Probability Weighting Functions*

Ali al-Nowaihi† **Sanjit Dhami**‡

12 June 2010

Abstract

Expected utility theory (EU) is unable to accommodate the observed non-linear weighting of probabilities. We outline three stylized facts on non-linear weighting that a theory of risk must ideally address. These are that people overweight small probabilities and underweight large ones (S1), do not choose stochastically dominated options when such dominance is obvious (S2) and ignore very small probabilities and code extremely large probabilities as one (S3). We then show that the concept of a probability weighting function (PWF) is crucial in addressing S1-S3. A PWF is not, however, in itself, a theory of risk. PWF's need to be embedded within some theory of risk in order to have significant predictive content. The two main alternative theories that are relevant in this regard are *rank dependent utility* (RDU) and *cumulative prospect theory* (CP). RDU and CP explain S1, S2 but not S3. We outline the recent proposal of [2] for *composite prospect theory* (CPP) that uses the *composite Prelec probability weighting function* (CPF). CPF is axiomatically founded, is flexible, and, is parsimonious. CPP can explain all three stylized facts S1, S2, S3.

---

*Invited, refereed contribution to the forthcoming "Wiley Encyclopedia of Operations Research and Management Science," John Wiley and Sons, New Jersey. Dr. Sanjit Dhami would like to acknowledge a one semester study leave provided by the University of Leicester. He also wishes to thank the International School of Economics (ISET) at Tbilisi State University, Georgia for their excellent hospitality and academic environment during the time that he was a visiting Professor in the spring/ summer of 2010.

†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.

‡Department of Economics, University of Leicester, University Road, Leicester. LE1 7RH, UK. Phone: +44-116-2522086. Fax: +44-116-2522908. E-mail: Sanjit.Dhami@le.ac.uk.

Under expected utility theory (EU) decision makers weight probabilities linearly. But the evidence strongly suggests non-linear weighting of probabilities. Consider the following example from [21], p.283. Suppose that one is compelled to play Russian roulette. One would be willing to pay much more to reduce the number of bullets from one to zero than from four to three. However, in each case, the reduction in the probability of a bullet firing is 1/6 and, so, under EU, the decision maker should be willing to pay the same amount. This suggests non-linear weighting of probabilities, which is also supported by the emerging neuroeconomic evidence; see, [9].

The main alternatives to EU, *rank dependent utility* (RDU), *prospect theory* (PT) and *cumulative prospect theory* (CP), incorporate non-linear weighting of probabilities using the device of a *probability weighting function* (PWF). Sections 3.1.1 - 3.1.5 of this Encyclopedia and, in particular, the five articles in section 3.1.1 discuss the axiomatic foundations of EU, RDU, PT and CP. They discuss the non-parametric restrictions which must be imposed on the PWF. The main restrictions are that the PWF is a continuous, increasing function. This, for instance, is the case in the axiomatizations of RDU by [1] and [44].

By contrast, the aim of this article is to explore the parametric PWF's. Our approach, therefore, complements the material elsewhere in the Encyclopedia, gives concrete guidance to the use of PWF's in practical applications and imposes parametric restrictions on the PWF which allows for sharper tests of the predictions.

1. Some basics

Let $X = \{x_1, x_2, ..., x_n\}$ be a fixed finite set of real numbers, which represents the possible monetary *outcomes/wealth levels*. Assume that $x_1 < x_2 < ... < x_n$. The decision maker has a set of feasible actions, A. Any action in A induces a probability distribution $(p_1, p_2, ..., p_n)$, $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$, over the outcomes. We then define a *first order lottery*, L, as

$$L = (x_1, p_1; x_2, p_2; ...; x_n, p_n), \tag{1.1}$$

with the interpretation that outcome x_i occurs with probability p_i. Higher order lotteries (lotteries of lotteries) can then be defined by recursion. Let $\mathcal{L}$ be the set of all lotteries. How should the decision maker choose an action in A? EU postulates that a decision maker has a complete transitive preference ordering, $\preceq$, on $\mathcal{L}$. Under well known assumptions (see e.g., [16]) each lottery is equivalent (under $\preceq$) to a first order lottery. Furthermore, $\preceq$ can be represented by an expected utility functional, $EU : \mathcal{L} \to \mathbb{R}$, which takes the form

$$EU(L) = \textstyle\sum_{i=1}^{n} p_i u(x_i), \tag{1.2}$$

where $u(x_i)$ is the utility of the outcome x_i and $(x_1, p_1; x_2, p_2; ...; x_n, p_n)$ is the first order lottery equivalent to L. Thus, $L_1 \preceq L_2 \Leftrightarrow EU(L_1) \leq EU(L_2)$. A key axiom used in the derivation of (1.2) is the *independence axiom*.

Definition 1 *(Independence axiom): Suppose that $\preceq$ is a preference relation defined over the set of lotteries. Then for all lotteries L_1, L_2, L, and all $p \in [0,1]$, $L_1 \preceq L_2 \Leftrightarrow (L_1, p; L, 1-p) \preceq (L_2, p; L, 1-p)$.*

The independence axiom is often violated by the evidence. Alternatives to EU mainly relax this axiom. Two main features of EU stand out. (1) There is additive separability across outcomes. (2) The objective function is linear in probabilities. Most alternatives to EU relax the second feature, i.e., linearity in probabilities.

Suppose that the decision maker weights outcome x_i with the weight π_i. Suppose that in place of (1.2) we have

$$V(L) = \sum_{i=1}^{n} \pi_i u(x_i). \tag{1.3}$$

Clearly, (1.2) is the special case of (1.3) with $\pi_i = p_i$. Early generalizations of EU took π_i to be a function of p_i. However, it is well known that such a point transformation of objective probabilities can lead the decision maker to choose stochastically dominated options (violation of monotonicity) even when such dominance is obvious (see, e.g., [39]).

2. Stylized facts on non-linear weighting of probabilities

Based on the brief discussion so far, we summarize two stylized facts, S1, S2.

- S1 Evidence suggests that decision makers weight probabilities in a non-linear manner. In particular, the evidence suggests that decision makers overweight low probabilities and underweight high probabilities (inverse S-shaped probability weighting).

- S2. Non-linear point transformation of probabilities, $\pi_i = \pi_i(p_i)$, can violate monotonicity even when such violation is obvious.

A third stylized fact of at least equal significance but that has received relatively less attention in theoretical work is the following.

- S3. For events close to the boundary of the probability interval, $[0,1]$, extensive empirical evidence suggests that decision makers (i) ignore events of extremely low probability and (ii) treat extremely high probability events as certain. In contrast to S1, S3 applies only to very low and very high probabilities.

We emphasize that we are not making a case for ditching S1. Indeed S1 is a robust empirical finding that allows one to resolve many important problems and puzzles in the literature. However, S1, in the absence of S3, leads to severe problems for theory in explaining facts concerning events with probabilities close to the endpoints of the probability interval $[0, 1]$. Indeed [2], [3] and [15] show that the set of such problems is large, and economically important. Hence, we make the case for a class of probability weighting functions that simultaneously incorporates both S1 and S3. Furthermore, our own view is that one cannot argue that either S1 or S3 is more important than the other.

S1 and S2 are well documented; see [20], [39] and Appendix A in [43]. However, S3 is less well documented and theoretical work that incorporates it explicitly and formally has only just begun to appear. We now briefly review the evidence for S3.

3. Evidence for S3

In their Noble prize winning work on *prospect theory* (PT), [21] were acutely aware of S3 and they explicitly choose to emphasize it's salience. In PT, there is a psychologically-rich *editing* phase followed by an *evaluation/decision* phase; the latter uses a rule of the form (1.3). From our perspective, the most critical aspect of the editing phase arises when decision makers decide which extreme probability events to ignore. This is reflected in the manner that $\pi(p)$ is drawn by [21]; see Figure 3.1.

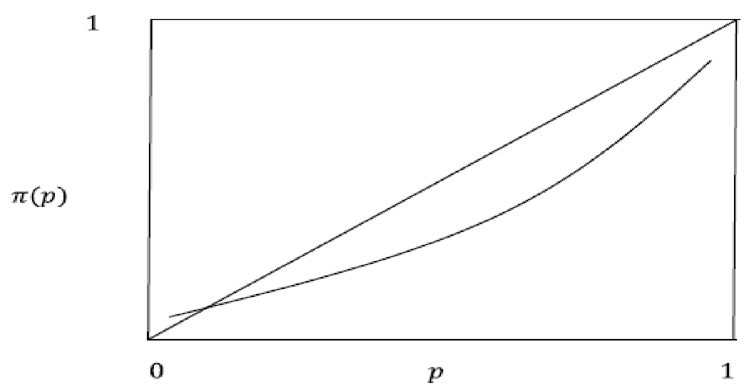

Figure 3.1: Ignorance at the endpoints. Source: [21], p. 283.

