 or \$2500 change in wealth are due to the fact that the ratio of probabilities of \$0 and \$2400 and \$2500 are not quite so extreme in the prior as in the likelihood function. More important is that the prior drew the probabilities of \$2400 and \$2500 closer together. The distortion of all probabilities other than certainty in this systematic way makes it plausible that individuals try to maximize expected utility yet still give answers that appear to contradict expected utility theory.

4.0.5 Problem 10

It appears that individuals have a hard time understanding compound lotteries. Consider the following problem due to Kahneman and Tversky [26], where individuals must make their choice before any stage is played.

> Problem10: Consider the following two stage game. In the rst stage there is a probability of .75 to end the game without winning anything, and a probability .25 to move into the second stage. If you reach the second stage you have a choice between
>
> A: \$4000 with probability .80, \$0 with probability .20   B: \$3000 with certainty.

By reducing these compound lotteries, A $= (4000, .20, 0)$ and B $= (3000, .25, 0)$. This is exactly the choice given in Kahneman and Tversky s problem 4. In that problem, 65% of 95 sunjects chose C, which corresponds to choice A here. Of 141 respondents 78% chose B in problem 10. This is similar to the response given in their problem 3, which is identical to the second stage of problem 10. In other words individuals ignored the rst stage when choosing between A and B. Responding as in problem 3 is in accordance with expected utility theory, however the differing response to problem 4 makes this a puzzle for non-expected utility.

Problem 10 is a special case because it involves a compound lottery. Using the learning model on the second-stage lottery A probabilities yields a posterior distribution as follows:

$$p_2(x) = \begin{cases} .7568 & \text{if} \quad x = 4000 \\ .2432 & \text{if} \quad x = 0. \end{cases} \tag{19}$$

Represent the posterior probability of reaching the second stage as ρ. Then the posterior utility of choice C is $EU = \rho(.7568 \times .7534 + .2432 \times .6321) + (1-\rho)(.6321) = .6321 + .0918\rho$. The expected utility of choice D is $EU =$

ρ.7275. This individual will choose D if $\rho \geq .9943$. This result lends some credence to the notion that individuals ignore (or massively distort) probabilities of intermediate stages in compound lotteries. This was hypothesized by Kahneman and Tversky, among others.

4.0.6 The Ellsberg Example

The Ellsberg Paradox speaks to the way in which individuals assign probabilities to events for which they have not been presented probabilities. This paradox was discovered by Ellsberg [13], and can be displayed in the following example.

Example 2 *Suppose I have two urns each containing 100 balls. These balls are either white or black. Urn 1 contains 49 white balls and 51 black balls. Urn 2 has an unspeci ed distribution of balls. The subject is told she will win $1000 if a white ball is drawn rst, and $1000 if a black ball is drawn second. The subject is then asked which earns should be used for the two draws.*

Using expected utility theory, the expected utility of urn 1 in the rst draw is $.49U(1000)$, and that of urn 2 is $p_w U(1000)$, where p_w is the subjective probability that a white ball is drawn from urn 2. Most subjects choose urn 1 for the rst draw, implying that $p_w < .49$. This means that the probability of drawing a black ball from urn 2, $p_b = 1 - p_w > .51$. If this is the case then individuals should request that the second ball be drawn from urn 2. This is overwhelmingly contradicted by the experimental evidence [13]. It appears from this paradox that individuals may not assign probability to events for which none has been speci ed, or rather that they may adjust their beliefs to the situation. Some have dubbed these sorts of problems *ambiguity.*

The problem of the Ellsberg example can be resolved by de ning the prior over winnings instead of number of balls. In each case $p(1000) = .4615$ (scaled so $p(0) + p(1000) = 1$). The weight implied by the distribution of balls is $R\left(\sigma_l^2, \sigma_p^2\right) = .3245$ and $L\left(\sigma_l^2, \sigma_p^2\right) = .6755$. Hence the choice of urn 1 in the rst draw can be represented as

$$p_2(x) = \begin{cases} .4807 & \text{if} \quad x = 1000 \\ .5193 & \text{if} \quad x = 0. \end{cases} \tag{20}$$

In the second draw the choice of urn 1 can be represented as

$$p_2(x) = \begin{cases} .4942 & \text{if} \quad x = 1000 \\ .5058 & \text{if} \quad x = 0. \end{cases} \tag{21}$$

The choice of urn 2 in either draw can be represented as

$$p(x) = \begin{cases} .4615 & \text{if} \quad x = 1000 \\ .5385 & \text{if} \quad x = 0. \end{cases}, \tag{22}$$

which is stochastically dominated by both of the posterior distributions for urn 1. This explanation of behavior relies on which cues the individual truly cares about. When making choices facing ambiguity, it is not likely the person has a feeling as to how many balls the experimenter wants to put in an urn of varying colors. It is much more likely that the individual has beliefs about how much money the experimenter is willing to give away, or rather the experimenters desire to lower costs.

4.0.7 Loomes New Violation

Loomes [30] found a new violation of the independence axiom. This violation is somewhat more severe in that no model proposed to date can account for this violation [9]. To illustrate this violation consider the following example based on Loomes [30].

Example 3 *In each of the next two problems A and B are amounts of money which you may choose, but which cannot sum to more than £20. Problem 1: There is a .60 probability of recieving amount A, and a .4 chance of recieving amount B. Problem 2: There is a .30 probability of recieving amount A, a .20 probability of recieving amount B and a .50 chance of recieving nothing.*

This problem is interesting because it is a test of probability ratio violations demonstrating that the behavioral problem is more complex than simple probability weighting can account for. Of 60 subjects, 31 placed a greater amount of money in A for problem 1 than for problem 2, 24 put equal amounts in each and 1 put less in A for problem 1. Suppose the individual behaves so as to maximize $\sum_i \pi(p_i) U(x_i)$, where p_i is the porbability of outcome x_i. This is a common class of models called probability weighting models. Expected utility is a special case of this model. It is commonly assumed that $\pi(.5) = .5$, that π is concave on $(0, .5)$ and convex on $(.5, 1)$. This means that individuals overweight low probabilities and underweight high probabilities. It is also commonly assumed that $\sum_i \pi(p_i) = 1$ for any particular problem. In either case the individual must then solve the following problem

$$\max_{A^i} \pi\left(p_A^i\right) U\left(A^i\right) + \pi\left(p_B^i\right) U\left(B^i\right) \text{ subject to } A^i + B^i = 20 \tag{23}$$

where I have assumed without loss of generality that $U(0) = 0$. The rst order conditions for this problem can be written as

$$\frac{U'(A^i)}{U'(B^i)} = \frac{1-\pi(p_A^i)-\pi(p_0^i)}{\pi(p_A^i)}, i = 1, 2 \tag{24}$$

$$A + B = 20. \tag{25}$$

This assumes an interior solution or equivalently a certain degree of risk aversion. Of the 24 who placed the same amount in each problem 3 placed all the money in A in both gambles. These were the only corner solutions observed in these problems suggesting that if our model is correct the individuals must be risk averse. Note that the probabilities in problem 2 are exactly half those in problem 1, meaning that

$$\frac{U'(A^1)}{U'(B^1)} = \frac{1-\pi(.6)}{\pi(.6)} \tag{26}$$

$$\frac{U'(A^1+\epsilon)}{U'(B^1-\epsilon)} = \frac{.5-\pi(.3)}{\pi(.3)}. \tag{27}$$

Expected utility is only consistent with $\pi(p) \equiv p$. Using the expected utility weighting function, these condiitons are obviously violated for any concave utility function. More can be said, however. If U is concave, the values for A must be higher than those for B (because A recieves more probability weight). This is true of the overwhelming majority of subjects. Thus it must be that $U'(B^1-\epsilon) < U'(B^1) < U'(A^1) < U'(A^1+\epsilon)$. This in turn implies that

$$\frac{.5-\pi(.3)}{\pi(.3)} > \frac{1-\pi(.6)}{\pi(.6)}, \tag{28}$$

which implies that $\pi(.3) < .5\pi(.6)$. This and other data suggest that probability weighting functions perform as if globally convex in this type of experiment. This is important because convex probability weighting functions would fail to explain the majority of data accumulated through choice experiments. Use of a prior could explain this violation of the independence axiom.

Using a prior that enters conditionally dependent upon prizes and distributions can also explain the independece violation that Loomes [30] discovered. If we modify the amount involved in his example, so that individuals can now allocate \$2000[3] instead of £20, we can observe that the speci cation

[3] The amount must be modi ed to t the computational resolution of my model, and computer.

of my model predicts the behavior Loomes discovered. None of the other problems dealt with amounts so small as £20, and infact the speci cation of the limited learning model in this paper always predicts allocating the full amount to choice A when amounts are so small. This is due to the small Arrow-Pratt absolute risk aversion of the utility function. Using the larger amount induces an interior solution. Using a more concave utility function would produce results similar to Loomes for smaller total gambles. It is somewhat difficult to solve the problem Loomes suggests using the limited learning model because the amounts in event A and B enter in to the variance functions. Taking derrivatives is difficult. Hence I used a computer simulation to locate the amount for A that maximizes expected utility according to the limited learning model. For problem 1, the model suggests $A = \$1117$. For problem 2 $A = 1002$. This is exactly the direction of bias discovered by Loomes. Hence this puzzle may be due to the limitations individuals have upon their understanding probability statements relative to their prior beliefs.

4.0.8 Expected Utilities

Table 1 displays results for all of the money value examples given by Kahneman and Tversky. By maximizing expected utility after generalized Bayesian updating, all of the modal values from their study are explained.

5 Prior and Probability Bias

In this section I will explore why the limited learning model is able to describe the behavior in the problems above, and give some rationale for why it is a reasonable model of decision under uncertainty. First, it will be important to introduce the Marschak-Machina triangle (also called the unit simplex) pictured in Figure 7. The unit simplex is used to graph indifference curves over lotteries all involving three possible outcomes (in this case \$6000, \$400 and \$0). Probability of the lowest value outcome is plotted on the horizontal axis, and probability of the highest outcome is plotted on the vertical axis. The remaining probability is given to the middle value outcome. Expected utility requires that all indifference curves be straight parallel lines. Many of the expected utility violations that are documented occur because individuals display indifference curves that are neither straight lines (called a violation

Problem	Dollar Amount	Scaled Prior[4]	Posterior	Likelihood	Expected Utility
Problem 1 A	$2500	.2795	.3444	.3300	.7103
	$2400	.2857	.6395	.6600	
	$0	.4348	.0161	.0100	
Problem 1 B	$2400	1	1	1	.7106
Problem 2 C	$2500	.3913	.3536	.3300	.6609
	$0	.6087	.6464	.6700	
Problem 2 D	$2400	.3966	.3615	.3400	.6605
	$0	.6034	.6385	.6600	
Problem 3 A	$4000	.3000	.5855	.8000	.7031
	$0	.7000	.4145	.2000	
Problem 3 B	$3000	1	1	1	.7275
Problem 3′ A	-$4000	.3000	.5855	.8000	.5262
	$0	.7000	.4145	.2000	
Problem 3′ B	-$3000	1	1	1	.5034
Problem 4 C	$4000	.3000	.2432	.2000	.6616
	$0	.7000	.7568	.8000	
Problem 4 D	$3000	.3636	.2940	.2500	.6602
	$0	.6364	.7060	.7500	
Problem 4′ C	-$4000	.3000	.2432	.2000	.5881
	$0	.7000	.7568	.8000	
Problem 4′ D	-$3000	.3636	.2940	.2500	.5943
	$0	.6364	.7060	.7500	
Problem 7 A	$6000	.1250	.1555	.4500	.6579
	$0	.8750	.8445	.5500	
Problem 7 B	$3000	.3636	.8617	.9000	.7143
	$0	.6364	.1383	.1000	
Problem 7′ A	-$6000	.1250	.1555	.4500	.5851
	$0	.8750	.8445	.5500	
Problem 7′ B	-$3000	.3636	.8617	.9000	.5212
	$0	.6364	.1383	.1000	
Problem 8 C	$6000	.1250	.0013	.0010	.6323
	$0	.8750	.9987	.9990	
Problem 8 D	$3000	.3636	.0022	.0020	.6323[5]
	$0	.6364	.9978	.9980	
Problem 8′ C	-$6000	.1250	.0013	.0010	.6317
	$0	.8750	.9987	.9990	
Problem 8′ D	-$3000	.3636	.0022	.0020	.6318
	$0	.6364	.9978	.9980	
Problem 11 A	$1000	.4615	.4875	.5000	.6492
	$0	.5385	.5125	.5000	
Problem 11 B	$500	1	1	1	.6501
Problem 12 C	-$1000	.4615	.4875	.5000	.6133
	$0	.5385	.5125	.5000	

of betweenness) or parallel (called a violation of independence). The slope of the indifference curves are related to the coefficient of absolute risk aversion (approximated by the three points on the utility curve) with a greater slope indicating greater risk aversion [31].

