University of Leicester

DEPARTMENT OF ECONOMICS

Composite Prospect Theory: A proposal to combine 'prospect theory' and 'cumulative prospect theory'

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

Working Paper No. 10/11
April 2010

Composite Prospect Theory: A proposal to combine 'prospect theory' and 'cumulative prospect theory'

Ali al-Nowaihi* **Sanjit Dhami†**

14 April 2010

Abstract

Evidence shows that (i) people overweight low probabilities and underweight high probabilities, but (ii) ignore events of extremely low probability and treat extremely high probability events as certain. The main alternative decision theories, *rank dependent utility* (RDU) and *cumulative prospect theory* (CP) incorporate (i) but not (ii). By contrast, *prospect theory* (PT) addresses (i) and (ii) by proposing an *editing phase* that eliminates extremely low probability events, followed by a *decision phase* that only makes a choice from among the remaining alternatives. However, PT allows for the choice of stochastically dominated options, even when such dominance is obvious. We propose to combine PT and CP into *composite cumulative prospect theory* (CCP). CCP combines the editing and decision phases of PT into one phase and does not allow for the choice of stochastically dominated options. This, we believe, provides the best available alternative among decision theories of risk at the moment. As illustrative examples, we also show that CCP allows us to resolve three paradoxes: the insurance paradox, the Becker paradox and the St. Petersburg paradox.

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.

JEL Classification: C60(General: Mathematical methods and programming), D81(Criteria for decision making under risk and uncertainty).

*Department of Economics, University of Leicester, University Road, Leicester. LE1 7RH, UK. Phone: +44-116-2522898. Fax: +44-116-2522908. E-mail: aa10@le.ac.uk.

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

"... 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, 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).

"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 $[0, 1]$, extensive evidence, that we review below, suggests the following. Decision makers (i) ignore events of extremely low probability and, (ii) treat extremely high probability events as certain.

Events close to the boundary of the interval $[0, 1]$ are of great significance in economic as well as non-economic domains. Consider some examples, which we discuss in Section 3. Should one attempt to cross the road when there is a small probability that one might meet an accident? Should one buy non-mandatory insurance for low probability events, such as earthquakes, floods and other natural hazards? Should one run red traffic lights when there is a small probability of an accident? Should one take-up breast cancer examination when there could be a small probability of having cancer? Should one wear seat belts when there is a small probability of an accident? What should one make of a fundamental proposition in law and economics, due to Becker (1968): "it is optimal to hang offenders with probability zero"? This, of course, only touches the domain of such low probability events that are of potential interest.

So what decision theory should one choose if one is interested in S1, S2, in addition to other possible desirable features? EU, because it weights probabilities linearly, fails to explain both, S1 and S2. The leading alternatives to EU are Quiggen's (1982, 1993)

[1] The evidence for stylized fact S1 is well documented; see, for instance, Kahneman and Tversky (1979), Kahneman and Tversky (2000) and Starmer (2000).

rank dependent utility (RDU), Kahneman and Tversky's (1979) *prospect theory* (PT) and Tversky and Kahneman's (1992) *cumulative prospect theory* (CP).[2]

PT and CP have proved enormously successful in explaining a wide range of human behavior in economics, psychology, political science, sociology and other disciplines. Indeed, some would argue that for situations of risk (i.e., known probabilities but unknown outcomes), PT and CP provide the most complete and satisfactory descriptions, among the available alternatives.[3] Indeed, PT and CP even have bite in the context of uncertainty.[4] In particular, because PT and CP incorporate several psychologically rich features (e.g., reference dependence, loss aversion, richer attitudes to risk), PT and CP can explain everything that RDU can, but the converse is not true.

Of the existing decision theories, and as we shall explain in greater detail, below, only PT addresses S1 and S2. By contrast, RDU and CP address S1 but not S2. However, PT is subjected to other problems. Notably, that decision makers can choose stochastically dominated options, even when such dominance is obvious. Furthermore, despite being extremely psychologically-rich, the treatment of S2 in PT is informal and heuristic in a manner that does not lend itself easily to formal analysis. Given the empirical importance of S2, this would seem to be an unsatisfactory and unresolved state of affairs in decision theory.

Ideally one would like a theory of decision making under risk that could (a) incorporate the psychological-richness of PT, using rigorous, formal, analytical foundations to address S1 and S2, and (b) yet, like CP not allow decision makers to choose stochastically dominated options when such dominance is obvious. In this paper we show that such a theory that combines PT and CP can be constructed. We call this theory *composite cumulative prospect theory* (CCP). We also show how RDU can be modified to simultaneously take account of S1 and S2. We call the result of this endeavour as *composite rank dependent utility theory* (CRDU). We argue that CCP is more satisfactory as compared to RDU. Furthermore, we apply CCP to show how it can resolve existing puzzles from concrete applications such as insurance, law and economics and the St. Petersburg paradox.

The schematic outline of the paper is as follows. Section 2 gives a heuristic discussion of PT and CP, with particular emphasis on stylized facts S1 and S2. It then outlines the proposal for CCP. Section 3 describes human behavior for low probability events from

[2]PT and CP are sometimes known, respectively, as first and second generation prospect theories.

[3]See, for instance, Kahneman and Tversky (1979), Kahneman and Tversky (2000), Starmer (2000), Barberis and Thaler (2003).

[4]Tversky and Koehler (1994), Rottenstreich and Tversky (1997) outline 'support theory', an influential approach to uncertainty. In essence, they propose axioms which allow, in the first stage, for a situation of uncertainty to be transformed into a situation of risk. Then, in the second stage, decision makers simply use PT or CP to choose among risky prospects. This, once again, allows for a central role for PT and CP, even in situations of uncertainty.

a wide variety of contexts that is relevant for stylized fact S2. Section 4 discusses non-linear weighting of probabilities and, in particular, the Prelec (1998) probability weighting function. Section 5 introduces the composite Prelec weighting function (CPF). Section 6 gives the axiomatic derivation of the CPF. Composite cumulative prospect theory (CCP) is introduced in section 7. Three applications of CCP to unresolved problems in economics are given in section 8. Section 9 argues that CCP is possibly the best among the alternative decision theories under risk. Brief conclusions are give in section 10. All proofs are collected in Appendix-A, while Appendix-B gives some useful results on Cauchy's algebraic functional equations.

2. PT, CP, and CCP: A heuristic discussion of the issues

2.1. Prospect Theory (PT)

PT, which was the outcome of many years of experiments conducted by Kahneman, Tversky, and others, is a psychologically rich theory.[5] The psychological foundations of PT rest, in an important manner, on the distinction between an *editing* and an *evaluation/decision* phase.

