Probability, Expected Utility, and the Ellsberg Paradox

Thomas S. Coleman
Close Mountain Advisors LLC

19 May 2011

The Ellsberg paradox (Ellsberg 1961) is often cited as evidence for unknowable "ambiguity" versus computable "risk", and a refutation of the Savage axioms regarding expected utility maximization and the program for revealing "subjective" or "belief-type" probabilities. This note argues that researchers have been too quick to embrace the Ellsberg critique as a refutation of standard expected utility theory. First, Ellsberg performed no actual experiments, and in fact recent empirical evidence on the Ellsberg paradox argues against ambiguity. Second, simple explanations for the paradox deserve as much attention as theories that introduce a new concept such as "ambiguity." One such simple explanation is to consider the Ellsberg thought experiment as part of a "meta-experiment" that includes repeated draws, in which case the choices described by Ellsberg are consistent with expected utility theory.

Introduction

Daniel Ellsberg (1961) discussed some thought experiments regarding choice under uncertainty. The resulting pattern of choices are said to present a paradox. (Note that, although popularized by Ellsberg and commonly going by his name, a version of this paradox was apparently noted by Keynes, 1921.) Ellsberg's paper is often cited as evidence for the distinction between unknowable *ambiguity* or *uncertainty* versus computable *risk*, and a refutation of the Savage axioms regarding expected utility maximization and the program for revealing "subjective" or "belief-type" probabilities. It is often taken for granted that Ellsberg's paper demonstrates severe difficulties with deriving subjective probabilities from a decision-maker's preferences (to paraphrase Halevy 2007). There has been work over the years to extend or modify the concept of expected utility to accommodate the Ellsberg paradox and formally model the concept of ambiguity. (See Epstein 1999, Schmeidler 1989.) In fact, the premise that Ellsberg's work provides support for uncertainty or ambiquity requires more critical attention then it has received. Empirical evidence from Halevy (2007) shows that the Ellsberg paradox does not support the existence of ambiguity separate from computable risk. Furthermore, there are alternative explanations for the behavior described by Ellsberg, explanations that are fully consistent with utility maximization.

Ellsberg presented "thought experiments" rather than actual experiments. Since Ellsberg's original paper, researchers have collected experimental evidence on the paradox, with Halevy (2007) being a recent and careful example. Interestingly, the evidence actually shows that subjects do not treat "ambiguity" differently from computable risk. More specifically, the data in Halevy (2007) show that subjects do not treat an unknown distribution differently from mixture distributions (compound lotteries). Since mixture distributions or compound lotteries are computable they incorporate no uncertainty or ambiguity, and the conclusion must be that the Ellsberg experiments provide no support for ambiguity. Nonetheless, the experiments do raise a paradox, since most subjects treat mixtures differently from a simple distribution that, objectively, provides the same payout. In other words, the experiments raise the paradox “why do subjects view mixtures differently than simple distributions in the Ellsberg experiments?” even though the data provide no support for the commonly-quoted paradox “why do subjects view ambiguity differently than risk?”

Although the evidence shows that the Ellsberg paradox does not relate to ambiguity or uncertainty, the Ellsberg paradox is still compelling. Why are the “more certain” bets preferred? One explanation is that subjects are simply mistaken, misinformed, or unable to correctly calculate probabilities. The data hint that this may be true for some subjects and this is no doubt part of the explanation. I believe, however, that the issue goes deeper. Indeed, upon introspection I believe I myself would choose the more certain bets and apparently violate the equivalence between apparently identical outcomes. There is, however, a simple explanation for why a risk-averse expected utility maximizing agent might prefer a simple distribution. I argue that a minor, and psychologically compelling, modification to the experiments can make Ellsberg’s observations consistent with expected utility maximization.

Let us turn to Ellsberg’s original thought-experiments. The experiments, described in detail below, consist of drawing balls from urns and receiving payouts based on the color of a ball. The subject of the experiment is given various choices on payouts. Subjects are told a single ball is to be drawn from an urn, although the subject is presented with multiple choices of colors and urns. In the context of a single draw experiment, the behavior described by Ellsberg is indeed inconsistent with subjective probability assessments and expected utility. With just a slight modification of the experimental set-up, however, the behavior becomes not only consistent with, but implied by risk-averse expected utility maximization. Consider the Ellsberg thought experiments in the context of a larger “meta-experiment”:

- x% probability of single draw (original Ellsberg thought experiment)
- 1-x% probability of repeated draws

Within the meta-experiment the choices described by Ellsberg become consistent with standard risk and expected utility calculations. The central point is that for a single draw (the Ellsberg case) the distribution of winnings are identical across urns and the pattern described by Ellsberg is indeed a paradox. In the meta-experiment with a finite probability of repeated draws, however, the distributions for the different urns are different and the pattern of choices described by Ellsberg is consistent with expected utility theory and risk aversion.

The crux of my argument is that, unconsciously, a subject may consider the meta-experiment even when instructed to consider only the single-draw experiment, and thus answer a question subtly different from that originally asked. Psychologically it may be difficult to ignore the meta-experiment (repeated draws), for two reasons. First, repeated events are common in life and so it may be hard to assign zero probability to the possibility of repeated draws, even when instructed to do so by a trusted experimenter. Second, the subject may not fully trust that the experimenter will only perform a single draw, and may assign a non-zero probability to the event that the experimenter “cheats” by subsequently announcing repeated draws. This is akin to the common “deceit-aversion” argument cited against the Ellsberg paradox, but with the additional force of specifying the exact mechanism by which the subject could be deceived. (For the standard deceit-aversion argument and the single-draw experiment it is hard to see how the experimenter could cheat since the subject gets to choose the color for payout and it would be easy to hide that choice from the experimenter.)

According to this line of argument the choices that appear paradoxical are simply the result of the subject assigning, either unconsciously or intentionally, a small but finite probability to repeated draws. From this it follows that, in such a case, the experiments are not paradoxical but are fully consistent with expected utility theory and risk aversion.

Ellsberg’s Thought Experiment

Statement of Example 1

Ellsberg (1961) presents two experiments, the first consisting of two urns with red and black balls, the second a single urn with combinations of red, black, yellow balls. I describe the first here, and the second in the appendix. They both show, essentially, the same results.

Ellsberg urn experiment 1:

- Urn 1 - 100 balls, 50 red, 50 black
- Urn 2 - 100 balls, red and black with proportions not specified

(Note that the order of urns is reversed from Ellsberg’s paper. I have found in discussions that many people seem to put the “more certain” urn first.)

Payoffs defined as:

I. “Payoff on Red_1”: Draw from Urn 1, receive \$1 if Red, \$0 if Black
II. “Payoff on Black_1”: Draw from Urn 1, receive \$0 if Red, \$1 if Black
III. “Payoff on Red_2”: Draw from Urn 2, receive \$1 if Red, \$0 if Black
IV. “Payoff on Black_2”: Draw from Urn 2, receive \$0 if Red, \$1 if Black

I use the term “payoff” rather than the original, and technically correct, “bet,” because bet may have the misleading implication that one must pay upfront, whereas these are pure payoffs without any upfront cost.