[21] wrote the following (p. 282-83) to cogently summarize the evidence on the endpoints of the probability interval $[0, 1]$. "*The sharp drops or apparent discontinuities of* $\pi(p)$ *at the end-points are consistent with the notion that there is a limit to how small a decision weight can be attached to an event, if it is given any weight at all. A similar quantum of doubt could impose an upper limit on any decision weight that is less than unity... On the other hand, the simplification of prospects can lead the individual to discard events of extremely low probability and to treat events of extremely high probability as if they were certain. Because people are limited in their ability to comprehend and evaluate extreme*

probabilities, highly unlikely events are either ignored or overweighted, and the difference between high probability and certainty is either neglected or exaggerated. Consequently $\pi(p)$ *is not well-behaved near the end-points.*"

[2], [3] and [15] review evidence strongly supportive of the claim of [21], from many disparate contexts. We now draw on these papers to briefly, and selectively, summarize the evidence.

3.1. Insurance for low probability events

The insurance industry is of tremendous economic importance. Yet, despite impressive progress, existing theoretical models are unable to explain the stylized facts on the take-up of insurance for low probability events. The seminal study by [24] provides striking evidence of individuals buying inadequate, non-mandatory, insurance against low probability events (e.g., earthquake, flood and hurricane damage in areas prone to these hazards).

EU predicts that a risk-averse decision maker facing an actuarially fair premium will, in the absence of transactions costs, buy *full* insurance for all probabilities, however small. To test this prediction, [24] (chapter 7) presented subjects with varying potential losses with various probabilities, keeping the expected value of the loss constant. Subjects faced actuarially fair, unfair or subsidized premiums. In each case, they found that there is a point below which the take-up of insurance drops dramatically, as the probability of the loss decreases (and as the magnitude of the loss increases, keeping the expected loss constant). These results were shown to be robust to a very large number of perturbations and factors that include generous government incentives to reduce premiums to below market rates, controlling for moral hazard etc.; see [3] for the details.

Arrow's own reading of the evidence in [24] summarized in the foreword to [24] is that: *Obviously in some sense it is right that he or she [the insuree] be less aware of low probability events, other things being equal; but it does appear from the data that the sensitivity goes down too rapidly as the probability decreases.*

[24] write: "*Based on these results, we hypothesize that most homeowners in hazard-prone areas have not even considered how they would recover should they suffer flood or earthquake damage. Rather they treat such events as having a probability of occurrence sufficiently low to permit them to ignore the consequences.*" These results are borne out from further testing by [12] who conclude that: "*Individuals seem to buy insurance only when the probability of risk is above a threshold ...*" This behavior is in close conformity to the observations of [21] outlined above.

3.2. The Becker paradox

A celebrated result due to [8] and known as the Becker proposition states that the most efficient way to deter a crime is to impose the '*severest possible penalty with the lowest possible probability*'. By reducing the probability of detection and conviction, society can economize on the costs of enforcement such as policing and trial costs. But by increasing the severity of the punishment (e.g., fines) the deterrence effect of the punishment is maintained. Indeed, under EU, risk neutrality or risk aversion, and allowing for infinitely severe punishments, the Becker proposition implies that crime would be deterred completely, however small the probability of detection and conviction. Risk seeking behavior is not sufficient to overturn the Becker proposition either; see [15]. The Becker proposition is sometimes phrased as *it is efficient to hang offenders with probability zero*. Empirical evidence, summarized in [15], however, is not supportive of the Becker proposition. This, we call as the Becker paradox.

Remark 1 *: It is shown in [15], in detail, that 9 explanations that have been seriously proposed in the literature cannot explain the Becker paradox. These explanations include the following. Risk seeking behavior on the part of offenders, the ability to avoid severe fines by declaring bankruptcy, the need for differential punishments, type-I and type-II errors in conviction, rent seeking behavior in the presence of severe punishments, abhorrence of severe punishments, objectives other than deterrence, non-availability of severe punishments, and the psychological traits of offenders.*

The behavior of decision makers for low probability events is critical to understanding the Becker paradox. [15] show that if the probability weighting function respects stylized fact S1 but not S3, then the Becker paradox survives even under RDU and CP. However, once one allows for a probability weighting function that respects both stylized facts S1, S3, then in conjunction with the reference point feature of CP, [15] show that the Becker paradox is easily resolved. Indeed S3, turns out to be necessary condition for the resolution of the Becker paradox (and related paradoxes, see Remark 2. below).

Remark 2 *: The Becker paradox is a general paradox that applies to all situations where a decision maker is not deterred from some act, a, that would result in almost infinite punishment with small probability. In each of these cases, and under plausible assumptions on risk attitudes, EU predicts that the decision maker will be dissuaded from the act a. Under RDU and CP the standard probability weighting functions respect S1 but not S3. This acts to dissuade the decision maker even more powerfully from the act a, as compared to EU. Hence, as [2], [3], [15] show formally, standard theories with non-linear weights make the Becker paradox even worse. Furthermore, they show that the simultaneous* *incorporation of S1 and S3 into the same probability weighting function is a necessary condition for solving the class of Becker paradoxes. In the remaining part of this section we consider more examples of the Becker paradox. These include evidence from jumping red traffic lights, driving and talking on mobile phones, seat belt usage, and the take-up of breast cancer examinations.*

3.3. Evidence from jumping red traffic lights

In running red lights there is a *small probability* of an accident. However, the consequences are self inflicted and potentially have *infinite costs.* This act is, therefore, like the act a in Remark 2, above. Rephrased, running red traffic lights is similar to *hanging oneself with a very small probability*, which is similar to the Becker proposition.

It is estimated in [7] that there are approximately 260,000 accidents per year in the USA caused by red-light running with implied costs of car repair alone of the order of $520 million per year. Using Israeli data, [5] calculated that the expected gain from jumping one red traffic is, at most, one minute (the length of a typical light cycle). Given the known probabilities they find that: "*If a slight injury causes a loss greater or equal to 0.9 days, a risk neutral person will be deterred by that risk alone. However, the corresponding numbers for the additional risks of serious and fatal injuries are 13.9 days and 69.4 days respectively*". To this must be added other costs arising from an accident. It stretches plausibility to assume that these are simply mistakes.

[5], [6], [7] provide near decisive evidence that the alternative explanations listed in Remark 1 for the Becker paradox, above, cannot explain the decision to run red traffic lights. It is quite clear, that several of those explanations do not work because the punishment is self-inflicted; see [15] for a detailed discussion. A far more natural explanation is that the probability of an accident is low enough for stylized fact S3 to apply for many people. Thus, red traffic light running is simply caused by some individuals ignoring (or seriously underweighting) the very low probability of an accident. The *bimodal perception of risk* (see section 10 below) then ensures that a fraction of the drivers do not run traffic lights. If the probability weighting function incorporates S1 but not S3, then we should not observe any red traffic light running at all (see Remark 2 above), which is contrary to the evidence.

3.4. Driving and talking on car mobile phones

Consider the usage of mobile phones by drivers in moving vehicles. Such an individual faces *potentially infinite punishment* (e.g., loss of one's and/or the family's life) with *low probability*, in the event of an accident. The Becker proposition applied to this situation would suggest that drivers of vehicles will not use mobile phones while driving or perhaps use hands-free phones, and so, avoid the self inflicted punishment. Yet, evidence suggest that drivers use mobile phones in moving vehicles (see [31], [36]), hence, this problem belongs to the general class of problems in the Becker paradox (see Remark 2 above). Evidence also suggests that hands free phone equipment does not offer significantly more safety. A natural explanation of this class of problems is that individuals simply ignore or substantially underweight the low probability of an accident caused by driving and talking on mobile phones, as in stylized fact S3.

3.5. Other examples

People were reluctant to employ seat belts prior to their mandatory use despite publicly available evidence that seat belts can make the difference between surviving and dying in a car crash, or simply that they provide better safety. Prior to 1985, in the US, only 10-20% of motorists wore seat belts voluntarily, hence, denying themselves *self-insurance*; see [45]. *Potentially fatal* car accidents may be perceived by individuals as *low probability events*, particularly if they are overconfident of their driving abilities. Overconfidence is supported from a wide range of contexts; see for instance [2]. Reluctance to wear seat belts can, thus, be seen to be a special case of the general Becker paradox outlined in Remark 2 above, which requires the use of stylized fact S3 as a necessary condition for an explanation.

