Composite Prospect Theory*

Sanjit Dhami† Ali al-Nowaihi††

26 October 2010

Abstract

Evidence suggests three important stylized facts, S1, S2a, S2b. Low probabilities are overweighted, high probabilities are underweighted (S1). Some people ignore events of extremely low probability and treat events of extremely high probability as certain (S2a). Others focus greatly on the size of outcomes, even for extremely low/high probabilities (S2b). We propose composite-cumulative-prospect-theory (CCP) that accounts jointly for S1, S2a, S2b. We discuss several applications, including insurance behavior, the class of Becker-paradoxes, the Allais-paradox and the St. Petersburg-paradox. CCP explains everything that expected utility, rank dependent utility, prospect theory and cumulative prospect theory do; the converse is false.

Keywords: Decision making under risk; Composite Prelec probability weighting functions; Composite cumulative prospect theory; Composite rank dependent utility theory; Insurance; St. Petersburg paradox; Becker’s paradox; Bimodal perception of risks.

JEL Classification: D01 (Microeconomic Behavior: Underlying Principles); D03 (Behavioral Economics: Underlying Principles), D81 (Criteria for Decision Making under Risk and Uncertainty).

*We are very grateful for insightful comments made at the presentation of this paper at the 2010 FUR XIV conference in Newcastle and seminars at the University of Nottingham, University of Leicester and ISET in Georgia. In particular we would like to thank Horst Zank, Ludovic Renou, Chris Starmer, Robin Cubitt, Daniel Seidmann, and Martin Sefton. Dr. Sanjit Dhami would also like to acknowledge a one semester study leave from the University of Leicester.

†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.

††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.
“... people may refuse to worry about losses whose probability is below some threshold. Probabilities below the threshold are treated as zero.” Kunreuther et al. (1978, p. 182).

“Obviously in some sense it is right that he or she be less aware of low probability events, ... but it does appear from the data that the sensitivity goes down too rapidly as the probability decreases.” Kenneth Arrow in Kunreuther et al. (1978, p. viii).

“An important form of simplification involves the discarding of extremely unlikely outcomes.” Kahneman and Tversky (1979, p. 275).

“Individuals seem to buy insurance only when the probability of risk is above a threshold...” Camerer and Kunreuther et al. (1989, p. 570).

1. Introduction

In this paper we are interested in the best possible decision theory that can address the following two stylized facts on human behavior over the probability range \([0, 1]\).

S1. For probabilities in the interval \([0, 1]\), that are bounded away from the end-points, decision makers overweight small probabilities and underweight large probabilities.\(^1\)

S2. For events close to the boundary of the probability interval \([0, 1]\), many that are of enormous importance, evidence suggests two kinds of behaviors.
S2a: A fraction \(\mu \in [0, 1]\) of decision makers (i) ignore events of extremely low probability and, (ii) treat extremely high probability events as certain.\(^2\)
S2b: The remaining fraction \(1 - \mu\) places great salience on the size of the low probability outcome (particularly losses). These individuals behave ‘as if’ the magnitude of the outcomes is of paramount importance.

1.1. Stylized fact S1 in decision theory

Non-expected utility (non-EU) theories postulate a probability weighting function, \(w(p) : [0, 1] \to [0, 1]\), that captures the subjective weight placed by decision makers on the objective probability, \(p\). Such theories, e.g., rank dependent utility (RDU) and cumulative prospect theory (CP), account for S1 by incorporating a \(w(p)\) function that overweights low probabilities but underweights high probabilities. Such non-EU theories can account for S1 and S2b, but they cannot account for S2a. An example is the Prelec (1998) function: \(w(p) = e^{-\beta(-\ln p)^\alpha}\), \(\alpha > 0, \beta > 0\), which is parsimonious and has an axiomatic foundation.

\(^1\)The evidence for stylized fact S1 is well documented and we do not pursue it further; see, for instance, Kahneman and Tversky (1979), Kahneman and Tversky (2000) and Starmer (2000).

\(^2\)In the context of take-up of insurance for low probability natural hazards, one set of experiments by Kunreuther et al. (1978) are consistent with \(\mu = 0.8\).
If $\alpha < 1$, then this function overweights low probabilities and underweights high probabilities, so it conforms to S1 and S2b but not S2a. All non-EU theories, particularly RDU and CP assume a weighting function of this form. However, for $\alpha > 1$ the Prelec function is in conflict with S1 and S2b but respects S2a. The two cases $\alpha < 1$ and $\alpha > 1$ are plotted below for the case $\beta = 1$ and respectively $\alpha = 0.5$ and $\alpha = 2$.

**Remark 1**: To distinguish the two cases $\alpha < 1$ and $\alpha > 1$, we call the former ($\alpha < 1$) the standard Prelec function. This function infinitely overweights infinitesimal probabilities in the sense that $\lim_{p \to 0} w(p)/p = \infty$ and infinitely underweights near-one probabilities in the sense that $\lim_{p \to 1} \frac{1-w(p)}{1-p} = \infty$. Most weighting functions in use in RDU and CP have this property. We shall call these the standard probability weighting functions.

### 1.2. Some basics of non-linear probability weighting

Consider the lottery $(x_1, p_1; x_2, p_2)$ where outcome $x_i$ occurs with probability $p_i \geq 0$, $x_1 < x_2$, $p_1 + p_2 = 1$.\(^3\) Let $u(x)$ be the utility of $x$. The expected utility of this lottery is $p_1 u(x_1) + p_2 u(x_2)$. However, under non-EU theories, decision makers use decision weights, $\pi_i$, to evaluate the value of the lottery as

$$\pi_1 u(x_1) + \pi_2 u(x_2). \quad (1.1)$$

Suppose that the decision weights transform point probabilities, i.e., $\pi_i = \pi_i(p_i)$, as in Kahneman and Tversky’s prospect theory (PT). This can lead to the choice of stochastically dominated lotteries; see Starmer (2000). This problem was rectified by Quiggin’s (1982, 1993) rank dependent utility, RDU, by using cumulative transformations of probability.

**Remark 2 (RDU, monotonicity)**: Suppose that the decision maker follows RDU. Let $w(p) : [0, 1] \to [0, 1]$ be a (strictly increasing) probability weighting function. For the

\(^3\)Everything we say here can be extended to general lotteries.
lottery \((x_1, p_1; x_2, p_2)\), \(x_1 < x_2\), the decision weights are given by the following cumulative transformations of probabilities,

\[
\pi_1 = w(p_1 + p_2) - w(p_2), \quad \pi_2 = w(p_2).
\]  

(1.2)

Notice from (1.2) that one requires that the weighting function \(w(p)\) be defined at all points in the domain \([0, 1]\).

1.3. A discussion of the rationale for Stylized fact S2

The evidence for S2 comes from several sources, which we now briefly review.

1.3.1. Bimodal perception of risk

There is strong evidence of a bimodal perception of risks, see Camerer and Kunreuther (1989) and Schade et al. (2001). Some individuals do not pay attention to losses whose probability falls below a certain threshold (stylized fact S2a), while for others, the size of the loss is relatively more salient despite the low probability (stylized fact S2b). McClelland et al. (1993) find strong evidence of a bimodal perception of risks for insurance for low probability losses. Individuals may have a threshold below which they underweight risk and above which they overweight it. Furthermore, individuals may have different thresholds. Across any population of individuals, for any given probability, one would then observe a bimodal perception of risks; see Viscusi (1998). Kunreuther et al. (1988) argue that the bimodal response to low probability events is pervasive in field studies.

1.3.2. Prospect theory (PT)

The only decision theory that incorporates S2 (and S1) is the Nobel prize winning work of Kahneman and Tversky’s (1979) prospect theory, PT. PT makes a distinction between an editing and an evaluation/decision phase. Whilst there are other aspects of the editing phase, from our perspective the most important aspect takes place when decision makers decide which improbable events to treat as impossible and which probable events to treat as certain (stylized fact S2a). Kahneman and Tversky (1979) used the particular point transformation of probability, \(\pi_i = w(p_i)\). They drew \(\pi(p)\), as in Figure 1.1, which is undefined at both ends to reflect issues of S2a, S2b.

Kahneman and Tversky’s (1979, p.282-83) summarize the evidence for S2, as follows. “The sharp drops or apparent discontinuities of \(\pi(p)\) [in our terminology] at the endpoints 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...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 $\pi(p)$ [$w(p)$ in our terminology] is not well-behaved near the end-points.”

In this quote, Kahneman and Tversky are very explicit about human behavior near $p = 0$. Such events, in their words, are either ignored (S2a) or overweighted (S2b). However, the informal and heuristic approach of PT creates difficulties in formal applications. Following the editing phase, the decision maker evaluates lotteries as in (1.1), above.

A second difficulty with PT is that point transformations of probabilities can lead to a violation of monotonicity (see Remark 2 and Kahneman and Tversky, 1984). In response, Tversky and Kahneman (1992) modified PT to cumulative prospect theory (CP) by incorporating the insights from RDU (see Remark 2). Hence, they needed a probability weighting function, $w(p)$, that is defined for the entire probability domain $[0,1]$ (unlike Figure 1.1). This meant eliminating the psychologically rich editing phase, which by incorporating S2a created the gaps at the end-points. However, this implies that under CP, S2a cannot be taken into account.

### 1.3.3. The class of Becker paradoxes

An interesting class of problems where stylized fact S2 applies is the following. Consider the class of non-mandatory acts where the decision maker’s actions result with some small probability in very large loss (inflicted by nature, self inflicted or imposed by others). We call the resulting small-probability loss as a Becker-type punishment. Examples include making a decision to (i) commit a crime when almost infinite punishments (e.g., capital punishments) are available, (ii) not buy insurance against low-probability, large-losses, (iii) jump red traffic lights, (iv) drive and talk on mobile phones, (v) not use seat belts in moving vehicles, (vi) cross the road, etc.
In the context of (i) Becker (1968) showed that, under expected utility (EU), it is optimal to impose the severest possible penalty (to deter crime) with the lowest possible probability (to economize on costs of deterrence). This is clearly an example of a Becker type punishment (indeed Kolm rephrased it as: it is efficient to hang offenders with probability zero). Empirical evidence is not supportive of the Becker proposition. This we call as the Becker paradox. The Becker paradox is even more severe under RDU and CP because of overweighting of infinitesimal probabilities (Remark 1, Dhami and al-Nowaihi, 2010b).

Similar comments apply to the decisions in (ii)-(vi). In each case, under EU, RDU, CP, all decision makers should be dissuaded from Becker-type punishments. This we call the class of Becker propositions. The evidence is to the contrary. We call the set of such refutations as the class of Becker paradoxes. Modifying EU such that decision makers follow S1, S2b alone (as in RDU, CP) does not resolve the paradoxes. A coherent explanation requires, necessarily, that a decision theory must respect S1, S2a, S2b. These claims are discussed in Section 2; see also Dhami and al-Nowaihi (2010a,b).

1.4. Determinants of $\mu$

The bimodal perception of risks framework and PT do not specify the size of the fraction $\mu$ that follows S2a. $\mu$ is likely to depend on the context and the problem. We believe that the salience of low-probability large outcomes, particularly losses (and so the size of $\mu$) can be influenced by the media, family, friends, and public policy.

Consider two examples from the class of Becker paradoxes in Section 2. The take-up of free breast cancer examination rose greatly after the greatly publicized mastectomies of Betty Ford and Happy Rockefeller. Vivid public warnings of the fatal consequences of speeding (viewed by some as low probability events) can have a similar effect.

The size of $\mu$ can also be influenced by emotions, experience, time available to make a decision, bounded rationality, framing, incentive effects and so on.\(^4\) We take an agnostic position as to the source of $\mu$. For our purposes we shall simply take $\mu$ as given.

1.5. Composite Prelec functions (CPF) that address S1, S2a, S2b

Some would argue that for situations of risk RDU, PT and CP are the most satisfactory alternatives to EU. PT and CP also apply to uncertainty; see Tversky and Kahneman (1992) and Wakker (2010). However, no decision theory simultaneously accounts for S1, S2a, S2b in a satisfactory manner. A desirable theory of decision making should incorporate S1 and S2 and, like RDU and CP, respect monotonicity. We propose such a theory,

that we call *composite cumulative prospect theory* (CCP).\(^5\) We also show how RDU can be modified to produce *composite rank dependent utility theory* (CRDU) that also possesses these desirable features. However, we argue that CCP is more satisfactory. At the heart of the CCP is our proposed *composite Prelec weighting function* (CPF) that, by modifying the end-points of the standard Prelec function, can address S1 and S2a; see Figure 1.2 for the general shape of a CPF. In Figure 1.2, decision makers heavily underweight very low probabilities in the range \([0, p_1]\) in the following sense.