From our perspective, the most important and interesting aspect of the editing phase takes place when decision makers decide which improbable events to treat as impossible and which probably events to treat as certain.[6] This is exemplified in the quote from Kahneman and Tversky (1979, p.282): "*On the other hand, the simplification of prospects can lead the individual to discard events of extremely low probability and to treat events of extremely high probability as if they were certain. Because people are limited in their ability to comprehend and evaluate extreme probabilities, highly unlikely events are either ignored or overweighted...*"

Suppose that we have a lottery $(x, p; y, 1-p)$ where $x < y$ are outcomes, and p is a probability. Let $u(x)$ be the utility of the outcome x. The expected utility of this lottery is $pu(x) + (1-p)\,u(y)$. However, under PT and CP, decision makers use decision weights, $\pi(p)$, to evaluate the value of the lottery as $\pi(p)u(x) + \pi(1-p)u(y)$. In the editing phase, among other things, Kahneman and Tversky (1979) were interested in the decision weights, $\pi(p)$, assigned by individuals to very low and very high probability events. They drew $\pi(p)$ as in Figure 2.1.[7] This decision function is discontinuous at both ends, reflecting the vexed issue of how decision makers behave over these ranges of probabilities.

[5] Its importance in economics can be gauged from the fact that Kahneman and Tversky (1979) is the second most cited paper in all of economics. We are grateful to Peter Wakker for pointing this out to us.

[6] Kahneman and Tversky (1979) also identify the *isolation effect* as another important heuristic in the editing phase. This allows decision makers to cancel 'nearly common' components of two prospects before evaluating them. We also show how our theory can also capture some aspects of the isolation effect.

[7] Formal definitions of subjective weights and probability weighting functions are given below.

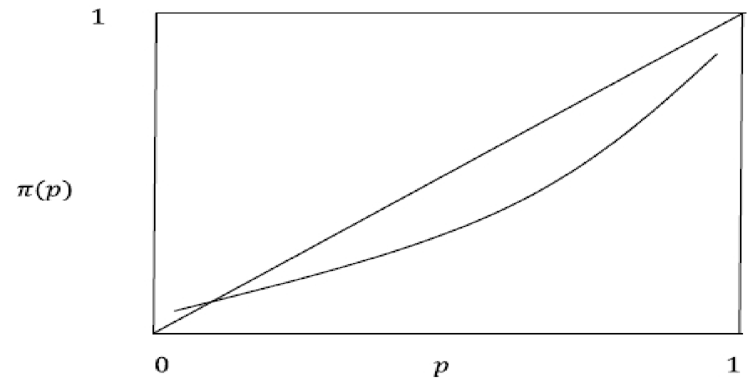


Figure 2.1: Ignorance at the endpoints. Source: Kahneman and Tversky (1979, p. 283)

Kahneman and Tversky's (1979) wrote the following (on p. 282-83) to summarize the evidence on the end-points of the probability interval $[0, 1]$. "*The sharp drops or apparent discontinuities of* $\pi(p)$ *at the end-points are consistent with the notion that there is a limit to how small a decision weight can be attached to an event, if it is given any weight at all. A similar quantum of doubt could impose an upper limit on any decision weight that is less than unity. This quantal effect may reflect the categorical distinction between certainty and uncertainty. On the other hand, the simplification of prospects can lead the individual to discard events of extremely low probability and to treat events of extremely high probability as if they were certain. Because people are limited in their ability to comprehend and evaluate extreme probabilities, highly unlikely events are either ignored or overweighted, and the difference between high probability and certainty is either neglected or exaggerated. Consequently* $\pi(p)$ *is not well-behaved near the end-points.*"

After the prospects are 'psychologically cleaned' in the *editing phase*, the decision maker chooses the prospect with the highest numerical value assigned by the *value function*.[8] This is the *decision or evaluation phase*. The point transformation of probabilities under PT is unsatisfactory, and undermined the theory. For example, it allows a decision maker to choose stochastically dominated options, even when such dominance is obvious.

Quiggin (1982, 1993), showed that these problems are solved if a cumulative transformation of probabilities is adopted rather than a point transformation. When EU is applied to the transformed probabilities, we get *rank dependent expected utility theory* (RDU).

Example 1 *Quiggin (1982, 1993): Let* $w(p)$ *be some suitably defined probability weighting function such that* $w(0) = 0$, $w(1) = 1$ *and* $w(p) : [0, 1] \to [0, 1]$ *is 1-1 and onto. In terms of the lottery* $(x, p; y, 1 - p)$ *where* $0 < x < y$, *the decision weights are derived as follows.*

[8] The construction of the value function is formally described in the paper, below.

$\pi(p) = w(p + 1 - p) - w(1 - p) = 1 - w(1 - p)$ *and* $\pi(1 - p) = w(1 - p)$*. This is the sense in which cumulative transformations of probability are used in both RDU and CP.*

2.2. Cumulative Prospect Theory (CP)

Tversky and Kahneman (1992) incorporated Quiggen's insight into PT, while retaining the other components of PT such as reference dependence and loss aversion. The result, *cumulative prospect theory* (CP), thus ensured (among other things) that stochastically dominated options would not be chosen. However, a heavy price had to be paid for this desirable feature. *CP dropped the editing phase altogether*, hence, also, giving up the psychological richness of PT.

In contrast to the non-continuous function in Figure 2.1, Tversky and Kahneman (1992) proposed a continuous decision function. This was achieved by postulating a continuous, 1-1 and onto, probability weighting function, $w(p)$ on $[0, 1]$(see Example 1). The decision weights in CP (as in RDU) are cumulative transformations of $w(p)$ as, for instance, in Example 1.[9] One such $w(p)$ function, that is consistent with the evidence on non-extreme probability events, and has axiomatic foundations, is the Prelec (1998) function, plotted in Figure 2.2.[10]

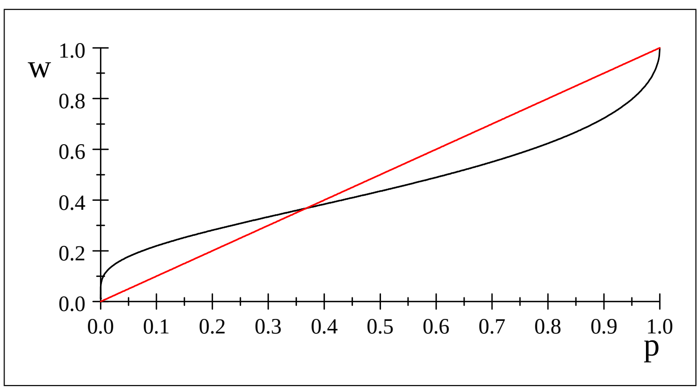


Figure 2.2: A plot of the Prelec (1988) function, $w(p) = e^{-(-\ln p)^{\frac{1}{2}}}$.