Gambles posed:

1. Which do you prefer: “Payoff on Red_1” vs. “Payoff on Black_1” (I vs. II or R vs. B within Urn 1)
2. Which do you prefer: “Payoff on Red_2” vs. “Payoff on Black_2” (III vs. IV or R vs. B within Urn 2)
3. Which do you prefer: “Payoff on Red_1” vs. “Payoff on Red_2” (I vs. III; i.e. R, but from either Urn 1 or Urn 2)
4. Which do you prefer: “Payoff on Black_1” vs. “Payoff on Black_2” (II vs. IV; i.e. B, but from either Urn 1 or Urn 2)

Note that you cannot choose both color and urn, and you are not necessarily told in advance all the possible gambles you will be able to choose from.

Results according to Ellsberg (apparently from introspection and non-experimental surveying of colleagues):

- Majority will be indifferent in (1) and (2), which indicates that subjective probabilities of Red versus Black are 50/50 for both Urn 1 and Urn 2.
- Majority prefers Red_1 in (3) and Black_1 in (4) - in other words most people prefer Urn 1 (known 50/50) over Urn 2 (unknown split between red and black).

The paradox is that indifference between red and black in (1) and (2) establishes 50/50 subjective probabilities for both urns, while preference for urn 1 in (3) and (4) contradicts that probabilities are 50/50 for both urns. In fact, the preference for urn 1 in both (3) and (4) implies that the total probability of urn 2 is less than 1, a clear inconsistency.

Ellsberg’s paper did not have controlled experiments, but I think the argument about the preference for the known 50/50 urn 1 distribution in (3) and (4) is convincing nonetheless. It is convincing because Ellsberg says he has surveyed various people, but more importantly from my own introspection - somehow the unknown distribution of urn 2 is “more risky” and I would prefer urn 1 in (3) and (4). Preference for urn 1 is commonly cited as evidence for ambiguity over risk - in other words the importance of non-computable uncertainty versus computable risks. (Frank Knight 1921 is usually cited as the originator of the distinction, although LeRoy and Singell (1987) claim that it more rightly belongs to Keynes 1921.)

Single versus Repeated Game

In the single-draw experiment the probabilities of outcomes should be the same for the two urns, but for a repeated-draw experiment they should be different. In essence the argument is that, psychologically, it may be hard to convince a subject (even oneself when introspecting regarding the thought experiment) that the experiment will be just a single draw. The knowledge that the urn 1 distribution is preferable in a repeated-draw experiment will unavoidably enter and be enough to push a risk-averse agent to choose urn 1 over urn 2. The argument comes down to this: What would be the rational response, to choose urn 2 which is equivalent to 1 in the single-draw case but worse in the repeated-draw case, or, for no extra cost, choose urn 1? The choice is obvious: as long as there is some non-zero chance that the experiment could involve repeated draws (and psychologically it is hard to ignore such a possibility) we should choose urn 1.

My development of the argument concerning the preference for urn 1 in a repeated-draw experiment is two-pronged. First, I set up a thought-experiment where the urn 2 distribution is specified and verifiable but is a mixture distribution or compound lottery rather than simply 50/50 red/black This removes any ambiguity or deceit-aversion regarding the true distribution of red versus black. The probabilities of outcomes for a single draw from urn 1 or urn 2 are now identical. Nonetheless, upon introspection I believe I would still choose urn 1 because the urn 2 mixture distribution somehow feels "more risky." Second, to understand why the urn 2 mixture distributions feels more risky we can extend the thought-experiment from a single draw to repeated draws from the same urn. In this thought experiment urn 2 is indeed more risky. The probabilities of outcomes from urn 1 and urn 2 do differ and the outcomes for urn 2 have more dispersion and are unambiguously more risky (the urn 1 distribution second-order stochastically dominates). Risk averse agents would choose urn 1. This preference for urn 1 over urn 2 in a repeated game extends to an unknown urn 2 distribution. I believe that it is hard to convince oneself that the repeated-draw distinction is irrelevant, and very difficult to design a single-draw experiment that would completely exclude consideration of the repeated-draw experiment.

Modification to Known Probabilities

Now let us turn to a simple modification of the Ellsberg urn problem, where we make the probabilities in urn 2 known, precise, and computable, thus removing any "ambiguity," but still not a simple 50/50 mix of red and black. Urn 1 remains the same 50/50 mix of red and black balls, but for urn 2 we substitute a choice between two urns. To summarize:

- Urn 1 **-** 100 balls, 50 red, 50 black
- Urn 2 **-** will be either Urn 2a or Urn 2b. Each urn has 100 balls, Urn 2a has 25 red, 75 black, Urn 2b has 75 red, 25 black. The urn actually used is chosen by flip of a fair coin, meaning that each urn has probability 1/2. The subject does not know a priori which urn will be used but can verify the mix of balls in each urn and can verify the coin and the fairness of the flip.

The result will be that the probability of drawing a single red from either urn 1 or urn 2 will be 50%. There is no ambiguity or non-computable risk since all the distributions are known and verifiable. Nonetheless the distributions are different: the urn 1 distribution remains a simple Bernoulli distribution while the urn 2 distribution will be a mixture of Bernoullis.

We will define payoffs the same as for original Ellsberg problem:

I. "Payoff on Red_1": Draw from Urn 2, receive \$1 if Red, \$0 if Black
II. "Payoff on Black_1: Draw from Urn 2, receive \$0 if Red, \$1 if Black
III. "Payoff on Red_2": Draw from Urn 1, receive \$1 if Red, \$0 if Black
IV. "Payoff on Black_2": Draw from Urn 1, receive \$0 if Red, \$1 if Black

Gambles posed also the same:

1. Which do you prefer: "Payoff on Red_1" vs. "Payoff on Black_1"
2. Which do you prefer: "Payoff on Red_2" vs. "Payoff on Black_2"
3. Which do you prefer: "Payoff on Red_1" vs. "Payoff on Red_2"

4. Which do you prefer: “Payoff on $Black_1$” vs. “Payoff on $Black_2$”

Remember that probabilities of a draw from urn 2 are now known to be 50% red, 50% black - there is no ambiguity or non-computable uncertainty. The probability for urn 2 is known with as much certainty as for urn 1, but it is a mixture rather than a simple probability. The actual urn may be 25% red or 75% red, but which urn is used is random with probability 1/2.

From introspection I believe I would still choose urn 1 in gambles (3) and (4), because urn 2 has “more risk”. My intuition is that most people presented with this problem would still be indifferent between red and black in (1) and (2) while most people would still prefer urn 1 in (3) and (4). In other words, even though there is now no “ambiguity” - the probabilities in urn 2 and urn 1 are equally computable - somehow urn 2 feels more risky.