**Remark 3**: For a CPF, \(\lim_{p \to 0} \frac{w(p)}{p} = 0\) and \(\lim_{p \to 1} \frac{1 - w(p)}{1 - p} = 0\).

Contrast the CPF with the standard Prelec function where \(\lim_{p \to 0} \frac{w(p)}{p} = 0\) (see Remark 1). Decision makers who use the CPF ignore very low probability events by assigning low subjective weights to them, hence, conforming to S2a. Events in the interval \([p_3, 1]\) are overweighted as suggested by Kahneman and Tversky (1979, p.282-83). In the middle segment, \(p \in [p, \overline{p}]\), the CPF is inverse-S shaped, identical to a standard Prelec function, and accounts for S1. We shall show that the CPF is *parsimonious, flexible* and has an *axiomatic derivation* and furthermore that the Prelec function is a special case of the CPF.

**1.6. Composite cumulative prospect theory (CCP)**

Under our proposed *composite prospect theory* (CCP) a fraction, \(\mu\), of individuals use CP but replace the standard weighting function (see Remark 1) with the CPF. Consequently, these individuals respect S1, S2a. The remaining fraction, \(1 - \mu\), use CP with its standard probability weighting function. Consequently, these individuals respect S1, S2b. In effect, this is what PT tries to achieve through it’s editing phase followed by an evaluation phase.

---

\(^5\)We shall abbreviate this to *composite prospect theory.*
However, and unlike PT, our analysis is completely formal and suitable for applications, axiomatically founded and, (like CP) respects stochastic dominance. Hence, CCP might be viewed as bridging the gap between PT and CP, yet enlarging the range of phenomena that can be explained, particularly those where $S_{2a}, S_{2b}$ coexist.

*Composite rank dependent utility* (CRDU) is similar to CCP except that RDU replaces CP. Unlike CRDU, CCP (like CP) incorporates reference dependence, loss aversion and richer attitudes towards risk. It explains everything that CRDU can, but the converse is false. Furthermore, CCP explains $S_1$, $S_{2a}$, $S_{2b}$, while CP (like RDU) only explains $S_1$, $S_{2b}$. Hence, CCP explains everything that CP can, but the converse is false.

1.7. Addressing potential criticisms of CCP

We would like to preempt the following potential criticisms of CCP.

1. *The argument about more parameters*: CCP has more parameters than CP, so one might argue that it is not a surprise that it performs better. However, it is often the case in science that if a theory B explains everything that theory A does and more, then theory B has more parameters.\(^6\) This potential criticism only makes sense if one could come up with a decision theory that explains everything that CCP does and more, yet has a smaller number of parameters. Clearly EU, RDU, and CP are not potential candidates for this exercise. Indeed for situations where $S_{2a}$ applies, EU, RDU and CP often make incorrect predictions, while CCP makes the correct predictions.

2. *The argument that CCP is not novel enough*: Except for probability weighting and mixtures of individual types, CCP shares its features with CP. It is also possibly true that, conceptually, the advance made by RDU over EU, or by CP over RDU is of a greater magnitude than that made by CCP over CP. However, in science, theories are judged by their explanatory power not by one’s subjective views about the magnitude of advance. CCP explains everything that CP, RDU, EU explain and more. The converse is false. Indeed the range of phenomena that CCP explains that other theories cannot, i.e., those associated with the simultaneous satisfaction of $S_1$, $S_{2a}$, $S_{2b}$, is very large and important, as we argue in this paper. Furthermore, all of the existing mainstream theories, such as EU, RDU, CP, are special cases of CCP. In that sense CCP is consistent with scientific progress.

\(^6\)For instance, EU has more parameters than expected value maximization, RDU has more parameters than EU, and CP has more parameters than RDU. Similarly, hyperbolic discounting has more parameters than exponential discounting, a range of other *regarding preferences models* have more parameters than the purely selfish agent model and so on.
3. The argument about the adequacy of the mixture of individuals proposed under CCP:
   Is the mixture of individuals with respective proportions, \( \mu, 1 - \mu \) rich enough?
   Why, for instance, are there not individuals with EU or RDU preferences among (or in addition to) this mixture? These arguments are easily answered. First, EU and RDU are special cases of CP, hence, individuals with such preferences are special cases of our theory. Second, the theory can also accommodate more categories in the mixture. For our purposes we have found the existing mixture in subsection 1.6 to be adequate and supported by the evidence.

1.8. Schematic outline of the paper

Section 2 gives the evidence for S2a. Section 3 discusses non-linear weighting of probabilities and, in particular, the Prelec (1998) function. Section 4 introduces the composite Prelec weighting function (CPF) while Section 5 gives its axiomatic derivation. Composite cumulative prospect theory (CCP) is introduced in Section 6. Section 7 gives applications of CCP, which include the economics of insurance and the economics of crime and punishment. We also comment in passing about how CCP can resolve the Allais paradox and the St. Petersbourg paradox and give a brief comparison with alternative decision theories under risk. Brief conclusions are given in Section 8. All proofs are collected in the Appendix.

2. The importance of low probability events and problems for existing theory: A discussion of S2a

In this section we present an overview of the evidence for stylized fact S2.

2.1. Insurance for low probability events

The total global gross insurance premiums for 2008 were 4.27 trillion dollars, which accounted for 6.18% of global GDP.\(^7\) The study of insurance is crucial in almost all branches of economics. Yet, despite impressive progress, existing theoretical models are unable to explain the stylized facts for the take-up of insurance for low probability events.

The seminal study of Kunreuther et al. (1978)\(^8\) 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). This study

\(^7\)See Plunkett’s Insurance Industry Almanac, 2010, Plunkett Research, Ltd., for the details.

\(^8\)In the foreword, Arrow (Kunreuther et al.,1978, p.vii) writes: “The following study is path breaking in opening up a new field of inquiry, the large scale field study of risk-taking behavior.” The reader can also consult similar but more recent evidence in Kunreuther and Pauly (2004, 2006).
had 135 expert contributors, involving samples of thousands. Survey data, econometric analysis and experimental evidence all gave identical results.

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. Kunreuther et al. (1978, chapter 7) presented subjects with varying potential losses with various probabilities, keeping the expected loss fixed. 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.9

The lack of interest in buying insurance for low probability events arose despite active government attempts to (i) provide subsidy to overcome transaction costs (ii) reduce premiums below their actuarially fair rates (iii) provide reinsurance for firms and (iv) provide relevant information. Hence, one can safely rule out these factors. Furthermore, insurees were aware of the losses (many overestimated them). Moral hazard issues (e.g., expectation of federal aid in the event of disaster) were not found to be important.

Arrow’s own reading of the evidence in Kunreuther et al. (1978) is that the problem is on the demand side rather than on the supply side. Arrow writes (Kunreuther et al., 1978, p.viii) “Clearly, a good part of the obstacle [to buying insurance] was the lack of interest on the part of purchasers.” Kunreuther et al. (1978, p. 238) 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.” This behavior is in close conformity to S2a.

2.2. The Becker (1968) paradox

We now develop further the arguments in Section 1.3.3. A celebrated result, the Becker (1968) proposition, states that the most efficient way to deter a crime is to impose the ‘severest possible penalty with the lowest possible probability’. This we refer to as a Becker-type punishment. 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, which is relatively less costly, the deterrence effect of the punishment is maintained. Indeed, under EU, risk-neutrality/aversion, potentially infinitely severe punishments10 and utility that is unbounded below, the Becker proposition implies that crime would be deterred completely, however small the probability of detection

---

9 These results were shown to be robust to changes in subject population, changes in experimental format, order of presentation, presenting the risks separately or simultaneously, bundling the risks, compounding over time and introducing ‘no claims bonuses’.

10 For instance, ruinous fines, slavery, torture, extraction of body parts (all of which have been historically important) and modern capital punishment.
and conviction. Kolm (1973) memorably phrased this proposition as *it is efficient to hang offenders with probability zero*. Empirical evidence strongly suggests that the behavior of a significant number of individuals does not conform to the Becker proposition. This is known as the **Becker paradox**.

Under RDU and CP, because decision makers heavily overweight the small probability of a punishment (see Remark 1), they are deterred by Becker type punishments even more as compared to EU and so they are unable to explain the Becker paradox.

Composite prospect theory (CCP) was outlined in subsection 1.6, above. Under CCP, a fraction $\mu$ of the population conforms to S1, S2a and heavily underweights very small probabilities (see Remark 3). These individuals are not deterred by Becker type punishments. The remaining fraction, $1 - \mu$, conforms to S1, S2b and use the standard CP model. They are deterred by Becker-type punishments. In conjunction, this explains the Becker paradox, i.e., why some individuals (the fraction $\mu$) is not deterred by Becker-type punishments while others (the fraction $1 - \mu$) is deterred. Again, this demonstrates the necessity of incorporating S1, S2a, S2b (see the simple crime example in Section 7.2).

These remarks apply to the *entire class of Becker paradoxes* (see subsection 1.3.3, above). This is the class of actions for decision makers where theory predicts that all rational individuals will be deterred by Becker-type punishments, but not all are deterred. Other members in this class are considered below. These include, red traffic light running, driving and talking on mobile phones, non-mandatory breast cancer examination etc.

Radelet and Ackers (1996) survey 67 of the 70 current and former presidents of three professional criminology organizations in the USA. Over 80% of the experts believe that existing research does not support the deterrence capabilities of capital punishment, as would be predicted by the Becker proposition. Levitt (2004) shows that the estimated contribution of capital punishment in deterring crime in the US over the period 1973-1991, was zero. History does not bear out the Becker proposition either. Since the late middle ages, the severity of punishments has declined while expenditures on enforcement have increased. Polinsky and Shavell (2007: 422-23) write that: "... substantial enforcement costs could be saved without sacrificing deterrence by reducing enforcement effort and simultaneously raising fines."

### 2.2.1. The competing explanations for the Becker paradox

The reader may, rightly, wonder if there are explanations for the class of Becker paradoxes other than some individuals simply ignoring low probabilities as under S2a (the explanation proposed in CCP). Dhami and al-Nowaihi (2010b) explore nine potential explanations

---

11The probability of capital punishment is not the issue here. If capital punishment cannot deter crime for a higher probability for some individuals then it cannot do so for near-zero probabilities either, which is the essence of the Becker proposition.
of the class of Becker paradoxes in the literature and show that none of these explanations suffice. It is beyond the scope of this paper to provide the details, so we merely list these potential explanations here: (1) Risk seeking behavior. (2) Bankruptcy issues. (3) Differential punishments. (4) Errors in conviction. (5) Rent seeking behavior. (6) Norms and abhorrence of severe punishments. (7) Objectives other than deterrence. (8) Incorporating the utility of offenders in the social welfare function. (9) Pathological traits of offenders.

In particular, all of these explanations are contradicted by the evidence from a member of the class of Becker paradoxes, i.e., jumping red traffic lights. We examine this next.

2.3. Evidence from jumping red traffic lights

Consider the act of running a red traffic light. There is (at least) a small probability of an accident with potentially infinite cost (e.g., loss of life). Hence, this belongs to the class of Becker paradoxes. Moreover, the fatal consequences are self-inflicted.

Bar-Ilan and Sacerdote (2001, 2004) estimate that there are 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. Clearly, this is an activity of economic significance and it is implausible to assume that running red traffic lights are simply ‘mistakes’. Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004) provide, to our minds, near decisive evidence that the explanations in subsection 2.2.1 above, singly or jointly, cannot provide a satisfactory explanation of the Becker paradox within an EU framework.\(^\text{12}\)

Using Israeli data, Bar-Ilan (2000) 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. But, the corresponding numbers for the additional risks of serious and fatal injuries are 13.9 days and 69.4 days respectively, which should completely deter the offense. However, evidence is hugely to the contrary.

Clearly EU combined with risk aversion would struggle to explain this evidence. Explanations 2-8 in section 2.2.1, are not applicable here, because the punishment is self-inflicted. In particular, one cannot argue along the lines of Explanation 6, that there are any particular norms or fairness considerations to jump red traffic lights. Explanation 9 is also inadequate, for Bar-Ilan and Sacerdote (2004) report “We find that red-light running decreases sharply in response to an increase in the fine ... Criminals convicted of violent offences or property offences run more red lights on average but have the same elasticity as drivers without a criminal record”. This leaves Explanation 1, i.e., global risk seeking. But that creates tremendous problems for almost all areas of economics.

A far more natural explanation for observed red traffic light running, is that stylized

\(^{12}\)More details can be found in Dhami and al-Nowaihi (2010b).
fact S2a applies for a fraction $\mu$ of individuals. Thus, red traffic light running is simply caused by some individuals ignoring (or seriously underweighting) the very low probability of an accident (for whatever reasons; see subsection 1.4, above).

