

# The Ellsberg paradox: A challenge to quantum decision theory?\*



**Ali al-Nowaihi**, University of Leicester  
**Sanjit Dhami**, University of Leicester

# The Ellsberg paradox: A challenge to quantum decision theory?<sup>\*</sup>

Ali al-Nowaihi<sup>†</sup>      Sanjit Dhami<sup>‡</sup>

11 July 2016

## Abstract

We set up a simple quantum decision model of the Ellsberg paradox. We find that the matching probabilities that our model predict are in good agreement with those empirically measured by Dimmock et al. (2015). Our derivation is parameter free. It only depends on quantum probability theory in conjunction with the heuristic of insufficient reason. We suggest that much of what is normally attributed to probability weighting might actually be due to quantum probability.

*Keywords:* Quantum probability; Ellsberg paradox; probability weighting; matching probabilities; projective expected utility; projective prospect theory.

*JEL Classification:* D03 (Behavioral microeconomics: Underlying principles).

---

\*We are grateful for valuable and critical comments from Jerome Busemeyer, Andrew Colman, Ehtibar Dzhafarov, Emmanuel Haven, Andrei Khrennikov, Jesse Matheson, Emi Mise, Briony Pulford, Sandro Sozzo, Peter Wakker, Mengxing Wei and two anonymous reviewers.

<sup>†</sup>Corresponding author. Department of Economics, University of Leicester, University Road, Leicester. LE1 7RH, UK. Phone: +44-116-2522898. Fax: +44-116-2522908. E-mail: aa10@le.ac.uk.

<sup>‡</sup>Department of Economics, University of Leicester, University Road, Leicester, LE1 7RH. Tel: +44-116-2522086. E-mail: sd106@le.ac.uk.

# 1 Introduction

Consider the following version of the Ellsberg experiment<sup>1</sup> due to Dimmock et al. (2015). This involves two urns: The known urn ( $K$ ) contains 100 balls of  $n$  different colors,  $1 < n \leq 100$ , the same number of balls for each color (for example, if  $n = 5$  then there are 5 different colors and 20 balls of each color in  $K$ ). The unknown urn ( $U$ ) also contains  $n$  balls of the same colors as urn  $K$  but in unknown proportions. The subject is asked to select one of the urns ( $K$  or  $U$ ). A ball is drawn at random from the urn chosen by the subject. There are two versions. In the low probability version, the subject wins a sum of money if the color of the ball drawn matches a preassigned color (which, however, could be chosen by the subject). In the high probability version, the subject wins the sum of money if the color of the randomly drawn ball matches any one of  $n - 1$  preassigned colors (again, these colors could be chosen by the subject). These two versions are, of course, equivalent if  $n = 2$ , but different for  $n > 2$ . The subject is also allowed to declare indifference between  $K$  and  $U$ .

If a subject prefers  $K$  to  $U$ , she is called *ambiguity averse*. If she prefers  $U$  to  $K$ , she is called *ambiguity seeking*. If she is indifferent between  $K$  and  $U$  she is called *ambiguity neutral*.

Dimmock et al. (2015) perform a second set of experiments. Here the ratio of the colors (whatever they are) in  $U$  were kept fixed. However, the ratio in  $K$  was varied until a subject declared indifference. This ratio is then called the *matching probability*. For example, in the low probability treatment, they found that for  $n = 10$  colors, subjects (on average) declared indifference between  $K$  and  $U$  when the new urn  $K$  contained 22 balls (out of 100) of the winning color. Hence the matching probability of 0.1 is  $m(0.1) = 0.22 > 0.1$ . Thus subjects exhibited ambiguity seeking for the low probability of 0.1. In the high probability treatment, they found that, again for  $n = 10$  colors, subjects (on average) declared indifference when the new urn  $K$  contained 69 balls of the winning colors. Hence, the matching probability of 0.9 is  $m(0.9) = 0.69 < 0.9$ . For  $n = 2$  colors, subjects (on average) declared indifference when the new urn  $K$  contained 40 balls of the winning color. Hence,  $m(0.5) = 0.4 < 0.5$ . Thus, subjects exhibited ambiguity aversion for medium and high probabilities but ambiguity seeking for low probabilities.

The reason why preferring  $K$  to  $U$  (or  $U$  to  $K$ ) was regarded as para-

---

<sup>1</sup>Keynes (1921) and Ellsberg (1961, 2001).

doxical<sup>2</sup> is as follows. Although experimental subjects know the proportion of colors in urn  $K$  (it contains exactly the same number of each color), they do not know the ratio in urn  $U$ . But they have no reason to believe that one color is more likely than another. Hence, by the heuristic of *insufficient reason* (or *equal a' priori probabilities*)<sup>3</sup>, they should assign the same probability to each color in urn  $U$ .<sup>4</sup> Hence, they should have no reason to prefer  $K$  to  $U$  or  $U$  to  $K$  on probabilistic grounds. Keynes (1921) pointed out that there is a difference in the strength or quality of the evidence. Subjects may reason that, although the assignment of the same probability to each color is sound, they are more confident in the correctness of this judgement in the case of  $K$  than in the case of  $U$ . Hence, they prefer  $K$  to  $U$ . Thus their preference works through the utility channel rather than the probability channel. However, this explanation appears to be contradicted by the evidence of Dimmock et al. (2015) that subjects are ambiguity seeking for low probabilities. Moreover, even when subjects are told that each color in  $U$  has the same probability, so that the heuristic of insufficient reason is not needed, they still exhibit a preference for  $K$  over  $U$  (Rode et al. 1999). Furthermore, because the probabilities in urn  $U$  have been revealed, the observed choice of  $K$  over  $U$  cannot be attributed to ambiguity aversion or differences in the strength or quality of the evidence.

The importance of the Ellsberg experiments is two fold. First, they provide tests for competing decision theories. Second, there are many real-world situations that appear similar to the Ellsberg paradox. One example is that of *home-bias* in investment (French and Poterba, 1991, Obstfeld and Rogoff, 2000). Investors are often observed to prefer investing in a domestic asset over a foreign asset with the same return and the same riskiness.

La Mura (2009) proposed to replace standard (Kolmogorov) probabilities in expected utility theory with quantum probabilities; and called the resulting decision theory *projective expected utility theory*. He gave an axiomatic foundation for this new theory and derived the equivalence of the preference representation and the utility representation. He applied the new theory to

---

<sup>2</sup>This was the situation before the advent of the source method, see subsection 3.5 below.

<sup>3</sup>*Insufficient reason* or *equal a' priori probabilities* in now commonly referred to as *indifference*. However, indifference has a well established alternative meaning in economics. To avoid confusion, we shall use the older terminology.

<sup>4</sup>The same reasoning can be repeated within any particular source in source dependent theory (see Section 3.5, below). So we have to take  $K$  and  $U$  as different sources.

explain the Allais paradox. He suggested it may explain the Ellsberg paradox.

Busemeyer and Bruza (2012, section 9.1.2) applied projective expected utility theory to explain the Ellsberg paradox. Their model has a free parameter,  $a$ . If  $a > 0$  we get ambiguity aversion, if  $a = 0$ , we get ambiguity neutrality, and if  $a < 0$  we get ambiguity seeking. However, it cannot explain the simultaneous occurrence in the same subject of ambiguity seeking (for low probabilities), ambiguity neutrality and ambiguity aversion (for medium and high probabilities), because  $a$  cannot be simultaneously negative, zero and positive. By contrast, our model (Section 5, below) provides a parameter-free derivation of quantum probabilities and can explain the simultaneous occurrence in the same subject of ambiguity seeking (low probabilities), ambiguity neutrality and ambiguity aversion (medium and high probabilities). Its predictions are in good agreement with the empirical evidence in Dimmock et al. (2015).

Busemeyer and Bruza (2012, section (9.1.2) conclude “In short, quantum models of decision making can accommodate the Allais and Ellsberg paradoxes. But so can non-additive weighted utility models, and so these paradoxes do not point to any unique advantage for the quantum model”. Note, however, that there is considerable arbitrariness in the choice of weights in weighted utility models. Hence they introduce flexibility at the cost of lower predictive power. In our model, we replace weights with quantum probabilities which are parameter-free. Thus, our application of projective expected utility theory has a clear advantage over all other decision theories. Furthermore, projective expected utility can be extended to include reference dependence and loss aversion, to yield *projective prospect theory*, where decision weights are replaced with quantum probabilities. This would have a clear advantage over all the standard (non-quantum) versions of prospect theory.

Aerts et al. (2014) formulate and study a quantum decision theory (QDT) model of the Ellsberg paradox. They consider one of the standard versions of the Ellsberg paradox. They consider a single urn with 30 red balls and 60 balls that are either yellow or black, the latter in unknown proportions. They use the heuristic of insufficient reason for the known distribution (red) but not for the unknown distribution (yellow or black). They prove that in their mode, the Ellsberg paradox reemerges if they use the heuristic of insufficient reason (or equal a' priori probabilities) for the unknown distribution. They, therefore, abandon this heuristic. They choose the ratio of yellow to black

to fit the evidence from their subjects.

Although abandoning the heuristic of insufficient reason gives models tremendous flexibility, it also reduces their predictive power. In both classical (Kolmogorov) probability theory and quantum probability theory any probabilities (provided they are non-negative and sum to 1) can be assigned to the elementary events. To make a theory predictive, some heuristic rule is needed to assign a' priori probabilities (we call this a heuristic because it does not follow from either classical or quantum probability theory). The heuristic commonly used is that of insufficient reason or equal a' priori probabilities.<sup>5</sup> This heuristic is crucial in deriving the Maxwell-Boltzmann distribution in classical statistical mechanics and the Bose-Einstein and Fermi-Dirac distributions in quantum statistical mechanics.<sup>6</sup> Furthermore, other theories can explain the Ellsberg paradox if we abandon *insufficient reason* (see Subsection 2.3). Thus, the explanation of Aerts et al. (2014) is not specifically quantum, although it is expressed in that language.

Khrennikov and Haven (2009) provide a general quantum-like framework for situations where Savage's sure-thing principle (Savage, 1954) is violated; one of these being the Ellsberg paradox. Their *quantum-like* or *contextual probabilistic (Växjö) model* is much more general than either the classical Kolmogorov model or the standard quantum model (see Khrennikov, 2010, and Haven and Khrennikov, 2013). On the other hand, our approach is located strictly within standard quantum theory. Furthermore, in their formulation, the Ellsberg paradox reemerges if one adopts (as we do) the heuristic of insufficient reason.<sup>7</sup>

We set up a simple quantum decision model of the Ellsberg paradox. We argue that our quantum decision model, in conjunction with the heuristic of insufficient reason, is in broad conformity with the evidence of Dimmock et al. (2015). In the table below, the second column gives the means across 666 subjects of the observed matching probabilities for 0.1 , 0.5 and 0.9. The third column gives the sample standard deviations. The fourth column gives the theoretical predictions of our model.

Our theoretical predictions of  $m(0.5)$  and  $m(0.9)$  are in excellent agreement with the average of observations. Our theoretical prediction of  $m(0.1)$

---

<sup>5</sup>To be sure, this heuristic is not without problems. See, for example, Gnedenko (1968), sections 5 and 6, pp 37 to 52.