Even as evidence accumulated about the dangers of breast cancer (which has a *low unconditional probability but fatal consequences*) women took up the offer of breast cancer examination, only sparingly. In the US, this only changed after the greatly publicized events of the mastectomies of Betty Ford and Happy Rockefeller; see [24] (p. xiii, p. 13-14). It is immediate that this too is a special case of the Becker paradox that requires an incorporation of S3 for its resolution.

3.6. Conclusion from these disparate contexts

[2], [3] and [15] draw two main conclusion from the evidence.

1. Human behavior for very low probability events cannot easily be explained by the existing mainstream theoretical models of risk. EU and the associated auxiliary assumptions are unable to explain the stylized facts. Furthermore, the leading non-expected utility alternatives such as *rank dependent utility* (RDU), and *cumulative prospect theory* (CP) make the problem even worse for those events where stylized fact S3 has bite. *Prospect theory* (PT) on the other hand is able to account for S1 and S3, however, it's treatment of S3 is at an informal/ heuristic level only; see particularly the criticisms in [10].

2. A natural explanation for these phenomena seems to be that individuals simply ignore or seriously underweight very low probability events (stylized fact S3).

4. Probability weighting functions (PWF)

Stylized facts S1, S2, S3 would seem, at least to us, to be the minimum requirements that a theory of risk should address. Most alternatives to EU that use non-linear weighting of probabilities, such as RDU, PT, CP invoke the concept of a *probability weighting function* (PWF) to incorporate stylized facts S1 and S2. None of these theories can incorporate all three stylized facts S1,S2,S3. We examine below emerging work due to [2], [3], [15] who propose *composite cumulative prospect theory* (CCP) that successfully addresses all three stylized facts S1, S2, S3.

Remark 3 *A PWF, by itself, is not a theory of risk. It needs to be embedded within the framework of a decision theory, say, RDU, PT, CP for it to have significant predictive content in concrete economic situations.*

Definition 2 *(Probability weighting function, PWF): By a PWF we mean a strictly increasing function* $w(p) : [0,1] \overset{onto}{\rightarrow} [0,1]$.

A simple proof, that we omit, can be used to demonstrate the following properties of a PWF; see [2] for the proofs.

Proposition 1 *: A probability weighting function has the following properties:*
(a) $w(0) = 0$, $w(1) = 1$. *(b)* w *has a unique inverse,* w^{-1}, *and* w^{-1} *is also a strictly increasing function from* $[0,1]$ *onto* $[0,1]$. *(c)* w *and* w^{-1} *are continuous.*

Definition 3 *: The function,* $w(p)$, *(a) infinitely-overweights infinitesimal probabilities, if* $\lim_{p\to 0} \frac{w(p)}{p} = \infty$, *and (b) infinitely-underweights near-one probabilities, if* $\lim_{p\to 1} \frac{1-w(p)}{1-p} = \infty$.

Definition 4 *: The function,* $w(p)$, *(a) zero-underweights infinitesimal probabilities, if* $\lim_{p\to 0} \frac{w(p)}{p} = 0$, *and (b) zero-overweights near-one probabilities, if* $\lim_{p\to 1} \frac{1-w(p)}{1-p} = 0$.

Definition 5 *: (a)* $w(p)$ *finitely-overweights infinitesimal probabilities, if* $\lim_{p\to 0} \frac{w(p)}{p} \in (1,\infty)$, *and (b)* $w(p)$ *finitely-underweights near-one probabilities, if* $\lim_{p\to 1} \frac{1-w(p)}{1-p} \in (1,\infty)$.

Definition 6 *: (a)* $w(p)$ *positively-underweights infinitesimal probabilities, if* $\lim_{p\to 0} \frac{w(p)}{p} \in (0,1)$, *and (b)* $w(p)$ *positively-overweights near-one probabilities, if* $\lim_{p\to 1} \frac{1-w(p)}{1-p} \in (0,1)$.

Some single parameter PWF's are: [13], [19], [22], [28], [35], [37]. Among the two-parameter PWF's are those by [17], [18], [25], [30], [34], [42]. The Prelec (1998) function appears to be the one with the strongest empirical support. To quote from [41], p.102: "the most predictive version of [cumulative prospect theory] has a power value curve, a single parameter risky weighting function due to [30] and a Logit stochastic process." It was also the first axiomatically derived PWF.

Definition 7 *([30]): By the Prelec function we mean the PWF* $w(p) : [0,1] \to [0,1]$ *given by*

$$w\left(0\right) = 0,\ w\left(1\right) = 1, \tag{4.1}$$

$$w\left(p\right) = e^{-\beta(-\ln p)^{\alpha}},\ 0 < p \leq 1,\ \alpha > 0,\ \beta > 0. \tag{4.2}$$

Definition 8 *: By the standard Prelec function we mean the Prelec function (Definition 7) with* $\alpha < 1$.

The following Proposition can be easily checked.

Proposition 2 *: The Prelec function (Definition 7) is a probability weighting function in the sense of Definition 2.*

We plot the Prelec PWF in Figure 4.1; this is a plot of $w\left(p\right) = e^{-(-\ln p)^{0.5}}$, i.e., $\beta = 1$ and $\alpha = 0.5$. Since $\alpha < 1$, the Prelec function plotted in Figure 4.1 is, in fact, a standard Prelec function (Definition 8).

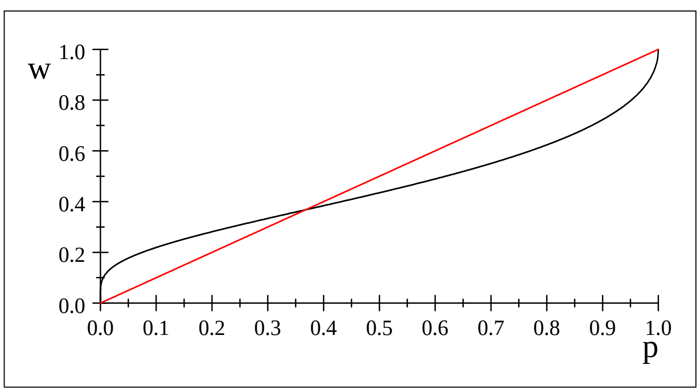

Figure 4.1: A plot of the Prelec (1988) function, $w\left(p\right) = e^{-(-\ln p)^{\frac{1}{2}}}$.

The parameter α controls the convexity/concavity of the Prelec function. Between the region of strict convexity ($w'' > 0$) and the region of strict concavity ($w'' < 0$), there is a point of inflexion ($w'' = 0$). The parameter β in the Prelec function controls the location of the inflexion point relative to the 45^0 line. Sometimes, the respective roles of α and

β are also referred to as the *curvature* and *elevation* properties of a PWF; see [18] and [23]. Not all PWF's allow for a clear separation between curvature and elevation. This is particularly the case for PWF's that involve one parameter rather than two parameters.

The full set of possibilities for the Prelec PWF, for different combinations of α, β is established in [2]. A simple proof leads to the following result.

Proposition 3 *([2]): (a) For $\alpha < 1$ the Prelec function, i.e., the standard Prelec function (Definitions 7, 8): (i) is inverse-S shaped, (ii) infinitely-overweights infinitesimal probabilities, i.e., $\lim_{p \to 0} \frac{w(p)}{p} = \infty$, and (iii) infinitely underweights near-one probabilities, i.e., $\lim_{p \to 1} \frac{1-w(p)}{1-p} = \infty$.*
(b) For $\alpha > 1$ the Prelec function: (i) is S shaped, (ii) zero-underweights infinitesimal probabilities, i.e., $\lim_{p \to 0} \frac{w(p)}{p} = 0$, and (iii) zero-overweights near-one probabilities, i.e., $\lim_{p \to 1} \frac{1-w(p)}{1-p} = 0$.

According to [30], p.505, the infinite limits in Proposition 3(a) capture the qualitative change as we move from certainty to probability and from impossibility to improbability. On the other hand, they contradict stylized fact S3, i.e., the observed behavior that people ignore events of very low probability and treat very high probability events as certain. These specific problems are avoided for $\alpha > 1$. However, for $\alpha > 1$, the Prelec function is S-shaped. This, however,

is in conflict with stylized fact S1.