2.4. Driving and talking on car mobile phones

Consider the usage of hand-held mobile phones in moving vehicles. A user of mobile phones 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.\(^\text{13}\) The Becker proposition applied to this situation suggests that drivers will not use mobile phones while driving. But evidence is to the contrary. Hence, this belongs to the class of Becker paradoxes.

Evidence in the UK indicates that up to 40 percent of individuals drive and talk on mobile phones; see, for example, the RSPA (2005). Pöystia et al. (2004) report that two thirds of Finnish drivers and 85% of American drivers use their phone while driving, which increases the risk of an accident by two to six fold. Hands-free equipment, although now obligatory in many countries, seems not to offer essential safety advantages.

None of the arguments in subsection 2.2.1 suffice to explain the facts. A more natural explanation is that S2a applies to a fraction $\mu$ of the population, that ignores the (perceived) very low probability of an accident.

2.5. Other examples from the class of Becker paradoxes

People were reluctant to use non-mandatory seat belts despite publicly available evidence that seat belts save lives. Prior to 1985, in the US, only 10-20% of motorists wore seat belts voluntarily; see Williams and Lund (1986). Car accidents may be perceived by individuals as *low probability events*, particularly if individuals are overconfident of their driving abilities\(^\text{14}\). But accidents are *potentially fatal*. Even as evidence accumulated on the dangers of breast cancer (*low probability, potentially fatal, event*\(^\text{15}\)) women took up the offer of breast cancer examination, only sparingly.\(^\text{16}\) Each of these examples is a member of the class of Becker paradoxes. In each case it seems that (for whatever reasons) a fraction $\mu$ of the population simply ignores low probability events.

---

\(^{13}\) Extensive evidence suggests that the perceived probability of an accident might be even lower than the actual probability because drivers are *overconfident* of their driving abilities. Taylor and Brown (1998) suggest that upto 90 percent of car accidents might be caused by overconfidence.

\(^{14}\) People assign confidence intervals to their estimates that are too narrow and 90% of those surveyed report that they have above average levels of intelligence and emotional ability. See Weinstein (1980). For further references and applications to finance, see Barberis and Thaler (2003).

\(^{15}\) We now know that the *conditional* probability of breast cancer if there is such a problem in close relatives is not low. However, we refer here to data from a time when such a link was less well understood.

\(^{16}\) In the US, this changed only after the greatly publicized events of the mastectomies of Betty Ford and Happy Rockefeller; see Kunreuther (1978, p.xiii,13-14). Hence, the fraction of individuals to whom stylized fact S2(a) applies, $\mu$, is not static.
2.6. Conclusion from these disparate contexts

Two main conclusions arise from the discussion in this Section. First, human behavior for low probability events cannot be easily explained by the existing mainstream theoretical models of risk. EU and the associated auxiliary assumptions are unable to explain the stylized facts. RDU and CP make the problem even worse (see Remark 1). Second, a natural explanation for these phenomena is that many individuals (the fraction, $\mu$) simply ignore or seriously underweight very low probability events, as in stylized fact S2a.

3. Non-linear transformation of probabilities

The main alternatives to EU in economics, i.e., RDU and CP, incorporate non-linear transformation of probability. We now introduce probability weighting functions with emphasis on the Prelec (1998) function, along with other useful concepts.

**Definition 1** (Probability weighting function)$^{17}$: By a probability weighting function we mean a strictly increasing function $w(p) : [0, 1] \to [0, 1]$.

**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 2**: $w(p)$ (i) infinitely-overweights infinitesimal probabilities, if for $\gamma > 0$, $\lim_{p \to 0} \frac{w(p)}{p^\gamma} = \infty$, and (ii) infinitely underweights near-one probabilities if $\lim_{p \to 1} \frac{1 - w(p)}{1 - p} = \infty$.$^{18}$

**Definition 3**: $w(p)$ (i) zero-underweights infinitesimal probabilities, if for $\gamma > 0$, $\lim_{p \to 0} \frac{w(p)}{p^\gamma} = 0$, and (ii) and zero underweights near-one probabilities in the sense that $\lim_{p \to 1} \frac{1 - w(p)}{1 - p} = 0$.

Data from experimental and field evidence typically suggest that decision makers exhibit an inverse-S shaped probability weighting over outcomes (stylized fact S1). For an example, see the plot of the Prelec function for $\alpha < 1$ in the introduction.$^{13}$

---

$^{17}$Sometimes it is not required that a probability weighting function be onto. We need it here, however, to guarantee the existence of its inverse, which is used in the proof of Proposition 10. We are grateful for discussions with Horst Zank which led to the clarification of this point.

$^{18}$In particular, for $\gamma = 1$, Definition 2(i) implies that $\lim_{p \to 0} \frac{w(p)}{p} = \infty$. Definition 2(ii) is needed for the explanation of stylized fact S2a(ii) (i.e., a fraction $\mu$ of decision makers treat extremely high probability events as certain) and for the explanation of the insurance puzzles when one uses RDU; see al-Nowaihi and Dhami (2010b). The same comment applies to Definition 3(ii), below.
Remark 4 (Standard probability weighting functions): A large number of weighting functions have been proposed, e.g., those by Tversky and Kahneman (1992), Prelec (1998), Gonzalez and Wu (1999) and Lattimore, Baker and Witte (1992). They all infinitely overweight infinitesimal probabilities. We shall call these the standard probability weighting functions. They are essential components of RDU and CP.

3.1. Prelec’s probability weighting function

In order to consider the implications of standard weighting functions for RDU and CP we now consider the most popular function, the Prelec (1998) function. The main comments in this Section apply to all standard weighting functions. The Prelec function is parsimonious and has an axiomatic foundation. Depending on parameter values, it is either consistent with S1, S2b or with S2a, but it is not consistent with all three together.

Definition 4 (Prelec, 1998): By the Prelec function we mean the probability weighting function \( w(p) : [0,1] \rightarrow [0,1] \) given by

\[
\begin{align*}
w(0) &= 0, \quad w(1) = 1, \\
w(p) &= e^{-\beta(-\ln p)^\alpha}, \quad 0 < p \leq 1, \quad \alpha > 0, \quad \beta > 0.
\end{align*}
\]

(3.1)

(3.2)

We make a distinction between the Prelec function and the standard Prelec function; Prelec (1998) prefers the latter for reasons we give below.

Definition 5 (Standard Prelec function): By the standard Prelec probability weighting function we mean the Prelec function, defined in (3.1) and (3.2), but with \( 0 < \alpha < 1 \).

Proposition 2: The Prelec function (Definition 4) is a probability weighting function in the sense of Definition 1.

We now highlight the roles of \( \alpha, \beta \) in the Prelec weighting function.

1. (Role of \( \alpha \)) The parameter \( \alpha \) controls the convexity/concavity of the Prelec function. It will be our main parameter of interest. If \( \alpha < 1 \), then the Prelec function is strictly concave for low probabilities but strictly convex for high probabilities. In this case, it is inverse-S shaped, as in \( w(p) = e^{-(-\ln p)^{\frac{1}{2}}} (\alpha = 0.5, \beta = 1) \), which is sketched as the thick curve in Figure 3.1. The converse holds if \( \alpha > 1 \), in which case the Prelec function is S shaped. An example is the curve \( w(p) = e^{-(-\ln p)^{2}} (\alpha = 2, \beta = 1) \), sketched in Figure 3.1 as the light curve (the straight line in Figure 3.1 is the 45° line).
Figure 3.1: Plots of $w(p) = e^{-(\ln p)^0.5}$ and $w(p) = e^{-(\ln p)^2}$.

2. (Role of $\beta$) Between the region of strict convexity ($w'' > 0$) and the region of strict concavity ($w'' < 0$), for each of the cases in 3.1, there is a point of inflexion ($w'' = 0$). The parameter $\beta$ controls the location of the inflexion point relative to the $45^0$ line. Thus, for $\beta = 1$, the point of inflexion is at $p = e^{-1}$ and lies on the $45^0$ line, as in Figure 3.1. However, if $\beta < 1$, then the point of inflexion lies above the $45^0$ line, as in $w(p) = e^{-\frac{1}{2}(\ln p)^2}(\alpha = 2, \beta = \frac{1}{2})$. For this example, the fixed point, $w(p^*) = p^*$, is at $p^* \approx 0.14$ but the point of inflexion, $w''(\tilde{p}) = 0$, is at $\tilde{p} \approx 0.20$.

**Proposition 3**: For $\alpha = 1$, the Prelec probability weighting function (Definition 4) takes the form $w(p) = p^\beta$, is strictly concave if $\beta < 1$ but strictly convex if $\beta > 1$. In particular, for $\beta = 1$, $w(p) = p$ (as under expected utility theory).

**Proposition 4**: Suppose $\alpha \neq 1$. Then the Prelec function has the following properties.

(a) It has exactly three fixed points at 0, $p^* = e^{-\frac{1}{2}(\ln \bar{p})^\alpha}$, and 1.
(b) It has a unique inflexion point, $\bar{p} \in (0, 1)$ at which $w''(\bar{p}) = 0$.
(c) It is strictly concave for $p < \bar{p}$ and strictly convex for $p > \bar{p}$ (inverse-S shaped) if $\alpha < 1$.
(d) It is strictly convex for $p < \bar{p}$ and strictly concave for $p > \bar{p}$ (S shaped) if $\alpha > 1$.
(e) The inflexion point, $\bar{p}$, lies above, on, or below, the $45^0$ line, respectively as $\beta$ is less than, equal to or greater than one.

Table 1, below, exhibits the various cases established by Proposition 4.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta &lt; 1$</th>
<th>$\beta = 1$</th>
<th>$\beta &gt; 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha &lt; 1$</td>
<td>inverse-S shape $\bar{p} &lt; w(\bar{p})$</td>
<td>inverse-S shape $\bar{p} = w(\bar{p})$</td>
<td>inverse-S shape $\bar{p} &gt; w(\bar{p})$</td>
</tr>
<tr>
<td>$\alpha = 1$</td>
<td>strictly concave $p &lt; w(p)$</td>
<td>$w(p) = p$</td>
<td>strictly convex $p &gt; w(p)$</td>
</tr>
<tr>
<td>$\alpha &gt; 1$</td>
<td>S shape $\tilde{p} &lt; w(\tilde{p})$</td>
<td>S shape $w(\tilde{p}) = \tilde{p}$</td>
<td>S shape $\tilde{p} &gt; w(\tilde{p})$</td>
</tr>
</tbody>
</table>

Table 2, graphs the Prelec function $w(p) = e^{-\beta(\ln p)^\alpha}$, for the cases in Table 1.
Table 2: Representative graphs of $w(p) = e^{-\beta(-\ln p)^{\alpha}}$.

Corollary 1: Suppose $\alpha \neq 1$. Then $\bar{p} = p^* = e^{-1}$ (i.e., the point of inflexion and the fixed point, coincide) if, and only if, $\beta = 1$. If $\beta = 1$, then:

(a) If $\alpha < 1$, then $w$ is strictly concave for $p < e^{-1}$ and strictly convex for $p > e^{-1}$ (inverse-S shape; see the thick curve in Figure 3.1).
(b) If $\alpha > 1$, then $w$ is strictly convex for $p < e^{-1}$ and strictly concave for $p > e^{-1}$ (S shape; see the light curve in Figure 3.1).

For the thick curve in Figure 3.1, where $\alpha < 1$, the slope of $w(p)$ becomes very steep near $p = 0$. By contrast, for the thin curve in Figure 3.1, where $\alpha > 1$, the slope of $w(p)$ becomes shallow near $p = 0$ and $p = 1$. 

<table>
<thead>
<tr>
<th>$\beta = \frac{1}{2}$</th>
<th>$\beta = 1$</th>
<th>$\beta = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ w(p) = e^{-\frac{1}{2}(-\ln p)^{\frac{1}{2}}} ]</td>
<td>[ w(p) = e^{-(\ln p)^{\frac{1}{2}}} ]</td>
<td>[ w(p) = e^{-2(\ln p)^{\frac{1}{2}}} ]</td>
</tr>
<tr>
<td>$\alpha = \frac{1}{2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ w(p) = p^{\frac{1}{2}} ]</td>
<td>$w(p) = p$</td>
<td>[ w(p) = p^2 ]</td>
</tr>
<tr>
<td>$\alpha = 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ w(p) = e^{-\frac{1}{2}(-\ln p)^{\frac{2}{2}}} ]</td>
<td>[ w(p) = e^{-(\ln p)^{\frac{2}{2}}} ]</td>
<td>[ w(p) = e^{-2(\ln p)^{\frac{2}{2}}} ]</td>
</tr>
<tr>
<td>$\alpha = 2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Proposition 5: (a) For $\alpha < 1$ the Prelec (1998) function: (i) infinitely-overweights infinitesimal probabilities, i.e., $\lim_{p \to 0} w(p) = \infty$, and (ii) 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) zero-underweights infinitesimal probabilities, i.e., $\lim_{p \to 0} w(p) = 0$, and (ii) zero-overweights near-one probabilities, i.e., $\lim_{p \to 1} \frac{1-w(p)}{1-p} = 0$.