<sup>6</sup>See Tolman, 1938, section 23, pp 59-62, for a good early discussion.

<sup>7</sup>Khrennikov and Haven (2009), subsection 4.6, p386.

| Matching probability | Mean | Standard deviation | Theoretical prediction |
|----------------------|------|--------------------|------------------------|
| $m(0.1)$             | 0.22 | 0.25               | 0.171 05               |
| $m(0.5)$             | 0.40 | 0.24               | 0.416 67               |
| $m(0.9)$             | 0.69 | 0.33               | 0.695 45               |

Table 1: Actual and predicted matching probabilities. Source: Dimmock et al. (2015), Table 4.3.

is not statistically significantly different from the average of observed values.<sup>8</sup>

Our model is more parsimonious than the alternatives. Unlike all other decision theory explanations of the Ellsberg paradox, our model is parameter free. Our results follow purely from quantum probability theory and the heuristic of insufficient reason. We think that this suggests that much of what is normally attributed to probability weighting might actually be due to quantum probability.

The fundamental difference between *quantum decision theory* (QDT) and all other decision theories is that events in the latter, but not the former, are distributive. Thus, in QDT the event “ $X$  and ( $Y$  or  $Z$ )” need not be equivalent to the event “( $X$  and  $Y$ ) or ( $X$  and  $Z$ )”. On the other hand, in all other decision theories, these two events are equivalent. This non-distributive nature of QDT is the key to its success in explaining paradoxes of behaviour that other decision theories find difficult to explain. For example, *order effects*, the *Linda paradox*, the *disjunction fallacy* and the *conjunction fallacy* (see Busemeyer and Bruza, 2012, for an introduction and review). For papers examining the limits of standard quantum theory when applied to cognitive psychology, see Khrennikov et al. (2014) and Basieva and Khrennikov (2015).

The rest of the paper is organized as follows. Section 2 gives the main stylized facts about the Ellsberg paradox and formulates a simple thought experiment that is used in the rest of the paper. It also discusses reduction of compound lotteries and the heuristic of insufficient reason. Section 3 reviews the leading explanations of the Ellsberg paradox. Section 4 reviews the elements of quantum probability needed for this paper. The main results of this paper are in Section 5. Section 6 summarizes and concludes.

---

<sup>8</sup>For  $m(0.1)$ ,  $z = \frac{0.22 - 0.17105}{0.25} = 0.1958 < 1.96$ . For such a large sample, the t-distribution is practically normal. Based on the normal test, the evidence does not reject the theoretical prediction  $m(0.1) = 0.17105$  at the 5% level of significance.

## 2 Stylized facts, a thought experiment, insufficient reason and the reduction of compound lotteries

### 2.1 Stylized facts

The following are the main stylized facts of Ellsberg experiments.

1. *Insensitivity*: Subjects are ambiguity averse for medium and high probabilities but ambiguity seeking for low probabilities (see Dimmock et al., 2015, for a recent survey and new experimental results).
2. *Exchangeability*: Subjects are indifferent between colors. Subjects are indifferent between being asked to choose a color first or an urn first (Abdellaoui et al., 2011).
3. *No error*: Suppose a subject chooses one urn ( $K$  or  $U$ ) over the other. It is then explained to the subject that, according to classical probability theory, she should have been indifferent. She is offered the chance to revise her assessment. Subjects usually decline to change their assessment (Curley et al., 1986).
4. *Salience*: Ambiguity aversion is stronger when the two urns are presented together than when they are presented separately. (Fox and Tversky, 1995, Chow and Sarin, 2001, 2002).<sup>9</sup>
5. *Anonymity* (or *fear of negative evaluation*): Ambiguity aversion does not occur if subjects are assured that their choice between urn  $U$  and urn  $K$  is anonymous (Curley et al., 1986 and Trautmann et al., 2008).

In the next section, we shall evaluate the main proposed explanations of the Ellsberg paradox in the light of stylized fact 1. They all satisfy stylized facts 2 and 3. It has been suggested several times in the literature that reference dependence might explain stylized fact 4 (see Chow and Sarin, 2001, 2002). None address stylized fact 5, nor do we.

---

<sup>9</sup>Even more strikingly, Fox and Tversky (1995) found that for probability  $\frac{1}{2}$ , subjects exhibited ambiguity aversion with the value of urn  $U$  remaining approximately the same but urn  $K$  revalued upwards. Chow and Sarin (2001, 2002) did not find this result, but did find that ambiguity aversion is more pronounced when subjects are presented with  $K$  and  $U$  together.

## 2.2 A thought experiment

Throughout the rest of the paper, we consider the following simplified version of the experiments in Dimmock et al. (2015). As far as we know, this simplified experiment has not been conducted. So, the following is merely a thought experiment.

We simplify the experiment in Dimmock et al. (2015) in several steps. First we replace colors by numerals (this is justified by stylized fact 2). Furthermore, we consider only two numerals: 1 and 2. The known urn  $K$  contains  $n$  balls,  $m$  of which are labeled “1” and  $n - m$  are labeled “2”. Thus, by the heuristic of insufficient reason, ball 1 is drawn with probability  $p = \frac{m}{n}$  and ball 2 is drawn with probability  $q = \frac{n-m}{n}$ .<sup>10</sup> In the main, we shall adopt the heuristic of insufficient reason. But, for purposes of comparison, we shall sometimes consider cases where the subject does not apply this heuristic. For example, the subject may be optimistic for low probabilities but pessimistic for high probabilities (Subsection 2.3.4). Our most drastic simplification will be to consider only two stages when constructing the unknown urn  $U$ . With two balls and two stages, we can do our work in a 4 dimensional space.

A subject is presented with two urns. The known urn ( $K$ ) contains exactly two balls, one labeled “1” and the other labeled “2”. Ball 1 is drawn from  $K$  with probability  $p$ . Ball 2 is drawn with probability  $q = 1 - p$ . To compare with the evidence reported in Dimmock et al. (2015), we are primarily interested in  $p = 0.1, 0.5$  and  $0.9$ .

Starting with urn  $K$ , construct urn  $U$  as follows. In the first round, draw a ball at random from  $K$  and place it in  $U$ . Replace that ball in  $K$  with an identically labeled ball. In the second round, draw a second ball at random from  $K$  and place it in  $U$ . Replace that ball in  $K$  with an identically labeled ball. Thus  $U$  contains two ball. Both could be labeled “1”, both could be labeled “2” or one could be labeled “1” and the other labeled “2”. A ball is drawn at random from  $U$ . The subject wins the sum of money  $v > 0$  if ball 1 is drawn but wins nothing if ball 2 is drawn.

For example, and as in Dimmock et al. (2015), assume urn  $K$  contains 100 balls of 10 different colors, 10 balls of each color. This is told to the subjects. The subjects are told that urn  $U$  also contains 100 balls of the same 10 colors as urn  $K$ , but in unknown proportions. However, “unknown proportions” is not defined any further for the subjects. We conjecture that subjects model

---

<sup>10</sup>This transformation is only for analytic convenience. In experiments, subjects are always presented with colored balls whose ratios match the probabilities.

"unknown proportions" in a simple way, for example, as described above.

Our theoretical predictions are in good agreement with the evidence in Dimmock et al. (2015). This could merely be an accident. However, perhaps there is a behavioral explanation. In particular, Dimmock et al. (2015), and many other experiments, describe the unknown urn ( $U$ ) as follows "The unknown urn ( $U$ ) contains  $n$  balls of the same colors as urn  $K$  but in unknown proportions". Maybe, this is too cognitively challenging. Maybe subjects do not consider all possible distributions of balls in urn  $U$ . Pulford and Colman (2008) provide strong evidence for this. There is also much evidence of such cognitive limitations from other areas. For example, in  $p$ -beauty contests, subjects think only up to level- $k$ , with low  $k$ , typically  $k = 2$  (Camerer, 2003). For a classically rational person,  $k$  should be infinite. Similar evidence of cognitive limitations comes from psychological games (Khalmetski et al., 2015). More generally, decision makers often simplify a problem before attempting to find a solution (Kahneman and Tversky, 1979, Thaler, 1999 and Hey et al., 2010).

## 2.3 Insufficient reason and the reduction of compound lotteries

According to Segal (1987, 1990), whether the results of Ellsberg experiments are paradoxical or not for a particular decision theory hinges on how compound lotteries are reduced in that theory. In our view, this is only partly correct. We shall argue that no decision theory that respects both the reduction of compound lotteries and the heuristic of insufficient reason can explain stylized fact 1 (insensitivity). However, relaxing one of these opens the way to explaining the Ellsberg paradox. Consider the thought experiment of Subsection 2.2.

### 2.3.1 The known urn

If the subject chooses urn  $K$ , she faces the simple lottery

$$\mathbf{K} = (p, v),$$

where she wins  $v > 0$  if ball 1 is drawn (with probability  $p$ ), or 0 (with probability  $q$ ), if ball 2 is drawn.

### 2.3.2 The unknown urn

By sketching the decision tree, and using the heuristic of insufficient reason, we can easily see that if the subject chooses urn  $U$ , then she faces the compound lottery

$$\mathbf{U} = \left( p, \left( p, (1, v); q, \left( \frac{1}{2}, v \right) \right); q, \left( p, \left( \frac{1}{2}v \right); q, (0, v) \right) \right).$$

Using the reduction of compound lotteries, we get

$$\mathbf{U} = (p, v).$$

Comparing with the case of the known urn ( $K$ ), we see that

$$\mathbf{U} = \mathbf{K}.$$

Thus, no decision theory that respects both the reduction of compound lotteries and the heuristic of insufficient reason can explain stylized fact 1.

### 2.3.3 Matching probabilities

Now keep the composition of urn  $U$  fixed but vary the probability with which ball 1 is drawn from urn  $K$  until the subject expresses indifference between urn  $K$  with its new composition and the old urn  $U$ . Suppose that indifference is reached when the probability with which ball 1 is drawn from the new urn  $K$  is  $P$  (hence, the probability with which ball 2 is drawn is  $1 - P$ ). Then  $P = m(p)$  is the matching probability. Note that the definition of the matching probability  $m(p)$  is operational and does not depend on the underlying decision theory.