Some effort has been made to document where the violations occur within the triangle (see [8, 9, 21] among others). Camerer [9] sums up the evidence by citing that most violations occur near the edges of the triangle (or outside the ellipse drawn in figure 8). This result has a high degree of relevance for the model I suggest. In particular, the ellipse drawn on figure 8 is an iso-variance curve. Within the ellipse, variance of outcomes is higher than outside of the ellipse. The fact that fewer violations occur where variance is higher, suggests that expected utility is more closely observed within this portion of the triangle. Outside of this ellipse, indifference curves are not parallel, or straight in a way that is significant. This suggests that in this region of the triangle, individuals are distorting probability in some way.

The hypothesis of the limited learning model, is that when individuals are faced with the lotteries presented commonly in experiments, they have a hard time believing in small variance gains, and hence will bias the small variance gambles *toward some prior beliefs.* While other models of choice under uncertainty have allowed some form of Baeyesian prior [10], I have found none that have recognized it as such. Further, most models make probability weights contingent on probabilities, and not the value of the prize itself. I will discuss why this is an unsavory property in the following section. In any case, it seems unreasonable that individuals might bias their beliefs toward some constant probabilities that ignore outcomes available. Much more reasonable is that this bias would be similar to those found in the judgment bias literature. If individuals are biasing their beliefs in this way, then manifestations of non-expected utility theory in the real world are already identified. Using most of the models supposed by non-expected utility, theorists have a hard time finding real world application or use. Figure 9 shows graphs of indifference curves under expected utility, variance and indifference curves under the limited learning model. The utility function and weighting functions used were those specified in the previous section. The indifference curves here have exactly the properties that Camerer [9] describes: the indifference curves reflect different levels of risk aversion, they fan in and out in a systematic way, they are not straight lines, they are more nearly parallel in the middle of the triangle, and curves for gains and losses reflect around the 45 degree line. Figure 9 above satisfies all of these conditions except re ection for gains and losses.

In order to accomplish this re ection affect all models up until now have assumed that individuals use different probability weighting functions over gains and losses (see for example [45]). Within the framework of the limited learning model, it may be accomplished simply by assuming a prior that decreases in probability density as possible gains (or losses) are further from zero. An illustration of this using the speci cation of the limited learning model in the previous section can be found in gure 10. These are very similar to the indifference curves estimated by Tversky and Kahneman [45].

Using the weighting of a prior has another desirable property that cannot be acheived using two weighting functions, one for losses, one for gains. It has been observed that the shape of indifference curves depend not only on the sign of the outcomes, but on the size of the outcomes as well [29, 21]. Further, probability weighting functions appear to change shape depending on the size of prizes that are being considered [29]. Figure 11 displays limited learning model probability weights for various prize levels. These properties cannot be explained by the theories that dissallow a prior type weighting function. In particular, prospect theory and rank dependent models do not allow for these shapes to change depending on the size of the prize (only on the rank of the prize). Further, updating with a prior using a constant weight results in straight line indifference curves, which can be rejected in statistical tests [7].

This rudimentary exploration of the properties of the limited learning model re ects favorably on the model. It appears that the model provides a clear and plausible psychological explanation of the facts that have been learned through non-expected utility experiments. It also provides a reasonable explanation of judgment biases and learning problems that individuals often display. Further, by allowing weights to depend upon other factors, such as presentation and elicitation, the limited learning model may be used to explain the individual variation in response to identical questions with the presentation altered. In these categories the limited learning model appears to be a reasonable alternative to the leading non-expected utility models. In the next section I will argue why any preference functional should have a form similar to the limited learning model.

6 Concavity

Most economists are now aware of the growing evidence against expected utility theory. Among the more recent objections are those criticizing concavity of the utility function as the only mechanism for risk aversion. Matthew Rabin [37, 38] is chief among the critics. While not the rst to notice a discrepancy (see Hansson [18]). Rabin points to the difference in estimated risk aversion levels observed over small gambles, and those over larger gambles. His argument stems from a calibration theorem, which he proves. A useful and illustrative corollary of this theorem is given below [38].

Corollary 4 *Suppose that for all* w, $U'(w) > 0$ *and* $U''(w) < 0$. *Suppose there exists* $g > l > 0$ *such that, for all* w, $.5U(w-l) + .5U(w+g) < U(w)$. *Then for all positive integers* k, *and all* $m < m(k)$, $.5(w-2kl) + .5U(w+mg) < U(w)$, *where*

$$m(k) = \begin{cases} \frac{\ln\left[1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i}\right]}{\ln\left(\frac{l}{g}\right)} - 1 & \textit{if}\ \ 1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i} > 0 \\ \infty & \textit{if}\ \ 1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i} \square\ 0. \end{cases} \quad (29)$$

For a statement of the theorem and a proof of both the theorem and the corollary presented above see the appendix of Rabin [38]. This corollary allows us to compare risk behavior over even-chance bets (commonly called 50-50 bets) under the assumption of risk aversion. For example, a person who will always turn down a lottery with a .5 probability of winning $110 and .5 probability of losing $100 (from now on (.5, 100, .5,-100)) will also always turn down the lottery represented by (.5, 2090, .5, -800). Expected utility assumes that concavity is the only explanator of risk attitude, and that individuals are approximately risk neutral for small gambles, meaning that individuals should accept fair bets if they are small enough. The problem with this assumption, is that when we observe someone turning down a small fair bet, we must assume that this is due to concavity of the utility function.

Unless the utility function changes from concave to convex as prizes get larger, outrageous behavior is implied. It is easy to con rm from the corollary above that if a global risk averter turns down (.5, 125, .5, -100), they will always turn down any bet with a .5 chance of losing $600, no matter how large of a gain may be had with the remaining .5 probability. Bernoulli introduced expected utility theory as a way to explain why individuals might not be willing to pay in nite amounts of money to play gambles with in- nite expected gains (e.g. the St. Petersburg Paradox). Variants on the St. Petersburg Paradox will sometimes yield an in nite certainty equivalent unless the utility function is bounded, implying eventual risk aversion for large enough gambles [33]. It is this same property, only in the small, that Rabin exploits in arguing the inconsistency of traditional expected utility theory. Even if there are non-convexities in the utility function, a bounded utility function will imply similarly ridiculous behavior if we compare to large enough positive gains.

To see this last point consider the lottery $(.5, x, .5, -100)$. Suppose without loss of generality, that $U(-100) = 0$, and $\lim_{x\to\infty} U(x) = 1$. The certainty equivalent of our lottery is given by $U(CE) = .5U(x) \square .5(0+1) = .5$. This means that an individual will turn down any bet with a 50% chance of losing \$100 for some xed amount of money, no matter what the possible gains. This also means that if you have wealth greater than $U^{-1}(.5)$ you will never choose to take any bet that involves a 50% chance of a loss of \$100.[7] If we limit ourselves to the notion that all risk aversion is due to concavity of utility, we must accept that either (1) more wealthy individuals are less willing to risk losses, or (2) individuals are willing to pay in nite amounts of money for some class of lotteries with in nite average payoffs (constructed similar to the St. Petersburg Paradox). I will call this the boundedness problem.

Alternatively, some theorists have pointed out that risk averse behavior may be due to misperceptions of probability, or transforming probabilities before expected utility optimization. Yaari [48] proposed a model where all risk behavior was due to a warping of probabilities rather than of monetary values. More commonly models are proposed that involve the optimization of some function

$$V(F) = \int \pi(p)\, U(x+w)\, dF(x), \tag{30}$$

where F is the distribution of possible payoffs, x are possible payoffs, w is current wealth, U is a utility function over payoffs and π represents a probability weighting function. This probability weighting function provides an alternate explanation for risk attitudes in the small. However, a probability

[7] Note, given that the graph of utility is connected, and that utility is bounded and monotone increasing, there must be some level of wealth, w, for which $U(w-100) = \frac{1}{2}\lim_{x\to\infty} U(w+x)$.

weighting function of this form cannot solve the problems that Rabin describes. For the remainder of this section I will prove an analog of Rabin s callibration theorem for the form (30) above, show that in order to avoid the boundedness problem we must introduce probability and outcome weights, or weights that are a function of x, and nally show that no analog of the callibration theorem can be proved for such a preference functional.

6.1 Probability Weights Can t Explain Risk Aversion Either

In generalizing Rabin s theorem, I wish to make it applicable to a wide range of models. I will use the notation $U(x|w)$ to mean the utility of a gain of x given the wealth level w. This allows for preference functionals such as those speci ed by Kahneman and Tversky [26]. The theorem can then be stated.

Theorem 5 *Suppose that for all* z, $U(z|w)$ *is strictly increasing and weakly concave. Suppose that there exists* $\overline{\phi} > \underline{\phi}$, $g > l > 0$ *such that for all* $\phi \in \left[\underline{\phi}, \overline{\phi}\right]$, $\pi(.5)\,U(z-l|w) + \pi(.5)\,U(z+g|w) < U(\phi|w)$ *where* $z = U^{-1}\left(\frac{U(\phi|w)}{2\pi(.5)}|w\right)$. *Then for all* $x \in U^{-1}\left(\frac{U\left(\left[\underline{\phi},\overline{\phi}\right]|w\right)}{2\pi(.5)}|w\right)$, *for all* $x > 0$

1. *If* $g < 2l$ *then* $U(z|w) - U(z-x|w) \geq$

$$\begin{cases} 2\sum_{i=1}^{k^*(x)} \left(\frac{g}{l}\right)^{i-1} r(z|w) & \text{if } \; z - \underline{z} + 2l \geq x \geq 2l \\ 2\sum_{i=1}^{k^*(z-\underline{z}+2l)} \left[\left(\frac{g}{l}\right)^{i-1} r(z|w)\right] & \text{if } \; z > z - \underline{z} + 2l. \end{cases} \tag{31}$$

2. $U(z+x|w) - U(z|w) \; \square$

$$\begin{cases} \sum_{i=0}^{k^{**}(x)} \left(\frac{l}{g}\right)^{i} r(z|w) & \text{if } \; x \; \square \; \overline{z} \\ \sum_{i=0}^{k^{**}(\overline{z})} \left(\frac{l}{g}\right)^{i} r(z|w) + \left[x - \overline{z}\right] \left(\frac{l}{g}\right)^{k^{**}(\overline{z})} r(z|w) & \text{if } \; x \geq \overline{z}, \end{cases} \tag{32}$$

where, $k^*(x) \equiv \left\lfloor \frac{x}{2l} \right\rfloor$, $k^{**}(x) \equiv \left\lfloor \frac{x}{g} + 1 \right\rfloor$, *and* $r(z|w) \equiv U(z|w) - U(z-l|w)$.

A proof of this theorem is straightforward given Rabin s theorem, and is provided in the appendix. A corollary similar to that of Rabin may also now be proven.