Remark 1 *(Infinite overweighting of infinitesimal probabilities): Several other weighting functions have been proposed in the literature and we discuss some below. However, like the Prelec function, they all have the feature that the decision maker (1) infinitely overweights infinitesimal probabilities in the sense that the ratio* $w(p)/p$ *goes to infinity as*

[9]This was the reason that CP was forced to drop the editing phase in PT. The editing phase creates discontinuities in the weighting function that is not admissible under CP.

[10]See Definition 7, below, for the formal definition of a Prelec weighting function. Figure 2.2 is a plot of $w(p) = e^{-(-\ln p)^{\frac{1}{2}}}$ where the probability $p \in [0, 1]$.

p goes to zero. (2) infinitely underweight near-one probabilities in the sense that the ratio $[1 - w(p)]/[1 - p]$ *goes to infinity as* p *goes to* 1.

Remark 2 *Using any of the standard probability weighting functions, CP (and RDU) can explain S1 but not S2.*

2.3. Composite cumulative Prospect theory (CCP)

We make the ambitious proposal of combining the psychological richness of PT with the more satisfactory cumulative transformation of probabilities in CP. In other words, we intend to combine PT and CP into a single theory, that we call *composite cumulative prospect theory, CCP*. If it aids intuition, CCP can be described as combining the editing and decision phases of PT into a single phase, while retaining cumulative transformations of probability, as in CP. CCP accounts for both stylized facts S1 and S2. Like CP, it does not allow for the choice of stochastically dominated options. Furthermore, it can explain everything that RDU and CP can, and more. For that reason, we believe that it is the most satisfactory theory of decision making under risk.

In order to implement CCP, we introduce a modification to the Prelec (1998) weighting function (see Figure 2.2). From remark 2, the Prelec weighting function explains S1 but fails on S2. Our suggestion modifies the end-points of the Prelec's weighting function in a manner that is consistent with the empirical evidence.[11] We call our suggested modification as *composite Prelec weighting function* (CPF). Figure 2.3 sketches the CPF, which can potentially address S1, S2.

In Figure 2.3, decision makers heavily underweight very low probabilities in the range $[0, p_1]$ (compare this to remark 1). Akin to Kahneman and Tversky's (1979) editing phase, decision makers who use the weighting function in Figure 2.3 would typically ignore very low probability events by assigning low subjective weights to them. Hence, in conformity with the evidence (see Section 3) they are unlikely to be dissuaded from low-probability high-punishment crimes, reluctant to buy insurance for very low probability events (unless mandatory), reluctant to wear seat belts (unless mandatory), reluctant to participate in voluntary breast screening programs (unless mandatory) and so on. Similar comments apply to the probability range $[p_3, 1]$ except that events with these probabilities are overweighted as suggested by the evidence; see Kahneman and Tversky's (1979, p.282-83) quote, above. In the middle segment, $p \in [p_1, p_3]$, the probability weighting function in Figure 2.3 is identical to the Prelec function, and so addresses stylized fact S1. Our proposed probability weighting function in Figure 2.3 is *axiomatic*, *parsimonious* and *flexible*, as we shall formally see, below.

[11] Hence, we eliminate the discontinuities at the end-points in Figure 2.1 with empirically-founded as well as axiomatically-founded behavior.

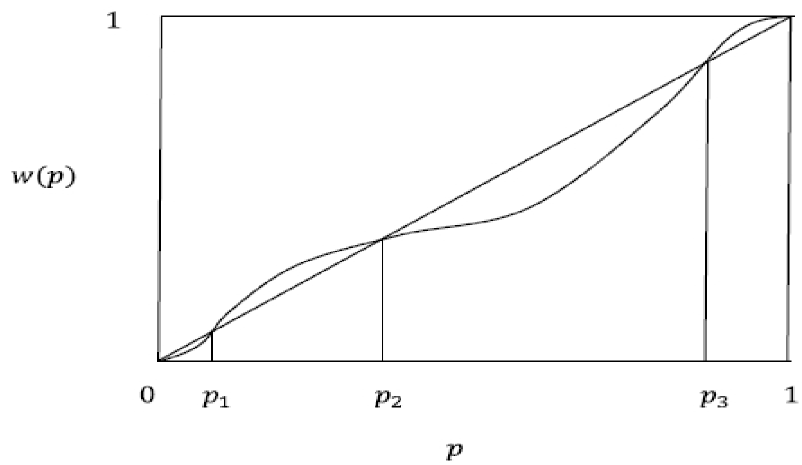


Figure 2.3: The composite Prelec weighting function (CPF).

Remark 3 *: We refer to otherwise standard CP, when combined with a CPF as composite cumulative prospect theory (CCP). Analogously, otherwise standard RDU, when combined with a CPF, will be referred to as composite rank dependent utility (CRDU). CCP and CRDU can explain S1 and S2. Because CCP, in addition, also incorporates reference dependence, loss aversion and richer attitudes towards risk, it can explain everything that CRDU can, but the converse is not true. Furthermore, because CCP incorporates both S1 and S2, while CP only incorporates S1, CCP can explain everything that CP can, but the converse is not true.*

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

In this section, we present the evidence for stylized fact S2 by examining human behavior for low probability events from several economic and non-economic contexts. This is not an exhaustive list of such events but one that should suffice.

3.1. Insurance for low probability events

The insurance industry is of tremendous economic importance. The total global gross insurance premiums for 2008 were 4.27 trillion dollars, which accounted for 6.18% of global GDP.[12] 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.

[12] See Plunkett's Insurance Industry Almanac, 2010, Plunkett Research, Ltd, for the details.

The seminal study of Kunreuther et al. (1978)[13] 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 was a major study, with 135 expert contributors, involving samples of thousands, survey data, econometric analysis and experimental evidence; all three methodologies gave the same conclusion.

Expected utility theory (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)[14] presented subjects with varying potential losses with various probabilities, keeping the expected value of the loss constant. Subjects faced actuarially fair, unfair or subsidized premiums. In each case, they found that there is a point below which the take-up of insurance drops dramatically, as the probability of the loss decreases (and as the magnitude of the loss increases, keeping the expected loss constant).[15]

Remarkably, the lack of interest in buying insurance 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 as contributing to the low take-up of insurance. Furthermore, insurees were aware of the losses (many overestimated them) and moral hazard issues (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 the observations of Kahneman and Tversky (1979) outlined in section 2.1 above.

[13]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."