### 2.3.4 Dropping insufficient reason or dropping reduction of compound lotteries?

Consider a decision maker who thinks that low probability events are more likely than what is justified by the heuristic of insufficient reason but that high probability events are less likely. However, assume that she respects reduction of compound lotteries. Suppose she faces the two lotteries

$$\begin{aligned} \mathbf{L} &= \left( \frac{1}{n}, \left( \frac{1}{n}, (1, v); \frac{n-1}{n}, \left( \frac{1}{2}, v \right) \right); \frac{n-1}{n}, \left( \frac{1}{n}, \left( \frac{1}{2}v \right); \frac{n-1}{n}, (0, v) \right) \right) \\ &= \left( \frac{1}{n}, v \right). \end{aligned}$$

$$\begin{aligned}\mathbf{H} &= \left( \frac{n-1}{n}, \left( \frac{n-1}{n}, (1, v) ; \frac{1}{n}, \left( \frac{1}{2}, v \right) \right) ; \frac{1}{n}, \left( \frac{n-1}{n}, \left( \frac{1}{2}v \right) ; \frac{1}{n}, (0, v) \right) \right) \\ &= \left( \frac{n-1}{n}, v \right).\end{aligned}$$

Lottery  $\mathbf{L}$  results in a win if the low probability event occurs ( $p = \frac{1}{n}$ ). Lottery  $\mathbf{H}$  results in a win if the high probability event occurs ( $p = \frac{n-1}{n}$ ). Assume that she assigns a probabilities of  $\frac{k}{n}$  and  $\frac{n-k}{n}$  to events whose probabilities according to insufficient reason are  $\frac{1}{n}$  and  $\frac{n-1}{n}$ , respectively, where  $1 < k < \frac{n}{2}$  so that

$$\frac{1}{n} < \frac{k}{n} < \frac{n-k}{n} < \frac{n-1}{n}. \quad (1)$$

For simplicity, assume that she assigns probabilities of  $0, \frac{1}{2}, 1$  to events whose probabilities according to insufficient reason are  $0, \frac{1}{2}, 1$ , respectively. Thus she codes lottery  $\mathbf{L}$  as

$$\begin{aligned}\mathbf{L}' &= \left( \frac{k}{n}, \left( \frac{k}{n}, (1, v) ; \frac{n-k}{n}, \left( \frac{1}{2}, v \right) \right) ; \frac{n-k}{n}, \left( \frac{k}{n}, \left( \frac{1}{2}v \right) ; \frac{n-k}{n}, (0, v) \right) \right) \\ &= \left( \frac{k}{n}, v \right).\end{aligned}$$

Analogously, she codes lottery  $\mathbf{H}$  as

$$\begin{aligned}\mathbf{H}' &= \left( \frac{n-k}{n}, \left( \frac{n-k}{n}, (1, v) ; \frac{k}{n}, \left( \frac{1}{2}, v \right) \right) ; \frac{k}{n}, \left( \frac{n-k}{n}, \left( \frac{1}{2}v \right) ; \frac{k}{n}, (0, v) \right) \right) \\ &= \left( \frac{n-k}{n}, v \right).\end{aligned}$$

Given (1), all decision theories in common use require that<sup>11</sup>

$$\left( \frac{1}{n}, v \right) \prec \left( \frac{k}{n}, v \right) \prec \left( \frac{n-k}{n}, v \right) \prec \left( \frac{n-1}{n}, v \right),$$

such a decision maker will exhibit ambiguity seeking for low probabilities and ambiguity aversion for high probabilities, in agreement with stylized fact 1. But such a theory would also be consistent with the reverse. Thus, it can

---

<sup>11</sup>  $a \prec b$  means lottery  $b$  is strictly preferred to lottery  $a$ .

accommodate the Ellsberg paradox at the expense of losing its predictive power.

We prefer to retain the heuristic of insufficient reason and, therefore, we have to modify or replace reduction of compound lotteries. But how? Or with what? Here we adopt quantum probabilities in place of standard (Kolmogorov) probabilities.

### 3 Classical (non-quantum) decision theories and the Ellsberg paradox

In this section, we review the main alternatives to QDT with respect to their success or failure to explain the stylized facts of the Ellsberg paradox.

In subsection 3.1, below, we give a brief review of standard (Kolmogorov) probability theory. We do this for two reasons. First, because it is fundamental to all decision theories. Second, to make clear the similarities and differences with quantum probability (section 4). Probabilities can be either *objective*, in the sense that they are the same for all decision makers, or they can be *subjective* in the sense that they can differ across decision makers. In the latter case, they can be elicited from a decision maker's observed choices, given the decision theory under consideration.

#### 3.1 Standard (Kolmogorov) probability theory

In the standard approach we have a non-empty set,  $\Omega$ , called the *sample space*, and a  $\sigma$ -algebra,  $S$ , of subsets of  $\Omega$ . The elements of  $S$  are called events.  $S$  has the following properties:  $\emptyset \in S$ ,  $X \in S \Rightarrow \Omega - X \in S$ ,  $\{X_i\}_{i=1}^{\infty} \subset S \Rightarrow \cup_{i=1}^{\infty} X_i \in S$ . Note that the distributive laws hold:  $X \cap (\cup_{j=1}^{\infty} Y_j) = \cup_{j=1}^{\infty} (X \cap Y_j)$  and  $X \cup (\cap_{j=1}^{\infty} Y_j) = \cap_{j=1}^{\infty} (X \cup Y_j)$ .

A *probability measure* is then defined as a function,  $P : S \rightarrow [0, 1]$  with the properties that  $P(\emptyset) = 0$ ,  $P(\Omega) = 1$  and if  $X_i \cap X_j = \emptyset$ ,  $i \neq j$ , then  $P(\cup_{i=1}^{\infty} X_i) = \sum_{i=1}^{\infty} P(X_i)$ .

Let  $X, Y \in S$ ,  $P(Y) \neq 0$ . Define  $P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$ . In particular, if  $X \subset Y$ , then  $P(X|Y) = \frac{P(X)}{P(Y)}$ .  $P(X|Y)$  is called the probability of  $X$  *conditional* on  $Y$ . Then  $P(.|Y)$  is a probability measure on the set  $\{X \in S : X = Z \cap Y, \text{ for some } Z \in S\}$ . From this we can derive Bayes law:  $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$  and its other equivalent forms. Importantly, the *law*

*of total probability* holds: Let  $X \in S$  and let  $\{Y_i\}_{i=1}^n$  be a partition of  $\Omega$ , so  $Y_i \in S$ ,  $Y_i \neq \emptyset$ ,  $\cup_{i=1}^n Y_i = \Omega$ ,  $Y_i \cap Y_j = \emptyset$  for  $i \neq j$ , then  $P(X) = \sum_{i=1}^n P(X|Y_i)P(Y_i)$ .

A *random variable* is a mapping,  $f : \Omega \rightarrow \mathbb{R}$  satisfying: For each  $r \in \mathbb{R}$ ,  $\{x \in \Omega : f(x) \leq r\} \in S$ . A random variable,  $f$ , is *non-negative* if  $f(x) \geq 0$  for each  $x \in \Omega$ . For two random variable,  $f, g$ , we write  $f \leq g$  if  $f(x) \leq g(x)$  for each  $x \in \Omega$ . A random variable,  $f$ , is *simple* if its range is finite. For any random variable,  $f$ , and any  $x \in \Omega$ , let  $f^+(x) = \max\{0, f(x)\}$  and  $f^-(x) = -\min\{0, f(x)\}$ . Then, clearly,  $f^+$  and  $f^-$  are both non-negative random variables and  $f(x) = f^+(x) - f^-(x)$ , for each  $x \in \Omega$ . We write this as  $f = f^+ - f^-$ .

Let  $f$  be a simple random variable with range  $\{f_1, f_2, \dots, f_n\}$ . Let  $X_i = \{x \in \Omega : f(x) = f_i\}$ . Then,  $X_i \in S$ ,  $X_i \cap X_j = \emptyset$  for  $i \neq j$  and  $\cup_{i=1}^n X_i = \Omega$ . The *expected value* of the simple random variable,  $f$ , is  $E(f) = \sum_{i=1}^n f_i P(X_i)$ . The *expected value* of the non-negative random variable,  $g$ , is  $E(g) = \sup\{E(f) : f \text{ is a simple random variable and } f \leq g\}$ . Note that  $E(g)$  may be infinite. If  $f = f^+ - f^-$  is an arbitrary random variable such that not both  $E(f^+)$  and  $E(f^-)$  are infinite, then the expected value of  $f$  is  $E(f) = E(f^+) - E(f^-)$ . Note that  $E(f)$  can be  $-\infty$ , finite or  $\infty$ . However, if both  $E(f^+)$  and  $E(f^-)$  are infinite then  $E(f)$  is undefined (because  $\infty - \infty$  is undefined).

### 3.2 Expected utility theory (EU)

It will be sufficient for our purposes to consider a partition of  $\Omega$  into a finite set of exhaustive and mutually exclusive events:  $\Omega = \cup_{i=1}^n X_i$ ,  $X_i \neq \emptyset$ ,  $X_i \cap X_j = \emptyset$  for  $i \neq j$ ,  $i = 1, 2, \dots, n$ . A decision maker can take an *action*  $a \in A$  that results in *outcome*  $o_i(a) \in O$  and *utility*  $u(o_i(a))$  if the event  $X_i$  occurs, where  $u : O \rightarrow \mathbb{R}$ . The decision maker chooses an action,  $a \in A$  before knowing which event,  $X_i$ , will occur or has occurred. Let  $p_i$  be the probability with which event  $X_i$  occurs. Then the decision maker's *expected utility* from choosing the action  $a \in A$  is  $Eu(a) = \sum_{i=1}^n p_i u(o_i(a))$ . The decision maker prefers action  $a \in A$  over action  $b \in A$  if  $Eu(a) \geq Eu(b)$ . The preference is strict if  $Eu(a) > Eu(b)$ . The decision maker is indifferent between  $a$  and  $b$  if  $Eu(a) = Eu(b)$ . The probabilities  $p_i$ ,  $i = 1, 2, \dots, n$ , can either be *objective* (the same for all decision makers, von Neumann and Morgenstern, 1947) or *subjective* (possibly different for different decision makers, Savage, 1954). In the latter case, it follows from Savage's axioms that these probabilities

can be uniquely elicited from the decision maker's behaviour. Note that the action  $a \in A$  results in the lottery  $(o_1(a), X_1; o_2(a), X_2; \dots; o_n(a), X_n)$ , i.e., the lottery that results in outcome  $o_i(a)$  if the event  $X_i$  occurs. In terms of probabilities this lottery can be written as  $(o_1(a), p_1; o_2(a), p_2; \dots; o_n(a), p_n)$ , i.e., the lottery that results in outcome  $o_i(a)$  with probability  $p_i$ . Sometimes it is more convenient to write the lottery explicitly rather than the action that gave rise to it.

**Example 1** (*Ellsberg paradox under expected utility*): *Normalize the subject's utility function so that  $u(0) = 0$ . Recalling subsections 2.3.1 and 2.3.2, straightforward calculations give  $Eu(\mathbf{K}) = Eu(\mathbf{U}) = pu(v)$ . Thus the subject is indifferent between urns K and U. Recalling subsection 2, these results are not consistent with stylized fact 1 (insensitivity).*

### 3.3 The smooth ambiguity model (SM)

The smooth ambiguity model (Klibanoff et al., 2005) is currently the most popular theory in economics for modelling ambiguity. It encompasses several earlier theories as special limiting cases. These include von Neumann and Morgenstern (1947), Hurwicz (1951), Savage (1954), Luce and Raiffa (1957), Gilboa and Schmeidler (1989) and Ghirardato et al. (2004). Conte and Hey (2013) find it provides the most satisfactory account of ambiguity<sup>12</sup>.