Modification to Repeated Game

To understand why urn 2 might feel like it has more risk, let us extend the thought experiment. Instead of a set of gambles and single draws, let us set up the following four “games”

1. Which do you prefer when playing 20 repeats: “Win on Red_1” vs. “Win on $Black_1$” (I vs. II or R vs. B within Urn 1)
2. Which do you prefer when playing 20 repeats: “Win on Red_2” vs. “Win on $Black_2$” (III vs. IV or R vs. B within Urn 2)
3. Which do you prefer when playing 20 repeats: “Win on Red_1” vs. “Win on Red_2” (I vs. III; i.e. R, but from either Urn 1 or Urn 2)
4. Which do you prefer when playing 20 repeats: “Win on $Black_1$” vs. “Win on $Black_2$” (II vs. IV; i.e. B, but from either Urn 1 or Urn 2)

Note that the choice of red / black or urn 1 / urn 2 for each “game” of 20 repeats is made up-front, so that in game 1 the choice is between 20 “Win on Red_1” vs. 20 “Win on $Black_1$”. Further, the urn 2 is not re-chosen; if it starts as urn 2a (25 red / 75 black) it remains as such for the 20 repeats.

Distribution of Winnings for Repeated Game

For each urn considered on its own the distribution of winnings will be symmetrical and the same for red and black. As a result the indifference between red and black in 1 and 2 (choosing between red versus black for the same urn) is maintained.

There is, however, now a clear and important difference between urn 1 and urn 2. As a result of this difference any risk-averse expected utility-maximizing agent should choose urn 1; i.e. we should expect to see the pattern of choices Ellsberg describes. Figure 1 shows the distribution for urn 1. It is very straight-forward - mean \$10 and binomial.

Figure 1 - Density Function for Winnings from Urn 1 - 20 repeats of “Win on Red_2” or “Win on $Black_2$”

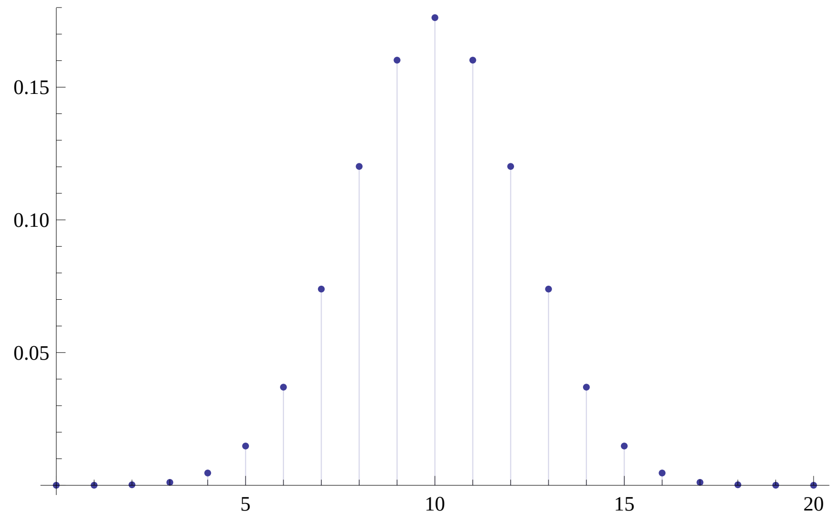
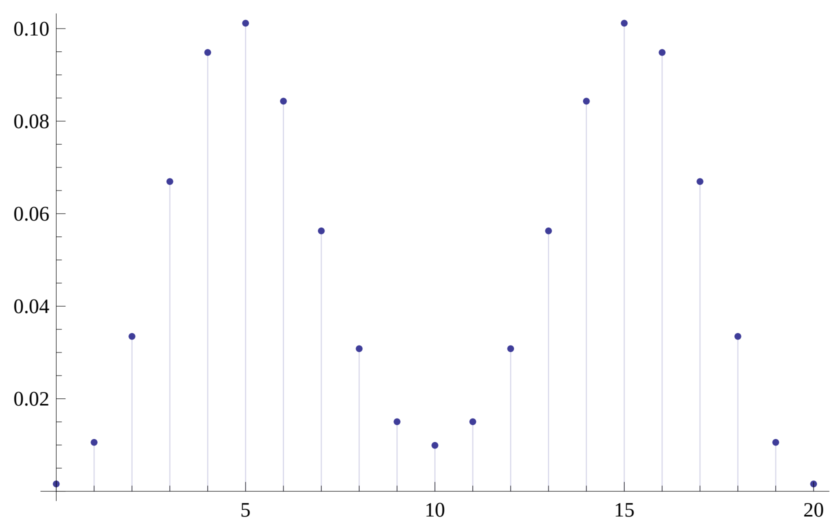


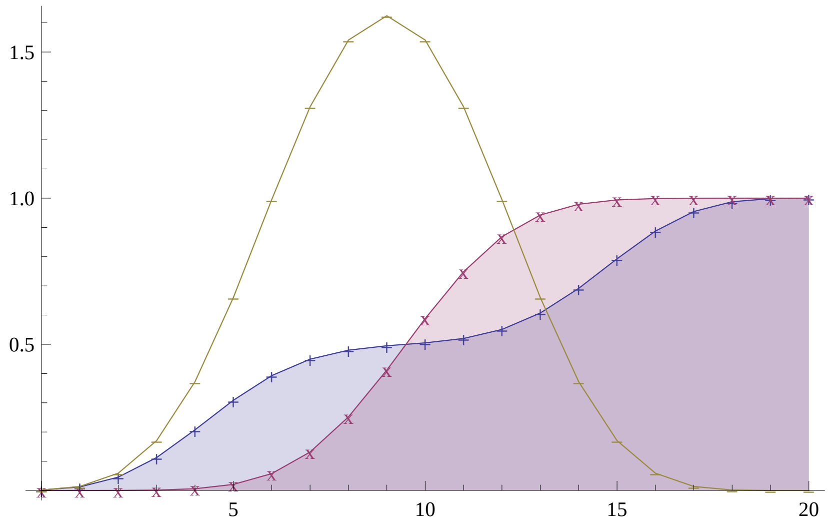
Figure 2 shows the distribution of winnings for urn 2. The distributions for red and black are the same so red / black indifference is ensured. The mean is \$10, like for urn 1, but the distribution is bimodal - it is a mixture of binomials. If the urn used turns out by chance to be urn 2a (25 red / 75 black) then the distribution will be binomial with mean \$5 while if it is urn 2b (75 red / 25 black) the distribution will be mean \$15. A priori the probabilities of the two urns are 1/2 each and so the mean is \$10. A priori the distribution is as shown in figure 2, but ex ante the distribution will be either the left-hand (lower) or right-hand (higher) distribution.

Figure 2 - Density Function for Winnings from Urn 2 - 20 repeats of “Win on Red$_1$” or “Win on Black$_1$”



There is a clear difference between the distributions for urn 1 and urn 2. The distribution of winnings for urn 1 has less dispersion than the distribution for urn 2. The expected winnings is the same in both cases but the expected utility from these distributions will clearly be different with urn 1 being higher utility for any risk-averse agent. The distribution from urn 1 will second-order stochastically dominate the distribution from urn 2 as shown in figure 3. (Remember that distribution F second-order stochastically dominates distribution G when the integral or sum of F-G is always positive. Any risk-averse agent (concave utility function) will prefer F to G.) There is no surprise that urn 1 is generally preferred to urn 2 in this repeated game situation. In fact it would be a surprise if it were not.