Corollary 6 *Suppose that for all z, $U'(z|w) > 0$ and $U''(z|w) < 0$. Suppose there exists $g > l > 0$ such that for any $\phi \in \left[\underline{\phi}, \overline{\phi}\right]$, $\pi(.5)\, U(z - l|w) + \pi(.5)\, U(z + g|w) < U(\phi|w)$ where $z = U^{-1}\left(\frac{U(\phi|w)}{2\pi(.5)}|w\right)$. Then for all positive integers k, $\forall m < m(k)$, $\pi(.5)\, U(z - 2kl|w) + \pi(.5)\, U(z + mg|w) < U(\phi|w)$, where*

$$m(k) \equiv \begin{cases} \frac{\ln\left[1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i}\right]}{\ln\frac{l}{g}} - 1 & \text{if } 1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i} > 0 \\ \infty & \text{if } 1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i} \;\square\; 0. \end{cases} \tag{33}$$

Again this corollary follows directly from Rabin s corollary, and a proof may be found in the appendix. Let us consider the lottery (.5,325,.5,100). If an individual would be willing to turn down this lottery for some certain amount $\phi = U^{-1}(2\pi(.5)\, U(200|w)\,|w)$, then the individual would turn down any bet with a 50% chance of winning \$100 for ϕ. If $\pi(p) = p$, then $\phi = 200$. For other functions π, ϕ will depend on both the value of $\pi(.5)$ and the shape of U. To see this relationship note that

$$U(\phi|w) = 2\pi(.5)\, U(z|w), \tag{34}$$

and by totally differentiating we nd that the following relationship holds

$$\frac{d\phi}{d\pi(.5)} = \frac{2U(z|w)}{U'(\phi|w)}. \tag{35}$$

This last expression requires that the certain amount ϕ depend positively on the weight of $\pi(.5)$. Most theories require that $\pi(.5) \;\square\; .5$ and then again ϕ is bounded above by 200. Alternatively, if $\pi(.5) > .5$, we can set an upper bound on ϕ, as all probability weighting functions require that $\pi(p) \;\square\; 1$. If $\pi(p) = 1$ then $U(\phi|w) = 2U(200|w)$, so that $\phi = U^{-1}(2U(200|w))$. We know that the inverse utility function is convex, hence we may place an upper bound on ϕ by using a Taylor expansion around ϕ. Thus $\widetilde{\phi} - \frac{\partial U^{-1}\left(U\left(\widetilde{\phi}\right)\right)}{\partial\widetilde{\phi}}\left(\frac{1}{2}U\left(\widetilde{\phi}\right)\right) = 200$ implies that $\phi - 200 \;\square\; \frac{U\left(\widetilde{\phi}\right)}{2U'\left(\widetilde{\phi}\right)}$. This limit is conceivably quite high. Most theories will require that $\pi(.5)$ be close enough to .5 that we need not consider this upper bound. In all cases we will nd that some xed amount of money will be preferred to any gamble with a .5 probability of a gain, a result nearly identical to that found by Rabin leading him to reject expected utility theory in favor of prospect theory. An extension of the above theorem to prospect theory (withought rank dependent probability weights) is straightforward.

6.2 Boundedness and Probability-outcome Weights

Expected utility theory can make outrageous predictions when utility is unbounded. Consider a gamble with payoffs having probability density function $f(x)$ for $x = 0, 1, 2, ...$ Hence, this is a lottery that can be respresented by the sequence $\{x_i, f(x_i)\}_{i=0}^{\infty}$. Expected utility of this gamble is $EU(\{x_i, f(x_i)\}_{i=0}^{\infty}) = \sum_{i=0}^{\infty} f(x_i)\, U(x_i)$. Suppose that $f(0) = 1 - \epsilon$, and $f(\widetilde{x}) = \epsilon$ with the probability assigned to every other outcome equal to 0. If $U(x)$ is assumed not to be bounded and monotonically increasing, then for any amount of money y, it is possible ot nd $\widetilde{x}$ such that such that $U(y) < (1 - \epsilon)\, U(0) + \epsilon U(\widetilde{x})$ no matter how small $\epsilon > 0$. This is at the heart of the problem with unbounded utility. To see how this causes problems let us consider a sequence of the following form, $f(0) = 1 - \epsilon$, $f(x_i) = \frac{1}{2^i}\epsilon$, $i \in \{1, 2, ...\}$. Select x_i such that $x_0 = 0$, and $U(x_i) \geq \frac{2^i}{\epsilon}$ for $i > 0$. Then the expected utility of this gamble would be $EU(\{x_i, f(x_i)\}_{i=0}^{\infty}) = \sum_{i=0}^{\infty} f(x_i)\, U(x_i) = (1 - \epsilon)\, U(0) + \frac{1}{2}\epsilon U(x_1) + \frac{1}{2^2}\epsilon U(x_2) + ... \geq (1 - \epsilon)\, U(0) + 1 + 1 + 1 + ... = \infty$. This just demonstrates that any unbounded utility function will imply that a gamble exists with as small probability as we would like on very large amounts such that the value of the gamble is in nite.

If we were to examine the same problem with a probability weighting function, the preference functional can be represented by $V(\{x_i, f(x_i)\}_{i=0}^{\infty}) = \sum_{i=0}^{\infty} \pi(f(x_i))\, U(x_i)$. First consider again the lottery $(.5, x, .5, -100)$. Suppose without loss of generality, that $U(-100) = 0$, and $\lim_{x\to\infty} U(x) = 1$. The certainty equivalent of our lottery is given by $U(CE) = \pi(.5)\, U(x)$ □ $\pi(.5)\, U(0 + 1) = \pi(.5)$. This means that an individual will turn down any bet with a 50% chance of losing \$100 for some xed amount of money, no matter what the possible gains. This also means that if you have wealth greater than $U^{-1}(\pi(.5))$ you will never choose to take any bet that involves a 50% chance of a loss of \$100. This is the same problem observed with expected utility. If we suppose that π is continuous,[8] monotonic and that $\pi(0) = 0$, $\pi(p) > 0$ for $p > 0$, then we can see that the Petersburg paradox is still in force when the utility function is unbounded.[9]

[8] Continuity is not necessary for the result, but allows for much simpler exposition. Alternatively, you could assume there was a second function that was continuous, positive, monotonic and everywhere less than or equal to π. Using this function in place of π in the following discussion would imply the result.

[9] Violating these assumptions on π causes extreme problems with non-monotonicity. For example there would be some ϵ, such that the lottery $(100, 1)$ would be prefered to

Suppose that $f(0) = 1-\epsilon$, and $f(\widetilde{x}) = \epsilon > 0$. The probability assigned to every other outcome equal to 0. If $U(x)$ is assumed not to be unbounded and monotonically increasing, then for any amount of money y, it is possible to find $\widetilde{x}$ such that such that $U(y) < \pi(1-\epsilon)U(0) + \pi(\epsilon)U(\widetilde{x})$ no matter how small $\epsilon > 0$. Consider a sequence of the following form, $f(0) = 1-\epsilon$, define ξ so that $\pi(1-\epsilon) = 1-\xi$. Then define $f(x_i) = \pi^{-1}\left(\frac{1}{2^i}\xi\right), i \in \{1, 2, ...\}$. This inverse must exist by continuity and monotonicity. Select x_i such that $x_0 = 0$, and $U(x_i) \geq \frac{2^i}{\xi}$ for $i > 0$. Then the expected utility of this gamble would be $V\left(\{x_i, f(x_i)\}_{i=0}^{\infty}\right) = \sum_{i=0}^{\infty} \pi(f(x_i))U(x_i) = (1-\xi)U(0) + \frac{1}{2}\xi U(x_1) + \frac{1}{2^2}\xi U(x_2) + ... \geq (1-\xi)U(0) + 1 + 1 + 1 + ... = \infty$. This again demonstrates that any unbounded utility function will imply that a gamble exists with as small probability as we would like on very large amounts such that the value of the gamble is infinite.

Finally consider a probability weighting function $\pi(p_i, x_i)$ increasing in p_i and decreasing in x_i (at least for large enough x_i)[10] such that for every sequence $\{x_i, p_i\}_{i=0}^{\infty}$ with p a proper density and $\lim_{i\to\infty} x_i = \infty$, there exists some k, $V\left(\{x_i, p_i\}_{i=0}^{\infty}\right) = \sum_{i=0}^{\infty} \pi(p_i, x_i)U(x_i) = k$. In order to find sufficient conditions for such a weighting function, let us arrange this sequence according to x_i in ascending order (note we can combine any identical amounts so that each x_i is distinct). We know (because $\sum_{i=0}^{\infty} p_i = 1$) that there exists p_n such that $p_i < p_n$ whenever $i > n$.[11] We know that for any fixed n, $\sum_{i=0}^{n} \pi(p_i, x_i)U(x_i)$ converges to some constant. Then we must only examine the properties of $\sum_{i=n}^{\infty} \pi(p_i, x_i)U(x_i)$, where $p_i < p_n$ and x_i increasing monotonically. We know this must converge if $\sum_{i=n}^{\infty} \pi(p_n, x_i)U(x_i)$ converges. Let $z_n = x_{n,}\ z_i = \max(x_i, z_{i-1}+1)$. One condition that accomplishes convergence requires that there exist some p such that for any $p_i < p$ there exists an x_i with $\pi(p_i, z(x)) < \frac{1}{z(x)^2 U(x)}$ whenever $x > x_i$. To see this note that if $\pi(p_i, z_i) \square \frac{1}{z_i^2 U(x_i)}$, then $\sum_{i=n}^{\infty} \pi(p_i, z_i)U(x_i) \square \sum_{i=n}^{\infty} \frac{1}{z_i^2}$ which converges. This sufficient condition only requires that for small enough probability weight larger prizes will induce some form of disbelief. For prizes that are large enough, the belief will decrease fast enough to disallow infinite expected utilities and hence there is no lottery inducing infinite willingness to pay.

$(100, 1-\epsilon, 10000, \epsilon)$

[10] Here I only consider positive outcomes. With negative outcomes a similar result can be proved.

[11] One of the basic requirements for convergence of a series is that $\lim_{i\to\infty} p_i = 0$.

6.3 Callibration Theorems and Probability-outcome Weights

Using a probability weighting function of this form also removes the problems Rabin points out. For instance, consider an individual with strictly concave utility function that will turn down the lottery (.5,125,.5,-100). The subjective expected utility would then be given by $\pi(.5,125)\,U(125,w)+\pi(.5,-100)\,U(-100,w)<U(0,w)$. Without loss of generality, let $U(-100,w)=0$, and $U(125,w)=1$. Then turning down the lottery requires only that $\pi(.5,125)<U(0,w)$. Concavity only requires that $U(0,w)\geq\frac{4}{9}$. If $\pi(.5,125)<\frac{4}{9}$ then we have placed no restrictions on the concavity of the utility function. By not requiring utility to be concave in these small gambles, we do not place any extra restrictions as in Rabin s corollary. For instance, in this case it may be that the utility curve is a straight line. Then any gamble, $(.5,x,.5,-y)$ will have expected value $\pi(.5,x)\frac{x+100}{225}+\pi(.5,-y)\frac{100-y}{225}$. The individual would then always accept the gamble if $x>\left(\frac{4}{9}+\frac{100-y}{225}\pi(.5,-y)\right)\frac{225}{\pi(.5,x)}-100$. Thus x will not be in nite if $\pi(.5,x)\geq 0$. This example demonstrates that a probability-outcome weighting function will not be susceptible to the same problems as expected utility and probability weighting functions. The arguments of this section are intended to make the case that a probability-outcome weighting function may be the only reasonable weighting function available for economic modeling. In the next section I will argue the empirical superiority of the limited learning model, a model employing a weighting function meeting the requirements of a probability-outcome weighting function.