For our purposes, it will be sufficient to consider the following special case of the smooth model. Recall that under expected utility theory (subsection 3.2), a decision maker chooses an action  $a \in A$  that results in the outcome,  $o_i(a)$ , with probability,  $p_i$ ,  $i = 1, 2, \dots, n$ . The outcome,  $o_i(a)$ , yields the utility  $u(o_i(a))$  to the decision maker. Hence, her expected utility is  $Eu(a) = \sum_{i=1}^n p_i u(o_i(a))$ . Now suppose that the decision maker is unsure of the probability  $p_i$  with which she believes action  $a$  will result in outcome  $o_i(a)$ . Furthermore, she believes that  $p_i$  will take the value  $p_{ij}$  with probability  $q_j$ ,  $j = 1, 2, \dots, m$ . Thus,  $\sum_{i=1}^n p_{ij} u(o_i(a))$  is the expected utility of action  $a \in A$  under the probability distribution  $(p_{1j}, p_{2j}, \dots, p_{nj})$ . Thus, as usual, the decision maker's attitude to risk is determined by  $u$ . To characterize the decision maker's attitude to ambiguity, a new function,  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ , is introduced and is assumed to be increasing. Then the decision maker's *expected utility* under the smooth model that results from choosing the action

---

<sup>12</sup>However, Kothiyal et al. (2014) disagree, see below.

$a \in A$  is  $Su(a) = \sum_{j=1}^m q_j \varphi(\sum_{i=1}^n p_{ij} u(o_i(a)))$ . The smooth model reduces to expected utility theory in the following two cases: (1),  $m = 1$ , so there is no ambiguity, (2),  $\varphi$  is positive affine.

Suppose  $m > 1$ , so we do have genuine ambiguity. If  $\varphi$  is strictly concave, then the smooth model can explain ambiguity aversion. It can explain ambiguity seeking, if  $\varphi$  is strictly convex. But it cannot explain 1 (insensitivity, i.e., ambiguity seeking for low probabilities and ambiguity aversion for high probabilities), because  $\varphi$  cannot be both strictly concave and strictly convex.

### 3.4 Rank dependent utility theory (RDU)

Expected utility theory (EU) is probably still the most popular decision theory in economics. The considerable refutations of EU have motivated many developments. One of the most popular of these is *rank dependent utility theory* (RDU). Recall that in EU (subsection 3.2) probabilities enter the objective function,  $Eu(a) = \sum_{i=1}^n p_i u(o_i(a))$ , linearly. However, in RDU, probabilities enter the objective function in a non-linear, though precise, way. We start with a *probability weighting function*, which is a strictly increasing function  $w : [0, 1] \xrightarrow{\text{onto}} [0, 1]$ , hence  $w(0) = 0$  and  $w(1) = 1$ . Typically, low probabilities are overweighted and high probabilities are underweighted. The probability weighting function is applied to the cumulative probability distribution. Hence, it transforms it into another cumulative probability distribution. Hence, we may view RDU as EU applied to the transformed probability distribution. The attraction of this is that the full machinery of risk analysis developed for EU can be utilized by RDU (Quiggin, 1982, 1993). We now give the details.

Consider a decision maker who can take an action,  $a \in A$ , that results in outcome,  $o_i(a) \in O$ , with probability  $p_i$ ,  $i = 1, 2, \dots, n$ ,  $p_i \geq 0$ ,  $\sum_{i=1}^n p_i = 1$ . The decision maker has a utility function,  $u : O \rightarrow \mathbb{R}$ . The decision maker has to choose her action before the outcome is realized. Order outcomes in increasing magnitude. Assuming an increasing utility function, this gives:  $u(o_1(a)) \leq u(o_2(a)) \leq \dots \leq u(o_n(a))$ . Define *decision weights*,  $\pi_i$ ,  $i = 1, 2, \dots, n$ , as follows.  $\pi_n = w(p_n)$ ,  $\pi_i = w\left(\sum_{j=i}^n p_j\right) - w\left(\sum_{j=i+1}^n p_j\right)$ ,  $i = 1, 2, \dots, n-1$ . The decision maker's rank dependent utility is then  $RDu(a) = \sum_{i=1}^n \pi_i u(o_i(a))$ . Expected utility theory (EU) is obtained by taking  $w(p) = p$ . Empirical evidence shows that typically  $w(p)$  is inverse-S shaped, so low probabilities are overweighted but high probabilities are underweighted.

Probabilities in the middle range are much less affected. It is important to note that this need not be because decision makers misperceive probabilities (although that does happen). Rather, the weights people assign to utilities are much more sensitive to probability changes near 0 and near 1 compared to probability changes in the the middle range.<sup>13</sup>

Applying RDU (with  $u(0) = 0$ ) to the lotteries  $\mathbf{K}$  and  $\mathbf{U}$  of subsections 2.3.1 and 2.3.2 we get  $RDU(\mathbf{U}) = RDU(\mathbf{K}) = w(p)u(v)$ . Hence, a decision maker obeying RDU will exhibit ambiguity neutrality.

Thus, just like EU, RDU is not consistent with stylized fact 1 (insensitivity).

Two important extensions of RDU that we do not review here are cumulative prospect theory (Tversky and Kahneman, 1992) and Choquet expected utility (Gilboa 1987, 2009, Schmeidler, 1989). Cumulative prospect theory extends RDU by including reference dependence and loss aversion from Kahneman and Tversky, (1979). Choquet expected utility extends RDU by replacing probability weighting functions with more general capacities (Choquet, 1953-1954). Like a probability measure, a capacity is defined on a  $\sigma$ -algebra of subsets of a set. However, unlike a probability measure, a capacity need not be additive. By contrast, the quantum probability measure is an additive measure but defined on the lattice of closed subspaces of a Hilbert space, rather than a  $\sigma$ -algebra of subsets of a set. Further extensions of both are reviewed in Wakker (2010). Despite their importance, these extensions are not immediately relevant to the results of this paper.

### 3.5 Source dependent probability theory (SDP)

Source dependent probability theory is probably the most satisfactory of classical (non-quantum) theories of ambiguity in general and the Ellsberg paradox in particular (Abdellaoui et al. 2011, Kothiyal et al. 2014 and Dimmock et al. 2015). Recall, from subsection 3.4, that RDU predicts ambiguity neutrality: For the lotteries  $\mathbf{K}$  and  $\mathbf{U}$  of subsections 2.3.1 and 2.3.2 we got  $RDU(\mathbf{U}) = RDU(\mathbf{K}) = w(p)u(v)$ . RDU can accommodate stylized fact 1 (subsection 2.1) if we introduce *source dependence* of the probability weighting function.<sup>14</sup> Specifically, let  $w_K(p)$  be the individual's probability weighting function when facing the known urn,  $K$ , and

---

<sup>13</sup>This feature enables RDU to account for the Allais paradox.

<sup>14</sup>More fully treated, source dependence is also introduced into prospect theory, see Wakker (2010). But RDU is sufficient for our purposes.

$w_U(p)$  her probability weighting function when facing the unknown urn  $U$ . If  $w_K(p) < w_U(p)$  for low  $p$  but  $w_K(p) > w_U(p)$  for high  $p$ , then the subject will exhibit ambiguity seeking for low probabilities but ambiguity aversion for high probabilities, in agreement with stylized fact 1. By the definition of matching probability,  $m(p)$ , we have  $w_K(m(p))u(v) = w_U(p)u(v)$  and, hence,  $m(p) = w_K^{-1}(w_U(p))$ . In the low probability treatment, Dimmock et al. (2015) found that for  $n = 10$  colors, subjects (on average) declared indifference between  $K$  and  $U$  when  $K$  contained 22 balls (out of 100) of the winning color. Hence,  $m(0.1) = 0.22 > 0.1$ . In the high probability treatment they found that, again for  $n = 10$  colors, subjects (on average) declared indifference when  $K$  contained 69 balls of the winning colors. Hence,  $m(0.9) = 0.69 < 0.9$ . For  $n = 2$  colors, subjects (on average) declared indifference when  $K$  contained 40 balls of the winning color. Hence,  $m(0.5) = 0.4 < 0.5$ . Thus, subjects exhibited ambiguity aversion for medium and high probabilities but ambiguity seeking for low probabilities.

Note that to measure  $m(p)$ , neither  $w_K(p)$  nor  $w_U(p)$  nor  $u(v)$  need be estimated. However, to make the theory predictive, we need to estimate  $w_K(p)$  and  $w_U(p)$ . For example, using Prelec (1998) probability weighting functions we have to estimate  $w_K(p) = e^{-\beta_K(-\ln p)^{\alpha_K}}$  and  $w_U(p) = e^{-\beta_U(-\ln p)^{\alpha_U}}$ . This involves estimating the four parameters  $\alpha_K, \beta_K, \alpha_U, \beta_U$ . By contrast, in section 5, we shall see that quantum probability gives a parameter-free prediction of  $m(p)$ ; and that this is close to the empirically observed values of  $m(p)$ . We shall see that the quantum predictions are  $m(0.1) = 0.171\,05$ ,  $m(0.5) = 0.416\,67$  and  $m(0.9) = 0.695\,45$ .

## 4 Elements of Quantum Probability Theory<sup>15</sup>

### 4.1 Vectors

For our purposes (as we shall show), it is sufficient to use a finite dimensional real vector space  $\mathbb{R}^n$  (in fact, with  $n = 2$  or  $n = 4$ ). A vector,  $\mathbf{x} \in \mathbb{R}^n$ , is represented by an  $n \times 1$  matrix ( $n$  rows, one column). Its *transpose*,  $\mathbf{x}^\dagger$ , is then the  $1 \times n$  matrix (one row,  $n$  columns) of the same elements but written as a

---

<sup>15</sup>See Busemeyer and Bruza (2012), Chapter 2 and Appendix B, for a more comprehensive but accessible introduction.