Figure 3 - Distributions Urn 1 and Urn 2, together with Sum (Urn1 - Urn2)



'+' = Urn 1, 'x' = Urn 2, '−' = Sum of Urn1 − Urn2

Essentially, with urn 2 a single draw has 50/50 probability of red, but if there are repeated draws then with urn 2 there is a 50% chance of bad luck, drawing the 25 red / 75 black urn 2a and suffering a string of 25% probability red draws. For a single draw urn 2 and urn 1 are (probabilistically) identical but for repeated draws they are different. Possibly the differentiation between urn 2 and urn 1 in the repeated game is the reason for the preference of urn 1 in the single game. In real life, repeated games are frequent and when I hear the non-repeated, single-shot experiment I may subconsciously extrapolate to a repeated game and reject urn 2 even though, for the single-shot experiment there should not be any difference. Indeed, in the original Ellsberg “experiment” the subject is not told the whole menu of gambles up-front. Since there is no cost for choosing urn 1 over urn 2, it would not be unreasonable for someone to project forward that a repeated game may arise and so choose urn 1.

Extension to Unknown Urn 2

We have just shown that an urn 2 mixture and urn 1 simple distribution should be treated the same for a single-shot experiment but differently for a repeated game. A mixture seems like a reasonable mental model for an unknown urn 2. The number of red / black could be anywhere from 100/0 to 0/100 and everything in between. A priori there seems to be no reason to think red should predominate over black, and so it seems reasonable to judge red and black as equally likely in a single draw. But by the same token there seems almost no reason for thinking that urn 2 is exactly 50 red / 50 black, and so almost zero probability that the a priori distribution for repeated draws will be a simple binomial. The a priori distribution for repeated draws should be a mixture of binomials, which will have greater dispersion than a simple binomial. Any mixture one would use as a mental model for an unknown urn 2 would be stochastically dominated by the 50/50 urn 1, and thus less preferred than urn 1 by risk-averse agents. Note that, according to this line of argument, people should treat an unknown urn and a mixture urn effectively the same.

Experimental Evidence

Ellsberg (1961) presented no empirical evidence to support the argument regarding choices over alternate urns, although personally I find Ellsberg’s original arguments persuasive. Since Ellsberg published his paper there have been empirical studies based on the thought-experiments; Halevy (2007) is recent and contains references to earlier work. The data, however, do not actually support the hypothesis of “ambiguity” separate from risk. Instead, the evidence shows that the majority of subjects view unknown or “ambiguous” similarly to mixture urns. The data also show, however, that a substantial fraction of subjects value mixture urns differently than a simple urn. The data reject the notion of ambiguity but raise the question of why so many subjects value mixture urns differently than a simple urn.

Halevy (2007) describes a careful set of experiments based on the Ellsberg thought-experiment discussed in the prior section. In his experiments, Halevy presents subjects with four urns (or boxes):

- Urn 1 - 10 balls; 5 red, 5 black
- Urn 2 - 10 balls; red and black with proportions not specified
- Urn 3 - will be one of 11 urns, chosen randomly (each has probability 1/11). Each urn has 10 balls. The first urn has 0 red, 10 black; the second 1 red, 9 black; …; the eleventh has 10 red, 0 black. The result is that urn 3 will be a mixture distribution of 11 Bernoulli random variables, with red probabilities ranging from 0.0, 0.1, up to 1.0.
- Urn 4 - will be one of two urns, chosen randomly (each has probability 1/2). Each urn has 10 balls. The first has 0 red, 10 black; the second has 10 red, 0 black. The result is that urn 4 will be a mixture distribution of two Bernoulli random variables, with red probabilities 0.0 or 1.0.

This experimental design closely corresponds to what I outlined above: There is an urn with a simple Bernoulli distribution (urn 1), an urn that is unknown or "ambiguous" (urn 2), and two urns with mixture distributions or compound lotteries (urns 3 and 4).

Halevy's experiment is designed to uncover subjects' reservation prices for each of the four urns. These reservation prices, V_1, V_2, V_3, V_4, approximate the certainty equivalent of the payoffs for the four urns. The most striking of Halevy's results is the strong association between how people view ambiguity relative to the simple urn (valuation of urn 2 versus urn 1, or V_2 versus V_1) and how they view mixtures or compound lotteries relative to the simple urn (V_3 and V_4 versus V_1). Those who are neutral regarding ambiguity (V_2=V_1) are overwhelmingly neutral regarding mixtures (V_3=V_4=V_1), and vice versa (those who love or hate ambiguity love or hate mixtures). In other words people view ambiguity and mixture distributions very much the same.

Halevy's table I, reproduced below, demonstrates the association between "ambiguity neutrality" (V_2=V_1) and neutrality regarding mixtures or compound lotteries (V_3=V_4=V_1). The table shows that virtually all subjects who were neutral regarding mixtures (23 who set V_3=V_4=V_1) were also neutral regarding ambiguity or the unknown urn (22 subjects valued V_3=V_4=V_1=V_2). Conversely, subjects who were not ambiguity neutral - valuing urn 2 less than urn 1 (ambiguity aversion, V_2<V_1) or more than urn 1 (ambiguity loving, V_2>V_1) - were also not neutral regarding mixtures or compound lotteries.

Table 1 - Halevy's Evidence on Association Between Attitudes to Ambiguity and Compound Objective Lotteries (Mixtures)

		Mixture rel to Simple		
Ambig rel to Simple		Diff	Neutral	Total
Diff	Count	113	1	114
	Expected	95.5	18.5	
Neutral	Count	6	22	28
	Expected	23.5	4.5	
Total		119	23	142

From Halevy (2007) Table I. The table combines the two samples reported by Halevy. "Ambig rel to Simple = Diff" means that subjects placed a different value on the ambiguous and simple urns (either V_2>V_1 or V_2<V_1). "Ambig rel to Simple = Neutral" means that subjects placed the same value on the ambiguous urn and the simple urn (V_2=V_1). "Mixture rel to Simple = Diff" means that subjects valued the mixture urns and the simple urns differently (V_3&V_4≠V_1) while "Mixture rel to simple = Neutral" means that subjects valued the mixture and the simple urns the same (V_3=V_4=V_1). Halevy reports that Fisher's exact test (2-sided) for the hypothesis that the association between valuation of ambiguity and mixture is random has a significance level of 2.3E-19, making it highly unlikely that the association observed in the table is purely random.

The most striking observation from table 1 is that it shows no indication of the existence of ambiguity as a separate concept. The primary signature of ambiguity should be that subjects treat ambiguity differntly than computable risk.

The mixture urns (urns 3 and 4) are computable, and for the single-draw experiment should be valued the same as the simple urn. There are 23 subjects who do value those urns the same (23 who set V_3=V_4=V_1). If the unknown urn were treated differently, then a reasonable number of those 23 subjects should treat it differently; in fact only one does so - 22 of those 23 subjects value the unknown urn the same as the others (they set V_3=V_4=V_1=V_2).