7 Estimation and Data

For estimation I used data gathered experimentally by Hey and Orme [20]. This data consists of responses of 80 individuals to a total of 200 questions. The subjects were asked 100 of the questions on each of two days. Each of the questions asked the individual to choose between two lotteries (denoted left anf right), or state that they were indifferent. One of the lotteries was chosen at random, ex post, to be played for real money. Money amounts involved were £0, £10, £20, £30. Probabilities were measured in eighths. This data set was chosen for its comprehensiveness, and because my results may be easily compared with those of Hey and Orme (among others who have used this data set).

Hey and Orme compared 10 different models of choice under uncertainty, estimating ordered probit models, and ranking based on the Akaike Information Criterion. They also compared nested models using likelihood ratio tests. They paid special consideration to the statement of indifference. Their model of choice was as follows

$$y = V(L,R) + \epsilon, \tag{36}$$

where $V(L,R)$ is the value of the left hand lottery, L, over the right hand lottery, R. This value will be positive if L is prefered and negative if R is prefered. The variable y is unobservable, but we observe the choice of lottery l, r, or c (for left, right and center, or indifference). So the response has the following form

$$X = \begin{cases} l & \text{if} \quad \tau < V(L,R) \\ c & \text{if} \quad -\tau < V(L,R) < \tau \\ r & \text{if} \quad V(L,R) < -\tau \end{cases}. \tag{37}$$

In this section I will use Hey and Orme s techniques and model to test the limited learning model against nine of the models used by Hey and Orme. I wish also to compare their model of indifference to a strict interpretation (i.e. that $\tau = 0$). The models used in Hey and Orme include risk neutrality (RN), expected utility (EU), disappointment aversion (DA), prospective reference (PR), quadratic utility (QU), regret theory with dependence (RD), regret theory with independence (RI), rank dependent weights with power functional form (RP), rank dependent weights with Quiggin s functional form (RQ), and Yaari s dual model (YD). I will not estimate RI, or YD because of poor performance in Hey and Orme s tests, and to reduce clutter. For a complete discussion I refer readers to Hey and Orme s section on these models. I use the following form for the limited learning model

$$R\left(\sigma_l^2\right) = b\left(\ln\left(\sigma_l^2 + 1\right)\right), \tag{38}$$

$$L\left(\sigma_l^2\right) = 1 - R\left(\sigma_l^2\right) \tag{39}$$

$$p_0(x) = e^{tx}. \tag{40}$$

Here the parameters to be estimated are b and t. In addition, I must also estimate the utility levels, $U(10)$, $U(20)$, $U(30)$. I have set $U(0) = 0$. It would require too much space to present all results of the tests within this paper, but in the following I will present many tests similar to those run by Hey and Orme.

	RN	EU	DA	PR	QU	RI	RP	RQ	WU	LLM
Data Set 1	7.51	3.20	3.69	4.05	5.70	7.35	3.21	2.04	4.55	3.70
Data Set 2	7.61	2.90	3.93	2.99	5.75	7.64	3.71	2.76	3.96	3.75
Data Set 3	8.48	4.11	4.42	3.16	4.66	5.98	3.75	3.05	3.31	4.06

Table 2: Average Ranking using AIC

	RN	EU	DA	PR	QU	RI	RP	RQ	WU	LLM
Data Set 1	8.00	3.00	3.00	2.50	6.00	8.00	3.00	2.00	5.00	4.00
Data Set 2	9.00	2.50	4.00	2.00	6.00	8.00	3.50	2.00	4.00	4.00
Data Set 3	9.00	5.00	4.50	3.00	6.00	7.00	4.00	3.00	3.00	4.00

Table 3: Median Ranking using AIC

7.1 Comparing to Hey and Orme

I estimated the 10 models using maximum likelihood, estimating a separate functional for each individual, just as Hey and Orme. This procedure was conducted for three separate data sets, responses to questions from the rst session (100), responses to questions in the second session (100), and responses from questions in both sessions (200). The Akaike Information Criterion (AIC) is a Bayesian approach to ordering models of this type. The speci c loss function used to construct the AIC in this context is [2]

$$AIC = -\frac{2 \log L(\widehat{\alpha})}{T} + \frac{2k}{T}, \tag{41}$$

where $L(\widehat{\alpha})$ is the likelihood function evaluated at it s optimum, T is the number of observations, and k is the number of parameters. The AIC was used to rank each model for each of the 80 individuals in the experiment. Table 2 contains the mean rank of each model for each data set, and 3 the median rank.

The LLM places near the middle using this type of ranking within each of the data sets, while RQ and PR appear to do the best. Hey and Orme also compare the models to expected utility using likelihood ratio tests. Table 4 contains the number of individuals for each of the repective models, for which the data reject the hypothesis of expected utility behavior in favor of each of other models at the .01 level. This test was only conducted using data set 3.

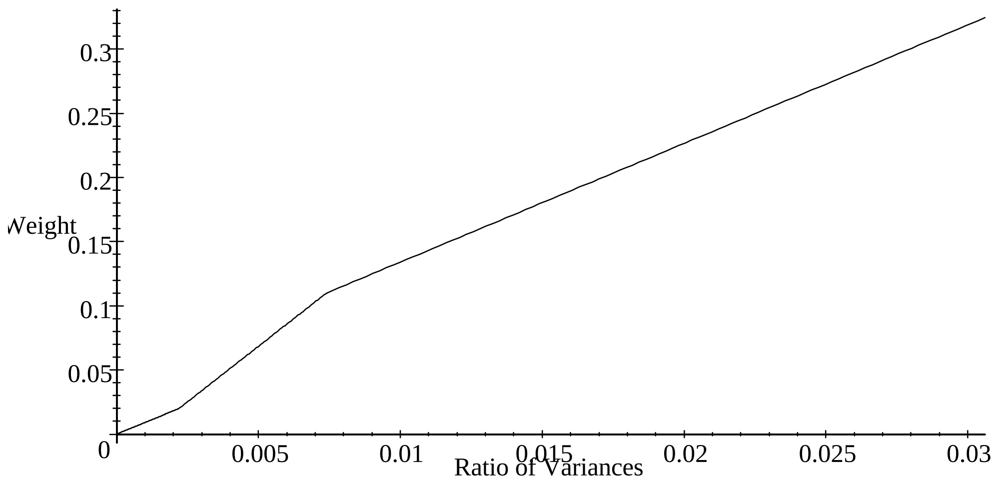

Figure 3: Weighting of the Prior Density

DA	PR	QU	RI	RP	RQ	WU	LLM
34	33	25	9	35	44	30	38

Table 4: The Number Rejecting EU using Likelihood Ratio at the .01 Level

	RN	EU	DA	PR	QU	RI	RP	RQ	WU	LLM
Data Set 1	1.41	1.12	1.10	2.47	1.15	1.39	1.09	1.05	1.12	1.08
	(.31)	(.41)	(.40)	(2.70)	(.36)	(.36)	(.40)	(.39)	(.37)	(.38)
Data Set 2	1.29	.94	.93	1.45	1.00	1.23	.95	.91	.93	.91
	(.31)	(.42)	(.38)	(1.81)	(.33)	(.35)	(.39)	(.38)	(.33)	(.36)
Data Set 3	1.36	1.04	.99	1.02	1.00	1.08	.99	1.01	.98	.99
	(.30)	(.45)	(.37)	(.38)	(.37)	(.39)	(.37)	(.40)	(.37)	(.39)

Table 5: Average AIC and Standard Deviation for each Decision Model

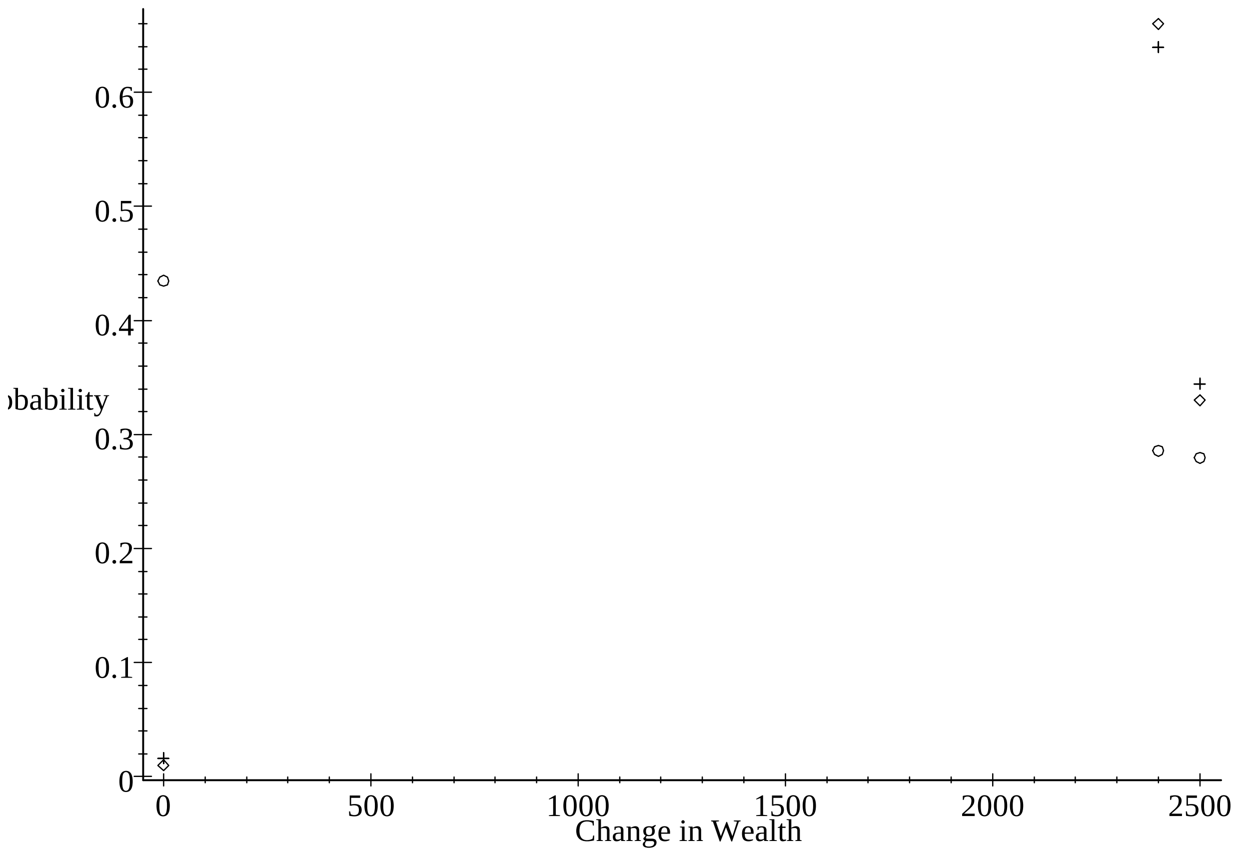

Figure 4: Problem 1 A: circles represent prior probabilities, diamonds represent the likelihood function, and crosses the posterior.