row.<sup>16</sup> The *zero vector*,  $\mathbf{0}$ , is the vector all of whose components are zero. Let  $r \in \mathbb{R}$  and  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  with components  $x_i$  and  $y_i$ , respectively. Then  $r\mathbf{x}$  is the vector whose components are  $rx_i$  and  $\mathbf{x} + \mathbf{y}$  is the vector whose components are  $x_i + y_i$ .  $\mathbf{y} \in \mathbb{R}^n$  is a *linear combination* of  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n$  if  $\mathbf{y} = \sum_{i=1}^m r_i \mathbf{x}_i$  for some real numbers  $r_1, r_2, \dots, r_m$ . The *inner product* of  $\mathbf{x}$  and  $\mathbf{y}$  is  $\mathbf{x}^\dagger \mathbf{y} = \sum_{i=1}^n x_i y_i$ , where  $x_i, y_i$  are the components of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively.<sup>17</sup> If  $\mathbf{x}^\dagger \mathbf{y} = 0$ , then  $\mathbf{x}$  is said to be *orthogonal* to  $\mathbf{y}$  and we write  $\mathbf{x} \perp \mathbf{y}$ . Note that  $\mathbf{x} \perp \mathbf{y}$  if, and only if,  $\mathbf{y} \perp \mathbf{x}$ . The *norm*, or *length*, of  $\mathbf{x}$  is  $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\dagger \mathbf{x}}$ .  $\mathbf{x}$  is *normalized* if  $\|\mathbf{x}\| = 1$ .<sup>18</sup>  $X \subset \mathbb{R}^n$  is a *vector subspace* (of  $\mathbb{R}^n$ ) if it satisfies:  $X \neq \emptyset$ ,  $\mathbf{x}, \mathbf{y} \in X \Rightarrow \mathbf{x} + \mathbf{y} \in X$  and  $r \in \mathbb{R}, \mathbf{x} \in X \Rightarrow r\mathbf{x} \in X$ . Let  $\mathcal{L}$  be the set of all vector subspaces of  $\mathbb{R}^n$ . Then  $\{\mathbf{0}\}, \mathbb{R}^n \in \mathcal{L}$ . Let  $X, Y \in \mathcal{L}$ . Then  $X \cap Y \in \mathcal{L}$  and  $X + Y = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in X, \mathbf{y} \in Y\} \in \mathcal{L}$ . If  $X_1, X_2, \dots, X_m \in \mathcal{L}$ , then  $\sum_{i=1}^m X_i = \{\sum_{i=1}^m \mathbf{x}_i : \mathbf{x}_i \in X_i\} \in \mathcal{L}$ . The *orthogonal complement* of  $X \in \mathcal{L}$  is  $X^\perp = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \perp \mathbf{x} \text{ for each } \mathbf{x} \in X\}$ . We have  $X^\perp \in \mathcal{L}$ ,  $(X^\perp)^\perp = X$ ,  $X \cap X^\perp = \{\mathbf{0}\}$ ,  $X + X^\perp = \mathbb{R}^n$ . Let  $\mathbf{z} \in \mathbb{R}^n$  and  $X \in \mathcal{L}$ , then there is a unique  $\mathbf{x} \in X$  such that  $\|\mathbf{z} - \mathbf{x}\| \leq \|\mathbf{z} - \mathbf{y}\|$  for all  $\mathbf{y} \in X$ .  $\mathbf{x}$  is called the *orthogonal projection of  $\mathbf{z}$  onto  $X$* . Let  $\delta_{ii} = 1$  but  $\delta_{ij} = 0$  for  $i \neq j$ .  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$  form an *orthonormal basis* for  $X \in \mathcal{L}$  if  $\mathbf{s}_i^\dagger \mathbf{s}_j = \delta_{ij}$  and if any vector  $\mathbf{x} \in X$  can be represented as a linear combination of the basis vectors:  $\mathbf{x} = \sum_{i=1}^m x_i \mathbf{s}_i$ , where the numbers  $x_1, x_2, \dots, x_m$  are uniquely determined by  $\mathbf{x}$  and  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$ . The choice of an orthonormal basis for a vector space is arbitrary. However, the inner product of two vectors is independent of the orthonormal basis chosen. We shall refer to a normalized vector,  $\mathbf{s} \in \mathbb{R}^n$ , as a *state vector*. In particular, if  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$  form an orthonormal basis for  $\mathbb{R}^n$ , then we shall refer to these as *eigenstates*. Note that if  $\mathbf{s} = \sum_{i=1}^n s_i \mathbf{s}_i$ , then  $\mathbf{s}$  is a state vector if, and only if,  $\|\mathbf{s}\| = 1$ , equivalently, if, and only if,  $\mathbf{s}^\dagger \mathbf{s} = \sum_{i=1}^n s_i s_i = 1$ . Let  $X \in \mathcal{L}$ . Let  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$  form an orthonormal basis for  $X$ . Extend  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$  to an orthonormal basis,  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m, \dots, \mathbf{s}_n$ , for  $\mathbb{R}^n$  (this can always be done). Then  $\mathbf{s}_{m+1}, \dots, \mathbf{s}_n$  form an orthonormal basis for the the orthogonal complement,  $X^\perp$ , of  $X$ . Let  $\mathbf{z} = \sum_{i=1}^n z_i \mathbf{s}_i \in \mathbb{R}^n$ . Then  $\sum_{i=1}^m z_i \mathbf{s}_i$  is the orthogonal projection of  $\mathbf{z}$  onto  $X$  and  $\sum_{i=m+1}^n z_i \mathbf{s}_i$  is the orthogonal projection of  $\mathbf{z}$  onto  $X^\perp$ .

---

<sup>16</sup>More generally, in  $\mathbb{C}^n$ ,  $\mathbf{x}^\dagger$  is the *adjoint*, of  $\mathbf{x}$ . For example, in  $\mathbb{C}^2$ , if  $\mathbf{x} = \begin{bmatrix} r_1 e^{i\theta_1} \\ r_2 e^{i\theta_2} \end{bmatrix}$ , where  $r_1, \theta_1, r_2, \theta_2$  are real and  $i = \sqrt{-1}$ , then  $\mathbf{x}^\dagger = \begin{bmatrix} r_1 e^{-i\theta_1} & r_2 e^{-i\theta_2} \end{bmatrix}$ .

<sup>17</sup>More generally, in  $\mathbb{C}^n$ ,  $\mathbf{x}^\dagger \mathbf{y} = \sum_{i=1}^n x_i^* y_i$ , where, if  $x = r e^{i\theta}$ ,  $r, \theta \in \mathbb{R}$ , then  $x^* = r e^{-i\theta}$ .

<sup>18</sup>In *Dirac notation*,  $\mathbf{x} = |x\rangle$ ,  $\mathbf{x}^\dagger = \langle x|$ ,  $\mathbf{x}^\dagger \mathbf{y} = \langle x|y\rangle$ ,  $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\dagger \mathbf{x}} = \sqrt{\langle x|x\rangle}$ .

We will represent the state of the known Ellsberg urn ( $K$ ) by a normalized vector in  $\mathbb{R}^2$  and the unknown Ellsberg urn ( $U$ ) by a normalized vector in  $\mathbb{R}^4$ .

## 4.2 State of a system, events and quantum probability measures

The state of a system (physical, biological or social) is represented by a normalized vector,  $s \in \mathbb{R}^n$ , i.e.,  $\|s\| = 1$ . The set of events is the set,  $\mathcal{L}$ , of vector subspaces of  $\mathbb{R}^n$ .  $\{\mathbf{0}\}$  is the impossible event and  $\mathbb{R}^n$  is the certain event.  $X^\perp \in \mathcal{L}$  is the complement of the event  $X \in \mathcal{L}$ . If  $X, Y \in \mathcal{L}$  then  $X \cap Y$  is the conjunction of the events  $X$  and  $Y$ ;  $X + Y$  is the event where either  $X$  occurs or  $Y$  occurs or both (if  $X, Y \in \mathcal{L}$  then, in general,  $X \cup Y \notin \mathcal{L}$ ). Recall that in a  $\sigma$ -algebra of subset of a set, the distributive law:  $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$ , and its dual<sup>19</sup>, holds. However, its analogue for  $\mathcal{L}$ :  $X \cap (Y + Z) = (X \cap Y) + (X \cap Z)$ , and its dual<sup>20</sup>, fails to hold in general. Consequently, the law of total probability also fails to hold in general. The failure of the distributive laws to hold in  $\mathcal{L}$  has profound consequences. This non-distributive nature of  $\mathcal{L}$  is the key to explaining many paradoxes of human behaviour.  $F : \mathcal{L} \rightarrow [0, 1]$  is *additive* if  $F(\sum_{i=1}^m X_i) = \sum_{i=1}^m F(X_i)$ , where  $X_i \in \mathcal{L}$  and  $X_i \cap X_j = \{\mathbf{0}\}$  for  $i \neq j$ . A *quantum probability measure* is an additive measure,  $P : \mathcal{L} \rightarrow [0, 1]$ ,  $P(\{\mathbf{0}\}) = 0$ ,  $P(\mathbb{R}^n) = 1$ . If a number can be interpreted as either a classical probability or a quantum probability, then we shall simply refer to it as a probability. Otherwise, we shall refer to it as either a classical probability or a quantum probability, whichever is the case.

## 4.3 Random variables and expected values

Let  $\mathcal{L}$  be the set of all vector subspaces of  $\mathbb{R}^n$ . A *random quantum variable* is a mapping,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  satisfying:  $\{\varphi \in \mathbb{R}^n : f(\varphi) \leq r\} \in \mathcal{L}$  for each  $r \in \mathbb{R}$ .

A random quantum variable,  $f$ , is *non-negative* if  $f(\varphi) \geq 0$  for each  $\varphi \in \mathbb{R}^n$ . For two random quantum variables,  $f, g$ , we write  $f \leq g$  if  $f(\varphi) \leq g(\varphi)$  for each  $\varphi \in H$ . A random quantum variable,  $f$ , is *simple* if its range is finite. For any random quantum variable,  $f$ , and any  $\varphi \in \mathbb{R}^n$ , let  $f^+(\varphi) =$

---

<sup>19</sup>  $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$

<sup>20</sup>  $X + (Y \cap Z) = (X + Y) \cap (X + Z)$

$\max\{0, f(\varphi)\}$  and  $f^-(\varphi) = -\min\{0, f(\varphi)\}$ . Then, clearly,  $f^+$  and  $f^-$  are both non-negative random quantum variables and  $f(\varphi) = f^+(\varphi) - f^-(\varphi)$ , for each  $\varphi \in \mathbb{R}^n$ . We write this as  $f = f^+ - f^-$ .

Let  $f$  be a simple random quantum variable with range  $\{f_1, f_2, \dots, f_n\}$ . Let  $X_i = \{\varphi \in \mathbb{R}^n : f(\varphi) = f_i\}$ . Then  $X_i \in \mathcal{L}$ ,  $X_i \cap X_j = \{\mathbf{0}\}$  for  $i \neq j$  and  $\sum_{i=1}^n X_i = \mathbb{R}^n$ . Then the *expected value* of the simple random quantum variable,  $f$ , is  $E(f) = \sum_{i=1}^n f_i P(X_i)$ . The *expected value* of the non-negative random quantum variable,  $g$ , is

$E(g) = \sup\{E(f) : f \leq g \text{ is a simple random quantum variable}\}$ . Note that  $E(g)$  may be infinite. If  $f = f^+ - f^-$  is an arbitrary random quantum variable such that not both  $E(f^+)$  and  $E(f^-)$  are infinite, then the expected value of  $f$  is  $E(f) = E(f^+) - E(f^-)$ . Note that  $E(f)$  can be  $-\infty$ , finite or  $\infty$ . However, if both  $E(f^+)$  and  $E(f^-)$  are both infinite then  $E(f)$  is undefined (because  $\infty - \infty$  is undefined).

## 4.4 Transition amplitudes and probabilities

Suppose  $\varphi, \chi \in \mathbb{R}^n$  are two states (thus, they are normalized:  $\|\varphi\| = \|\chi\| = 1$ ).  $\varphi \rightarrow \chi$  symbolizes the transition from  $\varphi$  to  $\chi$ . Then, by definition, the *amplitude* of  $\varphi \rightarrow \chi$  is given by  $A(\varphi \rightarrow \chi) = \varphi^\dagger \chi$ . Its quantum probability is  $P(\varphi \rightarrow \chi) = (\varphi^\dagger \chi)^2$ .<sup>21</sup>

Consider the state  $\varphi \in \mathbb{R}^n$  ( $\|\varphi\| = 1$ ). The occurrence of the event  $X \in \mathcal{L}$  causes a transition,  $\varphi \rightarrow \psi$ . The new state,  $\psi$  ( $\|\psi\| = 1$ ), can be found as follows. Let  $\pi$  be the orthogonal projection of  $\varphi$  onto  $X$  (recall subsection 4.1). Suppose that  $\pi \neq \mathbf{0}$  (if  $\pi = \mathbf{0}$ , then  $\pi$  and  $X$  are incompatible, that is, if  $X$  occurs then the transition  $\varphi \rightarrow \psi$  is impossible). Then  $\psi = \frac{\pi}{\|\pi\|}$  is the new *state conditional on  $X$* .