Setting $\gamma = 1$ in Proposition 5 gives the following corollary.

Corollary 2: Consider the the Prelec (1998) function. (a) For $\alpha < 1$, $\lim_{p \to 0} \frac{w(p)}{p} = \infty$. (b) For $\alpha > 1$, $\lim_{p \to 0} \frac{w(p)}{p} = 0$.

Recall from Definitions 4 and 5 (and Remark 1 in the introduction), that we make a distinction between the Prelec function ($\alpha > 0$) and the standard Prelec function ($0 < \alpha < 1$). Prelec (1998) himself favours the case $0 < \alpha < 1$ and it is this case which is standard in RDU and CP. According to Prelec (1998, p505), the infinite limit in Corollary 2a capture the qualitative change as we move from impossibility to improbability and from certainty to probability. On the other hand, they contradict stylized fact S2a, i.e., the observed behavior that many people ignore events of very low probability. These problems are avoided for $\alpha > 1$. However, for $\alpha > 1$, the Prelec function is S-shaped, see Proposition 4(d) and Figure 3.1, which conflicts with stylized fact S1 and S2b.

4. Composite Prelec probability weighting function

In this Section, we derive the composite Prelec probability weighting function (CPF) that was motivated in the introduction; see Figure 1.2. The CPF is able to simultaneously address the two stylized facts S1 and S2a. We now see how segments from 3 Prelec functions can be combined in order to construct a CPF using two illustrative numerical examples motivated by the empirical evidence from Kunreuther et al. (1978). The axiomatic derivation follows in Section 5, below.

4.1. First example: The farm experiments in Kunreuther (1978)

In their “farm” experiments Kunreuther et al. (1978, ch.7) report that the take-up of actuarially fair insurance declines if the probability of the loss (keeping the expected loss constant) goes below 0.05. To capture this finding, we postulate a CPF composed of segments from three Prelec functions, and given by

$$w(p) = \begin{cases} 
  e^{-0.19286(-\ln p)^2}, & \text{i.e., } \alpha = 2, \beta = 0.19286, \text{ if } 0 < p < 0.05 \\
  e^{-(\ln p)^2}, & \text{i.e., } \alpha = 0.5, \beta = 1, \text{ if } 0.05 \leq p \leq 0.95 \\
  e^{-86.081(-\ln p)^2}, & \text{i.e., } \alpha = 2, \beta = 86.081, \text{ if } 0.95 < p \leq 1 
\end{cases}$$ (4.1)
The CPF in (4.1) is plotted in Figure 4.1. For $0 \leq p < 0.05$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_0(-\ln p)\alpha_0}$, with $\alpha_0 = 2$, $\beta_0 = 0.19286$. For $0.05 \leq p \leq 0.95$, the CPF is identical to the inverse-S shaped Prelec function with $\alpha = 0.5$, $\beta = 1$. For $0.95 < p \leq 1$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_1(-\ln p)\alpha_1}$, with $\alpha_1 = 2$, $\beta_1 = 86.081$.

The first and the third segments of the CPF correspond to the light Prelec curve in Figure 3.1 (with $\alpha > 1$), while the second segment corresponds to the thick Prelec curve in Figure 3.1 (with $\alpha < 1$). $\beta_0$ is chosen to make $w(p)$ continuous at $p = 0.05$, while $\beta_1$ is chosen to make $w(p)$ continuous at $p = 0.95$.

**Remark 5** (Fixed points): This CPF (4.1) has five fixed points: $0$, $0.0055993$, $e^{-1}$, $0.98845$ and $1$. It is strictly concave for $0.05 < p < e^{-1}$ and strictly convex for $e^{-1} < p < 0.95$. It is strictly convex for $0 < p < 0.05$ and strictly concave for $0.95 < p < 1$.

**Remark 6**: (Underweighting and overweighting of probabilities): The CPF in Figure 4.1 overweights ‘low’ probabilities, in the range $0.0055993 < p < e^{-1}$ and underweights ‘high’ probabilities, in the range $e^{-1} < p < 0.98845$. Figures 4.2 and 4.3, below, respectively, magnify the regions near 0 and near 1. The CPF underweights ‘very low’ probabilities, in the range $0 < p < 0.0055993$ and overweights ‘very high’ probabilities, in the range $0.98845 < p < 1$. For $p$ close to zero, the CPF is nearly flat, thus capturing Arrow’s astute observation: “Obviously in some sense it is right that he or she 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.” (Kenneth Arrow in Kunreuther et al., 1978, p. viii). The CPF is also nearly flat near 1. These two segments, i.e., $p \in (0, 0.0055993) \cup (0.98845, 1)$ are able to address S2a.
Remark 7: For other data sets (see e.g., the urn experiments, below), one need not obtain the same, or similar, cut-off points in (4.1). The CPF allows for such flexibility.

Remark 8: In representing the data from Kunreuther et al. (1978) we have chosen the values of $a$ to give the correct shape of the CPF. We then chose the values of $b$ to ensure continuity across the three segments of the Prelec function. The converse, i.e., choosing $b$ values first and then fixing the $a$ values to ensure continuity is less attractive because the $a$ values control the degree of concavity/convexity of the Prelec function and, so, must ensure that the CPF is of the “correct shape” as in Figure 1.2.

4.2. Second example: The urn experiment in Kunreuther (1978)

In their “urn experiments”, Kunreuther et al. (1978, chapter 7) report that 80% of subjects facing actuarially fair premiums took up insurance against a loss with probability 0.25. But the take-up of insurance declined when the probability of the loss declined, keeping the expected loss constant. At a probability of 0.001, only 20% took up insurance. These results motivate the following CPF,

$$w(p) = \begin{cases} 
    e^{-0.61266(-\ln p)^2}, & \text{i.e., } a = 2, \beta = 0.61266, \quad \text{if } 0 \leq p < 0.25, \\
    e^{-(\ln p)^{\frac{1}{2}}}, & \text{i.e., } a = 0.5, \beta = 1, \quad \text{if } 0.25 \leq p \leq 0.75, \\
    e^{-6.4808(-\ln p)^2}, & \text{i.e., } a = 2, \beta = 6.4808, \quad \text{if } 0.75 < p \leq 1.
\end{cases}$$

(4.2)
The plot of the CPF in (4.2) and it’s properties are similar to the case in subsection 4.1. The interested reader can find the details and discussion in al-Nowaihi and Dhami (2010b).

4.3. A more formal treatment of the CPF

For the first segment of the CPF’s in (4.1), (4.2), the upper cutoff point, \( p \) (as in Figure 1.2), is respectively, at probabilities 0.05 and 0.25. Similarly, for the second segment of the CPF in (4.1), (4.2), the upper cutoff point, \( \overline{p} \) (as in Figure 1.2), is respectively, at probabilities 0.95 and 0.75. Now define,

\[
\overline{p} = e^{-\left(\frac{\beta}{\alpha_0}\right)\frac{1}{\alpha}}, \quad p = e^{-\left(\frac{\beta}{\alpha_1}\right)\frac{1}{\alpha}}. \tag{4.3}
\]

The CPF’s in (4.1), (4.2), suggest the following definition (a more general definition is given in Section 5, below).

**Definition 6** (Composite Prelec weighting function, CPF): By the composite Prelec weighting function we mean the probability weighting function \( w : [0, 1] \rightarrow [0, 1] \) given by

\[
w(p) = \begin{cases} 
0 & \text{if } p = 0 \\
 e^{-\beta_0(-\ln p)^\alpha_0} & \text{if } 0 < p \leq \overline{p} \\
 e^{-\beta(-\ln p)^\alpha} & \text{if } \overline{p} < p \leq \overline{\overline{p}} \\
 e^{-\beta_1(-\ln p)^\alpha_1} & \text{if } \overline{\overline{p}} < p \leq 1 
\end{cases} \tag{4.4}
\]

where \( \overline{p} \) and \( \overline{\overline{p}} \) are given by (4.3) and

\[
0 < \alpha < 1, \beta > 0; \alpha_0 > 1, \beta_0 > 0; \alpha_1 > 1, \beta_1 > 0, \beta_0 < 1/\beta \frac{\alpha_0 - 1}{1 - \alpha}, \beta_1 > 1/\beta \frac{\alpha_1 - 1}{1 - \alpha}. \tag{4.5}
\]

**Proposition 6**: The composite Prelec function (Definition 6) is a probability weighting function in the sense of Definition 1.

The restrictions \( \alpha > 0, \beta > 0, \beta_0 > 0 \) and \( \beta_1 > 0 \), in (4.5), are required by the axiomatic derivations of the Prelec function (see Prelec (1998), Luce (2001) and al-Nowaihi and Dhami (2006)). The restriction \( \beta_0 < 1/\beta \frac{\alpha_0 - 1}{1 - \alpha} \) and \( \beta_1 > 1/\beta \frac{\alpha_1 - 1}{1 - \alpha} \) ensure that the interval \((\overline{p}, \overline{\overline{p}})\) is not empty. The interval limits are chosen so that the CPF in (4.4) is continuous across them. Define \( p_1, p_2, p_3 \) (see Figure 1.2) as

\[
p_1 = e^{-\left(\frac{1}{\alpha_0}\right)\frac{1}{\alpha}}, \quad p_2 = e^{-\left(\frac{1}{\beta}\right)\frac{1}{\alpha}}, \quad p_3 = e^{-\left(\frac{1}{\alpha_1}\right)\frac{1}{\alpha}}. \tag{4.6}
\]

**Proposition 7**: (a) \( p_1 < p < p_2 < \overline{\overline{p}} < p_3 \). (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, a CPF overweights low probabilities in the range \((p_1, p_2)\) and underweights high probabilities in the range \((p_2, p_3)\), thus, accounting for S1. But, in addition, and unlike all the standard probability weighting functions, it underweights near-zero probabilities in the range \((0, p_1)\), and overweights near-one probabilities in the range \((p_3, 1)\) as required in S2a.

The restrictions \(\alpha_0 > 1, \alpha_1 > 1\) in (4.5) ensure that a CPF has the following properties, that help explain human behavior for extremely low probability event.

**Proposition 8**: For the CPF (4.4), \(\forall \gamma > 0\), the CPF (i) zero-underweights infinitesimal probabilities, i.e., \(\lim_{p \to 0} \frac{w(p)}{p^\gamma} = 0\), and (ii) zero-overweights near-one probabilities, i.e., \(\lim_{p \to 1} \frac{1-w(p)}{1-p} = 0\).

5. Axiomatic derivation of the CPF

We may start with any of the standard axiomatizations of preferences over lotteries that guarantee the existence of probability weighting functions and then ask what extra assumptions are needed to generate the CPF? All standard axiomatizations allow for a representation of preferences of a form that are additively separable in terms that are products of probability weights and utility from outcomes.\(^{19}\)

Wakker (2001), for instance, allows for a preference axiomatization of convex capacities. This allows for a weighting function that is either concave or convex throughout, hence, these global conditions cannot simultaneously address stylized facts S1, S2a, S2b. Diecidue et al. (2009) introduce local conditions that allow for probabilistic risk aversion to vary over the probability interval \([0, 1]\). For one set of parameter values, they can address S1 while for the complementary set of parameter values, they can address S2a. However, their proposal does not simultaneously address S1 and S2a, S2b.

We choose the more direct route and assume that preferences over lotteries is represented by a payoff function that is additively separable in terms that are products of probability weights and utilities.\(^{20}\) We then ask what axioms must be imposed directly on the probability weights that lead to the CPF. It will turn out that a modification of the axiom of power invariance (due to al-Nowaihi and Dhami, 2006), we call local power invariance, turns out to suffice.

**Remark 9**: al-Nowaihi and Dhami (2006) give their axiom of power invariance directly in terms of lotteries. Indeed in what follows we could have alternatively stated everthing in

\(^{19}\)For example Wakker (1994), Wakker and Zank (2002) and Abdellaoui (2002).

\(^{20}\)Such a representation is a consequence in any of the standard axiomatizations, e.g., Abdellaoui (2002) or Wakker (1994).
terms of lotteries. This would have required using any of the standard axioms on lotteries that represents preferences as additively separable in decision weights and utilities from outcome and impose, in addition, the axiom of local power invariance on lotteries instead of probability weights. Since all this is completely standard, we prefer our more direct and shorter approach.

Prelec (1998) gave an axiomatic derivation of the Prelec probability weighting function based on ‘compound invariance’, Luce (2001) provided a derivation based on ‘reduction invariance’ and al-Nowaihi and Dhami (2006) give a derivation based on ‘power invariance’. Since the Prelec function satisfies all three, ‘compound invariance’, ‘reduction invariance’ and ‘power invariance’ are all equivalent. These derivations do not put any restrictions on \( \alpha \) and \( \beta \) (see Definition 4) other than \( \alpha > 0 \) and \( \beta > 0 \).