Remark 4 *(Standard probability weighting functions): A large number of probability weighting functions have been proposed, e.g., those by [18], [25], [30] (with $\alpha < 1$, i.e., the standard Prelec function), [34], [42]. They all overweight infinitesimal probabilities. We shall call these the standard probability weighting functions. All these functions violate stylized fact S3.*

4.1. Axiomatic derivations of Prelec's PWF

Here, we overview three derivations of Prelec's function: [30], based on *compound invariance*, [26] based on *reduction invariance* and [4] based on *power invariance*.

We assume here that $0 \in X$, the set of outcomes, and we shall restrict ourselves to the special class of lotteries defined as follows.

Definition 9 *: Let $\mathbf{S} \subset \mathcal{L}$ be the subset of all lotteries of the forms (x), $(x,\ p_1)$, $((x,\ p_1),\ p_2)$ and $(((x,\ p_1),\ p_2),\ p_3)$, where $x \in X$ and $p_1, p_2, p_3 \in [0,1]$.*

To simplify notation, we shall refer to $((x,\ p_1),\ p_2)$ and $(((x,\ p_1),\ p_2),\ p_3)$ by $(x;\ p_1,\ p_2)$ and $(x;\ p_1,\ p_2,\ p_3)$, respectively. Thus, (x) is the lottery whose outcome is x for sure, $(x;\ p_1)$ is the lottery whose outcomes are x with probability p_1 and 0 with probability $1-p_1$, $(x;\ p_1,\ p_2)$ is the lottery whose outcomes are $(x;\ p_1)$ with probability p_2 and 0 with probability $1-p_2$ and $(x;\ p_1,\ p_2,\ p_3)$ is the lottery whose outcomes are $(x;\ p_1,\ p_2)$ with probability p_3 and 0 with probability $1-p_3$.

Given a strictly increasing function, $u: X \to \mathbb{R}$, and a PWF, w, we can extend u to a function, $U: \mathbf{S} \to \mathbb{R}$, by the following definition.

Definition 10 *:* $U(x) = u(x)$, $U(x;\ p_1) = w(p_1)U(x)$, $U(x;\ p_1,\ p_2) = w(p_2)U(x;\ p_1)$ *and* $U(x;\ p_1,\ p_2,\ p_3) = w(p_3)U(x;\ p_1,\ p_2)$.

Definition 11 *: Let* $\preceq$ *be the order on* $\mathbf{S}$ *induced by* U*, i.e., for all* $L_1, L_2 \in \mathbf{S}$, $L_1 \preceq L_2 \Leftrightarrow U(L_1) \leq U(L_2)$.

Definition 12 *([30]): The preference relation,* $\preceq$*, satisfies compound invariance if, for all outcomes* $x,\ y,\ x',\ y' \in X$*, probabilities* $p,\ q,\ r,\ s \in [0,1]$ *and integers* $n \geq 1$*, the following holds. If* $(x,p) \sim (y,q)$ *and* $(x,r) \sim (y,s)$*, then* $(x',p^n) \sim (y',q^n)$ *implies* $(x',r^n) \sim (y',s^n)$.

Definition 13 *([26]): The preference relation,* $\preceq$*, satisfies reduction invariance if, for all outcomes* $x \in X$*, probabilities* $p_1, p_2, q \in [0,1]$ *and* $\lambda \in \{2,3\}$*:* $(x;p_1,p_2) \sim (x;q) \Rightarrow \left(x;p_1^\lambda,p_2^\lambda\right) \sim \left(x;q^\lambda\right)$.

Definition 14 *([4]): The preference relation,* $\preceq$*, satisfies power invariance if, for all outcomes* $x \in X$*, probabilities* $p,q \in [0,1]$ *and* $\lambda \in \{2,3\}$*:* $(x;\ p,p) \sim (x;q) \Rightarrow \left(x;\ p^\lambda,p^\lambda\right) \sim \left(x;q^\lambda\right)$ *and* $(x;\ p,p,p) \sim (x;q) \Rightarrow \left(x;\ p^\lambda,p^\lambda,p^\lambda\right) \sim \left(x;q^\lambda\right)$.

It is easy to check that for $w(p) = p$ (i.e., the EU case) *compound invariance*, *reduction invariance* and *power invariance* are all satisfied. Hence, each of these constitutes a weakening (or generalization) of EU.

Proposition 4 *([4]): A PWF,* w*, is the Prelec PWF if, and only if, the induced preference relation,* $\preceq$*, satisfies either compound invariance, reduction invariance or power invariance.*

From Proposition 4, all three axioms, *compound invariance*, *reduction invariance* and *power invariance* are equivalent.

5. Addressing stylized fact S1

Addressing stylized fact S1 requires a probability weighting function that is inverse-S shaped, i.e., one that overweights low probabilities and underweights high probabilities. All the standard probability weighting functions have this feature. In particular, this is true for the Prelec probability weighting function if $\alpha < 1$, i.e., what we called the standard Prelec function (Proposition 3(a)).

6. Addressing stylized fact S2

There are two main ways of addressing S2. Either one uses *rank dependent expected utility theory* (RDU) or *cumulative prospect theory* (CP). Quiggen's main insight in [32], [33] was that it is not individual probabilities that should be transformed (which gave rise to the violation of monotonicity reported in S2) but, rather, *cumulative probabilities*. When EU is applied to the transformed cumulative probabilities we get what is now known as RDU.

In RDU, the decision maker uses (1.3) with the decision weights generated as follows.

Definition 15 *: Consider the lottery $(x_1, p_1; x_2, p_2; ...; x_n, p_n)$, where $x_1 < x_2 < ... < x_n$. Let w be the PWF. For RDU, the decision weights, π_i, are defined as follows.*
$\pi_n = w(p_n)$,
$\pi_{n-1} = w(p_{n-1} + p_n) - w(p_n)$,
...
$\pi_i = w\left(\Sigma_{j=i}^n p_j\right) - w\left(\Sigma_{j=i+1}^n p_j\right)$,
...
$\pi_1 = w\left(\Sigma_{j=1}^n p_j\right) - w\left(\Sigma_{j=2}^n p_j\right) = w(1) - w\left(\Sigma_{j=2}^n p_j\right) = 1 - w\left(\Sigma_{j=2}^n p_j\right)$.

From Definition 15, we get that,

$$\pi_j \geq 0 \text{ and } \sum_{j=1}^n \pi_j = 1. \tag{6.1}$$

Proposition 5 *([1], [32]): A decision maker who uses RDU, never chooses stochastically dominated options (i.e., does not violate monotonicity). In other words, for a RDU decision maker there is no problem with explaining S2.*

A result analogous to that in Proposition 5 also holds for the case of *cumulative prospect theory* (CP); see [27], [42] and [40]. In CP, the domain of outcomes is split into the domain of *gains* and the domain of *losses* by expressing each outcome relative to some *reference point*. We note here that the decision weights across the domain of gains and losses under CP *do not* necessarily add up to 1. This contrasts with the case of RDU, in which there is no conception of different domains of gains and losses, so the decision weights add up to one. Since CP uses the cumulative weighting machinery from RDU, the following Proposition holds.

Proposition 6 *: A decision maker who uses CP does not chooses stochastically dominated options. Hence, CP can address stylized fact S2.*

The PWF plays an important role in determining a rich set of attitudes towards risk under CP. Under EU, attitudes to risk are determined purely by the curvature of the utility function. Under CP, however, the utility function is concave in the domain of gains and convex in the domain of losses (see, [21]). Hence, it might be tempting to conclude that under CP the decision maker is risk averse in the domain of gains and risk loving in the domain of losses. This is not true because of the role played by the interaction of the PWF and the curvature of the utility function under CP, in determining attitudes to risk. The following four-fold pattern of risk preferences can be show under CP; see [20]. The decision maker is risk loving for small probabilities in the domain of gains and non-small probabilities in the domain of losses. He/she is also risk averse for non-small probabilities in the domain of gains and small probabilities in the domain of losses.

7. Addressing stylized fact S3

While RDU and CP in conjunction with, say, the standard Prelec function (Definitions 7, 8), are able to explain S1 and S2, they are unable to address stylized fact S3.

[2] make the ambitious proposal of combining the psychological-richness of PT (in accounting for S3 in the editing phase) with the more satisfactory cumulative transformation of probabilities in CP. They combine PT and CP into a single theory, that they call *composite cumulative prospect theory* (CCP), which essentially combines the editing and decision phases of PT into a single phase, while retaining cumulative transformations of probability, as in CP. CCP successfully accounts for all three stylized facts S1, S2 and S3. It can explain everything that RDU and CP can, but the reverse is false.