The data can be analyzed in more detail, using data on individual subjects' valuations that is reported in Halevy (2005, 2007b; see appendix below). There is considerable noise in individual valuations, but two conclusions stand out: First, a large fraction of subjects value the simple urn differently than the other urns, both the unknown and the mixture urns. This is obvious from table 1 but is reinforced by a more detailed analysis. The second conclusion - more important for the present argument - is that the majority of subjects treat the unknown or ambiguous urn the same as the mixture urns. In other words there is no evidence that subjects treat mixture and unknown urns differently. The data seem to tell us that the Ellsberg paradox is "why do subjects value mixture distributions (compound lotteries) differently than simple mixtures?" This does not require any conjectured "ambiguity." For this paradox the explanation may be two-fold: First, some subjects are simply mistaken in their calculation of probability. Second, as proposed above, some subjects may extend from a single draw to a repeated draw framework.

Conclusion

Personally, I find the behavior Ellsberg describes for his examples convincing, but I find the "paradox" unconvincing. I am not confident, as is Ellsberg, in concluding that the behavior violates the Savage axioms and requires a concept of non-computable "ambiguity" distinct from computable "risk". Ellsberg states, referring to those who follow the frequently-observed choices in the thought experiments above:

> It follows from the propositions above that for their behavior in the situations in question, the Bayesian or Savage approach gives wrong predictions and, by their lights, bad advice. They act in conflict with the axioms deliberately, without apology, because it seems to them the sensible way to behave. Are they clearly mistaken? (Ellsberg, 1961, p. 669)

I am not convinced that the thought experiments actually do conflict with the axioms. Maybe I (and others) unwittingly or unconsciously apply our knowledge about the differences between the distributions for repeated games, where standard risk aversion and the higher dispersion for winnings from urn 2 (in example 1) dictate that I shun those choices. I am not convinced that the paradoxes require "ambiguity" since standard "risk" and risk-aversion would explain the results - if I am inappropriately but unconsciously extending the thought experiments to repeated games. Furthermore, the experimental evidence provided by Halevy (2007) argues against a concept of ambiguity or non-computable uncertainty.

Appendixes

Appendix A - Analysis of Halevy Data

Halvey (2005 or 2007b) provide the data on individual subjects' valuations of the different urns. The experiments uncovered absolute valuations (certainty equivalents) for the four urns.[1] These asbolute valuations are problematic.

For example, for round one the majority of those subjects who were "ambiguity averse" (V_2<V_1) valued urn 1 considerably higher than the expected value of $1.00 - the average valuation was $1.17 and the median $1.10. Remember that this is the simple urn - no mixtures, no uncertainty versus risk. If we were to assume subjects are expected utilitiy maximizers, correctly calculate probabilities, and follow the experimental instructions, these high certainty equivalents seem to imply an unreasonably high degree of risk loving among the subjects. Across all round one subjects the average certainty equivalent was $1.06 (above the expected value) and only 32% of subjects exhibited risk aversion.

The relative valuations may be more reliable. It may be that subjects were better able to value urns relative to each other than absolutely. Thus I focus on the relative valutions, the valuation of urns 2, 3, and 4 relative to urn 1. For each of these urns we can categorize the data according to a whether subject has a high valuation for the simple urn (hate mixtures, $V_3<V_1$, $V_4<V_1$, or hate ambiguity, $V_2<V_1$), neutral valuation ($V_3=V_1$, $V_4=V_1$, or $V_2=V_1$), or low valuation (love mixtures, $V_3>V_1$, $V_4>V_1$, or love ambiguity, $V_2>V_1$). Such a partition of the data gives three categorical variables (C_3 for V_3 vs. V_1, C_4 for V_4 vs. V_1, C_2 for V_2 vs. V_1), with each categorical variable having three possible outcomes depending on the valuation relative to the simple urn: for example C_3 can be H (for high simple urn valuation or hating mixture urn 3, meaning $V_3<V_1$), N (for neutral, $V_3=V_1$), or L (for low simple urn valuation or loving mixture urn 3, $V_3>V_1$)

The beauty of Halevy's implementation of the Ellsberg experiments is that it allows us to separately study behvior regarding the mixture urns and the unknown urn. First, we can examine how subjects treat mixture distributions or compound lotteries by examining the relative valuations of urns 3 and 4. Second, we can examine how subjects treat uncertainty or ambiguity by examining the relative valuations of urn 2.

Table 2 summarizes the data for each of the urns separately (the marginal distributions).[2] First, we will focus on urns 3 and 4, the mixture or compound lottery urns. According to the rules of the experiments (most notably drawing a single ball) the distribution of payouts on urns 3 and 4 are identical to that for urn 1. Subjects should be neutral regarding these urns since, objectively, they provide the same distribution of payouts as urn 1. The data, however, show that only a small minority of subjects are neutral regarding urn 3 or urn 4. Most subjects hate the mixture urns, and a minority love them. This, to my mind, is the most surprising aspect of Halevy's Ellsberg experiments.

Table 2 - Summary of Valuation of Mixture and Unknown Urns Relative to Simple Urn

	C3 (urn 3 vs. urn 1)			C4 (urn 4 vs. urn 1)			C2 (urn 2 vs. urn 1)		
	Hate	Neut	Love	Hate	Neut	Love	Hate	Neut	Love
Count	75	32	35	67	47	28	88	28	26
Percent	52.8	22.5	24.6	47.2	33.1	19.7	62.	19.7	18.3

From Halevy (2005, 2007b). The table combines the two samples reported by Halevy. C_3 is the categorical variable for the valuation of urn 3 (mixture distribution or compound lottery) versus urn 1 (simple distribution). "Hate" means $V_3<V_1$ (the subject hates the mixture relative to the simple distribution, or values the simple urn higher than the mixture urn); "Neut" mean , $V_3=V_1$ (the subject is neutral with respect to the mixture); "Love" means , $V_3>V_1$ (the subject loves the mixture relative to the simple distribution, or values the simple urn lower than the mixture urn). The definitions for C_4 and C_2 are analogous.

Remember that urns 3 and 4 are constructed so that they will produce the same distribution of payoffs as urn 1 for a single draw so that, objectively, there is no difference between urns 1, 3, and 4. In spite of this only a small minority of subjects value urns 3 and 4 the same as urn 1 (are neutral). This is pretty extraordinary, and means that most subjects for some reason do not correctly calculate the probabilities or do not follow the experimental instructions. This is a paradox.

Next, we turn to urn 2, the unknown urn. If there were such a thing as ambiguity separate from risk, then we should see two things. First, subjects should value the unknown urn differently than the known urn. In other words, subjects should not be neutral regarding urn 2. Second, and what seems to be often ignored, is that we should see subjects treating the unknown urn differently than the known mixture urns, 3 and 4. We have already seen that most subjects are not neutral regarding the known mixtures, and to argue that ambiguity is something different than merely mixtures we require evidence that subjects treat mixtures and ambiguity differently.