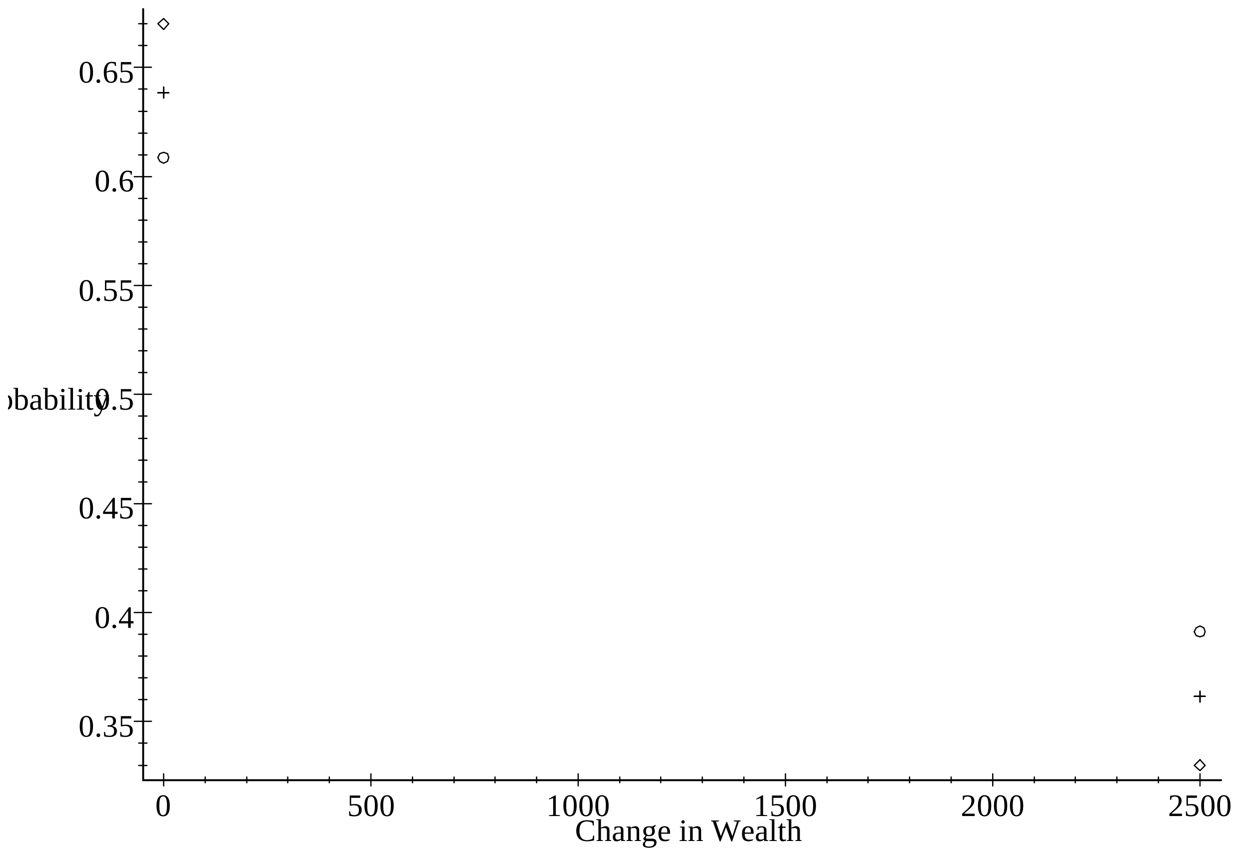

Figure 5: Problem 2 C: circles represent prior probabilities, diamonds represent the likelihood function, and crosses the posterior.

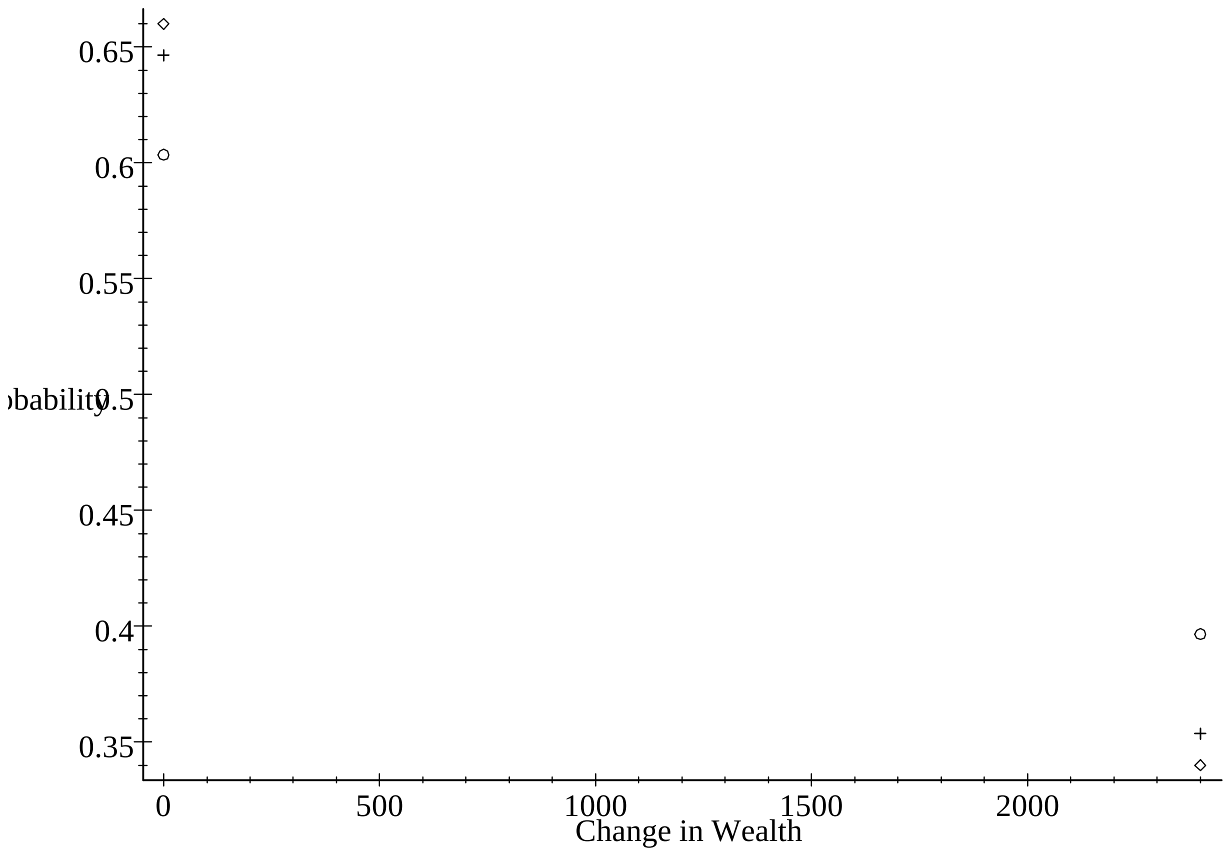

Figure 6: Problem 2 D: circles represent prior probabilities, diamonds represent the likelihood function, and crosses the posterior.

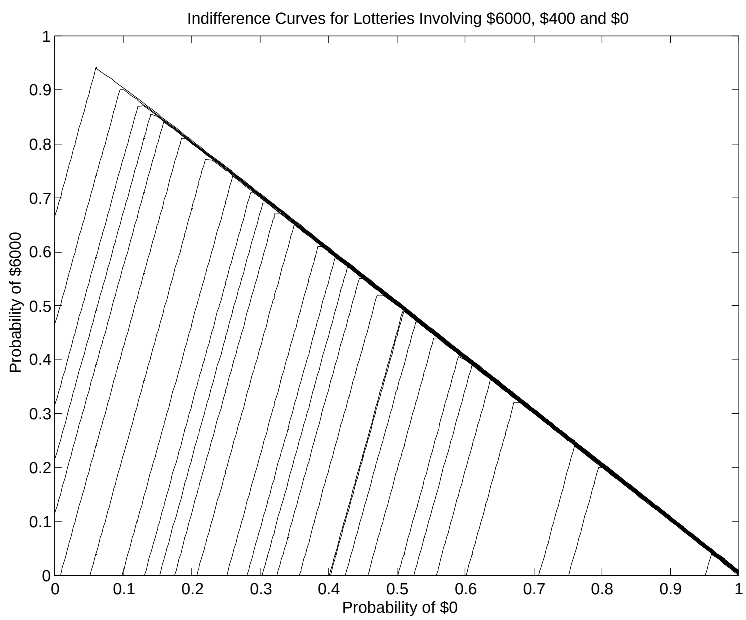

Figure 7: The Marschak-Machina Triangle

Table 5 presents the average AIC for each model, with the standard deviation in parenthises. The average AIC tells quite a different story. Here it appears that many of the models that rank highly, have large errors for many individuals. In particular, PR does poorly in the smaller samples. LLM, does well in this respect, having both a low average AIC, and little variation. The standard deviations for each model is large enough to suggest that it would be hard to statistically differentiate between many of these models. Hey and Orme also use a combination of the likelihood ratio test and the AIC to nd what they call the best model to explain each individuals behavior. For each individual if only one model is signi cantly better than EU based on the likelyhood ratio test, this model is designated best. If more than one is signi cantly better, then the model among those that are signi cantly better with the lowest AIC is called best. If none are signi cantly better, then the model with the lowest AIC is called best. Table 6 presents the number of individuals for which each of the models was best. Again the LLM does not appear to perform well using this measure, while PR and RQ appear to dominate.

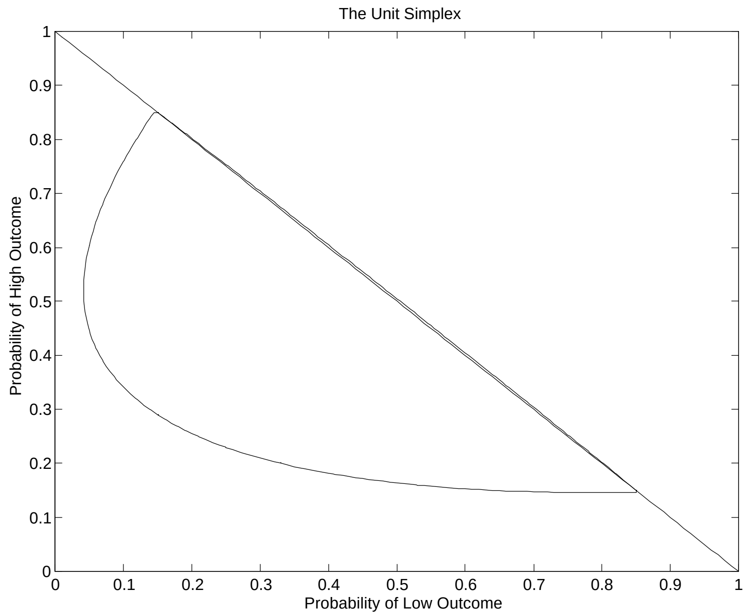

Figure 8: Violations are Concentrated Outside the Ellipse

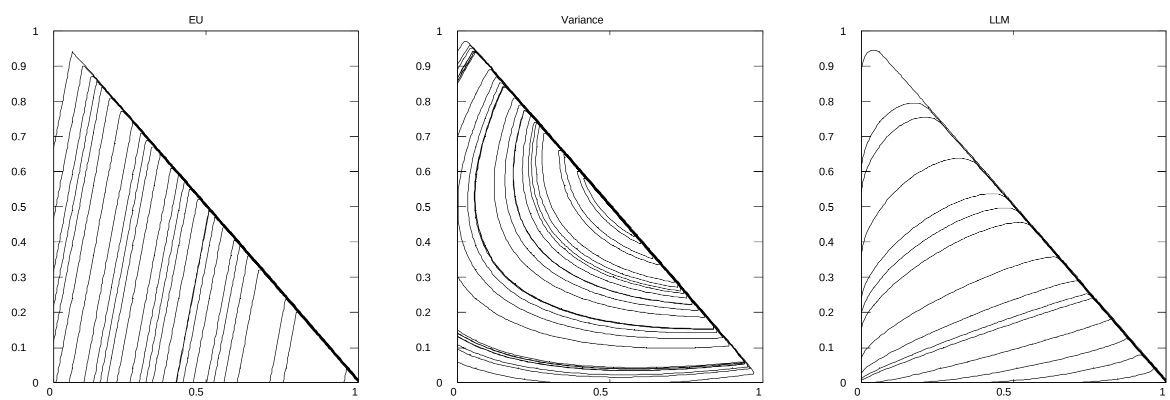

Figure 9: Marschak-Machina Triangle

DA	PR	QU	RI	RP	RQ	WU	LLM
9	20	8	2	12	15	6	8

Table 6: The Number of Individuals for which Each Model was Best

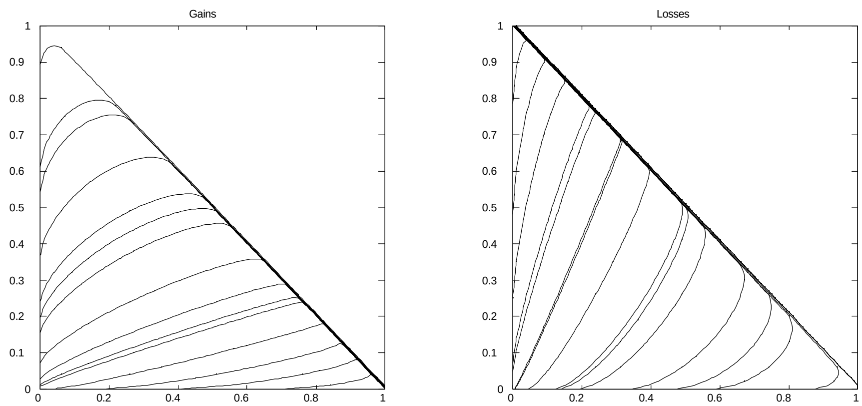

Figure 10: LLM for gains (or losses) of $600, $400, and $0.