An immediate implication of Proposition 3(a) is that the standard Prelec function ($\alpha < 1$) can explain S1 but not S3. From Proposition 3(b), we know that for $\alpha > 1$, the Prelec function is S-shaped, which explains S3 but contradicts S1. In order to explain S1, S2, S3 using CCP, [2] introduce a modification to the Prelec PWF in a manner that is consistent with the empirical evidence. They eliminate the discontinuities at the end-points in Figure 3.1 with empirically-founded as well as axiomatically-founded behavior. They call their suggested modification as *composite Prelec weighting function* (CPF). Figure 7.1 sketches the general shape of CPF.

In Figure 7.1, decision makers heavily underweight very low probabilities in the range $[0, p_1]$. Decision makers who use the weighting function in Figure 7.1 would ignore very low probability events by assigning low subjective weights to them. Hence, in conformity with the evidence (see [2], [3] and [15]) they are unlikely to be dissuaded from 'low-probability high-punishment' crimes, reluctant to buy insurance for very low probability events, reluctant to wear seat belts, reluctant to participate in voluntary breast screening programs, willing to run red traffic lights and so on. For an even fuller description of behavior that accounts for the observed pattern of heterogeneity one would need to incorporate further considerations such as those in section 10, above.

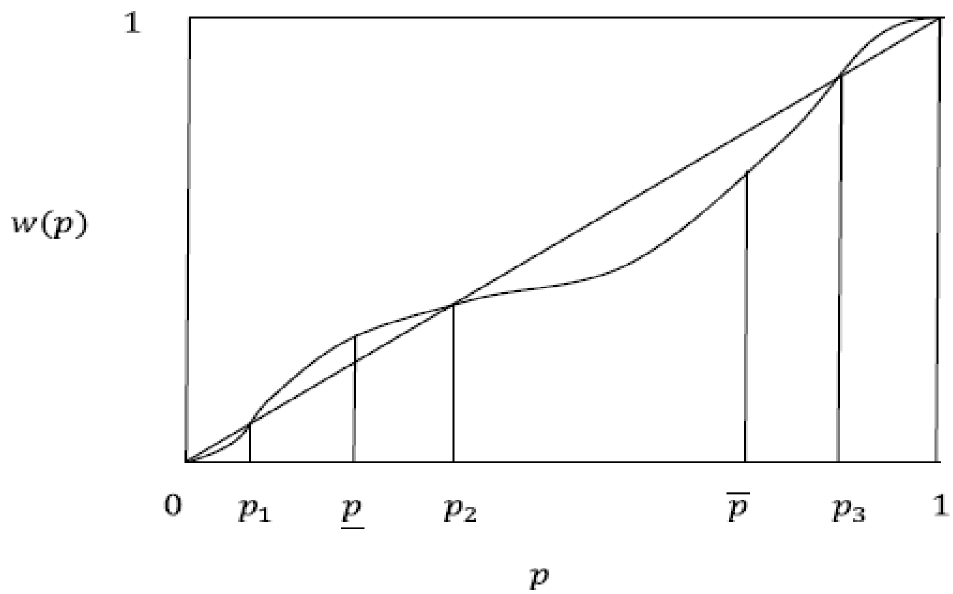

Figure 7.1: The composite Prelec weighting function (CPF).

In the probability range $[p_3, 1]$, events are overweighted as suggested by [21], p.282-83. In the middle segment, $p \in [p_1, p_3]$, the PWF is similar to the Prelec function with $\alpha < 1$, i.e., a standard Prelec function and so successfully addresses S1.

For examples of the CPF, fitted to actual data, see [2], [3]. Due to space limitations we restrict ourselves to a brief, formal, description of the CPF. This implements the general shape of the CPF in Figure 7.1. Define,

$$\underline{p} = e^{-\left(\frac{\beta}{\beta_0}\right)^{\frac{1}{\alpha_0 - \alpha}}}, \ \overline{p} = e^{-\left(\frac{\beta}{\beta_1}\right)^{\frac{1}{\alpha_1 - \alpha}}}. \tag{7.1}$$

Essentially, the CPF in Figure 7.1 is described by segments from three different Prelec PWF's. The first is defined over the range $0 \leq p \leq \underline{p}$, the second over the range $\underline{p} < p \leq \overline{p}$, and the third is defined over the range $\overline{p} < p \leq 1$. The choice of $\underline{p}$ and $\overline{p}$ will guarantee that the three segments are joined continuously.

Definition 16 *([2]): By the composite Prelec weighting function (CPF) we mean the PWF $w : [0, 1] \to [0, 1]$ given by*

$$w(p) = \begin{cases} 0 & if \quad p = 0 \\ e^{-\beta_0(-\ln p)^{\alpha_0}} & if \quad 0 < p \leq \underline{p} \\ e^{-\beta(-\ln p)^{\alpha}} & if \quad \underline{p} < p \leq \overline{p} \\ e^{-\beta_1(-\ln p)^{\alpha_1}} & if \quad \overline{p} < p \leq 1 \end{cases} \tag{7.2}$$

where $\underline{p}$ *and* $\overline{p}$ *are given by (7.1) and*

$$0 < \alpha < 1,\ \beta > 0;\ \alpha_0 > 1,\ \beta_0 > 0;\ \alpha_1 > 1,\ \beta_1 > 0, \beta_0 < 1/\beta^{\frac{\alpha_0 - 1}{1-\alpha}},\ \beta_1 > 1/\beta^{\frac{\alpha_1 - 1}{1-\alpha}}. \quad (7.3)$$

Note the following. The first segment ($0 \leq p \leq \underline{p}$) of the CPF is identical to the Prelec function $e^{-\beta_0(-\ln p)^{\alpha_0}}$, which is S-shaped because $\alpha_0 > 1$. The second segment ($\underline{p} < p \leq \overline{p}$) is identical to the standard Prelec function $e^{-\beta(-\ln p)^{\alpha}}$, which is inverse-S shaped because $\alpha < 1$. The third, and final, segment of the CPF ($\overline{p} < p \leq 1$) is identical to the Prelec function $e^{-\beta_1(-\ln p)^{\alpha_1}}$, which is S-shaped because $\alpha_1 > 1$.

Proposition 7 *([2]): The composite Prelec function, CPF, is a PWF.*

Define p_1, p_2, p_3 that correspond to the notation used for the general shape of a CPF in Figure 7.1.

$$p_1 = e^{-\left(\frac{1}{\beta_0}\right)^{\frac{1}{\alpha_0 - 1}}},\ p_2 = e^{-\left(\frac{1}{\beta}\right)^{\frac{1}{\alpha - 1}}},\ p_3 = e^{-\left(\frac{1}{\beta_1}\right)^{\frac{1}{\alpha_1 - 1}}}. \quad (7.4)$$

Proposition 8 *([2]): (a)* $p_1 < \underline{p} < p_2 < \overline{p} < p_3$. *(b)* $p \in (0, p_1) \Rightarrow w(p) < p$. *(c)* $p \in (p_1, p_2) \Rightarrow w(p) > p$. *(d)* $p \in (p_2, p_3) \Rightarrow w(p) < p$. *(e)* $p \in (p_3, 1) \Rightarrow w(p) > p$.

By Proposition 7, the CPF in (7.2), (7.3) is a PWF in the sense of Definition 2. By Proposition 8, a CPF overweights low probabilities, i.e., those in the range (p_1, p_2), and underweights high probabilities, i.e., those in the range (p_2, p_3). Thus, it accounts for stylized fact S1. But, in addition, and unlike all the standard PWF's, it underweights probabilities near zero, i.e., those in the range $(0, p_1)$, and overweights probabilities close to one, i.e., those in the range $(p_3, 1)$ as required in S3.