When we examine the data in table 2 we do in fact observe that most subjects are not neutral. Only 19.7% of subjects are neutral regarding urn 2, and a majority of 62% hate the unknown urn. This pattern, however, is similar to that observed for urns 3 and 4, and this similarity suggests that subjects treat the unknown urn similarly to the mixture urns. Nonetheless, the data in table 2 could be consistent with subjects treating the unknown urns differently. There are fewer subjects neutral regarding urn 2 than regarding urns 3 or 4, and it could be, for example that all 26 subjects who love the unknown urn (C_2=L or V_2>V_1) simultaneously hate mixtures (C_3 or C_4 = H) - they could treat mixtures and the unknown urn very differently. More generally, the joint probabilities for hating versus loving the mixture and unknown urns could be independent, even given the similar pattern in the marginals. More detailed analysis of the data, however, show pretty definitevly that subjects treat the unknown urn similarly to the mixture urns. It is to this analysis that we now turn.

To examine the joint distribution of the mixture and unknown urns we turn to a contingency table analysis. We have three variables (C_3, C_4, C_2) and each variables has three levels (*H*ate, *N*eutral, *L*ove). This is a three-dimensional categorical analysis: three variables, three categories each, or 3×3×3. We could perform a multidimensional contingency table analysis (see, for example, Goodman, 1970, 1971) but the analysis is more clear if we instead consider urn 2 vesus urns 3 and 4 separately: separate two-dimensional 3×3 contingency tables: (C_3 vs. C_2 and C_4 vs. C_2). The data are so conclusive that we gain considerably in clarity while losing little in rigor.

Before discussing urn 2, however, we will compare urn 3 versus urn 4. Objectively, subjects should treat both urn 3 and urn 4 as neutral, but since they do not we should test to see if they treat the two urns the same (relative to urn 1). If they do treat them the same then this is, at the least, suggestive that subjects generally treat mixtures similarly and different from simple distributions.

Table 3 is a contingency table for C_3 vs. C_4 and compares how subjects treat the two mixture or compound lottery urns. The first cell of the left panel of table 3 shows that the number of subjects who hated urn 3 (C_3=H or V_3<V_1) and simultaneously hated urn 4 was 47. The rest of the cells show the other eight possibilities. The middle panel of the table shows the expected number of subjects in each cell, conditional on the marginals and assuming that probabilities for the relative valuations of urn 3 and urn 4 were independent. The right panel of the table shows the actual minus expected (squared) divided by the expected number. These elements sum to give the standard contingency table test statistic that is, in large samples, distributed as a chi-squared variate. The large sample assumption is not necessarily appropriate here, but the individual cells serve to indicate where the joint distribution most deviates from independence. In other words, large values in the cells of this right panel indicate that the actual number in the left panel is either particularly greater or less than would be expected if subjects treated urn 3 and urn 4 independently. The right panel shows that it would be particularly unexpected (given the marginals and if subjects treated the urns independently) to observe the 23 subjects who treat both urn 3 and urn 4 as neutral .

Table 3 - Contingency Table Analysis of the Relative Valuation for Mixture Urn 3 (C_3) versus Mixture Urn 4 (C_4)

	Urn 4					Expected					(Act−Exp)^2/Exp			
Urn 3	Hate	Neut	Love	Total	Urn 3	Hate	Neut	Love	Total	Urn 3	Hate	Neut	Love	Total
Hate	47	17	11	75	Hate	35.4	24.8	14.8	75.	Hate	3.8	2.5	1.	
Neut	5	23	4	32	Neut	15.1	10.6	6.3	32.	Neut	6.8	14.5	0.8	
Love	15	7	13	35	Love	16.5	11.6	6.9	35.	Love	0.1	1.8	5.4	
Total	67	47	28	142	Total	67.	47.	28.	142.	Total				36.7
	Exact	$=5\times10^{-7}$									Asym=		$2.\times10^{-7}$	

Underlying data from Halevy (2005, 2007b). The table combines the two samples reported by Halevy. "Urn 3" is C_3, the categorical variable for the valuation of urn 3 (mixture distribution or compound lottery) versus urn 1 (simple distribution). "Hate" means V_3<V_1 (the subject hates the mixture relative to the simple distribution, or values the simple urn higher than the mixture urn); "Neut" mean , V_3=V_1 (the subject is neutral with respect to the mixture); "Love" means , V_3>V_1 (the subject loves the mixture relative to the simple distribution, or values the simple urn lower than the mixture urn). The definition for "Urn 4" is analogous. The left panel is the actual counts. "Exact" is Fisher's exact test (2-sided) showing the probability level for the hypothesis that the two categorical variables are independent. The middle panel ("Expected") is the expected counts conditional on the row and column counts, assuming independence of the two distributions. The right panel ("Act-Exp)^2/Exp") is the actual less expected (squared) divided by expected counts. This is summed to give the standard contingency table test statistic that is asymptotically distributed as χ^2. The table itself shows what cells contribute to the statistic. The asymptotic significance level of the statistic is shown as "Asym".

Although there is noise in the data, we can strongly reject the hypothesis that people treat urns 3 and 4 independently. Conditional on the marginals (i.e. conditional on how many people hate or love urn 4, and also how many people hate or love urn 3) the probability that we would observe the actual counts if subjects treated urns 3 and 4 independently would be very small: 5×10^{-7}. Subjects treat the two urns similarly, either hating or loving both simultaneously. The actual count on the diagonals, where subjects valued both in the same way, are higher than expected. The off-diagonals, where subjects view urns 3 and 4 differently, are consistently less than expected (conditional on the marginals and assuming independence). The conclusion is that subjects generally treat both mixtures the same (relative to the simple urn 1).

Turning to the unknown urn, table 4 shows the contingency table analysis for C_3 (mixture urn) vs. C_2 (unknown urn). Here we are looking for evidence that subjects treat the unknown urn differently from the mixture urn. To argue for a concept of ambiguity we would require some evidence of independence between how subjects treat urn 3 and urn 2. If subjects treat urn 2 the same as urn 3 then any argument in favor of amibguity must be based on something other than empirical evidence.

Table 4 - Contingency Table Analysis of the Relative Valuation for Mixture Urn 3 (C_3) versus Unknown Urn 2 (C_2)

	Urn 2					Expected					(Act−Exp)^2/Exp			
Urn 3	Hate	Neut	Love	Total	Urn 3	Hate	Neut	Love	Total	Urn 3	Hate	Neut	Love	Total
Hate	65	3	7	75	Hate	46.5	14.8	13.7	75.	Hate	7.4	9.4	3.3	
Neut	6	25	1	32	Neut	19.8	6.3	5.9	32.	Neut	9.6	55.4	4.	
Love	17	0	18	35	Love	21.7	6.9	6.4	35.	Love	1.	6.9	21.	
Total	88	28	26	142	Total	88.	28.	26.	142.	Total				118.
	Exact		$<2\times10^{-16}$								Asym=		0.	