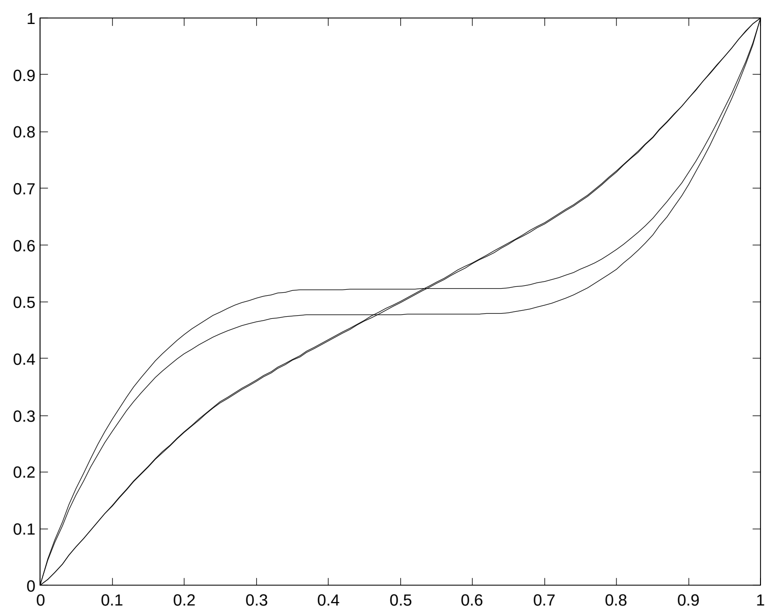

Figure 11: Probability Weighting Functions for Various Prize Amounts using LLM

	RN	EU	DA	PR	QU	RI	RP	RQ	WU	LLM
Parameters	2	4	5	5	10	7	5	5	6	6
Data Set 1	.64	.76	.77	.71	.79	.70	.77	.78	.78	.78
	(.09)	(.10)	(.10)	(.18)	(.09)	(.12)	(.10)	(.09)	(.09)	(.09)
Data Set 2	.67	.81	.81	.80	.82	.74	.81	.82	.81	.82
	(.13)	(.11)	(.11)	(.14)	(.09)	(.12)	(.11)	(.09)	(.09)	(.10)
Data Set 3	.64	.77	.77	.71	.79	.72	.77	.79	.78	.79
	(.11)	(.10)	(.10)	(.18)	(.09)	(.11)	(.10)	(.08)	(.09)	(.09)

Table 7: Prediction Rates for Each Decision Model

Within the rest of this subsection I will present other measures of performance that might be useful in assessing these models of decision under uncertainty. While the tests Hey and Orme use focus on the individual, it seems unlikely econometricians will be able to apply a separate functional for each individual in a data set. Further, even if different models could be used for each individual, it would be impossible to tell a priori which model t with which individual. This means pre-testing would need to be done, diminishing the ability to subsequently test any hypotheses. A different approach is to test the hypothesis that all individuals were conforming to one model of decision-making. I will rst present some summary statistics, and then conduct two tests.

Table 7 presents the prediction rates for each of the models with standard deviation of rates within parenthases. The prediction rates are slightly better for QU, RQ and LLM, but the standard deviations are such that it would be hard to distinguish between many of these models statistically. The prediction rates seem to be higher for the second data set, although the same questions were asked in each of the sessions (only with left and right switched). This might re ect some sort of learning taking place between the sessions.

Now consider the hypothesis that each individual is behaving according a speci c model, with individual speci c parameters. This may be tested by treating the data set as $200 \times 80 = 16000$ observations, and $80 \times k$ parameters to be estimated in each model. By assuming uninformative prior odds, it is possible to derive posterior odds for the model speci ed by all maximum-likelihood parameters of LLM versus each other model. The correct form for this is just the ratio of likelihood functions. The results are presented in table 8. Here ∞ denotes that the denomonator was small enough to cause the

EU	RN	DA	PR	QU	RI	RP	RQ	WU
∞	∞	3.1×10^{111}	∞	2.2×10^{6}	∞	1.3×10^{141}	1.8×10^{-6}	7.8×10^{180}

Table 8: Posterior Odds in Favor of LLM

EU	RN	DA	PR	QU	RI	RP	RQ	WU	LLM
1.09	1.36	1.02	2.45	1.02	1.26	1.03	.99	1.05	.99

Table 9: AIC for Models of All Individuals

odds to exceed the measurement capability of my computer. According to these numbers, the LLM is prefered to all but RQ. Care must be taken when interpreting these odds, as they do not account for the number of parameters in each model, treating each model as a xed stochastic process which might have generated the observations.

A better way to assess these behavioral models may be to use the AIC to rank the models ability to describe all individuals behavior. Here the AIC is given by the following formula

$$AIC = -\frac{2l}{16000} - \frac{80k}{16000}, \tag{42}$$

where l is the sum of log likelihood functions for all individuals evaluated at their respective maxima, and k is the number of parameters in the model describing each individual. The AIC for each model is displayed in table 9. When taking into account the number of parameters, I nd that LLM, and RQ perform the best. As was noted earlier, RQ is often violated when comparing choices of large outcome prospects to those of small. With data covering a wider range of outcomes, it may be easier to distinguish between these two models.

7.2 A Strict Interpretation

Maximum likelihood estimation is very sensitive to threshold estimates such as that of τ in the above estimations. Beyond this there are procedural questions about using responses of indifference. In particular, there is no way to give incentives to respond with indifference when the person is truly indifferent. For this reason it might be instructive to examine results from estimation using differing methods. In particular I compare four different models: the ordered probit model of Hey and Orme, one that assigns all

	RN	EU	DA	PR	QU	RI	RP	RQ	WU	LLM
Right	8.56	4.08	4.76	2.80	4.64	4.50	3.91	2.93	3.80	5.01
Left	8.51	4.08	4.83	2.88	4.86	4.36	3.98	2.81	4.06	4.64
Random	8.51	4.04	4.73	2.86	4.84	4.32	3.98	2.88	3.93	4.91
Dropped	8.54	3.83	4.81	2.81	4.98	4.48	3.81	3.08	3.91	4.76
Threshold	8.48	4.11	4.42	3.16	4.66	5.98	3.75	3.05	3.31	4.06

Table 10: Average Ranking using AIC for Several Indifference Methods

	RN	EU	DA	PR	QU	RI	RP	RQ	WU	LLM
Right	9.00	4.00	5.00	3.00	5.00	4.50	4.00	2.00	4.00	5.00
Left	9.00	4.00	5.00	3.00	6.00	5.00	4.00	2.00	4.00	5.00
Random	9.00	4.00	5.00	2.50	6.00	4.00	4.00	2.00	4.00	5.00
Dropped	9.00	3.00	5.00	3.00	6.00	4.50	4.00	2.50	4.00	5.00
Threshold	9.00	5.00	4.50	3.00	6.00	7.00	4.00	3.00	3.00	4.00

Table 11: Median Ranking using AIC for Several Indifference Methods

responses of indifference to the lottery presented to the right, one that assigns all responses of indifference to the lottery presented to the left, one that randomly assigns (.5 probability) to each of the lotteries presented, and one that drops all indifferent observations. In tables 10 and 11 I present the mean and median ranking according to AIC using each of these methods for data set 3. It appears that the median AIC is particularly sensitive to the use of τ. Perhaps further research should disallow indifference between two lottery choices in order to reduce estimation error.

7.3 Conclusion

I have used a combination of the methods employed by Hey and Orme, and other techniques to demonstrate the viability of LLM as a behavioral model. While the model performs about average using the techniques of Hey and Orme, their techniques are designed to select a different model for each individual, and then select as best the model that is chosen most often. This sort of voting can cause the same problems in model selection as individual voting can cause in welfare analysis. In other words it may ignore the model that explains everyone s behavior, only to select the model that best explains a minority of individuals behavior. Econometricians are unlikely to have the freedom this sort of testing implies in model selection. If non-expected utility models are to come into wide use, we need to have a single model that describes all individuals (or at least most) over wide ranges of events. It is unlikely that designing models to t experimental data alone will produce such a model. Tests for the best overall model suggest that both LLM and RQ models performed best describing behavior within this experiment. Experiments of this nature may be more helpful if the range of prizes is varied more widely, however, limitations on research funds almost guaruntee this will never happen. For this reason, it may be best to test models that have been designed to explain experimental observations outside of the laboratory. This will most certainly increase the power of the tests, and may make obvious which models more closely conform to actual behavior in the aggregate.

8 Applications

There are many important areas of agricultural economics dealing with uncertainty, making it likely that perception of uncertainty will play an important role in modeling many agricultural decisions [24]. In this section I will present two scenarios where I think the limited learning model may be of particular importance: crop insurance and production contracts. Within each of these models I will assume that actors have accurate perceptions of underlying uncertainty, until some change occurs. The reaction to the change will depend upon how well one is able to accurately update ones beliefs. This ability to make optimal use of information should be dependent upon the resources available to the rm in question [47]. The next two examples will illustrate the importance of human capital in this context.

8.1 Crop Insurance

Suppose that farmers have pro t function in any year given by

$$y = \overline{\pi} + g\epsilon, \tag{43}$$

where $\overline{\pi}$ is the mean pro t, g is a scale factor, and ϵ is some random disturbance term distributed normally with mean 0 and variance 1, independent accros time. The government offers the individual crop insurance that is actuarially fair. If $\epsilon < \overline{\epsilon}$, then the government will pay $\frac{I}{\Phi(\overline{\epsilon})}$, where I is the premium paid annually. The farmer can choose I each year, or choose not to insure ($I = 0$). The farmer that maximizes expected utility of wealth will face the problem

$$\max_{I \geq 0} \int_{\overline{\epsilon}}^{\infty} U\left(\overline{\pi} + g\epsilon - I\right)\phi\left(\epsilon\right)d\epsilon + \int_{-\infty}^{\overline{\epsilon}} U\left(\overline{\pi} + g\epsilon + I\left(\frac{1-\Phi\left(\overline{\epsilon}\right)}{\Phi\left(\overline{\epsilon}\right)}\right)\right)\phi\left(\epsilon\right)d\epsilon, \quad (44)$$

This model ignores the ability of farmers to defer risk through altering production practices, but may be instructive in describing other risk behavior. This model also assumes that the level of wealth remains stable each year. The rst order neccessary condition is

$$-\int_{\overline{\epsilon}}^{\infty} U'\left(\overline{\pi} + g\epsilon - I\right)\phi\left(\epsilon\right)d\epsilon + \left(1-\Phi\left(\overline{\epsilon}\right)\right)\int_{-\infty}^{\overline{\epsilon}} U'\left(\overline{\pi} + g\epsilon + I\left(\frac{1-\Phi\left(\overline{\epsilon}\right)}{\Phi\left(\overline{\epsilon}\right)}\right)\right)\phi\left(\epsilon\right)d\epsilon = 0. \quad (45)$$

Assuming that the individual has a concave utility function ensures an interior solution. Suppose now that the individual may not perceive g to be its true value. If the individual displays overcon dence, he will perceive g to be closer to 0. Let $\widetilde{g}$ be the perceived level of g. This new level will also effect the distribution used to calculate expectations. The individual will now believe a payment to occur whenever $\widetilde{g}\epsilon < g\overline{\epsilon}$, or when $\epsilon < \frac{g}{\widetilde{g}}\overline{\epsilon}$, the level set by the government. The rst order condition in this case will be

$$-\int_{\frac{g}{\widetilde{g}}\overline{\epsilon}}^{\infty} U'\left(\overline{\pi} + \widetilde{g}\epsilon - I\right)\phi\left(\epsilon\right)d\epsilon + \left(1-\Phi\left(\overline{\epsilon}\right)\right)\int_{-\infty}^{\frac{g}{\widetilde{g}}\overline{\epsilon}} U'\left(\overline{\pi} + \widetilde{g}\epsilon + I\left(\frac{1-\Phi\left(\overline{\epsilon}\right)}{\Phi\left(\overline{\epsilon}\right)}\right)\right)\phi\left(\epsilon\right)d\epsilon = 0. \quad (46)$$

Here the government has set a payment schedule, and, hence, the perception has no effect on payment level, but only on expectation (how often the payment will be given). It is now possible to conduct comparative statics on the rst order conditions, and nd the effect of perception of g on level of insurance. As expected, level of insurance increases with $\widetilde{g}$ (if the second derivative of U does not change too sharply). This is simply the result of a mean preserving spread.