Proposition 9 *([2]): The CPF (7.2):*
(a) zero-underweights infinitesimal probabilities, i.e., $\lim_{p \to 0} \frac{w(p)}{p} = 0$ *(Definition 4a),*
(b) zero-overweights near-one probabilities, i.e., $\lim_{p \to 1} \frac{1-w(p)}{1-p} = 0$ *(Definition 4b).*

Remark 5 *(Axiomatic foundations of CPF-I) The axiomatic foundations for the CPF in [2] rely on the axiom of* local power invariance*. This axiom, in turn, is a modification of the axiom of* power invariance *used in [4] to axiomatically derive the Prelec PWF.*

Remark 6 *(Axiomatic foundations of CPF-II) Several axiomatizations of RDU allow for additive separability between terms that multiply utility with probability weights; see [1], [43], [44]. These axiomatizations are discussed elsewhere in the Encyclopedia (see section 3.1.1) and lie outside the scope of this entry. However, [1], [44] assume increasing PWF's, thus, CPF is consistent with them. In order to ensure concavity or convexity of the weighting function, these contributions introduce additional global auxiliary conditions* *such as probabilistic risk aversion. See in particular, [43] who allows for a preference axiomatization of convex capacities. This allows for a weighting function that is either concave or convex throughout, hence, these cannot, simultaneously address stylized facts S1, S3. [14] introduce local conditions that allows for probabilistic risk aversion to vary over the probability interval* $[0,1]$*. For one set of parameter values, they can address S1 (in which case* $\lim_{p\to 0}\frac{w(p)}{p}=\infty$*) while for the complementary set of parameter values, they can address S3 (in which case* $\lim_{p\to 0}\frac{w(p)}{p}=0$*). However, their proposal does not simultaneously address S1 and S2.*

Remark 7 *(Axiomatic foundations of CPF-III) An advantage of the axiomatization of CPF by [2] is that from the class of all increasing PWF's, which is an assumption made in all axiomatizations, they choose a particular parametric member, the CPF. Thus, the CPF (and the associated axiomatization that relies on local power invariance) allows for a sharper prediction which turns out to be consistent with the evidence. Therefore, relative to the other axiomatizations, the work by [2] addresses the further question "what extra axioms are needed to get more specific (and, hence, more predictive) PWF's?"*

Definition 17 *([2]): Otherwise standard CP, when combined with a CPF is called composite cumulative prospect theory (CCP). Analogously, otherwise standard RDU, when combined with a CPF, is referred to as composite rank dependent utility (CRDU).*

[2] prove the following proposition, whose intuition would by now be largely clear to the reader from our discussion of the CPF. Because probabilities in the middle ranges are weighted as in [30], stylized fact S1 is explained. Because cumulative transformations of probability are undertaken in CCP and CRDU, S2 is explained. And because of the property of the CPF in Proposition 9, S3 is explained.

Proposition 10 *([2]): CCP and CRDU can explain S1, S2 and S3.*

[11] shows that the St. Petersburg paradox re-emerges under CP. [2] show that the St. Petersburg paradox can be resolved under CCP mainly through the role played by the CPF. [34] also propose a PWF that resolves the St. Petersburg paradox but it cannot explain stylized fact S3.

In comparison to CRDU, CCP, in addition, incorporates reference dependence, loss aversion and richer attitudes towards risk. Hence, it can explain everything that CRDU can, but the converse is false. Furthermore, because CCP explains S1, S2 and S3, while CP (and RDU) can only explain S1, S2, CCP can explain everything that CP (and RDU) can, but the converse is false. In light of these observations it is interesting to note the observation in [29] that "RDU is currently the most popular decision theory under risk."

8. The Allais paradox under CP and CCP

Consider the Allais paradox as a *common consequence effect*. We omit a discussion of the *common ratio form*, as similar comments apply in that case. Consider the following two pairs of lotteries where outcomes are presented relative to the status quo (so the reference point is zero):

$$a_1 = (1, 1)\,;\, a_2 = (0, 0.01; 1, 0.89; 5, 0.1) \tag{8.1}$$

$$a_3 = (0, 0.89; 1, 0.11);\, a_4 = (0, 0.9; 5, 0.10). \tag{8.2}$$

Allais (1953) conjectured that decision makers would prefer a_1 over a_2, while at the same time preferring a_4 over a_3 (this conjecture has been confirmed by a large number of experiments). As is well known, this pattern of preferences is violated under EU (on account of a violation of the independence axiom). Notice that all outcomes in the lotteries in (8.1), (8.2) are in the domain of gains. Let the utility function be of the usual power form, $u(x) = x^{\gamma}$, with $\gamma = 0.88$ as suggested by the experimental data; see [42]. Then, under CP (and also CCP), the stated pattern of preferences implies the following:

$$a_1 \succ a_2 \Leftrightarrow 1^{0.88} > [w(0.89 + 0.1) - w(0.1)]\, 1^{0.88} + w(0.01)5^{0.88} \tag{8.3}$$

$$a_4 \succ a_3 \Leftrightarrow w(0.1)5^{0.88} > w(0.11)1^{0.88} \tag{8.4}$$

Suppose that under CP, the weighting function is the Prelec function with parameters $\alpha = 0.5$ and $\beta = 1$ so, $w(p) = e^{-(-\ln p)^{0.5}}$. Thus, we can rewrite (8.3), (8.4) as:

$$a_1 \succ a_2 \Leftrightarrow 1^{0.88} > \left[e^{-(-\ln 0.99)^{0.5}} - e^{-(-\ln 0.1)^{0.5}}\right] 1^{0.88} + w(0.01)5^{0.88} \tag{8.5}$$

which is true.

$$a_4 \succ a_3 \Leftrightarrow e^{-(-\ln 0.1)^{0.5}} 5^{0.88} > e^{-(-\ln 0.11)^{0.5}} 1^{0.88} \tag{8.6}$$

which is also true. Hence, under CP, non-linear weighting of probabilities allows for an explanation of the Allais paradox (in the common consequence effect form) that could not be explained under EU.

Remark 8 : *Can the Allais paradox (in the common consequence and common ratio effects forms) also be explained under CCP? From Definition 16 and equation (7.2), we know that over the probability range* $\left[\underline{p}, \overline{p}\right]$*, CPF and CCP are identical. Hence, for lotteries that only have probabilities in the interval* $\left[\underline{p}, \overline{p}\right]$ *both theories make identical predictions. The intuition is that these probabilities are sufficiently non-extreme that only stylized fact S1 has any bite. Stylized fact S3 has bite only if probabilities lie outside the range* $\left[\underline{p}, \overline{p}\right]$*, say, in the two extreme probability intervals* $[0, p_1]$ *and* $[p_3, 1]$ *(see Figure 7.1, definition 16 and equation (7.2)). In these two intervals, different segments of the Prelec* *function apply (because the CPF is made up of 3 different segments of a Prelec function) and one simply needs to alter the relevant values of α, β in inequalities such as (8.5), (8.6) above and recheck them.*

In the light of Remark 8, it becomes important to compute $\underline{p}, \overline{p}$. In an ideal world one would obtain data from the relevant context and problem in order to elicit the values $\underline{p}, \overline{p}$. Here, we rely on the computation made by [2] of these values based on data in [24]. From two separate sets of data, the farm and the urn experiments, respectively, [2] derive the following two sets of values for the range $\left(\underline{p}, \overline{p}\right]$: $(0.05, 0.95]$ and $(0.25, 0.75]$.

The probabilities involved in the two comparisons in (8.5), (8.6) are $0.33, 0.34$ and 0.99. The last of these probabilities lies outside the upper limit of both ranges $(0.05, 0.95]$ and $(0.25, 0.75]$. This does not affect the comparison in (8.6) but does affect the comparison in (8.5). For the urn and the farm experiments, [2] show that the respective values of (α, β) for the upper segment of the Prelec function in the CPF are $(2, 6.4808)$ and $(2, 86.081)$. We now apply these values to the comparison in (8.5), sequentially, for the urn and farm experiments.

$$Urn : a_1 \succ a_2 \Leftrightarrow 1^{0.88} > \left[e^{-6.4808(-\ln 0.99)^2} - e^{-(-\ln 0.1)^{0.5}}\right] 1^{0.88} + w(0.01)5^{0.88}$$

which is true.

$$Farm : a_1 \succ a_2 \Leftrightarrow 1^{0.88} > \left[e^{-86.081(-\ln 0.99)^2} - e^{-(-\ln 0.1)^{0.5}}\right] 1^{0.88} + w(0.01)5^{0.88}$$

which is also true.

Hence, the Allais paradox in common consequence effect form also holds true for CCP when we use a CPF that is fitted to the data in [24]. We emphasize that we have not provided a proof that the Allais paradox will necessarily hold under CCP for all parameter values. Ideally one should have experimental data for the particular situation and then proceed along the lines of Remark 8 to check the resolution of the paradox.

9. A comparison of CCP with configural weights models

We now compare CCP with the class of *configural weight models*, which are reviewed elsewhere in the Encyclopedia (see section 3.1.1). So we shall be brief and keep references to a minimum. The interested reader can directly consult [10] for an exhaustive set of references and a recent survey of these models.