## 4.5 Born's rule

We can now give the empirical interpretation of the state vector. Consider a physical, biological or social system. On measuring a certain observable pertaining to the system, this observable can take the value  $v_i \in \mathbb{R}$  with probability  $p_i \geq 0$ ,  $\sum_{i=1}^n p_i = 1$ . To model this situation, let  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$  form an orthonormal bases for  $\mathbb{R}^n$ . Take  $\mathbf{s}_i$  to be the state (eigenstate) where the

---

<sup>21</sup>In  $\mathbb{C}^n$ ,  $P(\varphi \rightarrow \chi) = (\varphi^\dagger \chi)(\varphi^\dagger \chi)^*$ . However, as we are working in  $\mathbb{R}^n$ ,  $(\varphi^\dagger \chi) = (\varphi^\dagger \chi)^*$ .

observable takes the value (eigenvalue)  $v_i$  for sure. Consider the general state  $\mathbf{s} = \sum_{i=1}^n s_i \mathbf{s}_i$ . If the act of measurement gives the value  $v_i$  for the observable, then this implies that the act of measurement has caused the transition  $\mathbf{s} \rightarrow \mathbf{s}_i$ . The probability of the transition  $\mathbf{s} \rightarrow \mathbf{s}_i$  is  $P(\mathbf{s} \rightarrow \mathbf{s}_i) = (\mathbf{s}^\dagger \mathbf{s}_i)^2 = s_i^2 = p_i$ . Thus, in the representation of the state of the system by  $\mathbf{s} = \sum_{i=1}^n s_i \mathbf{s}_i$ ,  $s_i^2$  is the probability of obtaining the value  $v_i$  on measurement<sup>22</sup>

## 4.6 Feynman's first rule (single path)<sup>23</sup>

Let  $\varphi, \chi, \psi$  be three states.  $\varphi \rightarrow \chi \rightarrow \psi$  symbolizes the transition from  $\varphi$  to  $\chi$  followed by the transition from  $\chi$  to  $\psi$ . The amplitude of  $\varphi \rightarrow \chi \rightarrow \psi$  is then the *product*,  $A(\varphi \rightarrow \chi \rightarrow \psi) = A(\varphi \rightarrow \chi) A(\chi \rightarrow \psi) = (\varphi^\dagger \chi)(\chi^\dagger \psi)$ , of the amplitudes of  $\varphi \rightarrow \chi$  and  $\chi \rightarrow \psi$ . The quantum probability of the transition,  $\varphi \rightarrow \chi \rightarrow \psi$ , is then  $P(\varphi \rightarrow \chi \rightarrow \psi) = (A(\varphi \rightarrow \chi \rightarrow \psi))^2 = ((\varphi^\dagger \chi)(\chi^\dagger \psi))^2 = (\varphi^\dagger \chi)^2 (\chi^\dagger \psi)^2$ , i.e., the product of the respective probabilities. This can be extended to any number of multiple transitions along a single path.

## 4.7 Feynman's second rule (multiple indistinguishable paths)

Suppose that the transition from  $\varphi$  to  $\psi$  can follow any of two paths:

$\varphi \rightarrow \chi_1 \rightarrow \psi$  or  $\varphi \rightarrow \chi_2 \rightarrow \psi$ . Furthermore, and this is crucial, assume that which path was followed is *not* observable. First, we calculate the amplitude of  $\varphi \rightarrow \chi_1 \rightarrow \psi$ , using Feynman's first rule. We also calculate the amplitude of

$\varphi \rightarrow \chi_2 \rightarrow \psi$ , using, again, Feynman's first rule. To find the amplitude of  $\varphi \rightarrow \psi$  (via  $\chi_1$  or  $\chi_2$ ) we *add the two amplitudes*. The amplitude of  $\varphi \rightarrow \psi$  is then  $(\varphi^\dagger \chi_1)(\chi_1^\dagger \psi) + (\varphi^\dagger \chi_2)(\chi_2^\dagger \psi)$ . Finally, the probability of the transition  $\varphi \rightarrow \psi$  (via  $\chi_1$  or  $\chi_2$ ) is  $((\varphi^\dagger \chi_1)(\chi_1^\dagger \psi) + (\varphi^\dagger \chi_2)(\chi_2^\dagger \psi))^2 = (\varphi^\dagger \chi_1)^2 (\chi_1^\dagger \psi)^2 + (\varphi^\dagger \chi_2)^2 (\chi_2^\dagger \psi)^2 + 2((\varphi^\dagger \chi_1)(\chi_1^\dagger \psi)(\varphi^\dagger \chi_2)(\chi_2^\dagger \psi))$ .

---

<sup>22</sup>More generally, if we use  $\mathbb{C}^n$ , then  $s_i^* s_i$ , is the probability of obtaining the value  $v_i$  on measurement.

<sup>23</sup>See Busemeyer and Bruza (2012), section 2.2, for the Feynman rules.

## 4.8 Feynman's third rule (multiple distinguishable paths)

Suppose that the transition from  $\varphi$  to  $\psi$  can follow any of two paths:

$\varphi \rightarrow \chi_1 \rightarrow \psi$  or  $\varphi \rightarrow \chi_2 \rightarrow \psi$ . Furthermore, and this is crucial, assume that which path was followed *is* observable (although it might not actually be observed). First, we calculate the quantum probability of  $\varphi \rightarrow \chi_1 \rightarrow \psi$ , using Feynman's first rule. We also calculate the quantum probability of  $\varphi \rightarrow \chi_2 \rightarrow \psi$ , using, again, Feynman's first rule. To find the total quantum probability of  $\varphi \rightarrow \psi$  (via  $\chi_1$  or  $\chi_2$ ) we *add the two probabilities*. The quantum probability of  $\varphi \rightarrow \psi$  is then  $(\varphi^\dagger \chi_1)^2 (\chi_1^\dagger \psi)^2 + (\varphi^\dagger \chi_2)^2 (\chi_2^\dagger \psi)^2$ .

Comparing the last expression with its analogue for Feynman's second rule, we see the absence here of the term  $2((\varphi^\dagger \chi_1)(\chi_1^\dagger \psi)(\varphi^\dagger \chi_2)(\chi_2^\dagger \psi))$ . This is called the *interference term*. Its presence or absence has profound implications in both quantum physics and quantum decision theory.

The Feynman rules play a role in quantum probability theory analogous to the rule played by Bayes' law and the law of total probability in classical theory.

## 4.9 An illustration

We give a simple example where it is clear which Feynman rule should be used. Consider an Ellsberg urn containing two balls. One ball is marked 1 and the other ball is marked 2. If a ball is drawn at random then, in line with the heuristic of insufficient reason, we assign probability  $\frac{1}{2}$  to ball 1 being drawn and probability  $\frac{1}{2}$  to ball 2 being drawn. Call this initial state  $s$ . Let the state where ball 1 is drawn be  $s_1$  and let  $s_2$  be the state if ball 2 is drawn. Now, suppose a ball is drawn but *returned* to the urn. This should not change the initial state of the urn. Both classical reasoning and quantum reasoning should yield this result.

### 4.9.1 Classical treatment

Consider the transition  $s \rightarrow s$ . This can occur via one of the two paths:  $s \rightarrow s_1 \rightarrow s$  or  $s \rightarrow s_2 \rightarrow s$ : either ball 1 is drawn then returned to the urn or ball 2 is drawn then returned to the urn. The classical treatment gives a probability  $\frac{1}{2}$  to the transition  $s \rightarrow s_1$ . Since returning ball 1 restores the original state of the urn, the classical probability of the transition  $s_1 \rightarrow s$  is 1. Hence, the classical probability of the transition  $s \rightarrow s_1 \rightarrow s$  is  $\frac{1}{2} \times 1 = \frac{1}{2}$ .

Similarly, the classical probability of the transition  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}$  is also  $\frac{1}{2}$ . Hence, the classical probability of the transition  $\mathbf{s} \rightarrow \mathbf{s}$  via either paths  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}$  or  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}$  is  $\frac{1}{2} + \frac{1}{2} = 1$ .

#### 4.9.2 Quantum treatment

We use  $\mathbb{R}^2$ . Let  $\mathbf{s}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  be the state if ball 1 is drawn and let  $\mathbf{s}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$  be the state if ball 2 is drawn. Take the initial state of the urn be  $\mathbf{s} = \sqrt{\frac{1}{2}}\mathbf{s}_1 + \sqrt{\frac{1}{2}}\mathbf{s}_2 = \begin{bmatrix} \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} \end{bmatrix}$ . Let us check to see if this is a reasonable assignment.  $\mathbf{s}_1$  and  $\mathbf{s}_2$  form an orthonormal basis for  $\mathbb{R}^2$ .  $\|\mathbf{s}\| = \sqrt{\mathbf{s}^\dagger \mathbf{s}} = 1$ . Hence,  $\mathbf{s}$  is a state vector. The amplitude of the transition  $\mathbf{s} \rightarrow \mathbf{s}_1$  is  $\mathbf{s}^\dagger \mathbf{s}_1 = \sqrt{\frac{1}{2}}$ . The amplitude of the transition  $\mathbf{s}_1 \rightarrow \mathbf{s}$  is  $\mathbf{s}_1^\dagger \mathbf{s} = \sqrt{\frac{1}{2}}$ . Hence, by Feynman's first rule (single path), the amplitude of the transition  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}$  is  $A(\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}) = A(\mathbf{s} \rightarrow \mathbf{s}_1)A(\mathbf{s}_1 \rightarrow \mathbf{s}) = \frac{1}{2}$ , in agreement with our intuitive reasoning. Similarly, the amplitude of the transition  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}$  is  $A(\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}) = \frac{1}{2}$ . We now compare the results from applying Feynman's second rule with the results from applying Feynman's third rule. Since  $P(\mathbf{s} \rightarrow \mathbf{s}) = (A(\mathbf{s} \rightarrow \mathbf{s}))^2 = (\mathbf{s}^\dagger \mathbf{s})^2 = (1)^2 = 1$ , the correct rule is the one that gives this result.

**Feynman's second rule (multiple indistinguishable paths)** Here we add the amplitudes of the transitions  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}$  and  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}$  to get the amplitude of the transition  $\mathbf{s} \rightarrow \mathbf{s}$ :  $A(\mathbf{s} \rightarrow \mathbf{s}) = A(\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}) + A(\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}) = \frac{1}{2} + \frac{1}{2} = 1$ . Hence, the quantum probability of the transition  $\mathbf{s} \rightarrow \mathbf{s}$ , through all paths, is  $P(\mathbf{s} \rightarrow \mathbf{s}) = (A(\mathbf{s} \rightarrow \mathbf{s}))^2 = (1)^2 = 1$ , in agreement with our intuitive analysis.