Underlying data from Halevy (2005, 2007b). The table combines the two samples reported by Halevy. "Urn 3" is C_3, the categorical variable for the valuation of urn 3 (mixture distribution or compound lottery) versus urn 1 (simple distribution). "Hate" means $V_3<V_1$ (the subject hates the mixture relative to the simple distribution, or values the simple urn higher than the mixture urn); "Neut" mean , $V_3=V_1$ (the subject is neutral with respect to the mixture); "Love" means , $V_3>V_1$ (the subject loves the mixture relative to the simple distribution, or values the simple urn lower than the mixture urn). The definition for "Urn 2" is analogous. The left panel is the actual counts. "Exact" is Fisher's exact test (2-sided) showing the probability level for the hypothesis that the two categorical variables are independent. The middle panel ("Expected") is the expected counts conditional on the row and column counts, assuming independence of the two distributions. The right panel ("Act-Exp)^2/Exp") is the actual less expected (squared) divided by expected counts. This is summed to give the standard contingency table test statistic that is asymptotically distributed as χ^2. The table itself shows what cells contribute to the statistic. The asymptotic significance level of the statistic is shown as "Asym".

The conclusion here is definitive - the distributions are not independent and subjects treat the unknown urn like mixture urn 3. The probability that we would observe the actual counts in the left panel if the joint probabilities of C_3 and C_2 were independent is less than 2×10^{-16}- a very small number. The right panel shows the contributions to the asumptotic χ^2-variate and we see that the most striking observation is the number of subjects who are neutral. Almost all the subjects who are neutral to either urn are in fact neutral to both the mixture urn and the unknown urn (25 are neutral to both). If the distributions were independent, then we would expect many fewer subjects who were neutral to both (only 6.3, conditional on the large number of subjects who hate urn 3 or urn 2). The data unambiguously reject the notion that subjects treat the unknown urn 2 differently than the mixture urn 3.

The final set of data compare the relative valuation of urn 4 and urn 2. Table 5 shows that, again, the data reject the hypothesis that subjects treat the mixture urn 4 and the unknown urn 2 differently.

Table 5 - Contingency Table Analysis of the Relative Valuation for Mixture Urn 4 (C_4) versus Unknown Urn 2 (C_2)

	Urn 2			
Urn 4	Hate	Neut	Love	Total
Hate	57	3	7	67
Neut	17	24	6	47
Love	14	1	13	28
Total	88	28	26	142
	Exact		$=7\times10^{-12}$	

	Expected			
Urn 4	Hate	Neut	Love	Total
Hate	41.5	13.2	12.3	67.
Neut	29.1	9.3	8.6	47.
Love	17.4	5.5	5.1	28.
Total	88.	28.	26.	142.

	(Act−Exp)^2/Exp			
Urn 4	Hate	Neut	Love	Total
Hate	5.8	7.9	2.3	
Neut	5.	23.4	0.8	
Love	0.6	3.7	12.1	
Total				61.6
		Asym=	$1.\times10^{-12}$	

Underlying data from Halevy (2005, 2007b). The table combines the two samples reported by Halevy. “Urn 4” is C_4, the categorical variable for the valuation of urn 4 (mixture distribution or compound lottery) versus urn 1 (simple distribution). “Hate” means $V_4<V_1$ (the subject hates the mixture relative to the simple distribution, or values the simple urn higher than the mixture urn); “Neut” mean , $V_4=V_1$ (the subject is neutral with respect to the mixture); “Love” means , $V_4>V_1$ (the subject loves the mixture relative to the simple distribution, or values the simple urn lower than the mixture urn). The definition for “Urn 2” is analogous. The left panel is the actual counts. “Exact” is Fisher’s exact test (2-sided) showing the probability level for the hypothesis that the two categorical variables are independent. The middle panel (“Expected”) is the expected counts conditional on the row and column counts, assuming independence of the two distributions. The right panel (“Act-Exp)^2/Exp”) is the actual less expected (squared) divided by expected counts. This is summed to give the standard contingency table test statistic that is asymptotically distributed as χ^2. The table itself shows what cells contribute to the statistic. The asymptotic significance level of the statistic is shown as “Asym”.

The data show definitively that subjects do not treat the unknown urn differently from the mixture urns. It is true that not all subjects treat the unknown urn exactly like either urn 3 or urn 4, but then we would hardly expect them to. First, as shown in table 3, not all subjects treat urns 3 and 4 the same (and remember that, objectively, urns 3 and 4 do provide identical payouts for a single draw); we should hardly expect more consistency for the unknown urn. In fact, examining tables 3 through 5 indicates more consistency for the unknown urn than between the two mixture urns themselves. (The significance level for rejecting independence between urns 3 and 4 - table 3 - is higher than the significance level for rejecting independence between the unknown urn and either urn 3 - table 4 - or urn 4 - table 5). Second, the unknown urn is unknown, and we should expect heterogeneity among subjects - we could hardly expect all subjects to treat it as exactly the same mixture.

One observation that might seem anomalous in tables 4 and 5 is the number of subjects who hate urn 2 but love either urn 3 (17 subjects shown in table 4) or urn 4 (14 subjects in table 5). These actual counts are less than would be expected given independence (17 versus 21.7 for urn 3 from table 4 and 14 versus 17.4 from table 5) but intuitively it seems odd that so many people hate urn 2 but love the mixture urns. Part of the answer lies in examining the joint distribution in more detail, where we see that there is little overlap across those subjects. Some subjects who hate urn 2 also hate urn 3 but love urn 4 or vice versa, but few subjects love both urn 3 and urn 4. Of the 17 subjects who hate urn 2 and love urn 3, only four love both urn 3 and urn 4 and 10 in fact hate urn 4.

The data show that most subjects treat the unknown urn as a mixture urn, but that still leaves the question of why subjects treat the mixture urns as they do. According to the rules of the experiment, the mixture urns have the same value as the simple urn 1. All subjects should be neutral and no subjects should either hate or love the mixture urns, urns 3 and 4. In fact, only a small minority of subjects are neutral. There seem two possible explanations for why most subjects are not neutral:

- Subjects miscalculate, misunderstand, or otherwise are mistaken in their treatment of probability
- Subjects consider the “meta-experiment” and assign a non-zero probability to repeated draws

Both of these explanations seem plausible. It may be that some of the subjects were simply not very good at using or thinking about probability (and the data on absolute valuations certainly lend credence to this idea). This is hardly surprising. Working with and thinking about probability is often difficult. Gigerenzer (2002, p. 44 ff) discusses how trained medical professionals commonly make mistakes in dealing with probability problems that often arise in the medical world. Feller (1968, pp. 86-88) discusses how intuition can often mislead even trained probabilists. Probabilistic thinking is often difficult, but that does not make it any less useful as a model of the world.

Alternatively, some of Halevy's subjects may consider the “meta-experiment” and assign a non-zero probability to repeated draws. I argued above that a risk averse expected-utility maximizer who does consider repeated draws would choose a simple urn over both a computable mixture and an unknown urn. (The argument could be extended to consider a risk-loving agent, who would prefer a mixture and an unknown urn. It could also be extended to alternative multiple-draw scenarios where a risk averse agent would actually prefer a mixture urn. In other words it is possible to produce agents who love mixtures and unknown urns.) The important point is that the meta-experiment approach predicts agents will generally treat both mixtures and unknown urns the same. This is exactly what the data do show.