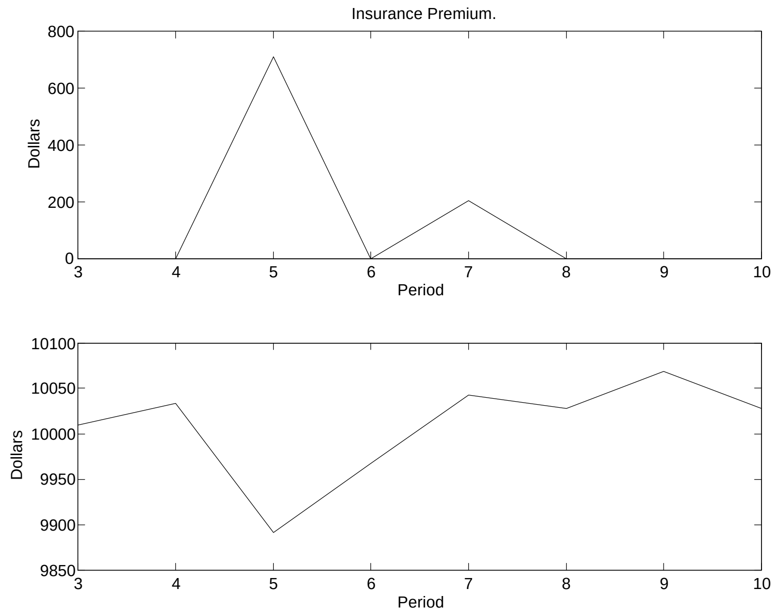

Figure 12: Simulated Crop Insurance with Limited Learning.

The limited learning model predicts that perception of both the mean and variance of the distribution will depend too heavily on the immediately preceding period s pro ts. Let us now consider a simulation of this crop insurance decision over several periods. Let $U(x) = 1 - e^{-.00001x}$, $\overline{\pi} = 10000$, $g = 10000$, $\overline{\epsilon} = -2$.[12] I will employ the weighting function displayed in (16). In each period, the farmer observes the pro t levels of 5 identical (but independent) regions and updates using the normal distribution with mean and variance equal to the sample values as the likelihood function. With correct perception, the optimal level I is about \$950. I endowed the farmer with correct perception in the rst period. A graphical plot of I and average ϵ for periods 3 through10 is displayed in gure 12.This simulation paints the picture of a farmer that learns too much from realized pro ts (or yields) in any given year. He will only insure his crops using actuarily fair insurance after a bad year, or after a large variance year. When weather returns to normal, he

[12] Estimates of true parameters of risk aversion all assume that individuals have perfect information. Without a reasonable estimate for this parameter, using true values for other parameters adds no degree of realism.

will cease to insure, forgetting how common catastrophic outcomes may be. This is exactly the behavior described by Glauber and Collins [15] describe following a low yield year in 1982. Following a year of extremely low yields, U. S. crop insurees increased sharply. After a year of normal yields, the number of enrollees was again reduced to the 1982 level. Since there were no changes in the crop insurance program over these years, Glauber and Collins attribute this to the especially low yields of 1982. In this case it appears that farmers are learning more than they should from an outlying event.

Roberts [39] describes a similar phenomenon. He nds that farmers adjust their planting (which is somewhat related to insurance) each year in response to the previous years yields. Roberts suggests their behavior is due to problems in learning, and assigning too much weight to current yeild levels when planning for the next year. While the supporting evidence given here is anecdotal (at best), it is compelling. Further research could include estimation of the LLM in farm management. It would be useful for governments to be able to predict (and react to) wide shifts in program participation due to learning problems in outlying years. There is also the possibility that behavior we have previously assumed was due to curvature of a utility function (or some other theoretically imovable object) may be due to poor training, or education (see Just [24]). Government programs that deal with risk should take these psychological factors into account, possibly by providing training with the program.

8.2 Production Contracts

Another area where perceptions may cause some testable distortion, is within agricultural production contract relationships. Typically large processing plants or shippers contract each year with farmers individually, although many of the contracts are designed by commodity associations. Presumably, farmers have fewer resources and less human capital than the large processing plants they contract with. Contracts are typically described by economists as a way to overcome asymmetric information, or to share risk. Incorrect perception may compound (or mitigate) the problem of assymetric information. Put simply, if the principal knows the tails of the distribution better than the agent, he may be able to design contracts that favor the principal in outlying events. Here, I will illustrate this point with a formal model following Hueths and Ligon s [23] application of the standard moral hazard model to the case of fresh tomato contracts.

In this case the farmer can take unobservable actions that affect the quality of his output. The amount of output is assumed to be xed, but the price of output to the shipper (principal), p, is random and conditioned on observed quality, q. The shipper contracts with the farmer, specifying a wage, $w(p)$, that is a function of the realized price for the goods. Due to the time involved in shipping, the shipper cannot directly observe quality. Represent the distribution of prices as $h(p|q)$, and the distribution perceived by the farmer as $g(p|q)$. The farmer wishes to maximize his expected utility given the wage schedule offered by the shipper, or

$$\max_q \int U(w(p))\, g(p|q)\, dp - z(q), \tag{47}$$

where $z(q)$ is the cost of producing at quality level q (this assumes separability of revenues and costs in the utility of wealth). Assuming the utility function is concave enough, and that the production costs are well behaved, the farmer will behave so that

$$\int U(w(p))\, g_q(p|q)\, dp - z_q(q) = 0. \tag{48}$$

Employing this assumption is suitably called the rst-order approach (see [32]). The shipper offers the contract that solves

$$\begin{array}{lll} \max_{q,\{w(p)\}} & \int V(p - w(p))\, h(p|q)\, dp & \text{subject to} \\ & \int U(w(p))\, g(p|q)\, dp - z(q) \geq \underline{U}, & \text{and} \\ & \int U(w(p))\, g_q(p|q)\, dp - z_q(q) = 0, & \end{array} \tag{49}$$

where $\underline{U}$ is the reservation level of utility. The solution to this problem is characterized by the following expression

$$\frac{V'(p - w(p))}{U'(w(p))} = \theta \frac{g(p|q)}{h(p|q)} + \mu \frac{g_q(p|q)}{h(p|q)}, \tag{50}$$

where θ and μ are the Lagrangian multipliers of the individual rationality and incentive compatibility constraints respectively.

It will be helpful here to contrast these conditions from the case of truthful perception. The incentive compatibility constraint will not bind if quality requirements are low, or incentives of the farmer happen to match those of the shipper. From (50), if $h(p|q) = g(p|q)$, then a non-binding incentive compatability constraint will result in a xed wage. However, suppose $g(p|q)$ has thinner tails than $h(p|q)$. Let p_m be a price falling in the center of the distributions (assuming they are close to one another), and p_t be a price falling in the tail of the distribution. Then with a non-binding incentive compatability constraint, $\frac{V'(p_m - w(p_m))}{U'(w(p_m))} > \frac{V'(p_t - w(p_t))}{U'(w(p_t))}$. Assuming that U is concave, and that V is not convex implies that $w(p_t) < w(p_m)$. Hence the wage will not be xed, even if incentives are not con icting.

Next let us examine the second term on the right hand side of (50). If $h(p|q) = g(p|q)$ then this term may be called the likelihood ratio, and is commonly interpreted as the ability to infer q from p. The ability to create incentives for greater q will be greater if the likelihood ratio is larger. Hence $w(p)$ will vary more widely with p. If again we assume that perceptions differ from the true distribution, then this likelihood ratio is an interaction of how much information the farmer believes is carried in the price signal and the true density. If the farmer believes that more information is carried in the signal, (a higher $g_q(p|q)$) then he will require a higher wage for the higher prices.

While it is difficult to run simulations with this model, it is easy to see some implications of the limited learning model in this context. As we have seen earlier, the farmer will tend to narrow the distribution, except following years with especially outlying realizations of price, or years with widely varied prices realized. After years with outlying realized prices, the contracts will perform more like the optimal contracts described in the literature. Most years, the principal will be able to take advantage of misperceptions of the distributions by increasing wages for middle prices and decreasing those for outlying prices. This prediction is not too different from the observed contract in Hueth and Ligon s research [23]. Their Figure 2 demonstrates that the observed contract appears to pay more than that considered optimum with truthful perceptions over some range of prices toward the center of the range of prices, and pay less than optimal toward the edges of the range of prices. If distorted perceptions are truly the cause of this deviation in pay schedule, the limited learning model would suggest that prices must be skewed downward, since super optimal pay occurs to the right of the center of the distribution of prices. As a symptom of the changing perceptions, the limited learning model also implies that the required quality may also change from year to year. Incorrect perceptions, and lack of informational resources might be one reason many contracts are negotiated with commodity associations. Farmers may be able to pool their resources and allow a better informed negotiator to help design the contract. There is a question then, however, of whether the farmer is able to recognize when a better contract has been negotiated.

9 Conclusions

Within this paper I present a model of learning based on the hypothesis that individuals have a limited ability to learn or store diffuse information. This model not only appears to describe behavior consistent with the literature on judgment bias, but also the violations of expected utility theory commonly observed. Standard empirical measures of performance using experimental data suggest that this model is comparable to the leading models, but several questions are raised regarding the usefulness of these measures in determining the best preference functional describing decision-making under uncertainty. The purest advantage in using the model I have described is its wide applicability. The model appears to (at least anecdotally) describe behavior in widely varying situations, including large nancial markets, and small experimental choices. I demonstrate here the plausibility of predictions within crop insurance and production contract negotiation.

Further research needs to be done to nd reasonable functional forms for the limited learning model, and to test for its implications in various settings. The model may be most applicable in the areas of nance, and agricultural marketing. The limited learning model makes several predictions that seem not only novel, but reasonable.Underlying the parameters of this model are the human capital and ability of the decision-maker. It is not clear that these parameters are xed, and monility of these parameters may be an important issue for policy-makers. When policy is designed to alter risk, perceptions need to be accounted for when predicting reactions, and changes in welfare. Further, it is not clear that policymakers have correct perceptions of the world when prescribing risk policy. In fact, there may be little incentive for policymakers to know relevant distributions beyond those that affect their own district. Researchers may need to account for low incentives when publishing research for use by policy-makers. By lowering the level of effort needed to understand research, it may be that our work will have wider applicability. Personal incentives of economic and other researchers, however, may con ict with the goal of applicability, as academic prestige seldom comes with low level work.