**Feynman's third rule (multiple distinguishable paths)** Here we calculate the quantum probabilities of the transitions  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}$  and  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}$ . This gives  $P(\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}) = (A(\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}))^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$  and  $P(\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}) = (A(\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}))^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$ . Then we add these quantum probabilities to get  $P(\mathbf{s} \rightarrow \mathbf{s}) = P(\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{s}) + P(\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{s}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ . This is a contradiction, since  $P(\mathbf{s} \rightarrow \mathbf{s}) = (A(\mathbf{s} \rightarrow \mathbf{s}))^2 = (\mathbf{s}^\dagger \mathbf{s})^2 = 1$ .

$$(1)^2 = 1.$$

## 5 Quantum decision theory and the Ellsberg paradox

We now give a quantum treatment of the thought experiment of subsection 2.2

### 5.1 Known urn

We have an urn,  $K$ , with two balls, one labeled “1” and the other labeled “2”.

The observable here is the label on the ball.

Let  $\mathbf{b}_1$  be the state where ball 1 is drawn (so label 1 is observed for sure).

Let  $\mathbf{b}_2$  be the state where ball 2 is drawn (so label 2 is observed for sure).

A particularly simple representation of  $\mathbf{b}_1$  and  $\mathbf{b}_2$  is (there are others, of course)

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Clearly,  $\{\mathbf{b}_1, \mathbf{b}_2\}$  forms an orthonormal basis for  $\mathbb{R}^2$ ,  $(\mathbf{b}_j^\dagger \mathbf{b}_k = \delta_{jk})$ , where  $\delta_{jj} = 1$  and  $\delta_{jk} = 0$  for  $j \neq k$ .<sup>24</sup>

Suppose ball 1 is drawn from  $K$  with probability  $p$ , so ball 2 is drawn with probability  $q = 1 - p$ . By Born’s rule (subsection 4.5), the initial state of urn  $K$  is given by<sup>25</sup>

$$\mathbf{s} = \sqrt{p}\mathbf{b}_1 + \sqrt{q}\mathbf{b}_2.$$

### 5.2 Unknown urn

Starting with urn  $K$ , construct urn  $U$  as follows. In the first round draw a ball at random from  $K$  and place it in  $U$ . Replace that ball in  $K$  with an

---

<sup>24</sup>  $\{\mathbf{b}_1, \mathbf{b}_2\}$  also forms an orthonormal basis for  $\mathbb{C}^2$ .

<sup>25</sup> It can be easily verified that the more general specification  $\mathbf{s} = \sqrt{\frac{1}{n}}e^{i\theta_1}\mathbf{b}_1 + \sqrt{\frac{n-1}{n}}e^{i\theta_2}\mathbf{b}_2$ , where  $\theta_1$  and  $\theta_2$  are real and  $i = \sqrt{-1}$ , changes none of our results. So we have elected to simplify the exposition by working with  $\mathbb{R}^n$  rather than  $\mathbb{C}^n$ .

identically labeled ball. In the second round draw a second ball at random from  $K$  and place it in  $U$ . Thus  $U$  contains two balls. Both could be labeled “1”, both could be labeled “2” or one could be labeled “1” and the other labeled “2”. A ball is drawn at random from  $U$ . If ball 1 is drawn, then the subject wins the sum  $v > 0$ . But wins nothing if ball 2 is drawn.

Let  $\mathbf{s}_1$  be the state where ball 1 is drawn in round one and ball 1 is drawn again in round two (each with probability  $p$ ).

Let  $\mathbf{s}_2$  be the state where ball 1 is drawn in round one (probability  $p$ ) then ball 2 is drawn in round two (probability  $q$ ).

Let  $\mathbf{s}_3$  be the state where ball 2 is drawn in round one (probability  $q$ ) then ball 1 is drawn in round two (probability  $p$ ).

Let  $\mathbf{s}_4$  be the state where ball 2 is drawn in round one then ball 2 is drawn again in round two (each with probability  $q$ ).

Urn  $U$  contains two balls labeled 1 if it is in state  $\mathbf{s}_1$ . It contains one ball labeled 1 and the other labeled 2 if it is either in state  $\mathbf{s}_2$  or in state  $\mathbf{s}_3$ . In state  $\mathbf{s}_4$  both balls are labeled 2. We represent these states by:

$$\mathbf{s}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{s}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{s}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{s}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Clearly,  $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4\}$  forms an orthonormal basis for  $\mathbb{R}^4$ .<sup>26</sup>

Let  $\mathbf{s}$  give the initial state of urn  $U$  (unknown composition). Then Born’s rule leads to:<sup>27</sup>

$$\mathbf{s} = p\mathbf{s}_1 + \sqrt{pq}\mathbf{s}_2 + \sqrt{pq}\mathbf{s}_3 + q\mathbf{s}_4.$$

Suppose the ball 1 was drawn from urn  $U$ . To find the state of urn  $U$  conditional on this information, we first project  $\mathbf{s}$  onto the subspace spanned by  $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ , then normalize. This gives

$$\mathbf{w} = \sqrt{\frac{p}{p+2q}}\mathbf{s}_1 + \sqrt{\frac{q}{p+2q}}\mathbf{s}_2 + \sqrt{\frac{q}{p+2q}}\mathbf{s}_3.$$

---

<sup>26</sup>In the language of tensor products,  $\mathbf{s}_1 = \mathbf{b}_1 \otimes \mathbf{b}_1$ ,  $\mathbf{s}_2 = \mathbf{b}_1 \otimes \mathbf{b}_2$ ,  $\mathbf{s}_3 = \mathbf{b}_2 \otimes \mathbf{b}_1$ ,  $\mathbf{s}_4 = \mathbf{b}_2 \otimes \mathbf{b}_2$ .

<sup>27</sup>Again, It can be easily verified that the more general specification  $\mathbf{s} = \frac{1}{n}e^{i\theta_1}\mathbf{s}_1 + \frac{\sqrt{n-1}}{n}e^{i\theta_2}\mathbf{s}_2 + \frac{\sqrt{n-1}}{n}e^{i\theta_3}\mathbf{s}_3 + \frac{n-1}{n}e^{i\theta_4}\mathbf{s}_4$ , where  $\theta_i$  are real and  $i = \sqrt{-1}$ , changes none of our results.

To arrive at the state where ball 1 is drawn, we must follow one of the three paths:

1.  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{w}$ ,
2.  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{w}$ .
3.  $\mathbf{s} \rightarrow \mathbf{s}_3 \rightarrow \mathbf{w}$ .

The relevant transition amplitudes are:

$$\begin{aligned} A(\mathbf{s} \rightarrow \mathbf{s}_1) &= \mathbf{s}_1^\dagger \mathbf{s}_1 = p, \quad A(\mathbf{s}_1 \rightarrow \mathbf{w}) = \mathbf{s}_1^\dagger \mathbf{w} = \sqrt{\frac{p}{p+2q}}, \quad A(\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{w}) = \\ &A(\mathbf{s} \rightarrow \mathbf{s}_1) A(\mathbf{s}_1 \rightarrow \mathbf{w}) = \sqrt{\frac{p^3}{p+2q}} \text{ (Feynman's first rule, single path).} \\ A(\mathbf{s} \rightarrow \mathbf{s}_2) &= \mathbf{s}_2^\dagger \mathbf{s}_2 = \sqrt{pq}, \quad A(\mathbf{s}_2 \rightarrow \mathbf{w}) = \mathbf{s}_2^\dagger \mathbf{w} = \sqrt{\frac{q}{p+2q}}, \quad A(\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{w}) = \\ &A(\mathbf{s} \rightarrow \mathbf{s}_2) A(\mathbf{s}_2 \rightarrow \mathbf{w}) = \sqrt{\frac{pq^2}{p+2q}} \text{ (Feynman's first rule, single path).} \\ A(\mathbf{s} \rightarrow \mathbf{s}_3) &= \mathbf{s}_3^\dagger \mathbf{s}_3 = \sqrt{pq}, \quad A(\mathbf{s}_3 \rightarrow \mathbf{w}) = \mathbf{s}_3^\dagger \mathbf{w} = \sqrt{\frac{q}{p+2q}}, \quad A(\mathbf{s} \rightarrow \mathbf{s}_3 \rightarrow \mathbf{w}) = \\ &A(\mathbf{s} \rightarrow \mathbf{s}_3) A(\mathbf{s}_3 \rightarrow \mathbf{w}) = \sqrt{\frac{pq^2}{p+2q}} \text{ (Feynman's first rule, single path).} \end{aligned}$$

We shall treat the paths  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{w}$  and  $\mathbf{s} \rightarrow \mathbf{s}_3 \rightarrow \mathbf{w}$  as indistinguishable from each other but both distinguishable from path  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{w}$ . Our argument for this is as follows. The path  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{w}$  results in urn  $U$  containing two balls labeled 1. This is clearly distinguishable from paths  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{w}$  and  $\mathbf{s} \rightarrow \mathbf{s}_3 \rightarrow \mathbf{w}$ , each of which result in urn  $U$  containing one ball labeled 1 and one ball labeled 2. From examining urn  $U$ , it is impossible to determine whether this arose by selecting ball 1 first (path  $\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{w}$ ), then ball 2 (path  $\mathbf{s} \rightarrow \mathbf{s}_3 \rightarrow \mathbf{w}$ ), or the other way round.

To find the amplitude of the transition  $\mathbf{s} \rightarrow \mathbf{w}$ , via  $\mathbf{s}_2$  or via  $\mathbf{s}_3$ , we add the amplitudes of these two paths. Thus,  $A(\mathbf{s} \rightarrow \mathbf{w})$ , via  $\mathbf{s}_2$  or  $\mathbf{s}_3$  is  $A(\mathbf{s} \rightarrow \mathbf{s}_2 \rightarrow \mathbf{w}) + A(\mathbf{s} \rightarrow \mathbf{s}_3 \rightarrow \mathbf{w}) = 2\sqrt{\frac{pq^2}{p+2q}}$  (Feynman's second rule, multiple indistinguishable paths). The probability of this transition is  $\left(2\sqrt{\frac{pq^2}{p+2q}}\right)^2 = \frac{4pq^2}{p+2q}$ . The probability of the transition  $\mathbf{s} \rightarrow \mathbf{s}_1 \rightarrow \mathbf{w}$  is  $\left(\sqrt{\frac{p^3}{p+2q}}\right)^2 = \frac{p^3}{p+2q}$ . To get the total probability of the transition  $\mathbf{s} \rightarrow \mathbf{w}$ , via all paths, we add these two. This gives  $P(\mathbf{s} \rightarrow \mathbf{w}) = \frac{p^3}{p+2q} + \frac{4pq^2}{p+2q} = \frac{5p^3 - 8p^2 + 4p}{2-p}$ .