Appendix B - Ellsberg Example 2

Statement of Example 2

Ellsberg presented two versions of his urn experiment. The Ellsberg urn experiment version 2 is:

- Single urn - 90 balls, 30 red and 60 some combination of black and yellow with proportions not specified

The "experiment" is to draw a single ball, with payoffs defined as:

I. "Payoff on Red": Receive $1 if Red, $0 if Black or Yellow
II. "Payoff on Black": Receive $0 if Red, $1 if Black, $0 if Yellow
III. "Payoff on Red or Yellow": Receive $1 if Red, $0 if Black, $1 if Yellow
IV. "Payoff on Black or Yellow": Receive $0 if Red, $1 if Black or Yellow

This can be represented as

	30	60	
	Red	Black	Yellow
I	$1	$0	$0
II	$0	$1	$0
II	$1	$0	$1
IV	$0	$1	$1

Gambles posed:

1. Which do you prefer: "Payoff on Red" vs. "Payoff on Black" (I vs. II)
2. Which do you prefer: "Payoff on Red/Yellow" vs. "Payoff on Black/Yellow" (III vs. IV)

Ellsberg says a frequent pattern of response is I preferred to II, IV preferred to III. Roughly speaking I preferred to II means $P[B] < 1/3$ while IV preferred to III means $P[B] > 1/3$ - inconsistent.

Modification to Repeated Game

The pattern of responses Ellsberg discusses is inconsistent for a single-shot experiment. But once again if we extend to a repeated game the pattern becomes not only reasonable but required for consistency with expected utility maximization (for a risk-averse agent). Again, let us start with a two-point mixture; alter the single urn to be:

- Single urn - either Urn a or Urn b. Each urn has 90 balls, 30 Red and 60 Black / Yellow. For Urn a the proportions of Black / Yellow are 15 Black / 45 Yellow and for Urn b they are 45 Black / 15 Yellow. The urn actually used is chosen by flip of a fair coin, meaning that each urn has probability 1/2. The subject does not know a priori which urn will be used but can verify the mix of balls in each urn and can verify the coin and the fairness of the flip.

The payoffs are the same, simply re-titled "Win" for the repeated game:

I. "Win on Red": Receive $1 if Red, $0 if Black or Yellow
II. "Win on Black": Receive $0 if Red, $1 if Black, $0 if Yellow
III. "Win on Red or Yellow": Receive $1 if Red, $0 if Black, $1 if Yellow
IV. "Win on Black or Yellow": Receive $0 if Red, $1 if Black or Yellow

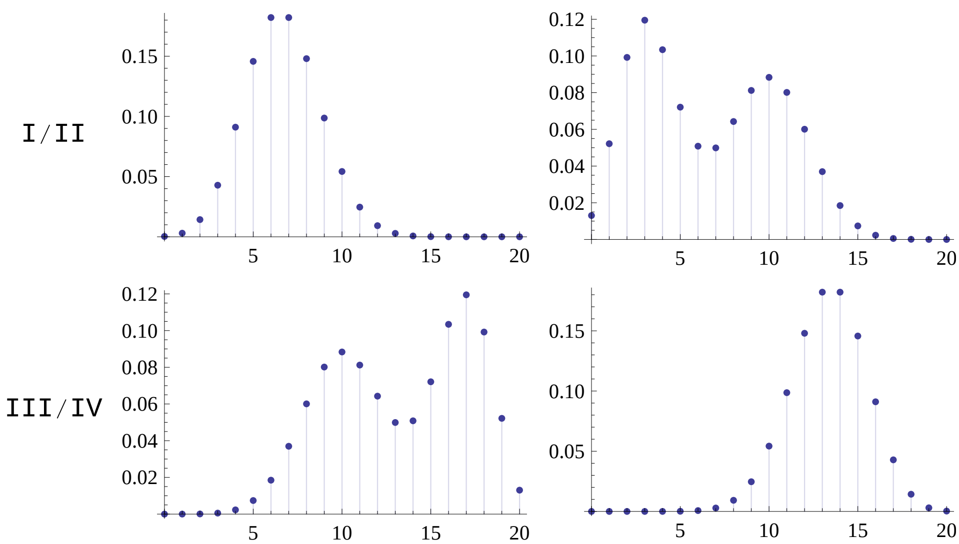
The two comparisons are now repeated “games”

1. Which do you prefer when playing 20 repeats: “Win on Red” vs. “Win on Black” (I vs. II)
2. Which do you prefer when playing 20 repeats: “Win on Red/Yellow” vs. “Win on Black/Yellow” (III vs. IV)

Distribution of Winnings for Repeated Game

The distributions for the four possibilities (I, II, III, IV) now show that a risk-averse agent should prefer I to II and IV to III. There is no inconsistency in these choices, it simply reflects the greater dispersion of the densities for II and III. Once again, anyone who subconsciously extended the single-shot experiment to a repeated game would naturally (and uncontroversially) show the pattern of preferences Ellsberg discussed for the single-shot experiment. Given that there is no cost to choosing I over II or IV over III, it hardly seems surprising that most people would exhibit these preferences.

Densities for Example 2, 20 repeats




```
Mean winnings for 20 repeats
       I/II   6.7   6.7
     III/IV  13.3  13.3

Standard Deviation for 20 repeats
          I/II   2.1  3.9
        III/IV   3.9  2.1
```

References

Ellsberg, Daniel (1961), “Risk, Ambiguity, and the Savage Axioms,” *The Quarterly Journal of Economics*, vol. 75 no. 4 (November 1961)

Epstein, Larry G. (1999), “A Definition of Uncertainty Aversion,” *The Review of Economic Studies*, vol. 66 no. 3 (July 1999)

Feller, William (1968), *An Introduction to Probability Theory and Its Applications, Volume 1*, 3rd ed. John Wiley & Sons, New York

Halevy, Yoram (2005), “Ellsberg Revisited: An Experimental Study,” (July 17, 2005). Available at SSRN: http://ssrn.com/abstract=770964

Halevy, Yoram (2007), “Ellsberg Revisited: An Experimental Study,” *Econometrica*, vol. 75, no. 2, pp. 503-536 (March 2007)

Halevy, Y. (2007b): “Supplement to ‘Ellsberg Revisited: An Experimental Study’ - Instructions and Data,” *Econometrica Supplementary Material*, 75, (www.econometricsociety.org, then search for “Ellsberg”)

Keynes, John Maynard (1921), *A Treatise on Probability*, Macmillan, London

Knight, Frank (1921), Risk, *Uncertainty, and Profit*, Houghton Mifflin Co., Boston

LeRoy, Stephen F., and Larry D. Singell Jr. (1987), “Knight on Risk and Uncertainty," *Journal of Political Economy*, vol. 95 no. 2 (April 1987)

Schmeidler, David (1989), “Subjective Probability and Expected Utility Without Additivity,” *Econometrica*, vol. 57 no. 3 (May 1989)

Notes

1 Halevy performed two rounds of experiments that differed primarily in the size of the stakes - the second round being ten times larger - and some details of recruitment and training of subjects. These difference mean the absolute valuations of the second round are more likely to be reliable. Nonetheless, the pattern of relative valuations discussed in the tables above is consistent across both rounds.

2 This is for both rounds combined. Analysis of rounds 1 and 2 separately lead to the same conclusions.