Here we introduce a version of power invariance that we call local power invariance. On the basis of this behavioral property, we derive the CPF.

**Definition 7** (Local power invariance): Let \( 0 = p_0 < p_1 < ... < p_n = 1 \). A probability weighting function, \( w \), satisfies local power invariance if, for \( i = 1, 2, ..., n \), and \( \forall p, q \in (p_{i-1}, p_i) \), \( (w_i(p))^\mu = w_i(q) \) and \( p^\lambda, q^\lambda \in (p_{i-1}, p_i) \) imply \( (w(p^\lambda))^\mu = w(q^\lambda) \).

**Definition 8** (Composite Prelec function, CPF): By the composite Prelec function we mean the function \( w : [0, 1] \rightarrow [0, 1] \) given by

\[
w(p) = \begin{cases} 
0 & \text{if } p = 0 \\
e^{-\beta_i(-\ln p)^{\alpha_i}} & \text{if } p_{i-1} < p \leq p_i, i = 1, 2, ...n,
\end{cases}
\]  

(5.1)

where \( \alpha_i > 0 \), \( \beta_i > 0 \), \( p_0 = 0 \), \( p_n = 1 \), and

\[e^{-\beta_i(-\ln p_i)^{\alpha_i}} = e^{-\beta_{i+1}(-\ln p_i)^{\alpha_{i+1}}} \], \( i = 1, 2, ...n - 1 \).  

(5.2)

The restriction (5.2) is needed to insure that \( w \) is continuous.

**Proposition 9**: The composite Prelec functions (Definition 8) are probability weighting functions in the sense of Definition 1.

**Definition 9** (Useful notation)\(^{21}\): Let \( 0 = p_0 < p_1 < ... < p_n = 1 \). Define \( P_1 = (0, p_1] \), \( P_n = [p_{n-1}, 1) \) and \( P_i = [p_{i-1}, p_i] \), \( i = 2, 3, ..., n - 1 \). Given \( p \in P_i \), \( i = 1, 2, ..., n \), define \( \Lambda_i \) as follows. \( \Lambda_1 = [\ln p_1, \infty) \), \( \Lambda_n = (0, \ln p_{n-1}] \), \( \Lambda_i = [\ln p_i, \ln p_{i-1}] \), \( i = 2, 3, ..., n - 1 \).

\(^{21}\)The problem with the notation \( [\ln p_i, \ln p_{i-1}] \) is that it is not defined for the two special cases \( i = 1 \), \( p = p_0 = 0 \) and \( i = n \), \( p = p_n = 1 \). We have, therefore, introduced Definition 9 to avoid this, by excluding the points 0 and 1.
Proposition 10 (CPF representation): The following are equivalent.

(a) The probability weighting function, $w$, satisfies local power invariance.

(b) There are functions, $\varphi_i : \Lambda_i \to \mathbb{R}_{++}$, such that $\varphi_i$ is $C^1$ on $\left(\frac{\ln p_i}{\ln p}, \frac{\ln p_i - 1}{\ln p}\right)$, $i = 1, 2, \ldots, n$, where $0 = p_0 < p_1 < \ldots < p_n = 1$, and, $\forall p \in P_i$, $\forall \lambda \in \Lambda_i$, $w(p^\lambda) = (w(p))^{\varphi_i(\lambda)}$. Moreover, for each $i = 1, 2, \ldots, n$, $\exists \alpha_i \in (0, \infty)$, $\varphi_i(\lambda) = \lambda^{\alpha_i}$.

(c) $w$ is a composite Prelec function (Definition 8).

Remark 10: Notice that the definition of a CPF allows for $n \geq 1$ segments. For $n = 1$, the CPF reduces to the Prelec function, local power invariance reduces to power invariance, and Proposition 10 reduces to Theorem 2 of al-Nowaihi and Dhami (2006). For $n = 3$ we get the three piece CPF of Definition 6 and illustrated in Figure 1.2 of the introduction.

6. Composite cumulative prospect theory (CCP)

The essence of composite cumulative prospect theory, CCP, has already been described in subsection 1.6 in the introduction. In CCP, all decision makers use standard cumulative prospect theory, CP, with the following essential difference. A fraction $\mu$ uses the CPF (for which $\lim_{p \to 0} \frac{w(p)}{p} = 0$, $\gamma > 0$). Hence, the behavior of these individuals conforms to S1, S2a. The remaining fraction $1 - \mu$ uses the standard Prelec probability weighting function (for which $\lim_{p \to 1} \frac{1 - w(p)}{1 - p} = 0$). Hence, the behavior of these individuals conforms to S1, S2b.

We give a brief formal statement of CCP. Consider a lottery in the form

$$L = \left(y_{-m}; p_{-m}; \ldots; y_0; p_0; y_1; p_1; \ldots; y_n; p_n\right),$$

where $y_{-m} < \ldots < y_0 < \ldots < y_n$ are the wealth outcomes and $p_{-m}, \ldots, p_0, \ldots p_n$ are the corresponding probabilities, so $\sum_{i=-m}^{n} p_i = 1$ and $p_i \geq 0$. In CCP, as in CP, decision makers derive utility from wealth relative to a reference point, $y_0$.\footnote{\textit{y}_0$ could be initial wealth, status-quo wealth, state-dependent wealth, average wealth, desired wealth, rational expectations of future wealth etc. depending on the context. For a discussion, see Kahneman and Tversky (2000), Koszegi and Rabin (2006) and Schmidt et al (2008).}

Definition 10 (Lotteries in incremental form) Let $x_i = y_i - y_0, i = -m, -m + 1, \ldots, n$ be the increment in wealth relative to $y_0$ and $x_{-m} < \ldots < x_0 = 0 < \ldots < x_n$. Then, a lottery in incremental form is given by $L = (x_{-m}; p_{-m}; \ldots; x_0; p_0; x_1; p_1; \ldots; x_n; p_n)$.

Definition 11 (Set of incremental Lotteries, $\mathcal{L}$): $\mathcal{L}$ is the set of all lotteries of the form given in Definition 10.

Definition 12 (Domains of losses and gains): The decision maker is said to be in the domain of gains if $x_i \geq 0$ and in the domain of losses if $x_i < 0$.\footnote{\textit{y}_0$ could be initial wealth, status-quo wealth, state-dependent wealth, average wealth, desired wealth, rational expectations of future wealth etc. depending on the context. For a discussion, see Kahneman and Tversky (2000), Koszegi and Rabin (2006) and Schmidt et al (2008).}
The utility function under CCP is defined, as in CP, over the set of outcomes, \( \{x_m, ..., x_n\} \).

**Definition 13** (Tversky and Kahneman, 1979). A utility function, \( v(x) \), is a continuous, strictly increasing, map \( v : \mathbb{R} \to \mathbb{R} \) that satisfies:

1. \( v(0) = 0 \) (reference dependence).
2. \( v(x) \) is concave for \( x \geq 0 \) (declining sensitivity for gains).
3. \( v(x) \) is convex for \( x \leq 0 \) (declining sensitivity for losses).
4. \( -v(-x) > v(x) \) for \( x > 0 \) (loss aversion).

Tversky and Kahneman (1992) propose the following utility function:

\[
v(x) = \begin{cases} 
  x^\gamma & \text{if } x \geq 0 \\
  -\lambda (-x)^\theta & \text{if } x < 0 
\end{cases}
\]  

(6.1)

where \( \gamma, \theta, \lambda \) are constants. The coefficients of the power function satisfy \( 0 < \gamma < 1 \), \( 0 < \theta < 1 \). \( \lambda > 1 \) is known as the coefficient of loss aversion.\(^{23}\) Tversky and Kahneman (1992) estimated that \( \gamma \simeq \theta \simeq 0.88 \) and \( \lambda \simeq 2.25 \). The properties in Definition 13, are illustrated in figure 6.1, which is a plot of (6.1) for \( \lambda = 2.5, \gamma = \theta = 0.5 \).

![Figure 6.1: The utility function under CCP](image)

Let \( w(p) \) be the CPF in Definition 6.\(^{24}\)

**Definition 14** (Tversky and Kahneman, 1992). For CCP, the decision weights, \( \pi_i \), are defined as in CP, and are as follows:

<table>
<thead>
<tr>
<th>Domain of Gains</th>
<th>Domain of Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_n = w(p_n) )</td>
<td>( \pi_{-m} = w(p_{-m}) )</td>
</tr>
<tr>
<td>( \pi_{n-1} = w(p_{n-1} + p_n) - w(p_n) )</td>
<td>( \pi_{-m+1} = w(p_{-m} + p_{-m+1}) - w(p_{-m}) )</td>
</tr>
<tr>
<td>( \pi_i = w(\sum_{j=i}^n p_j) - w(\sum_{j=i+1}^n p_j) )</td>
<td>( \pi_j = w(\sum_{j=-m}^i p_j) - w(\sum_{j=-m+1}^i p_j) )</td>
</tr>
<tr>
<td>( \pi_1 = w(\sum_{j=1}^n p_j) - w(\sum_{j=2}^n p_j) )</td>
<td>( \pi_{-1} = w(\sum_{j=-m}^{i-1} p_i) - w(\sum_{j=-m}^{i-2} p_i) )</td>
</tr>
</tbody>
</table>

\(^{23}\)Tversky and Kahneman (1992) assert (but do not prove) that the axiom of preference homogeneity \( ((x, p) \sim y \Rightarrow (kx, p) \sim ky, k \geq 0) \) generates this value function. al-Nowaihi et al. (2008) give a formal proof, as well as some other results (e.g., that \( \gamma \) is necessarily identical to \( \theta \)). See also Wakker and Zank (2002) for another axiomatization of the power form of utility.

\(^{24}\)Weighting functions for the domain of gains, \( w^+(p) \), and losses, \( w^-(p) \), could differ. However, we make the empirically founded assumption that \( w^+(p) = w^-(p) \); see Prelec (1998).
Note that the decision weights across the domain of gains and losses do not necessarily add up to 1. In applying Definition 14, note that in CCP, for the fraction $\mu$ which uses the CPF, one must take into account that the CPF is a three piece Prelec function. Hence, one must use the correct weighting function for each segment.

**Definition 15** (The value function under CCP) The value of the lottery, $L \in \mathcal{L}$, to the decision maker under CCP is given by following well defined value function\(^{25}\)

$$V(L) = \sum_{i=-\infty}^{\infty} \pi_i v(x_i).$$ (6.2)

### 7. Some Applications of CCP

We now give three applications of CCP. The first two applications, to insurance and the Becker paradox respectively, demonstrate that CCP can explain the stylized facts from these problems while EU, RDU and CP fail to do so. We then show that CCP can explain the Allais paradox and the St. Petersburg paradox. Finally, we evaluate CCP against some alternatives.

#### 7.1. Insurance

Suppose that a decision maker can suffer the loss, $L > 0$, with probability $p \in (0,1)$. He/She can buy coverage, $C \in [0, L]$, at the cost $rC$, where $r \in (0,1)$ is the premium rate (which is actuarially fair if $r = p$). Hence, with probability, $1 - p$, the decision maker’s wealth is $W - rC$, and with probability $p$, her wealth is $W - rC - L + C \leq W - rC$. Suppose that the reference point of the individual is the status quo, i.e., $y_0 = W$.

In incremental form (see Definition 10), the decision maker under CCP faces the lottery $-L + C - rC, p; -rC, 1 - p)$. Thus, in both states, the outcomes are in the domain of loss. Recall our discussion of CCP in Section 6, and the value function in (6.2). Thus, the value function from the act of purchasing the level of coverage $C \in [0, L]$ is

$$V_1(C) = w(p) v(-L + C - rC) + [1 - w(p)] v(-rC).$$ (7.1)

Since $V_1(C)$ is continuous on the non-empty compact interval $[0, L]$, an optimal level of coverage, $C^*$, exists, which gives the payoff, $V_1(C^*)$. From Definition 13(3), the utility function is convex for losses, thus, the optimal coverage is either zero or full coverage\(^{26}\):

$$C^* = 0 \text{ or } C^* = L.$$ (7.2)

\(^{25}\)Note that we have not defined $\pi_0$. But since $\pi_0 v(x_0) = \pi_0 v(0) = 0$, we can give $\pi_0$ any value we desire, for example $\pi_0 = 1 - \sum_{i=-\infty}^{\infty} \pi_i - \sum_{i=1}^{\infty} \pi_i$.