[14]Chapter 7 was written by Paul Slovic, Baruch Fischoff, Sarah Lichtenstein, Bernard Corrigan and Barbara Combs; see note page 186.

[15]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'.

3.2. Becker (1968) Paradox

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*'. 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 not costly, the deterrence effect of the punishment is maintained. Indeed, under EU, risk-neutrality and infinitely severe punishments,[16] the Becker proposition implies that crime would be deterred completely, however small the probability of detection and conviction. Kolm (1973) memorably phrased this proposition as *it is efficient to hang offenders with probability zero.*

Empirical evidence is not supportive of the Becker proposition. For example, 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, is zero. History does not bear out the Becker proposition either. Since the late middle ages, the severity of punishments has been declining while expenditures on enforcement have been increasing. In their review, 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."

Under RDU and CP, because decision makers heavily overweight the small probability of a punishment (see Remark 1, above), Becker's proposition can be shown to hold with even greater force. But this contradicts the empirical evidence that the Becker proposition does not hold (which is known as the *Becker paradox*). Under CCP, on the other hand, because very small probabilities are heavily underweighted (Figure 2.3), Becker's paradox can be explained. Again, these results hinge on stylized fact S2.[17]

3.2.1. The competing explanations for the Becker paradox

The reader may, rightly, wonder if there are explanations other than simply ignoring low probabilities that might account for Becker's paradox? Dhami and al-Nowaihi (2010h) explore nine other possible explanations of the Becker paradox and show that none of these explanations suffice. It is beyond the scope of this paper to provide the details, so we simply list these potential explanations here for the interested reader. (1) Risk seeking

[16] For instance, ruinous fines, slavery, torture, extraction of body parts (all of which have been historically important), and modern capital punishment.

[17] See Dhami and al-Nowaihi (2010h) for these claims.

behavior. (2) Bankruptcy issues. (3) Differential punishments. (4) Errors in conviction. (5) Rent seeking behavior. (6) Abhorrence of severe punishments. (7) Objectives other than deterrence. (8) Risk aversion. (9) Pathological traits of offenders.

In particular, all of these explanations are contradicted by the evidence from jumping red traffic lights, that we examine in section 3.3 below.

3.3. Evidence from jumping red traffic lights

Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004) provide near decisive evidence that the nine explanations in subsection 3.2.1 cannot explain the Becker paradox. They estimate that there are approximately 260,000 accidents per year in the USA caused by red-light running with implied costs of car repair alone of the order of $520 million per year. It stretches plausibility to assume that these are simply mistakes. In running red lights, there is a *small probability* of an accident, but, the consequences are self inflicted and potentially have infinite costs. Rephrased, running red traffic lights is similar to *hanging oneself with a very small probability*, which is similar to the Becker proposition.

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. However, the corresponding numbers for the additional risks of serious and fatal injuries are 13.9 days and 69.4 days respectively".[18]

Clearly EU combined with risk aversion cannot explain this evidence. Potential explanations 2-8 in section 3.2.1, are not applicable here, because the punishment is self inflicted. One also cannot argue along the lines of explanation 6, in section 3.2.1, 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...". This leaves explanation 1 in section 3.2.1. Unfortunately, the authors do not report the risk-attitudes of offenders. It is clear that offenders do have car-insurance, but it is not reported whether this is mandatory. If it turns out that red-light runners also voluntarily take up insurance of any sort (such as extended warranties, extra life cover etc.), then the explanation based on EU with risk seeking behavior, would not be tenable.

A far more natural explanation, along the lines of our framework, is that stylized fact S2 applies. Thus, red traffic light running is simply caused by some individuals ignoring

[18] To these, should be added the time lost due to police involvement, time and money lost due to auto-repairs, court appearances, fines, increase in car-insurance premiums and the cost and pain of injury and death.

(or seriously underweighting) the very low probability of an accident.[19]

3.4. Driving and talking on car mobile phones

Consider the usage of 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. The Becker proposition applied to this situation would suggest that drivers of vehicles will not use mobile phones while driving or perhaps use hands-free phones, and so, self-insure to avoid the self inflicted punishments. However, the evidence is to the contrary.[20] None of the explanations in section 3.2.1 applies (simply use the arguments in section 3.3). A more natural explanation is the individuals simply ignore or substantially underweight the low probability of an accident as in stylized fact S2.

3.5. Other examples

People were reluctant to use seat belts prior to their mandatory use despite publicly available evidence that seat belts save lives. Prior to 1985, in the US, only 10-20% of motorists wore seat belts voluntarily, hence, denying themselves *self-insurance*; see Williams and Lund (1986). Car accidents may be perceived by individuals as *low probability events*, particularly if they are overconfident of their driving abilities.[21]

Even as evidence accumulated about the dangers of breast cancer (*low probability event*[22]) women took up the offer of breast cancer examination, only sparingly[23]

3.6. 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

[19] A more complete model is needed to address all the relevant stylized facts for this problem but this is beyond the scope of our paper. For instance, some individuals could be so law abiding that they will not run red traffic lights under any circumstances. For others, with a lower level of civic responsibility, and in conjunction with stylized fact S2, ocassionally running red lights could be optimal, which explains the stylized facts.

[20] Various survey evidence indicates that up to 40 percent of individuals drive and talk on mobile phones; see, for example, the Royal Society for the Prevention of Accidents (2005). In surveying the evidence, Pöystia et al. (2004) report that two thirds of Finnish drivers and 85% of American drivers use their phone while driving. Mobile phone usage, while driving, 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 advantage over hand-held units.

[21] 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).

[22] 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.

[23] In the US, this changed after the greatly publicised events of the mastectomies of Betty Ford and Happy Rockefeller; see Kunreuther (1978, p. xiii). See also p. 13-14 in Kunreuther (1978)

models of risk. EU and the associated auxiliary assumptions are unable to explain the stylized facts, however, in the light of Remark 1, RDU and CP make the problem even worse. Second, a natural explanation for these phenomena seems to be that individuals simply ignore or seriously underweight very low probability events, as shown in Figure 2.3.

4. Non-linear transformation of probabilities

The main alternatives to EU, i.e., RDU and CP, introduce non-linear transformation of probability. In this section we introduce the concept of a probability weighting function with particular emphasis on the Prelec function. We also introduce some concepts which are crucial for the rest of the paper.

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

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

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

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

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

Data from experimental and field evidence typically suggest that decision makers exhibit an inverse S-shaped probability weighting over outcomes (stylized fact S1). See Figure 2.2 for an example. Tversky and Kahneman (1992) propose the following probability weighting function, where the lower bound on γ comes from Rieger and Wang (2006).