Configural weight models, like several other alternatives to EU, non-linearly weight probabilities of outcomes. In these models, one first writes down the decision tree corresponding to the problem. Consider some lottery, L, that has n possible consequences that arise from the different *branches* of the tree, ordered as, $x_n \leq x_{n-1} \leq ... \leq x_2 \leq x_1$

with associated probabilities $p_n, p_{n-1}, ..., p_2, p_1$. Outcome x_j is then said to have *rank* j, $j = 1, 2, ..., n$ (One could use the convention $x_1 \leq x_2 \leq ... \leq x_{n-1} \leq x_n$ but in this case the rank of the j^{th} number in the list (counting from the left) will be $n - j$).

9.1. The rank affected multiplicative weights model (RAM)

Using equation (7) in [10], in the *rank affected multiplicative weights model* (RAM), the utility of lottery, L, is given by:

$$RAM(L) = \frac{\sum_{i=1}^{n} a(i,n,s)t(p_i)u(x_i)}{\sum_{j=1}^{n} a(j,n,s)t(p_j)} \tag{9.1}$$

Indices i, j denote the *rank* of the outcome (defined above). s is the *augmented sign* of the branch's consequence (positive, zero or negative). $t(p)$ is an increasing function of p, that typically takes the form

$$t(p) = p^{\gamma}; 0 < \gamma < 1. \tag{9.2}$$

The utility function $u(x)$ typically takes the power form:

$$u(x) = x^{\beta}; 0 < \beta < 1. \tag{9.3}$$

$a(i,n,s)$ is the *rank and sign augmented branch weight* corresponding to the outcome that has rank i. For two branch gambles, e.g., the lottery $L = (0, 0.5; 100, 0.5)$, the branch corresponding to the lower outcome ($0) is associated with $a = 2$ and the higher outcome is associated with $a = 1$. For m branch gambles, such that $m \geq 3$ the weights are simply the ranks of the outcomes for that branch. For example, for the set of of outcomes, $x_m \leq x_{m-1} \leq ... \leq x_1$ the weights (the $a's$) are m, $m-1$, $m-2$,...,1. Thus, lower outcomes are given higher weights, which is one of the main features of the model.

From (9.1) the weight associated with outcome x_i is:

$$w(p_i) = \frac{a(i,n,s)t(p_i)}{\sum_{j=1}^{n} a(j,n,s)t(p_j)}.$$

Using (9.2), this can be written as:

$$w(p_i) = \frac{1}{1 + \sum_{j \neq i} \frac{a(j,n,s)}{a(i,n,s)} \left(\frac{p_j}{p_i}\right)^{\gamma}}. \tag{9.4}$$

From (9.4) $w(0) = 1$ and $w(1) < 1$. So this does not satisfy the assumptions of a PWF. In particular, from (9.4):

$$\frac{w(p_i)}{p_i^{\gamma}} = \frac{1}{p_i^{\gamma} + \sum_{j \neq i} \frac{a(j,n,s)}{a(i,n,s)} p_j^{\gamma}}.$$

Since $a(i,n,s)$ is non-zero and $\frac{a(j,n,s)}{a(i,n,s)}$ is finite, so

$$\lim_{p_i \to 0} \frac{w(p_i)}{p_i^\gamma} = \frac{1}{\sum_{j \neq i} \frac{a(j,n,s)}{a(i,n,s)} p_j^\gamma},$$

which is finite. Hence, extremely low probabilities are not ignored, which contradicts stylized fact S3.

9.2. Transfer of attention in exchange (TAX) models

In the *transfer of attention in exchange models* (TAX), the decision maker is assumed to possess a fixed amount of limited attention. Due to this, the decision maker transfers weight from higher ranked to lower ranked outcomes. Consider a *three-branch* lottery, L, given by $L = (x_1, p_1; x_2, p_2; x_3, p_3)$. Then, from equation (9) in [10] we get that the utility of this lottery to the decision maker under TAX is given by

$$TAX(L) = \frac{w_1 u(x_1) + w_2 u(x_2) + w_3 u(x_3)}{w_1 + w_2 + w_3}, \tag{9.5}$$

where,

$$w_1 = t(p_1) - \frac{2}{4}\delta t(p_1); w_2 = t(p_2) - \frac{1}{4}\delta t(p_2) + \frac{1}{4}\delta t(p_1); w_3 = t(p_3) + \frac{1}{4}\delta t(p_1) + \frac{1}{4}\delta t(p_2). \tag{9.6}$$

$t(p)$ continues to be given in (9.2) and $\delta > 0$ is some constant. In this version of the TAX model, also known as the *special TAX model*, one transfers a fixed amount of probability weight from a higher outcome, proportionately, to lower outcomes. Thus, $w_1 + w_2 + w_3 = t(p_1) + t(p_2) + t(p_3)$. From w_1 in (9.6) it follows that the weight on the highest outcome, x_1, is reduced by $\frac{2}{4}\delta t(p_1)$ and each of the lower two outcomes gains one half of this weight. Similarly, the middle ranked outcome, x_2, looses weight $\frac{1}{4}\delta t(p_2)$ which is gained by the worse outcome, x_3.

From (9.6), (9.2) we can compute the ratio of weights to probabilities as:

$$\frac{w_1}{p_1^\gamma} = 1 - \frac{1}{2}\delta; \frac{w_2}{p_2^\gamma} = 1 - \frac{1}{4}\delta + \left(\frac{p_1}{p_2}\right)^\gamma; \frac{w_3}{p_3^\gamma} = 1 + \frac{1}{4}\delta\left(\frac{p_1}{p_3}\right)^\gamma + \frac{1}{4}\delta\left(\frac{p_2}{p_3}\right)^\gamma.$$

We assume that $0 < \delta \leq 2$ so that all weights are positive. There are three possibilities. (1) p_1 is the lowest probability. (2) p_2 is the lowest probability. (3) p_3 is the lowest probability. Since $t(p)$ is given in (9.2), it follows that in these three cases, respectively

$$\lim_{p_1 \to 0} \frac{w_1}{p_1^\gamma} = 1 - \frac{1}{2}\delta \in (0,1]; \ \lim_{p_2 \to 0} \frac{w_2}{p_2^\gamma} = \infty; \ \lim_{p_3 \to 0} \frac{w_3}{p_3^\gamma} = \infty. \tag{9.7}$$

From (9.7), we see again that stylized fact S3 is violated for this class of theories.

9.3. The gains decomposition utility (GDU) model

This model is set out as equations (10)-(12) in [10]. The idea is to reduce multi-branch lotteries to simple, two branch, lotteries. For instance, consider the binary lottery $L_2 = (x_1, p; x_2, 1-p)$ such that $x_1 \leq x_2$. Then, under the *gains decomposition utility model* (GDU), the utility of this binary lottery is:

$$GDU(x_1, p, x_2) = w(p)u(x_1) + (1 - w(p))\, u(x_2). \tag{9.8}$$

Notice that the decision weights are formed exactly as under RDU. Three-branch lotteries such as $L_3 = (x_1, p_1; x_2, p_2; x_3, p_3)$ are then evaluated as follows. First, a choice is made between the lowest outcome x_3 and the binary lottery that comprises of the remaining two outcomes. Probabilities are renormalized so that they add up to one in the binary lottery. In other words, L_3 is treated as:

$$L_3 = \left(\left(x_1, \frac{p_1}{p_1+p_2}; x_2, 1 - \frac{p_1}{p_1+p_2}\right); x_3, 1 - p_1 - p_2\right) \Leftrightarrow L_3 = (L_2, p_1 + p_2; x_3, 1 - p_1 - p_2)\,. \tag{9.9}$$

Thus, L_3 is transformed into a binary lottery which can be evaluated using (9.8) as follows

$$GDU(L_3) = w(p_1 + p_2) GDU\left(x_1, \frac{p_1}{p_1+p_2}, x_2\right) + [1 - w(p_1 + p_2)]\, u(x_3).$$

In terms of the weighting of outcomes this is exactly like RDU. The weighting functions used in this theory are the standard weighting function (e.g., the standard Prelec function is used by [10]). Thus, $\lim_{p \to 0} \frac{w(p)}{p^\gamma} = \infty$ and, so, stylized fact S3 cannot be addressed by this theory either.