\(^{26}\)For the formal proof, see Proposition 6a in al-Nowaihi and Dhami (2010b).
If the decision maker does not buy any insurance coverage, i.e., \( C = 0 \), recalling that \( v(0) = 0 \), the value function is \( V_{NI} = w(p) v(-L) \). For the decision maker to buy any level of coverage \( C^* > 0 \), the participation constraint, \( V_{NI} \leq V_I(C^*) \), must be satisfied. Using (7.2), for \( C^* = L \), the following participation constraint is necessary and sufficient

\[
w(p) v(-L) \leq v(-rL).
\]

Furthermore, and only to facilitate exposition, assume that we have actuarially fair premium, \( r = p \). Then (7.3) becomes:

\[
w(p) v(-L) \leq v(-pL).
\]

For the utility function (6.1), which is axiomatically founded and is consistent with the evidence, this becomes, \( -\lambda w(p) L^\gamma \leq -\lambda p^\gamma L^\gamma \). Simplifying, we get

\[
\frac{w(p)}{p^\gamma} \geq 1.
\]

Thus, if (7.5) holds, then the decision maker will fully insure, \( C^* = L \), otherwise the decision maker will not insure at all, \( C^* = 0 \). In particular, if \( \lim_{p \to 0} \frac{w(p)}{p^\gamma} = \infty \), then the decision maker will fully insure against all losses of sufficiently small probability. On the other hand, if \( \lim_{p \to 0} \frac{w(p)}{p^\gamma} = 0 \), the decision maker will not buy any insurance against any event of sufficiently low probability. These two cases are true, respectively for a fraction \( 1 - \mu \) and \( \mu \) of the decision makers under CCP (recall the introductory discussion in Section 6).

**Remark 11**: Two kinds of insurance behavior are observed for low probability losses. Kunreuther et al. (1978) report little or no take-up of insurance for many individuals. However, Sydnor (2010) reports overinsure for modest risks for many individuals. Under CCP this can be accounted for, due to the presence of the two types of individuals with respective fractions \( \mu \) and \( 1 - \mu \) and the fact that the evidence suggests that \( \mu \) depends on the problem, frame, and context.

The following examples compare CP and CCP using insurance data from Kunreuther et al. (1978).

**Example 1**: Suppose that a decision maker faces a loss, \( L \), with probability \( p = 0.001 \). Suppose that the utility function of the decision maker is (6.1). The experimental value for \( \gamma \) suggested by Kahneman and Tversky (1979) is \( \gamma = 0.88 \). We check if it is optimal for a decision maker to insure under, respectively, CP and CCP.

(a) Decision maker uses CP: In particular, the decision maker’s probability weighting
function is the standard Prelec function with $\beta = 1$ and $\alpha = 0.50$, i.e., $w(p) = e^{-(\ln p)^{0.50}}$. In this case (7.5) requires that
\[
e^{-\left(\ln 0.001\right)^{0.50}} \geq 1 \iff 31.518 \geq 1, \text{ which is true.}
\]
So, a decision maker using CP will fully insure. But, as noted in section 4.2, Kunreuther’s (1978) data shows that only 20% of the decision makers insure in this case.

(b) Decision maker uses CCP: Decision makers who belong to the fraction $1 - \mu$ use CP and so, as in (a), will fully insure. Now consider those who belong to the fraction $\mu$. They use the composite Prelec function given in (4.2). Since $p = 0.001 \in (0, 0.25)$, (4.2) gives $w(0.001) = e^{-0.61266(-\ln 0.001)^2}$. In this case (7.5) requires that
\[
e^{-0.61266(-\ln 0.001)^2} \geq 1 \iff 8.7838 \times 10^{-11} \geq 1, \text{ which is not true.}
\]
Hence, a decision maker under CCP who belongs to the fraction $\mu$ of individuals who respect $S2a$ will not insure. Thus, Kunreuther’s (1978) data can be explained by CCP when $\mu = 0.8$ and $1 - \mu = 0.2$.

**Example 2**: We continue to use the set-up of Example 1. However, let the probability of the loss be $p = 0.25$ (instead of $p = 0.001$). Since $p = 0.25 \in [0.25, 0.75]$, (4.2) gives $w(0.25) = e^{-(\ln 0.25)^{0.50}}$. In this case (7.5) requires that
\[
e^{-\left(\ln 0.25\right)^{0.50}} \geq 1 \iff 1.0434 \geq 1, \text{ which is true.}
\]
Hence, such a decision maker will insure fully against any loss whose probability of occurrence $p \geq 0.25$. Kunreuther’s (1978) data shows that 80% of the experimental subjects took up insurance in this case. Hence, for losses whose probability is bounded well away from the end-points, the predictions of, both, CP and CCP are in close (but not perfect) conformity with the evidence.

In conjunction, Examples 1 and 2 illustrate how CCP can account well for the evidence for events of all probabilities while CP’s predictions for low probability events are incorrect.

In actual practice, the insurance problem is quite complicated.\textsuperscript{27} In particular, there are considerations of fixed costs of insurance, and insurance premiums might not be actuarially

---
\textsuperscript{27}Much of the economics of insurance literature operates under an EU framework. However, EU is unable to explain many important stylized facts in insurance. First, it does not explain the lack of insurance for very low probability events. Second, EU is unable to explain why many people simultaneously gamble and insure (PT, CP, CCP easily explain this). Third, EU recommends probabilistic insurance which is contradicted by the experimental evidence (Kahneman and Tversky, 1979: 269-271).
A full treatment is beyond the scope of the current paper but can be found in al-Nowaihi and Dhami (2010b); the central insight of Examples 1 and 2 survives even in these more complicated settings. The low take-up of insurance for very low probability disasters can be explained by CCP (and CRDU) but not EU, RDU or CP.

7.2. Is it optimal to “hang offenders with probability zero”? 

We now examine the Becker proposition that we outlined in section 2.2. Suppose that an individual receives income \( y_0 \geq 0 \) from being engaged in some legal activity and income \( y_1 \geq y_0 \) from being engaged in some illegal activity. Hence, the benefit, \( b \), from the illegal activity is \( b = y_1 - y_0 \geq 0 \). If engaged in the illegal activity, the individual is caught with some probability \( p \), \( 0 \leq p \leq 1 \). If caught, the individual is asked to pay a fine \( F \),

\[
b \leq F \leq F_{\text{max}} \leq \infty.
\]

(7.6)

Thus, it is feasible to levy a fine that is at least as great as the benefit from crime, \( b \). Society also imposes an upper limit on the fine, \( F_{\text{max}} \). Given the enforcement parameters \( p, F \) the individual makes only one choice: to commit the crime or not.\(^{28}\)

We consider a hyperbolic punishment function which encapsulates in a simple manner, the idea that \( p, F \) are substitutes in deterrence,\(^{29}\)

\[
F = \varphi(p) = b/p.
\]

(7.7)

Notice that in (7.7), fines vary continuously with \( p \).

7.2.1. Illustration of the Becker proposition under EU

Consider an individual with continuously differentiable and strictly increasing utility of income, \( u \). So, from the activity ‘no-crime’, his payoff, \( U_{NC} \), is \( U_{NC} = u(y_0) \). His expected utility from the activity ‘crime’, \( EU_C \), is given by \( EU_C = pu(y_1 - F) + (1 - p) u(y_1) \). It is not worthwhile to engage in crime if the no-crime condition (NCC), \( EU_C \leq U_{NC} \), is satisfied. Substituting \( U_{NC}, EU_C \) in the NCC we get

\[
p u(y_1 - F) + (1 - p) u(y_1) \leq u(y_0).
\]

(7.8)

A simple proof (see Dhami and al-Nowaihi, 2010b, Proposition 1) gives the following result.

\(^{28}\)Dhami and al-Nowaihi (2010b) show that this basic framework nests many important cases in the economics of crime. These include, but are not restricted to, theft/robbery, tax evasion, violation of pollution abatement laws etc.

\(^{29}\)Dhami and al-Nowaihi (2010a,b) show that the hyperbolic punishment function is optimal for a wide class of cost of deterrence and damage from crime functions. Furthermore, it provides an upper bound on punishments for a large and sensible class of cost and damage functions. Thus, if the hyperbolic punishment function cannot deter crime, then neither can the optimal punishment function, if different from the hyperbolic.
Proposition 11 (Becker Proposition; Becker, 1968): Under EU, if the individual is risk neutral or risk averse, then the hyperbolic punishment function $\varphi(p) = \frac{b}{p}$ will deter crime. It follows that given any probability of detection and conviction, $p > 0$, no matter how small, crime can be deterred by a sufficiently large punishment.

7.2.2. The Becker paradox under CP and CCP

Recall from section 2.2 that the empirical evidence is not supportive of the Becker proposition for all individuals (Becker paradox). We now show how the Becker paradox can be resolved. Suppose that the reference incomes for the two activities, crime and no-crime, are, respectively, $y_R$ and $y_r$. Then, since carriers of utility in CP, CCP are outcomes relative to the reference point (see Section 6), the payoff from not committing crime is

$$V_{NC} = v(y_0 - y_r). \quad (7.9)$$

In CP, CCP, the outcome space is split into the domains of gains and losses (Definition 12). If an individual commits a crime he gets income $y_1$, and if caught (with probability $p$), he pays the fine, $F$. We assume that the outcome ($y_1 - F$) is in the domain of losses (i.e. $y_1 - F - y_R < 0$). On the other hand, if he commits a crime and is not caught, then the outcome, $y_1$, is assumed to be in the domain of gains (i.e. $y_1 - y_R > 0$). Thus, we have one outcome each in the domain of losses and gains, with respective decision weights equal to $w(p)$ and $w(1-p)$ (see Definition 14). Then, under CP (or CCP), the individual’s payoff from committing a crime is given by the value function (see Definition 15):

$$V_C = w(p) v(y_1 - F - y_R) + w(1 - p) v(y_1 - y_R). \quad (7.10)$$

Definition 16 (Elation): $v(y_1 - y_R)$ is the elation from committing a crime and getting away with it.

Substituting (7.9), (7.10) into the ‘no crime condition’ ($NCC$), $V_C \leq V_{NC}$, we get

$$w(p) v(y_1 - F - y_R) + w(1 - p) v(y_1 - y_R) \leq v(y_0 - y_r). \quad (7.11)$$

The $NCC$ in (7.11) depends on the two reference points, $y_r$ and $y_R$. The recent literature has suggested that the reference point should be the rational expectation of income.\footnote{See, for instance, Koszegi and Rabin (2006). The act dependence of the reference point that we use is also consistent with the third generation prospect theory of Schmidt et al (2008).} Our model is consistent with a perfect foresight rational expectations path. Hence, the rational expectation of income from any activity is the expected income from that activity. Thus,

$$y_r = y_0; \; y_R = y_1 - p \varphi(p). \quad (7.12)$$
Substituting (7.12) in (7.11) and recalling that \( v(0) = 0 \), the NCC (7.11) becomes

\[
\begin{align*}
w(p) v(-(1-p)\varphi(p)) + w(1-p) v(p\varphi(p)) & \leq 0. \\
(7.13)
\end{align*}
\]

For the power function form of utility (6.1) with \( \vartheta = \gamma \) (which is axiomatic and consistent with the evidence), and noting that \( \varphi(p) > 0 \), the NCC (7.13) becomes,

\[
\begin{align*}
\frac{w(p)}{p^\vartheta} & \geq \frac{w(1-p)}{\lambda(1-p)^\vartheta}. \\
(7.14)
\end{align*}
\]

**Proposition 12**: As the probability of detection approaches zero, a decision maker using CP or CCP who (i) faces a strictly positive punishment, i.e., \( \varphi(p) > 0 \), (ii) satisfies preference homogeneity\(^{31}\) and (iii) whose reference points are given by (7.12),

(a) does not engage in the criminal activity if

\[
\lim_{p \to 0} \frac{w(p)}{p^\vartheta} > \frac{1}{\lambda}. \\
(7.15)
\]

(b) On the other hand, the same individual engages in crime if

\[
\lim_{p \to 0} \frac{w(p)}{p^\vartheta} < \frac{1}{\lambda}. \\
(7.16)
\]

From Proposition 12(a), the Becker paradox reemerges under CP for any standard probability weighting function, for which \( \lim_{p \to 0} \frac{w(p)}{p^\vartheta} = \infty \) (see Section 6). However, under CCP for the fraction \( \mu \) of decision makers, we get that \( \lim_{p \to 0} \frac{w(p)}{p^\vartheta} = 0 \) (see Section 6). The Becker paradox can then be explained under CCP by using Proposition 12(b). Individuals who are deterred by Becker-type punishments under CCP belong to the fraction \( 1 - \mu \) (whose behavior is as in CP), while the remaining fraction \( \mu \) is not deterred, consistent with the evidence.

Several possible extensions are worthwhile. One could allow for issues of heterogeneity of reference points and psychologically richer reference points that allow for regret or social responsibility. It is also desirable to explicitly model the costs of deterrence and damages from crime functions explicitly. All these extensions are considered in Dhami and al-Nowaihi (2010b) but these do not alter the results of the simple model.