Definition 6 *: The Tversky and Kahneman probability weighting function is given by*

$$w(p) = \frac{p^{\gamma}}{\left[p^{\gamma} + (1-p)^{\gamma}\right]^{\frac{1}{\gamma}}},\ 0.5 \le \gamma < 1,\ 0 \le p \le 1. \quad (4.1)$$

Proposition 2 *: The Tversky and Kahneman (1992) probability weighting function (4.1) infinitely overweights infinitesimal probabilities and infinitely underweights near-one probabilities, i.e.,* $\lim_{p\to 0}\frac{w(p)}{p}=\infty$ *and* $\lim_{p\to 1}\frac{1-w(p)}{1-p}=\infty$*, respectively.*

Remark 4 *(Standard probability weighting functions): A large number of other probability weighting functions have been proposed, e.g., those by Gonzalez and Wu (1999) and Lattimore, Baker and Witte (1992). Like the Tversky and Kahneman (1992) function, they all infinitely overweight infinitesimal probabilities and infinitely underweight near-one probabilities. We shall call these as the standard probability weighting functions.*

4.1. Prelec's probability weighting function

The Prelec (1998) probability weighting function has the following merits: parsimony, consistency with much of the available empirical evidence (in the sense of stylized fact S1) and an axiomatic foundation.

Definition 7 *(Prelec, 1998): By the Prelec function we mean the probability weighting function* $w(p):[0,1]\to[0,1]$ *given by*

$$w(0)=0,\; w(1)=1, \tag{4.2}$$

$$w(p)=e^{-\beta(-\ln p)^{\alpha}},\; 0<p\leq 1,\; \alpha>0,\; \beta>0. \tag{4.3}$$

Proposition 3 *: The Prelec function (Definition 7) is a probability weighting function in the sense of Definition 1.*

Remark 5 *(Axiomatic foundations): Prelec (1998) gave an axiomatic derivation of (4.2) and (4.3) 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. Note, in particular, that these derivations do not put any restrictions on* α *and* β *other than* $\alpha>0$ *and* $\beta>0$*.*

1. (Role of α) The parameter α controls the convexity/concavity of the Prelec function. If $\alpha<1$, then the Prelec function is strictly concave for low probabilities but strictly convex for high probabilities, i.e., it is *inverse S-shaped*, as in $w(p)=e^{-(-\ln p)^{\frac{1}{2}}}$ ($\alpha=\frac{1}{2}$, $\beta=1$), which is sketched in Figure 2.2, above. The converse holds if $\alpha>1$. The Prelec function is then strictly convex for low probabilities but strictly concave for high probabilities, i.e., it is *S-shaped*. An examples is the curve $w(p)=e^{-(-\ln p)^{2}}$ ($\alpha=2$, $\beta=1$), sketched in Figure 4.1 as the light, lower, curve (the straight line in Figure 4.1 is the 45^{o} line).

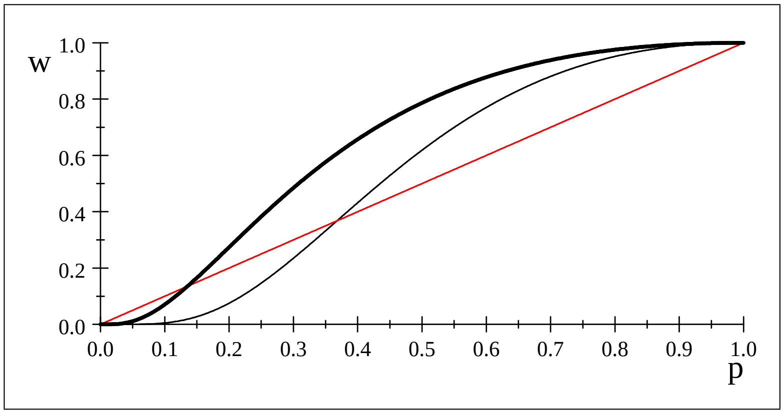


Figure 4.1: A plot of $w(p) = e^{-\frac{1}{2}(-\ln p)^2}$ and $w(p) = e^{-(-\ln p)^2}$.

2. (Role of β) Between the region of strict convexity ($w'' > 0$) and the region of strict concavity ($w'' < 0$), for each of the cases in Figures 2.2 and 4.1, there is a point of inflexion ($w'' = 0$). The parameter β in the Prelec function controls the location of the inflexion point relative to the 45^0 line. Thus, for $\beta = 1$, the point of inflexion is at $p = e^{-1}$ and lies on the 45^0 line, as in Figures 2.2 and 4.1 (light curve), above. However, if $\beta < 1$, then the point of inflexion lies above the 45^0 line, as in $w(p) = e^{-\frac{1}{2}(-\ln p)^2}(\alpha = 2, \beta = \frac{1}{2})$, sketched as the thicker, upper, curve in Figure 4.1. For this example, the fixed point, $w(p^*) = p^*$, is at $p^* \simeq 0.14$ but the point of inflexion, $w''(\widetilde{p}) = 0$, is at $\widetilde{p} \simeq 0.20$.

The full set of possibilities is established by the following two propositions.

Proposition 4 : *For $\alpha = 1$, the Prelec probability weighting function (Definition 7) 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).*

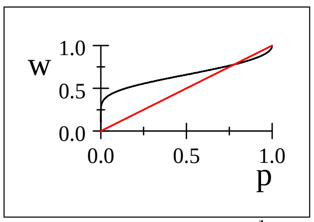
Proposition 5 : *Suppose $\alpha \neq 1$. Then:*
(a) The Prelec function (Definition 7) has exactly three fixed points, at respectively, 0, $p^* = e^{-\left(\frac{1}{\beta}\right)^{\frac{1}{\alpha-1}}}$ and 1.
(b) The Prelec function has a unique inflexion point, $\widetilde{p} \in (0,1)$ at which $w''(\widetilde{p}) = 0$.
(c) If $\alpha < 1$, the Prelec function is strictly concave for $p < \widetilde{p}$ and strictly convex for $p > \widetilde{p}$ (inverse S-shaped).
(d) If $\alpha > 1$, then the converse holds: The Prelec function is strictly convex for $p < \widetilde{p}$ and strictly concave for $p > \widetilde{p}$ (S-shaped).
(e) If $\beta < 1$, then the inflexion point, $\widetilde{p}$, lies above the 45^0 line ($\widetilde{p} < w(\widetilde{p})$).
(f) If $\beta = 1$, then the inflexion point, $\widetilde{p}$, lies on the 45^0 line ($\widetilde{p} = w(\widetilde{p})$).
(g) If $\beta > 1$, then the inflexion point, $\widetilde{p}$, lies below the 45^0 line ($\widetilde{p} > w(\widetilde{p})$).