Remark 9 : *A great deal of the differences in prediction between CP (and by implication, CCP) and the configural weights models arise in the case of* event splitting*; see [10]. This occurs when there are two identical outcomes in a lottery. However, for the examples listed in section 3, event splitting does not occur; see [2], [3], [15].*

Remark 10 : *For lotteries of the form $(x, p; y, 1-p)$, where $x < 0 < y$, [10] reports that CP and TAX give close results. So, for TAX to be different, we need lotteries with, at least, three outcomes. More generally, it is difficult to distinguish between CP and TAX within the probability triangle. These observations are important. First, they show that where CP is successful (the experiments reported by [21], [42] and, more generally, events that can be represented in the probability triangle), TAX is also successful. But for the "new" paradoxes, [10] argues that TAX succeeds where CP fails. In many applications that involve use of stylized fact S3, configural weights models are likely to fail as do EU, RDU, and CP, while CCP is likely to succeed in these cases. In these cases, the configural weights models can benefit by incorporating stylized fact S3 through using a composite Prelec PWF.*

10. A possible formalization of the bimodal perception of risk

There is evidence of a *bimodal perception of risks*; see [12]. For the evidence, see [38]. Some individuals focus more on the probability while others on the size of the loss. The former do not pay attention to losses that fall below a certain probability threshold, while for the latter, the size of the loss is relatively more salient. Hence, S3 applies particularly to the former set of individuals.

In line with the evidence on bimodal perception of risks, one might envisage a heterogenous population such that for a fraction of the population, $\lim_{p\to 0}\frac{w(p)}{p}=0$ holds while for others $\lim_{p\to 0}\frac{w(p)}{p}=\infty$, holds. For the former set of individuals the size of the loss is salient because they place extremely high probability weights on it (even for very low probabilities) while for the latter, the magnitude of the probability is the key salient feature. One could also imagine that the former set of individuals use the Prelec weighting function while the latter set uses the CPF. This would then explain, for instance, why the Becker proposition holds for some individuals and not others or why some individuals will buy insurance against low probability hazards while others won't. Hence, this framework provides one possible underpinning to the empirical evidence on bimodal perception of risks.

References

[1] Abdellaoui, M. A Genuine Rank-Dependent Generalization of the Von Neumann-Morgenstern Expected Utility Theorem. Econometrica. 2002. 70: 717-736.

[2] al-Nowaihi, A, Dhami, S. Composite prospect theory: A proposal to combine prospect theory and cumulative prospect theory. University of Leicester Discussion Paper. (2010a).

[3] al-Nowaihi, A, Dhami, S. Insurance behavior for low probability events. University of Leicester Discussion Paper. (2010b)

[4] al-Nowaihi, A, Dhami, S. A simple derivation of Prelec's probability weighting function. *The Journal of Mathematical Psychology*. 2006. 50: 521-524.

[5] Bar Ilan, A. The Response to Large and Small Penalties in a Natural Experiment. Department of Economics. University of Haifa. 2000. No. 31905.

[6] Bar Ilan, A, Sacerdote, B. The response to fines and probability of detection in a series of experiments. National Bureau of Economic Research Working Paper. 2001. No. 8638.

[7] Bar Ilan, A, Sacerdote, B. The response of criminals and noncriminals to fines" *Journal of Law and Economics*. 2004. 47: 1-17.

[8] Becker, G. Crime and Punishment: an Economic Approach. *Journal of Political Economy*. 1968. 76: 169-217.

[9] Berns, GS, Capra, CM, Chappelow, Moore, J, and Noussair, C. Nonlinear neurobiological probability weighting functions for aversive outcomes. *Neuroimage*. 2008. 39: 2047-57.

[10] Birnbaum, M. H. New paradoxes of risky decision making. Psychological Review. 2008. 115: 463-501.

[11] Blavatskyy, PR. Back to the St. Petersburg paradox? *Management Science*. 2005. 51: 677-678.

[12] Camerer, C, and Kunreuther, H. Experimental markets for insurance. Journal of Risk and Uncertainty. 1989. 2: 265-300.

[13] Currim, IS, and Rakesh, KS. Prospect Versus Utility. *Management Science*. 1989. 35: 22-41

[14] Diecidue, E, Ulrich, S, and Horst, Z. Parametric Weighting Functions. *Journal of Economic Theory*. 2009. 144: 1102-1118.

[15] Dhami, S, and al-Nowaihi, A. The Becker paradox reconsidered through the lens of behavioral economics. University of Leicester Discussion Paper. 2010.

[16] Fishburn, PC. The foundations of expected utility. D. Reidel Publishing Company. Dordrecht: Holland/Boston: U.S.A. London: England. 1982.

[17] Goldstein, WM, and Einhorn, HJ. Expression Theory and the Preference Reversal Phenomena. *Psychological Review*. 1987. 94: 236-254.

[18] Gonzalez, R, Wu, G. On the shape of the probability weighting function. *Cognitive Psychology*. 1999. 38: 129-166.

[19] Hey, JD, and Orme, C. Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica*. 1994. 62: 1291-1326.

[20] Kahneman, D, and Tversky, A. Choices, values and frames. 2000. Cambridge University Press, New York.

[21] Kahneman, D, and Tversky, A. Prospect theory : An analysis of decision under risk. *Econometrica*. 1979. 47: 263-291.

[22] Karmarkar, US. Subjectively Weighted Utility and the Allais Paradox. *Organizational Behavior and Human Performance*. 1979. 24: 67-72.

[23] Kilka, M, and Weber, M. What Determines the Shape of the Probability Weighting Function under Uncertainty? *Management Science*. 2001. 47: 1712-1726.

[24] Kunreuther, H, Ginsberg, R, Miller, L, Sagi, P, Slovic, P, Borkan, B, Katz, N. Disaster insurance protection: Public policy lessons. 1978. Wiley, New York.

[25] Lattimore, JR, Baker, JK, Witte, AD. The influence of probability on risky choice: A parametric investigation. *Journal of Economic Behavior and Organization*. 1992. 17: 377-400.

[26] Luce, RD. Reduction invariance and Prelec's weighting functions. *Journal of Mathematical Psychology*. 2001. 45: 167-179.

[27] Luce, RD, Fishburn, PC. Rank- and sign-dependent linear utility models for finite first-order gambles. Journal of Risk and Uncertainty. 1991. 4: 29-59.

[28] Luce, RD, Mellers, BA, and Chang, Shi-Jie. Is Choice the Correct Primitive? On Using Certainty Equivalents and Reference Levels to Predict Choices among Gambles. *Journal of Risk and Uncertainty*. 1993. 6: 115-143.

[29] Machina, MJ. Non-expected utility theory. In Durlaf, SN, and Blume, LE. eds. The New Palgrave Dictionary of Economics. 2nd Edition. 2008. Macmillan, Basingstoke and New York.

[30] Prelec, D. The probability weighting function. *Econometrica*. 1998. 60: 497-528.

[31] Pöystia, L, Rajalina, S, and Summala, H. Factors influencing the use of cellular (mobile) phone during driving and hazards while using it. *Accident Analysis & Prevention*. 2005. 37: 47-51.

[32] Quiggin, J. A theory of anticipated utility. *Journal of Economic Behavior and Organization*. 1982. 3: 323-343.

[33] Quiggin, J. Generalized Expected Utility Theory. 1993. Kluwer Academic Publishers.

[34] Rieger, MO, Wang, M. Cumulative prospect theory and the St. Petersburg paradox. Economic Theory. 2006. 28: 665-679.

[35] Röell, A. Risk Aversion in Quiggin and Yaari's Rank-Order Model of Choice under Uncertainty. 1987. *Economic Journal.* 97: 143-160.

[36] *The Royal Society for the Prevention of Accidents*. The risk of using a mobile phone while driving. 2005. Full text of the report can be found at www.rospa.com.

[37] Safra, Z, and Segal, U. Constant Risk Aversion. 1998. *Journal of Economic Theory.* 83. 19-42.

[38] Schade, C, Kunreuther, H, and Kaas, KP. 2001. Low-Probability Insurance: Are Decisions Consistent with Normative Predictions? mimeo. University of Humboldt. 2001.

[39] Starmer, C. Developments in Non-expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature*. 2000. 38: 332-382.

[40] Starmer, C, and Sugden, R. Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect. *Journal of Risk and Uncertainty.* 1989. 2: 159-178.

[41] Stott, HP. Choosing from cumulative prospect theory's functional menargerie. *Journal of Risk and Uncertainty.* 2006. 32: 101-130.

[42] Tversky, A, Kahneman, D. Advances in prospect theory : Cumulative representation of uncertainty. *Journal of Risk and Uncertainty.* 1992. 5: 297-323.

[43] Wakker, P. Testing and characterizing properties of nonadditive measures through violations of the sure-thing principle. Econometrica. 2001. 69: 1039-1059.

[44] Wakker, P. Separating Marginal Utility and Probabilistic Risk Aversion. Theory and Decision. 1994. 36: 1-44.

[45] Williams, A, and Lund, A. Seat belt use laws and occupant crash protection in the United States. *American Journal of Public Health*. 1986. 76: 1438-42.