### 7.3. The Allais and the St. Petersburg paradoxes

It is well known that the Allais paradox is explained under CP. Since CP and CCP are identical for the fraction, \( 1 - \mu \), of individuals whose behavior conforms to S2b, hence, for

\(^{31}\)This ensures that the value function under CP and CCP has the power form in (6.1); see al-Nowaihi et al. (2008).
these individuals, the Allais paradox is also explained by CCP. The remaining fraction, \( \mu \), whose behavior conforms to S2a, use the composite Prelec function (CPF) rather than the standard Prelec function under CP (and RDU). From examples (4.1), (4.2) and Definition 6, the CPF and the standard Prelec functions are identical for the middle range of probabilities in \([p, \bar{p}]\). So for any lottery whose probabilities lie in this interval they make identical predictions. However, if some probabilities lie outside this range, then one must use the appropriate segment of the CPF in Definition 6. The interval size \([p, \bar{p}]\) is context/frame/problem dependent. In the working paper version, al-Nowaihi and Dhami (2010a) use the CPF in (4.1), (4.2) to show that the Allais paradox is explained under CCP.

Rieger and Wang (2006) show that the St. Petersburg paradox re-emerges under CP. al-Nowaihi and Dhami (2010a) show that this paradox can also be resolved under CCP.

7.4. A brief comparison of CCP with some alternatives

EU is refuted by a large body of literature\(^{32}\), so we focus on other theories. One could consider composite rank dependent utility, CRDU, which is otherwise standard RDU but combined with CPF (see subsection 1.6). However, CRDU, unlike CCP, rules out the psychologically powerful notions of loss aversion and reference dependence, which have strong explanatory power in a variety of contexts.\(^{33}\)

Regret theory tries to incorporate regret and rejoice, into decision theory.\(^{34}\) However, there is no notion of loss aversion and reference dependence; problems that it shares with RDU and CRDU. Furthermore, regret theory is not able to account for the important stylized facts S1 and S2. Hence, it would, for instance, find it hard to solve the applications that we discuss in Section 7.

Birnbaum (2008) has criticized CP (and so, by implication, CCP). The configural weights models, proposed by Birnbaum are considered in detail in the working paper version, al-Nowaihi and Dhami (2010a), but these models are unable to incorporate stylized fact S2. Furthermore, it is shown that for the applications in this paper, the configural weights models do not provide any advantages.

8. Conclusions

Kahneman and Tversky’s (1979) prospect theory (PT) revolutionized decision theory. It was a psychologically rich, largely empirically corroborated and rigorous account of decision

\(^{32}\)See, for instance, Kahneman and Tversky (1979), Kahneman and Tversky (2000), Starmer (2000), and Barberis and Thaler (2003).

\(^{33}\)See, for instance, the applications discussed in Camerer (2000), Kahneman and Tversky (2000) and Barberis and Thaler (2003).

\(^{34}\)We refer the reader to the well known references in Starmer (2000).
making under risk. The psychological richness stemmed, partly, from a distinction between an editing phase and an evaluation/decision phase. An important part of the editing phase was to determine which low probability events to ignore and which high probability events to treat as certain. But this was not rigorously formalized and violations of monotonicity, even when obvious, were allowed.

Tversky and Kahneman (1992) proposed cumulative prospect theory (CP), which eliminated the editing phase altogether but ensured, using insights introduced by Quiggen’s (1982) rank dependent utility (RDU), that decision makers would not choose stochastically dominated options. Thus, the gains from introducing CP were somewhat diminished by substantial loss in psychological realism.

In this paper, we combine PT and CP into composite prospect theory (CCP). In CCP, the editing and decision phases are combined into a single phase using (1) the probability weighting function and (2) two types of decision makers with respective fractions \( \mu \), \( 1 - \mu \). Thus, we are able to combine the psychological richness of PT with the more satisfactory attitudes towards stochastic dominance under CP. CCP is consistent with the evidence, which shows that (i) people overweight low probabilities and underweight high probabilities, but that (ii) many people ignore events of extremely low probability and treat extremely high probability events as certain. We also provide applications of CCP to outstanding problems in economics: insurance behavior towards low probability events and the class of Becker paradoxes (subsection 1.3.3 and Section 2).

9. Appendix: Proofs

Proof of Proposition 1: These properties follow immediately from Definition 1. ■.

Proof of Proposition 2: Straightforward from Definition 4. ■.

Proof of Proposition 3: Obvious from Definition 4. ■.

The following lemmas will be useful.

Lemma 1: For \( \alpha \neq 1 \), the Prelec function (Definition 4) has exactly three fixed points, at 0, \( p^* = e^{-\beta(\frac{1}{\alpha})} \) and 1. In particular, for \( \beta = 1 \), \( p^* = e^{-1} \).

Proof of Lemma 1: From Propositions 1a and 2 it follows that 0 and 1 are fixed points of the Prelec function. For \( \alpha \neq 1 \) and \( p^* \in (0, 1) \), a simple calculation shows that

\[
e^{-\beta(-\ln p^*)^\alpha} = p^* \iff p^* = e^{-\beta(\frac{1}{\alpha})}.\]

In particular, \( \beta = 1 \) gives \( p^* = e^{-1} \). ■.

Lemma 2: Let \( w(p) \) be the Prelec function (Definition 4) and let

\[
f(p) = \alpha \beta (-\ln p)^\alpha + \ln p + 1 - \alpha, \quad p \in (0, 1),
\]

(9.1)
then
\[ f'(p) = \frac{1}{p} \left[ 1 - \alpha^2 \beta (-\ln p)^{\alpha-1} \right], \quad p \in (0, 1), \quad (9.2) \]
\[ w''(p) = \frac{w'(p) - f'(p)}{p(-\ln p)}, \quad p \in (0, 1), \quad (9.3) \]
\[ w''(p) \leq 0 \iff f'(p) \leq 0, \quad p \in (0, 1). \quad (9.4) \]

**Proof of Lemma 2**: Differentiate (9.1) to get (9.2). Differentiate (3.2) twice and use (9.1) to get (9.3). \(-\ln p > 0\), since \(p \in (0, 1)\). \(w'(p) > 0\) follows from Definitions (1) and (4) and Proposition (2). (9.4) then follows from (9.3). \(\blacksquare\).

**Lemma 3**: Let \(w(p)\) be the Prelec function (Definition 4). Suppose \(\alpha \neq 1\). Then
(a) \(w''(\tilde{p}) = 0\) for some \(\tilde{p} \in (0, 1)\) and, for any such \(\tilde{p}\):
(i) For \(\alpha < 1\): \(p < \tilde{p} \Rightarrow w''(p) < 0, \quad p > \tilde{p} \Rightarrow w''(p) > 0 \quad (9.5)\)
(ii) For \(\alpha > 1\): \(p < \tilde{p} \Rightarrow w''(p) > 0, \quad p > \tilde{p} \Rightarrow w''(p) < 0 \quad (9.6)\)
(b) The Prelec function has a unique inflexion point, \(\tilde{p} \in (0, 1)\), and is characterized by
\(f(\tilde{p}) = 0\), where \(f(p)\) is defined in (9.1) i.e., \(\alpha \beta (-\ln \tilde{p})^\alpha + \ln \tilde{p} + 1 - \alpha = 0\).
(c) \(\beta = 1 \Rightarrow \tilde{p} = e^{-1}\).
(d) \(\frac{\partial \tilde{p}}{\partial \beta} = \frac{\alpha \tilde{p} (-\ln \tilde{p})^{1+\alpha}}{(\alpha-1)(\alpha-1)}\).
(e) \(\frac{\partial \ln(\tilde{p})}{\partial \beta} = \frac{(\tilde{p} (-\ln \tilde{p})^{1+\alpha}}{(\alpha-1)(\alpha-1)} \left( e^{\frac{1-\alpha}{\alpha}} \tilde{p}^{\frac{1-\alpha}{\alpha}} - 1 \right)\).
(f) \(\tilde{p} \leq \frac{1}{w(\tilde{p})} \iff \beta \leq \frac{1}{\alpha}\).

**Proof of Lemma 3**: (a) Suppose \(\alpha < 1\). From (9.1) we see that
\[ \lim_{p \to 1} f(p) = 1 - \alpha > 0 \quad \text{lim}_{p \to 0} f(p) = \left[ \frac{\alpha \beta}{(-\ln p)^{1-\alpha}} + 1 \right] \ln p + 1 - \alpha = -\infty. \]
Since \(f\) is continuous, it follows that \(f(\tilde{p}) = 0\), for some \(\tilde{p} \in (0, 1)\). From (9.4), it follows that \(w''(\tilde{p}) = 0\). Since \(\alpha < 1\), (9.1) gives \(\alpha \beta (-\ln \tilde{p})^\alpha + \ln \tilde{p} < 0\) and, hence,
\[ \tilde{p} < e^{-(\alpha \beta)^{-1}}. \quad (9.7) \]
Consider the case, \(p < \tilde{p}\). From (9.7) it follows that \(p < e^{-(\alpha \beta)^{-1}}\) and, hence, \(1 - \frac{\alpha^2 \beta}{(-\ln p)^{1-\alpha}} > 1 - \alpha > 0\). Thus, from (9.2), \(f'(p) > 0\). Since \(f(\tilde{p}) = 0\), it follows that \(f(p) < 0\). Hence, from (9.4), it follows that \(w''(p) < 0\). This establishes the first part of (9.5). The derivation of the second part of (9.5) is similar. The case \(\alpha > 1\), (9.6), is similar.
(b) follows from (a) and (9.1), (9.4).
Lemma 4: Let \( p_i, P_i \) and \( \Lambda_i \) be as in Definition 9. Then,

\[
\lambda \in \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right),
\]

\[p \in (p_{i-1}, p_i) \Rightarrow \lambda \in \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right),
\]

\[
\lambda \notin \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right)
\]

furthermore, \( 0 < \frac{\ln p_i}{\ln p} < 1 < \frac{\ln p_{i-1}}{\ln p} \).

\[p \in P_i \Rightarrow \lambda \in \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right),
\]

\[p^\lambda \in P_i \Leftrightarrow \lambda \in \Lambda_i.
\]

(c) Since \( f(e^{-1}) = 0 \) for \( \beta = 1 \), it follows from (b) that \( \bar{p} = e^{-1} \) in this case.

(d) Differentiating the identity \( f(\bar{p}) = 0 \) with respect to \( \beta \) gives

\[
\frac{\partial f}{\partial \beta} = \frac{\alpha}{\beta} \left( -\frac{\ln \bar{p}}{\bar{p}} \right),
\]

then using \( f(\bar{p}) = 0 \) gives

\[
\frac{\partial f}{\partial \beta} = \frac{\alpha}{\beta} \left( -\frac{\ln \bar{p}}{\bar{p}} \right).
\]

(e) Differentiate \( w(\bar{p}) - \bar{p} = e^{-\beta(\ln \bar{p})^\alpha} - \bar{p} \) with respect to \( \beta \), and use (d) and \( f(\bar{p}) = 0 \), to get

\[
\frac{\partial [w(\bar{p}) - \bar{p}]}{\partial \beta} = \frac{\alpha}{\beta} \left( -\frac{\ln \bar{p}}{\bar{p}} \right).
\]

(f) Assume \( \alpha < 1 \). For \( \beta = 1 \), \( \bar{p} = e^{-1} \) and, hence, \( e^{\frac{1}{\alpha}} - (\frac{1}{\alpha}) \alpha = e^{\frac{1}{\alpha}} (e^{-1})^{\frac{1}{\alpha}} - \alpha = 1 - \alpha > 0 \). Since \( \frac{\partial f}{\partial \beta} < 0 \) for \( \alpha < 1 \), it follows that \( e^{\frac{1}{\alpha}} - (\frac{1}{\alpha}) \alpha > 0 \) for \( \beta \leq 1 \). Hence, \( \frac{\partial [w(\bar{p}) - \bar{p}]}{\partial \beta} < 0 \) for \( \beta \leq 1 \). We have \( w(\bar{p}) - \bar{p} = w(e^{-1}) - e^{-1} = w(p^*) - p^* = 0 \) for \( \beta = 1 \) (recall part c and Lemma 1). Hence, \( w(\bar{p}) > \bar{p} \) for \( \beta < 1 \). The case \( \beta \geq 0 \) is similar. The case \( \alpha > 1 \) is similar.

Proof of Proposition 4: (a) is established by Lemma 1. (b) is established by Lemma 3b. (c) follows from Lemma 3a(i). (d) follows from Lemma 3a(ii). (e) follows from Lemma 3f.

Proof of Corollary 1: Immediate from Proposition 4.