Table 1, below, exhibits the various cases established by Proposition 5.

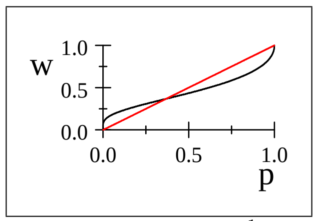
	$\beta < 1$	$\beta = 1$	$\beta > 1$
$\alpha < 1$	inverse S-shape $\widetilde{p} < w(\widetilde{p})$	inverse S-shape $\widetilde{p} = w(\widetilde{p})$	inverse S-shape $\widetilde{p} > w(\widetilde{p})$
$\alpha = 1$	strictly concave $p < w(p)$	$w(p) = p$	strictly convex $p > w(p)$
$\alpha > 1$	S-shape $\widetilde{p} < w(\widetilde{p})$	S-shape $w(\widetilde{p}) = \widetilde{p}$	S-shape $\widetilde{p} > w(\widetilde{p})$

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

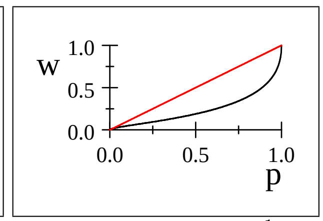
$\beta = \frac{1}{2}$ $\beta = 1$ $\beta = 2$



$w(p) = e^{-\frac{1}{2}(-\ln p)^{\frac{1}{2}}}$

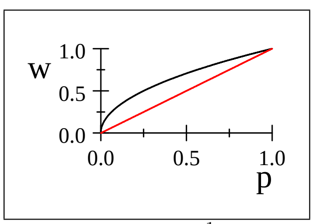


$w(p) = e^{-(-\ln p)^{\frac{1}{2}}}.$

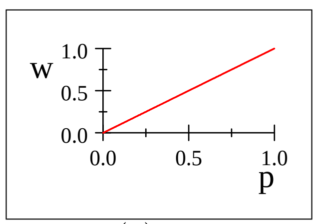


$w(p) = e^{-2(-\ln p)^{\frac{1}{2}}}$

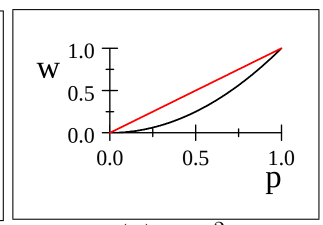
$\alpha = \frac{1}{2}$



$w(p) = p^{\frac{1}{2}}$

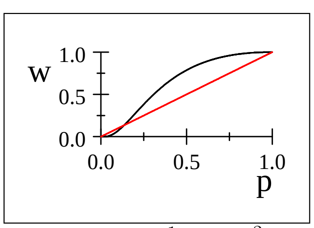


$w(p) = p$

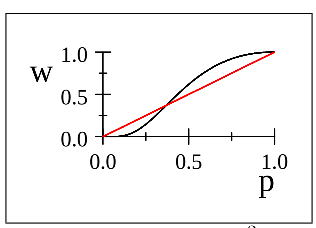


$w(p) = p^2$

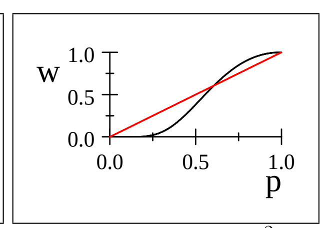
$\alpha = 1$



$w(p) = e^{-\frac{1}{2}(-\ln p)^2}.$



$w(p) = e^{-(-\ln p)^2}.$



$w(p) = e^{-2(-\ln p)^2}$

$\alpha = 2$

Table 2: Representative graphs of $w(p) = e^{-\beta(-\ln p)^{\alpha}}$.

Corollary 1 : *Suppose $\alpha \neq 1$. Then $\widetilde{p} = p^* = e^{-1}$ (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 Figure 2.2).
(b) If $\alpha > 1$, then w is strictly convex for $p < e^{-1}$ and strictly concave for $p > e^{-1}$ (S shape, see Figure 4.1).

In Figure 2.2 (and first row in Table 2), where $\alpha < 1$, note that the slope of $w(p)$ becomes very steep near $p = 0$. By contrast, in figure 4.1 (and last row in Table 2), where $\alpha > 1$, the slope of $w(p)$ becomes very gentle near $p = 0$. This is established by the following proposition, which will be important for us.

Proposition 6 *: (a) For $\alpha < 1$ the Prelec function (Definition 7): (i) infinitely-overweights infinitesimal probabilities, i.e., $\lim\limits_{p\to 0} \frac{w(p)}{p} = \infty$, and (ii) infinitely underweights near-one probabilities, i.e., $\lim\limits_{p\to 1} \frac{1-w(p)}{1-p} = \infty$ (Prelec, 1998, p505); see Definition 2 and Figure 2.2.*
(b) For $\alpha > 1$ the Prelec function: (i) zero-underweights infinitesimal probabilities, i.e., $\lim\limits_{p\to 0} \frac{w(p)}{p} = 0$, and (ii) zero-overweights near-one probabilities, i.e., $\lim\limits_{p\to 1} \frac{1-w(p)}{1-p} = 0$; see Definition 3 and figure 4.1.

According to Prelec (1998, p505), the infinite limits in Proposition 6a capture the qualitative change as we move from certainty to probability and from impossibility to improbability. On the other hand, they contradict stylized fact S2, i.e., the observed behavior that people ignore events of very low probability and treat very high probability events as certain; see, e.g., Kahneman and Tversky (1979). These specific problems are avoided for $\alpha > 1$. However, for $\alpha > 1$, the Prelec function is S-shaped, see Proposition 5(d) and Figure 4.1. This, however is in conflict with stylized fact S1.

5. Composite Prelec Weighting Function

We now make progress towards deriving the *composite Prelec probability weighting function* (CPF) that was motivated in section 2.3; see Figure 2.3. The CPF is able to simultaneously address the two stylized facts S1 and S2 outlined in the introduction. We begin by providing two numerical examples of CPF, motivated by the empirical evidence from Kunreuther (1978). The axiomatic derivation is in section 6, below.

5.1. Two numerical examples of CPF

5.1.1. The urn experiment in Kunreuther (1978)

An obvious solution that addresses S1, S2 is to adopt segments from 3 Prelec functions, as outlined in Figure 2.3, above. Figure 5.1, below gives a numerical example of such a composite Prelec function (CPF).