References

Alpert, M. and H. Raiffa. *Judgement Under Uncertainty: Heuristics and Biases*, chapter A Progress Report on the Training of Probability Assessors, 294 305. Cambridge: Cambridge University Press, 1982.

Amemiya, T. Selection of Regressors, *International Economic Review*, *2*:331 354 (1980).

Arrow, K. J. *Essays in the Theory of Risk Bearing*, chapter 3, 90 133. Chicago, IL: Markham Publishing Co., 1971.

Arrow, Kenneth. Risk Perception in Psychology and Economics, *Economic Inquiry*, *20*:1 9 (1982).

Buschena, D. and D. Zilberman. Predictive Value of Incentives, Decision Difficulty, and Expected Utility Theory for Risky Choices, *Journal of Risk and Uncertainty* (forthcoming).

Camerer, C. F. Do Biases in Probability Judgement Matter in Markets? Experimental Evidence, *American Economic Review*, *77*:981 97 (1987).

Camerer, C. F. The Rationality of Prices and Volume in Experimental Markets, *Organizational Behavior and Human Decision Processes*, *51*:237 72 (1992).

Camerer, C. F. and T.-H. Ho. Violations of the Betweenness Axiom and Nonlinearity in Probability, *Journal of Risk and Uncertainty*, *8*:167 96 (1994).

Camerer, Colin F. *The Handbook of Experimental Economics*, chapter Individual Decision Making, 587 704. Princeton, New Jersey: Princeton University Press, 1995.

Chew, S. H. A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox, *Econometrica*, *51*:1065 92 (1983).

DeBondt, Werner F. M. and Richard Thaler. Does the Stock Market Overeact?, *Journal of Finance*, *40*:793 805 (1985).

Edwards, Ward. *Judgement Under Uncertainty: Heuristics and Biases*, chapter Conservatism in Human Information Processing, 355 358. Cambridge University Press, 1982.

Ellsberg, D. Risk, Ambiguity, and the Savage Axioms, *Quarterly Journal off Economics*, *75*:643 69 (1961).

Fox, C. R. and J. R. Irwin. The Role of Context in the Communication of Uncertain Beliefs, *Basic and Applied Social Psychology*, *20*:57 70 (1998).

Glauber, Joseph W. and Keith J. Collins. *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, chapter The Case for Government Safety Nets and Disaster Payments. New York: Kluwer Academic Press, 2001.

Grether, D. M. Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence, *Journal of Economic Behavior and Organization*, *17*:31 57 (1990).

Grether, David M. Bayes Rule as a Descriptive Model: The Representative Heuristic, *Quarterly Journal of Economics*, *95*:537 557 (1980).

Hansson, B. *Decision, Probability, and Utility*, chapter Risk Aversion as a Problem of Cojoint Measurement, 136 158. New York: Cambridge University Press, 1988.

Haruvy, Ernan, et al. The Medium Prizes Paradox: Evidence from a Simulated Casino, (2001).

Hey, J. D. and Chris Orme. Investigating Generalizations of Expected Utility Theory Using Experimental Data, *Econometrica*, *62*:1291 1326 (1994).

Hey, J. D. and Elisabetta Strazzera. Estimation of Indifference Curves in the Marschak-Machina Triangle: A Direct Test of the Fanning Out Hypothesis, *Journal of Behavioral Decision Making*, *2*:239 60 (1989).

Hogarth, R. M. and H. J. Einhorn. Order Effects in Belief Updating: The Belief Adjustment Model, *Cognitive Psychology*, *24*:1 55 (1992).

Hueth, Brent and Ethan Ligon. Producer Price Risk and Quality Measurement, *American Journal of Agricultural Economics*, *81*:512 524 (1999).

Just, David R. *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, chapter Information, Processing Capacity, and Judgment Bias in Risk Assessment. New York: Kluwer Academic Press, 2001.

Just, David R. *Learning and Information*. PhD dissertation, University of California, Berkeley, in progress.

Kahneman, D. and A. Tversky. Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, *47*:263 92 (1979).

Kahneman, Daniel, et al., editors. *Judgement under Uncertainty: Biases and Heuristics*. Cambridge, New York: Cambridge University Press, 1982.

Kahneman, Daniel and Amos Tversky. *Judgement under Uncertainty: Heuristics and Biases*, chapter Subjective Probability: A Judgement of Representativeness, 32 47. Cambridge University Press, 1982.

Lattimore, P. K., et al. The In uence of Probability on Risky Choice: A Parametric Examination, *The Journal of Economics Behavior and Organization*, *17*:377 400 (1992).

Loomes, G. Evidence of a New Violation of the Independence Axiom, *Journal of Risk and Uncertainty*, *4*:91 108 (1991).

Machina, M. J. Expected Utility Analysis Without the Independence Axiom, *Econometrica*, *50*:277 323 (1982).

Macho-Stadler, Ines and David Perez-Castrillo. *An Introduction to the Economics of Information: Incentives and Contracts*. New York: Oxford University Press, 1997.

Markowitz, Harry. The Utility of Wealth, *The Journal of Political Economy*, *60*:151 158 (1952).

Neuman, J. Von and O. Morgenstern. *Theory of Games and Econoimic Behavior*. Princeton, NJ: Princeton University Press, 1944.

Oskamp, Stuart. *Judgement under Uncertainty: Heuristics and Biases*, chapter Overcon dence in Case-Study Judgements, 287 293. Cambridge University Press, 1982.

Quiggin, J. A Theory of Anticipated Utility, *The Journal of Economic Behavior and Organization*, *3*:323 43 (1982).

Rabin, Matthew. Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion. 2000.

Rabin, Matthew. Risk Aversion and Expected-Utility Theory: A Calibration Theorem. 2000.

Roberts, Michael J. and Nigel Key. *A Comprehensive Analysis of the Role of Risk in US Agriculture*, chapter Does Liquidity Matter to Agricultural Production? How Transitory Yield Shocks In uence Subsequent Plantings. New York: Kluwer Academic Press, 2001.

Rubinstein, A. Similarity and Decision-Making Under Risk: Is There a Utility Theory Resolution to the Allais Paradox?, *Journal of Economic Theory*, *46*:145 53 (1988).

Schultz, T. The Value of the Ability to Deal with Disequilibria, *Journal of Economic Literature*, *13*:827 37 (1975).

Simon, H. A. A Behavioral Model of Rational Choice, *Quarterly Journal of Economics*, *69*:99 118 (1955).

Simon, H. A. Theories of Decision-Making in Economics and Behavioral Science, *American Economic Review*, *49*:253 283 (1959).

Simon, H. A. Rationality as Process and as Product of Thought, *American Economic Review*, *68*:1 16 (1978). Richard T. Ely Lecture, American Economic Association Annual Meetings December 1977.

Tversky, A. and D. Kahneman. Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty*, *5*:297 323 (1992).

Tversky, Amos and Daniel Kahneman. *Judgement under Uncertainty: Heuristics and Biases*, chapter Belief in the Law of Small Numbers, 23 31. Cambridge University Press, 1982.

Wolf, S., et al. Between Data and Decisions: The Organization of Agricultural Economic Information Systems, *Research Policy* (forthcoming).

Yaari, M. E. The Dual Theory of Choice under Risk, *Econometrica*, *55*:95 115 (1987).

Zellner, A. Information Processing and Bayes Theorem. In progress., 2000.

A Mathematical Proofs

These proofs follow (nearly verbatim) those provided by Rabin [38], I have only modi ed the proofs where necessary to include a weighting function, and reference point.

Proof of Theorem 1, part 1: Let $r(z|w) \equiv U(z|w) - U(z-l|w) = 1$. Then by concavity of U$(\cdot|w)$, $U(z-l|w) - U(z-2l|w) \geq 1$. The assumption $2l > g > l$ implies that $z-2l+g \in (z-l, z)$. By concavity of $U(\cdot|w)$ we know that $U(z-2l+g|w) - U(z-l|w) \geq \frac{g-l}{l} = \frac{g}{l} - 1$. Thus, $U(z-2l+g|w) - U(z-2l|w) \geq \frac{g}{l} - 1 + 1 = \frac{g}{l}$. So, if $z - 2l \geq U^{-1}\left(\frac{U(\underline{\phi}|w)}{2\pi(.5)}|w\right)$, then $U(z-2l|w) - U(z-3l|w) \geq \frac{g}{l}$ since $U(z-2l-l|w) + U(z-2l+g|w) < \frac{1}{\pi(.5)} U\left(\widetilde{\phi}|w\right) = 2U(z-2l|w)$, where $\widetilde{\phi} = U^{-1}(2\pi(.5) U(z-2l|w)|w)$, by assumption. By concavity, we know $U(z-3l|w) - U(z-4l|w) \geq \frac{g}{l}$. We can extend this result to the general case, if $z - 2kl \geq U^{-1}\left(\frac{U(\underline{\phi}|w)}{2\pi(.5)}|w\right)$, then $U(z-(2k-1)l|w) - U(z-2kl|w) \geq U(z-2(k-1)l|w) - U(z-(2k-1)l)$, which implies that

$$U(z-2kl+g|w) - U(z-2kl|w) \geq \frac{g}{l}\left[U(z-2(k-1)l|w) - U(z-(2k-1)l|w)\right]. \quad (51)$$

This last inequality implies that

$$U\left(z-2kl|w\right)-U\left(z-\left(2k+1\right)l|w\right)\geq \frac{g}{l}\left[U\left(z-2\left(k-1\right)l|w\right)-U\left(z-\left(2k-1\right)l|w\right)\right]. \quad (52)$$

This expression yields a lower bound on marginal utilities, which, in turn, provides a lower bound on utility $U\left(z|w\right)-U\left(z-x|w\right)$, by summing over intervals. Note that the weighting function precludes this proof from applying to prospect theory. I wished to use the more widely used weighting function. A proof using the weighting function in Kahneman and Tversky [26] is much simpler.

Proof of Theorem 1, part 2: Let $r\left(z|w\right)\equiv U\left(z|w\right)-U\left(z-l|w\right)=1$. Then $U\left(z+g|w\right)-U\left(z|w\right)\;\square\;1$. By the concavity of $U\left(\cdot|w\right)$, $U\left(z+g|w\right)-U\left(z+g-l\right)\;\square\;\frac{l}{g}$. If $z+g\;\square\;U^{-1}\left(\frac{U\left(\overline{\phi}|w\right)}{2\pi(.5)}|w\right)$, then $U\left(z+2g|w\right)-U\left(z+g|w\right)\;\square\;\frac{l}{g}$ (see the proof to part 1) by assumption. Again it is possible to generalize this result, if $z+mg\;\square\;U^{-1}\left(\frac{U\left(\overline{\phi}|w\right)}{2\pi(.5)}|w\right)$, then $U\left(z+mg+g|w\right)-U\left(z+mg|w\right)\geq\frac{l}{g}\left[U\left(z+mg|w\right)-U\left(z+mg-g|w\right)\right]$. This inequality implies upper bounds on marginal utilities, and, hence, on utility $U\left(z+x|w\right)-U\left(z|w\right)$ in part 2 of the theorem.

Proof of Corollary: From the theorem, we know $U\left(z|w\right)-U\left(z-2kl|w\right)\geq 2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i-1}r\left(z|w\right)$ and $U\left(z+mg|w\right)-U\left(z|w\right)\;\square\;\sum_{i=0}^{m+1}\left(\frac{l}{g}\right)^{i}r\left(z|w\right)$. Thus, if $U\left(z|w\right)-U\left(z-2kl|w\right)<U\left(z+mg|w\right)-U\left(z|w\right)$, then $2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i-1}<\sum_{i=0}^{m+1}\left(\frac{l}{g}\right)^{i}$. Solving this inequality for m yields the formula. If $g>2l$, we need only that $U\left(z|w\right)-U\left(z-2kl|w\right)\geq 2k\left(U\left(z|w\right)-U\left(z-1|w\right)\right)$ to obtain the desired result.