Thus, if the probability of drawing ball 1 from the known urn  $K$  is  $p$ , then the quantum probability of drawing ball 1 from the unknown urn  $U$  is

$$Q(p) = \frac{5p^3 - 8p^2 + 4p}{2 - p}.$$

In particular, we get

$$\begin{aligned} Q(0.1) &= 0.17105, \\ Q(0.5) &= 0.41667, \\ Q(0.9) &= 0.69545. \end{aligned}$$

The following results are easily established.

$$Q(0) = 0, Q(1) = 1.$$

$$Q(p) + Q(1-p) < 1 \text{ for all } 0 < p < 0.$$

$$\lim_{p \rightarrow 0} Q(p) = 0, \lim_{p \rightarrow 0} \frac{Q(p)}{p} = 2, \lim_{p \rightarrow 1} \frac{Q(p)}{p} = 1.$$

$$\begin{aligned} p < 0.4 &\Rightarrow Q(p) > p : \text{ambiguity seeking}, \\ p = 0.4 &\Rightarrow Q(p) = p : \text{ambiguity neutral}, \\ p > 0.4 &\Rightarrow Q(p) < p : \text{ambiguity averse}. \end{aligned}$$

### 5.3 Quantum probabilities are matching probabilities

If  $p$  is the probability of drawing ball 1 from the known urn  $K$ , then  $Q(p)$  is the quantum probability of drawing ball 1 from the unknown urn  $U$ . Let  $u$  be the utility function of a subject participating in the Ellsberg thought experiment outlined in subsection 2.2. Normalize  $u$  so that  $u(0) = 0$ . She wins the sum of money,  $v > 0$ , if ball 1 is drawn from the unknown urn  $U$ , but zero if ball 2 is drawn from that same urn. Hence, her projective expected utility (in the sense of La Mura, 2009) is

$$Q(p)u(v). \tag{2}$$

Now construct a new known urn  $K_1$  from which ball 1 is drawn with probability  $Q(p)$ . Her projective expected utility is

$$Q(p) u(v). \quad (3)$$

Hence, from (2) and (3),  $Q(p)$  is the matching probability for  $p$  in our thought experiment (recall subsection 2.3.3).

## 5.4 Evidence

Dimmock et al. (2015) report the results of their experiments outlined in the Introduction, Table 1. In that table, the second column gives the means across 666 subjects of the observed matching probabilities for 0.1, 0.5 and 0.9. The third column gives the sample standard deviations. The fourth column gives the theoretical predictions of our model.

Our theoretical predictions for  $m(0.5)$  and  $m(0.9)$  are in excellent agreement with the average of observations. Our theoretical prediction for  $m(0.1)$  is not statistically significantly different from the average of observed values. For  $m(0.1)$ ,  $z = \frac{0.22 - 0.17105}{0.25} = 0.1958 < 1.96$ . For such a large sample, the t-distribution is practically normal. Based on the normal test, the evidence does not reject the theoretical prediction  $m(0.1) = 0.17105$  at the 5% level of significance.

Thus expected utility theory can explain stylized fact 1 (insensitivity: ambiguity seeking for low probabilities but ambiguity aversion for high probabilities) if we replace classical probabilities with quantum probabilities to get what La Mura (2009) called *projective expected utility*. Similarly, prospect theory can explain insensitivity if we take the reference point to be the subject's wealth just before the experiment (a common choice), so that the subject is always in the domain of gains, and if we replace decision weights with quantum probabilities.

## 6 Summary and conclusions

We have set up a simple quantum decision model of the Ellsberg paradox. We found its predictions to be in excellent agreement with the evidence. Our derivation is parameter free. It only depends on quantum probability theory in conjunction with the heuristic of insufficient reason. To our mind, this suggests that much of what is normally attributed to probability weighting might actually be due to quantum probability. In particular, many apparent paradoxes may be explained by projective expected utility theory (where

classical probabilities are replaced with quantum probabilities), or explained by prospect theory if we replace decision weights with quantum probabilities. Even when no paradox is involved, *projective prospect theory* (if we may use this expression when decision weights are replaced with quantum probabilities) may provide a more parsimonious representation than the standard forms of prospect theory.

We have modelled the known urn,  $K$ , in  $\mathbb{R}^2$  and the unknown urn,  $U$ , separately in  $\mathbb{R}^4$ . We can model them together in the tensor product of  $K$  and  $U$ ,  $K \otimes U$ , which will be in  $\mathbb{R}^8$ . We can then reproduce the work in this paper in  $K \otimes U$ . More interestingly, we could use an entangled state vector. The entanglement could come from a subject choosing one of the urns as the reference point. This idea has been suggested several times in the literature but has not been carried out as far as we know (Chow and Sarin 2001, 2002) and not in a quantum context. This could shed light on stylized facts 4 (salience) and 5 (anonymity). According to *anonymity* (Curley et al., 1986 and Trautmann et al., 2008) ambiguity aversion does not occur if subjects are assured that their choice is anonymous. Here, psychological game theory may help. People are known to care about the opinion of others, in particular, they fear negative evaluation (Khalmetski et al., 2015).

## References

- [1] Abdellaoui M. A., Baillon A., Placido L. and Wakker P. P. (2011). The rich domain of uncertainty: source functions and their experimental implementation. *American Economic Review* **101**(2). 695–723.
- [2] Aerts D., Sozzo S. and Tapia J. (2014). Identifying Quantum Structures in the Ellsberg Paradox. *International Journal of Theoretical Physics* **53**(10). 3666-3682.
- [3] Basieva I. and Khrennikov A. (2015). On the possibility to combine the order effect with sequential reproducibility for quantum measurements. *Foundations of Physics* **45**(10). 1379-1393.
- [4] Busemeyer J. R. and Bruza P. D. (2012). *Quantum Models of Cognition and Decision*. Cambridge University Press.
- [5] Camerer C. (2003). *Behavioral Game Theory*. Princeton University Press.

- [6] Choquet G. (1953-1954). Theory of Capacities. *Annales de le'Institut Fourier*, **5** (Grenoble): 131-295.
- [7] Chow C. C. and Sarin R. K. (2001). Comparative Ignorance and the Ellsberg Paradox. *Journal of Risk and Uncertainty* **22**. 129–139.
- [8] Chow C. C. and Sarin R. K. (2002). Known, Unknown and Unknowable Uncertainties. *Theory and Decision* **52**. 127–138.
- [9] Conte A. and Hey, J.D. (2013). Assessing multiple prior models of behaviour under ambiguity. *Journal of Risk and Uncertainty* **46(2)**. 113–132.
- [10] Curley S.P., Yates J.F. and Abrams R.A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes* **38**. 230–256.
- [11] Dimmock S. G., Kouwenberg R. and Wakker, P. P. (2015). Ambiguity attitudes in a large representative sample. *Management Science*. Article in Advance. Published Online: November 2, 2015. 10.1287/mnsc.2015.2198
- [12] Ellsberg D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics* **75**. 643-669.
- [13] Ellsberg D. (2001). *Risk, Ambiguity and Decision*. Garland Publishers, New York. Original Ph.D. dissertation: Ellsberg D. 1962. Risk, Ambiguity and Decision. Harvard, Cambridge, MA.
- [14] Fox C. R. and Tversky A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics* **110(3)**. 585–603.
- [15] French K. R. and Poterba J. M. (1991). Investor diversification and international equity markets. *American Economic Review* **81(2)**. 222-226.
- [16] Ghirardato P., Maccheroni F. and Marinacci M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory* **118(2)**. 133–173.
- [17] Gilboa I. (1987). Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics* **16**. 65-88.

- [18] Gilboa I. (2009). *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.
- [19] Gilboa I. and Schmeidler. (1989). Maximin Expected Utility with a Non-Unique Prior. *Journal of Mathematical Economics* **18**. 141-153.
- [20] Gnedenko B. V. (1968). *The theory of probability*. Forth edition. Chelsea Publishing Company, New York, N. Y.
- [21] Haven E. and Khrennikov A. (2013). *Quantum Social Science*. Cambridge University Press.
- [22] Hey J. D., Lotito, G. and Maffioletti A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of Risk and Uncertainty* **41**(2). 81–111.
- [23] Hurwicz L. (1951). Some specification problems and applications to econometric models. *Econometrica* **19**. 343–344.
- [24] Kahneman D. and Tversky A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* **47**. 263-291.
- [25] Keynes J. M. (1921). *A treatise on probability*. London: Macmillan Co.
- [26] Khalmetski K., Ockenfels A. and Werner P. (2015). Surprising gifts: Theory and laboratory evidence. *Journal of Economic Theory* **159**. 163-208.
- [27] Khrennikov A. (2010). *Ubiquitous Quantum Structure: From Psychology to Finance*. Springer.
- [28] Khrennikov A. and Haven E. (2009). Quantum mechanics and violations of the sure-thing principle: The use of probability interference and other concepts. *Journal of Mathematical Psychology* **53**. 378-388.
- [29] Khrennikov A., Basieva I., Dzhafarov E. N. and Busemeyer J. R. (2014). Quantum models for psychological measurement: An unsolved problem. *PLoS ONE* **9**(10) e110909.
- [30] Klibanoff P., Marinacci M. and Mukerji S. (2005). A smooth model of decision making under ambiguity. *Econometrica* **73**(6). 1849–1892.

- [31] Kothiyal A., Spinu, V. and Wakker P. P. (2014). An experimental test of prospect theory for predicting choice under ambiguity. *Journal of Risk and Uncertainty* **48**(1). 1–17.
- [32] La Mura P. (2009). Projective expected utility. *Journal of Mathematical Psychology* **53**(5), 408-414.
- [33] Luce R. D. and Raiffa H. (1957). *Games and Decisions*. New York: Wiley.
- [34] Obstfeld M. and Rogoff K. (2000). The six major puzzles in international economics: Is there a common cause? *NBER Macroeconomic Annual* **15**(1). 339-390.
- [35] Prelec D. (1998) The probability weighting function. *Econometrica* **60**. 497-524.
- [36] Pulford B. D. and Colman A. M. (2008). Ambiguity aversion in Ellsberg urns with few balls. *Experimental Psychology* **55**(1). 31-37.
- [37] Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3**. 323-343.
- [38] Quiggin, J. (1993). *Generalized Expected Utility Theory - The Rank-Dependent Model*. Kulwar Academic Publishers, Boston/Dordrecht/London.
- [39] Rode C., Cosmides L., Hell W. and Tooby J. (1999). When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition* **72**. 269-304.
- [40] Savage L. J. (1954). *The Foundations of Statistics*. New York: Wiley and Sons.
- [41] Schmeidler D. (1989). Subjective probability and expected utility without additivity. *Econometrica* **57**. 571-587.
- [42] Segal U. (1987). The Ellsberg paradox and risk aversion: an anticipated utility approach. *International Economic Review* **28**(1). 175–202.

- [43] Segal U. (1990). Two-stage lotteries without the reduction axiom. *Econometrica* **58**(2). 349–377.
- [44] Thaler R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making* **12**. 183-206.
- [45] Tolman R. C. (1938). *The principles of statistical mechanics*. Oxford University Press, Oxford.
- [46] Trautmann S. T., Vieider F.M. and Wakker P. P. (2008). Causes of ambiguity aversion: known versus unknown preferences. *Journal of Risk and Uncertainty* **36**(3). 225–243.
- [47] Tversky A. and Kahneman D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5**. 297-323.
- [48] von Neumann J. and Morgenstern O. (1947). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- [49] Wakker P. P. (2010). *Prospect Theory for Risk and Ambiguity*. Cambridge: Cambridge University Press.