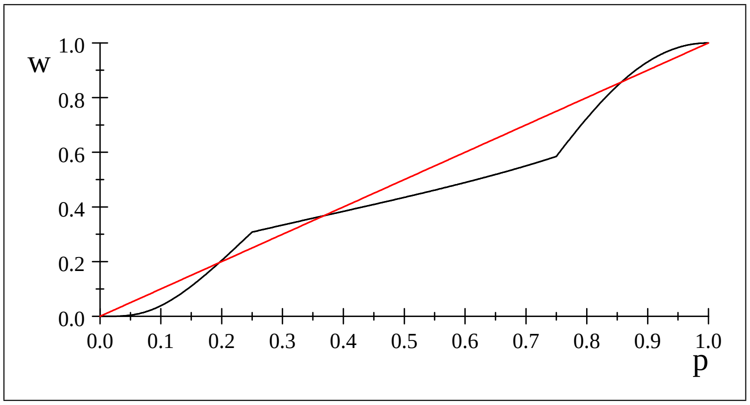


Figure 5.1: The composite Prelec function.

The CPF in Figure 5.1 is composed of segments from three Prelec functions, and is given by

$$w(p)=\begin{cases} e^{-0.61266(-\ln p)^2} & i.e.\ \alpha=2,\beta=0.61266 \quad if \quad 0\leq p<0.25 \\ e^{-(-\ln p)^{\frac{1}{2}}} & i.e.\ \alpha=0.5,\beta=1 \quad if \quad 0.25\leq p\leq 0.75 \\ e^{-6.4808(-\ln p)^2} & i.e.\ \alpha=2,\beta=6.4808 \quad if \quad 0.75<p\leq 1 \end{cases} \tag{5.1}$$

The three segments of the CPF in (5.1) are described as follows.

1. For $0 \leq p < 0.25$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_0(-\ln p)^{\alpha_0}}$, with $\alpha_0 = 2$, $\beta_0 = 0.61266$. β_0 is chosen to make $w(p)$ continuous at $p = 0.25$.

2. For $0.25 \leq p \leq 0.75$, the CPF is identical to the inverse S-shaped Prelec function of Figure 2.2 ($\alpha = 0.5, \beta = 1$).

3. For $0.75 < p \leq 1$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_1(-\ln p)^{\alpha_1}}$, with $\alpha_1 = 2$, $\beta_1 = 6.4808$. β_1 is chosen to make $w(p)$ continuous at $p = 0.75$.

Remark 6 *(Fixed points, concavity, convexity): The CPF in Figure 5.1 has five fixed points*: 0, 0.19549, $e^{-1} = 0.36788$, 0.85701 *and* 1. *It is strictly concave for* $0.25 < p < e^{-1}$ *and strictly convex for* $e^{-1} < p < 0.75$ *(a bit hard to see in Figure 5.1).*[24] *The CPF is strictly convex for* $0 < p < 0.25$ *and strictly concave for* $0.75 < p < 1$.

Remark 7 *(Underweighting and overweighting of probabilities): The CPF in Figure 5.1 overweights 'low' probabilities, in the range* $0.19549 < p < e^{-1}$ *and underweights 'high' probabilities, in the range* $e^{-1} < p < 0.85701$. *These regions, therefore, address stylized*

[24] For $\alpha \in \left(0, \frac{1}{2}\right)$ or $\alpha \in \left(\frac{1}{2}, 1\right)$ the concavity/convexity is even milder than for $\alpha = \frac{1}{2}$, with the slope being less steep for $\alpha \in \left(0, \frac{1}{2}\right)$ but more steep for $\alpha \in \left(\frac{1}{2}, 1\right)$. In fact, $w'(p) \to 0$ as $\alpha \to 0$ and $w'(p) \to 1$ as $\alpha \to 1$.

fact S1. Furthermore, the CPF underweights 'very low' probabilities, in the range $0 < p < 0.19549$ *and overweights 'very high' probabilities, in the range* $0.85701 < p < 1$*. 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). Note that this probability weighting function is also nearly flat near 1. These two segments, i.e.,* $p \in (0, 0.19549) \cup (0.85701, 1)$ *are able to address stylized fact S2.*

The parameters in (5.1) were chosen primarily to clarify the properties in Remarks 6 and 7. The cutoff points 0.25 and 0.75 in (5.1) and Figure 5.1 were motivated by actual evidence. Kunreuther et al. (1978, chapter 7) report that in one set of their experiments (the "urn" experiments) 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). When the probability of the loss reached 0.001, only 20% of the subjects took up insurance (although the premiums were fair). Thus, although Figure 5.1 was chosen primarily for illustrative purposes, its qualitative features do match the evidence reported in Kunreuther et al. (1978).

5.1.2. The farm experiments in Kunreuther (1978)

In a second set of experiments, the "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. This is captured by Figure 5.2.

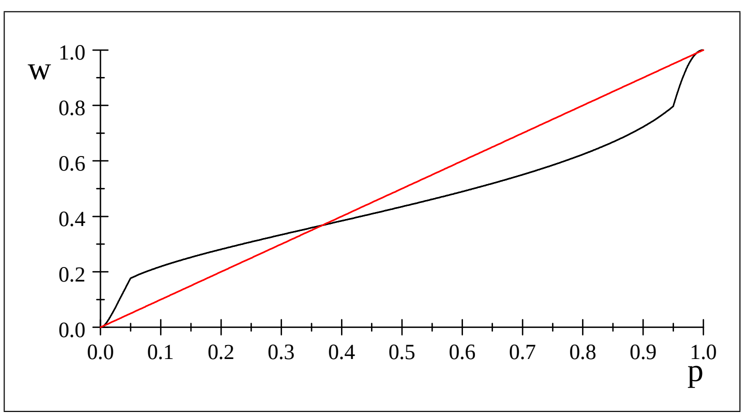


Figure 5.2: The composite Prelec function.

The CPF in Figure 5.2 is composed of segments from three Prelec functions, and is given by

$$w(p) = \begin{cases} e^{-0.19286(-\ln p)^2} & i.e.\ \alpha = 2, \beta = 0.19286 \quad if \quad 0 < p < 0.05 \\ e^{-(-\ln p)^{\frac{1}{2}}} & i.e.\ \alpha = 0.5, \beta = 1 \quad if \quad 0.05 \le p \le 0.95 \\ e^{-86.081(-\ln p)^2} & i.e.\ \alpha = 2, \beta = 86.081 \quad if \quad 0.95 < p \le 1 \end{cases} \tag{5.2}$$

The three segments of the CPF in (5.2) are described as follows.

1. For $0 \le p < 0.05$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_0(-\ln p)^{\alpha_0}}$, with $\alpha_0 = 2$, $\beta_0 = 0.19286$. β_0 is chosen to make $w(p)$ continuous at $p = 0.05$.