Proof of Proposition 5: Since \( p^{-\gamma} = e^{-\gamma \ln p} \), (3.2) gives

\[
\frac{w(p)}{p} = e^{-\gamma \ln p} \left( \gamma - \frac{\beta}{(-\ln p)^{1-\alpha}} \right).
\]

Note that \( \lim_{p \to 0} (-\ln p) = \infty \).

(a) For \( 0 < \alpha < 1 \), we get \( \lim_{p \to 0} (-\ln p)^{1-\alpha} = \infty \) and, hence, \( \lim_{p \to 0} \frac{\beta}{(-\ln p)^{1-\alpha}} = 0 \). Hence, since \( \gamma > 0 \), we get \( \lim_{p \to 0} \frac{w(p)}{p} = \lim_{p \to 0} e^{-\gamma \ln p} = \lim_{p \to 0} p^{-\gamma} = \infty \). Part (ii) follows from the following result. If \( p \to 1 \), then \( w(p) \to w(1) = 1 \). By L’Hospital’s rule,

\[
\frac{1-w(p)}{1-p} \to \frac{d(1-w(p))}{dp} = w(p).
\]

Taking the derivative of the Prelec function at \( p = 1 \) then gives the result.

(b) For \( \alpha > 1 \), we get \( \lim_{p \to 0} (-\ln p)^{1-\alpha} = 0 \) and, hence, \( \lim_{p \to 0} \frac{\beta}{(-\ln p)^{1-\alpha}} = \infty \). Hence, since \( \gamma > 0 \), we get \( \lim_{p \to 0} \frac{w(p)}{p} = \lim_{p \to 0} e^{-\gamma \ln p} \left( \gamma - \frac{\beta}{(-\ln p)^{1-\alpha}} \right) = 0 \). Part b(ii) follows along the same lines as a(ii).

Proof of Proposition 6: Straightforward from Definition 6.

Proof of Proposition 7: Follows by direct calculation from (4.4) and (4.5).

Proof of Proposition 8: The result follows from Proposition 5(b) and by noting, from Definition 6, that for the CPF, \( \alpha_0 > 1 \) and \( \alpha_1 > 1 \).

Proof of Proposition 9: Straightforward from Definitions 1 and 8.
Proof of Lemma 4: Straightforward from Definition 9. ■.

Lemma 5 (Cauchy’s 3rd algebraic functional equation): Let \( G : (A, B) \rightarrow \mathbb{R}_{++} \) be continuous or monotonic, \( 0 < A < 1 < B \), \( \forall X, Y \in (A, B) \), s.t., \( XY \in (A, B) \), \( G(XY) = G(X)G(Y) \). Then \( \exists c \in \mathbb{R}, \forall X \in (A, B), G(X) = X^c \).


Proof of Proposition 10 (CPF representation): (a)⇒(b). Suppose the probability weighting function, \( w \), satisfies local power invariance.

Let
\[
f(x, \lambda) = w \left( \left( w^{-1} \left( e^{-x} \right) \right)^{\lambda} \right), \quad x, \lambda \in \mathbb{R}_{++},
\]
and
\[
\varphi(\lambda) = -\ln f(1, \lambda) = -\ln w \left( \left( w^{-1} \left( e^{-1} \right) \right)^{\lambda} \right), \quad \lambda \in \mathbb{R}_{++}.
\]
Obviously,
\[
\varphi \text{ maps } \mathbb{R}_{++} \text{ into } \mathbb{R}_{++}.
\]
Since \( w^{-1}(e^{-1}) \in (0,1) \), it follows that \( (w^{-1}(e^{-1}))^\lambda \) is a strictly decreasing function of \( \lambda \), and so are \( w \left( \left( w^{-1}(e^{-1}) \right)^{\lambda} \right) \) and \( \ln w \left( \left( w^{-1}(e^{-1}) \right)^{\lambda} \right) \). Hence, from (9.12),
\[
\varphi \text{ is a strictly increasing function of } \lambda.
\]

From (9.11) we get
\[
f(-\mu \ln w (p), \lambda) = w \left( \left( w^{-1} \left( (w(p))^\mu \right) \right)^{\lambda} \right), \quad p \in (0,1), \lambda, \mu \in \mathbb{R}_{++}.
\]
Let \( 0 = p_0 < p_1 < \ldots < p_n = 1 \). Let
\[
p, q \in (p_{i-1}, p_i), \quad (w(p))^\mu = w(q), \quad p^\lambda, q^\lambda \in (p_{i-1}, p_i).
\]
From (9.16) we get
\[
q = w^{-1}((w(p))^\mu).
\]
From (9.16) and local power invariance, we get
\[
(w(p^\lambda))^\mu = w(q^\lambda).
\]
Substituting for \( q \) from (9.17) into (9.18), we get
\[
(w(p^\lambda))^\mu = w \left( \left( w^{-1} \left( (w(p))^\mu \right) \right)^{\lambda} \right), \quad p, p^\lambda \in (p_{i-1}, p_i).
\]
From (9.19) and (9.15) we get
\[ f(-\mu \ln w(p), \lambda) = \left(w(p^\lambda)\right)^\mu, \quad p, p^\lambda \in (p_{i-1}, p_i). \]  \tag{9.20}

In particular, for \( \mu = 1 \), (9.20) gives

\[ f(-\ln w(p), \lambda) = w(p^\lambda), \quad p, p^\lambda \in (p_{i-1}, p_i). \]  \tag{9.21}

From (9.21) we get

\[ (f(-\ln w(p), \lambda))^{\mu} = (w(p^\lambda))^{\mu}, \quad p, p^\lambda \in (p_{i-1}, p_i). \]  \tag{9.22}

From (9.20) and (9.22) we get

\[ f(-\mu \ln w(p), \lambda) = (f(-\ln w(p), \lambda))^{\mu}, \quad p, p^\lambda \in (p_{i-1}, p_i). \]  \tag{9.23}

Put

\[ z = -\ln w(p). \]  \tag{9.24}

From (9.23) and (9.24) we get

\[ f(\mu z) = (f(z, \lambda))^\mu, \quad p, p^\lambda \in (p_{i-1}, p_i). \]  \tag{9.25}

From (9.12) and (9.25) we get

\[ f(\lambda) = (f(1, \lambda))^\mu = e^{-\mu \varphi(\lambda)}, \quad p, p^\lambda \in (p_{i-1}, p_i), \]  \tag{9.26}

and, hence,

\[ f(-\ln w(p), \lambda) = (w(p))^\varphi(\lambda), \quad p, p^\lambda \in (p_{i-1}, p_i). \]  \tag{9.27}

From (9.21) and (9.27) we get

\[ w(p^\lambda) = (w(p))^\varphi(\lambda), \quad p, p^\lambda \in (p_{i-1}, p_i), \]  \tag{9.28}

from which we get,

\[ \varphi(\lambda) = \frac{\ln w(p^\lambda)}{\ln w(p)}, \quad p, p^\lambda \in (p_{i-1}, p_i). \]  \tag{9.29}

Let \( p, p^\lambda, p^\mu, p^{\lambda\mu} \in (p_{i-1}, p_i) \). From (9.28) and (9.29) we get

\[ \varphi(\lambda\mu) = \frac{\ln w(p^{\lambda\mu})}{\ln w(p)} = \frac{\ln w(p^\lambda)}{\ln w(p)} \cdot \frac{\ln w(p^\mu)}{\ln w(p)} = \frac{\varphi(\lambda) \ln [w(p)^\varphi(\mu)]}{\ln w(p)} = \frac{\varphi(\lambda)}{\ln w(p)} \cdot \frac{\ln w(p)^{\varphi(\mu)}}{\ln w(p)} = \varphi(\lambda) \varphi(\mu), \quad \text{i.e.,} \]

\[ \varphi(\lambda\mu) = \varphi(\lambda) \varphi(\mu), \quad p, p^\lambda, p^\mu, p^{\lambda\mu} \in (p_{i-1}, p_i). \]  \tag{9.30}

From (9.8), (9.9), and (9.30) we have: \( \varphi \) is strictly increasing on \( \left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right) \).
0 < \frac{\ln p_i}{\ln p} < 1 < \frac{\ln p_{i-1}}{\ln p}, \forall \lambda, \mu \in \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right), \text{ s.t., } \lambda \mu \in \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right), \varphi(\lambda \mu) = \varphi(\lambda) \varphi(\mu). \text{ Hence, Lemma 5 gives:}

\exists \alpha_i \in \mathbb{R}, \forall \lambda \in \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right), \varphi(\lambda) = \lambda^{\alpha_i}. \quad (9.31)

But, by (9.14), \varphi is a strictly increasing function of \lambda. Hence,

\alpha_i > 0. \quad (9.32)

Let P_1 and \Lambda_i be as in Definition 9. Let p \in P_i. Define \varphi_i : \Lambda_i \rightarrow \mathbb{R}_{++} by \varphi_i(\lambda) = \lambda^{\alpha_i}. Then, clearly, \varphi_i is C^1 on \left( \frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p} \right). A simple calculation verifies that \forall p \in P_i, \forall \lambda \in \Lambda_i, w(p^\lambda) = (w(p))^{\varphi_i(\lambda)}. This completes the proof of part (b).

(b)⇒(c). Since e^{-1} \in (0, 1), e^{-1} \in P_i for some i = 1, 2, ..., n. We first establish the result for P_i, then we use induction, and the continuity conditions (5.2), to extend the result to P_{i+1}, P_{i+2}, ..., P_n and \forall p \in P_{i-1}, P_i, P_{i-2}, ..., P_1. Let \beta_i = -\ln w(e^{-1}). Then w(e^{-1}) = e^{-\beta_i}.

Let p \in P_i. Let \lambda = -\ln p. Then p = e^{-\lambda}. Hence w(p) = w(e^{-\lambda}) = w\left(\left(e^{-1}\right)^{\lambda} = (w\left(e^{-1}\right))^{\varphi_i(\lambda)} = e^{-\beta_i \lambda^{\alpha_i}} = e^{-\beta_i (-\ln p)^{\alpha_i}}. Thus we have shown

w(p) = e^{-\beta_i (-\ln p)^{\alpha_i}}, p \in P_i. \quad (9.33)

Let p \in P_i. Let \lambda = \frac{\ln p}{\ln p_i}. Then p = p_i^\lambda. Hence, w(p) = w(p_i^\lambda) = (w(p_i))^{\varphi_{i+1}(\lambda)} = (w(p_i))^{\varphi_{i+1}(\lambda)} = (e^{-\beta_i (-\ln p_i)^{\alpha_i}})^{\lambda^{\alpha_i+1}} = e^{-\beta_{i+1} (-\ln p_i)^{\alpha_i+1}} = e^{-\beta_{i+1} (-\ln p_i)^{\alpha_i+1}}. Thus we have shown

w(p) = e^{-\beta_{i+1} (-\ln p)^{\alpha_i+1}}, p \in P_{i+1}. \quad (9.34)

Let p \in P_{i-1}. Let \lambda = \frac{\ln p}{\ln p_{i-1}}. Then p = p_{i-1}^\lambda. Hence, w(p) = w(p_{i-1}^\lambda) = (w(p_{i-1}))^{\varphi_i(\lambda)} = (w(p_{i-1}))^{\varphi_i(\lambda)} = (e^{-\beta_{i-1} (-\ln p_{i-1})^{\alpha_i}})^{\lambda^{\alpha_i-1}} = e^{-\beta_{i-1} (-\ln p_{i-1})^{\alpha_i-1}} = e^{-\beta_{i-1} (-\ln p_{i-1})^{\alpha_i-1}}. Thus we have shown

w(p) = e^{-\beta_{i-1} (-\ln p)^{\alpha_i-1}}, p \in P_{i-1}. \quad (9.35)

Continuing the above process, we get

w(p) = e^{-\beta_i (-\ln p)^{\alpha_i}}, p \in P_i, i = 1, 2, ..., n, \quad (9.36)

which establishes part (c).

Finally, a simple calculation shows that (c) implies (a). ■

Proof of Proposition 11 (Becker Proposition): If \varphi(p) = \frac{b}{p} then the NCC holds for concave u and, hence, the result follows. ■
Proof of Proposition 12: First, note that $\lim_{p \to 0} \frac{w(1-p)}{(1-p)^p} = 1$ because $w(1) = 1$. If (7.15) holds, then the NCC (7.14) will hold with strict inequality in some non-empty interval $(0, p_1)$.\footnote{For the Prelec weighting function, for all suitably high values of $1-p$, $w(1-p) < 1-p$. However as $p \to 0$ and so $1-p \to 1$, $w(1) = 1.$} Hence, no crime will occur if $p \in (0, p_1)$. If (7.16) holds, then the converse of the NCC (7.14) holds with strict inequality in some non-empty interval $(0, p_2)$. Hence, for punishment to deter in this case, we must have $p > p_2$. ■

References