2. For $0.05 \le p \le 0.95$, the CPF is identical to the inverse S-shaped Prelec function of Figure 2.2 ($\alpha = 0.5$, $\beta = 1$).

3. For $0.95 < p \le 1$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_1(-\ln p)^{\alpha_1}}$, with $\alpha_1 = 2$, $\beta_1 = 86.081$. β_1 is chosen to make $w(p)$ continuous at $p = 0.95$.

Remark 8 *(Fixed points): This CPF 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 9 *(Underweighting and overweighting of probabilities): The CPF overweights 'low' probabilities, in the range* $0.0055993 < p < e^{-1}$ *and underweights 'high' probabilities, in the range* $e^{-1} = 0.36788 < p < 0.98845$. *This accounts for stylized fact S1. Behavior near* 0, *and near* 1, *is not obvious from Figure 5.2. So, Figures 5.3 and 5.4, below, respectively, magnify the regions near* 0 *and near* 1.

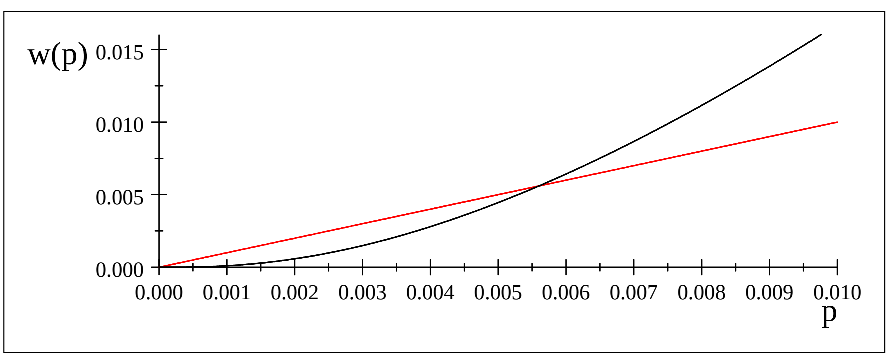


Figure 5.3: Behaviour of Figure 5.2 near 0.

From Figure 5.3, we see that (5.2) underweights very low probabilities, in the range $0 < p < 0.0055993$. For p close to zero, we see that this probability weighting function is nearly flat, thus, again capturing Arrow's observation "...it does appear from the data that the sensitivity goes down too rapidly as the probability decreases." From Figure 5.4, we see that (5.2) overweights very high probabilities, in the range $0.98845 < p < 1$. For p close to one, we see that this probability weighting function is nearly flat.

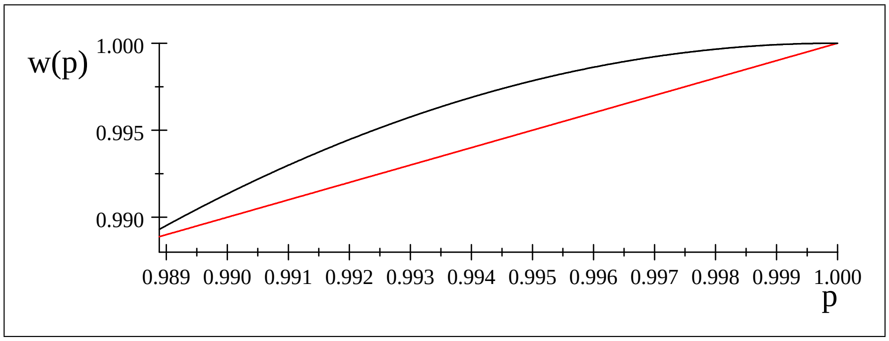


Figure 5.4: Behaviour of Figure 5.2 near 1.

5.2. A more formal treatment of the CPF

Notice that the upper cutoff points for the *first segment* of the CPF's in Figures 5.1 and 5.2 are respectively, at probabilities 0.25 and 0.05. Denote this cutoff point as $\underline{p}$. Similarly, the upper cutoff point for the *second segment* of the CPF in Figures 2.3 and 5.2 are respectively, at probabilities 0.75 and 0.95. Denote this cutoff point as $\overline{p}$. Now define,

$$\underline{p} = e^{-\left(\frac{\beta}{\beta_0}\right)^{\frac{1}{\alpha_0 - \alpha}}};\ \overline{p} = e^{-\left(\frac{\beta}{\beta_1}\right)^{\frac{1}{\alpha_1 - \alpha}}} \tag{5.3}$$

The CPF's in (5.1), (5.2) and their graphs, Figures 5.1, 5.4, suggest the following definition.

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

$$w(p) = \begin{cases} 0 & if \quad p = 0 \\ e^{-\beta_0(-\ln p)^{\alpha_0}} & if \quad 0 < p \leq \underline{p} \\ e^{-\beta(-\ln p)^{\alpha}} & if \quad \underline{p} < p \leq \overline{p} \\ e^{-\beta_1(-\ln p)^{\alpha_1}} & if \quad \overline{p} < p \leq 1 \end{cases} \tag{5.4}$$

where $\underline{p}$ *and* $\overline{p}$ *are given by (5.3) and*

$$0 < \alpha < 1,\ \beta > 0;\ \alpha_0 > 1,\ \beta_0 > 0;\ \alpha_1 > 1,\ \beta_1 > 0, \beta_0 < 1/\beta^{\frac{\alpha_0 - 1}{1-\alpha}},\ \beta_1 > 1/\beta^{\frac{\alpha_1 - 1}{1-\alpha}}. \tag{5.5}$$

Proposition 7 *: The composite Prelec function (Definition 8) 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 (5.5), are required by the axiomatic derivations of the Prelec function (see Prelec (1998), Luce (2001) and al-Nowaihi and Dhami (2006)). The restriction $\beta_0 < 1/\beta^{\frac{\alpha_0-1}{1-\alpha}}$ guarantees that the first segment of the CPF, $e^{-\beta_0(-\ln p)^{\alpha_0}}$, crosses the 45^0 to the left of $\underline{p}$ and the restriction $\beta_1 > 1/\beta^{\frac{\alpha_1-1}{1-\alpha}}$ guarantees that the third segment of the CPF, $e^{-\beta_1(-\ln p)^{\alpha_1}}$, crosses the 45^0 degree line to the right of $\overline{p}$. Together, they imply that the second segment of CPF, $e^{-\beta(-\ln p)^{\alpha}}$, crosses the 45^0 between these two limits. It follows that the interval $\left(\underline{p}, \overline{p}\right)$ is not empty. These interval limits are chosen so that the CPF in (5.4) is continuous across them. These observations lead to the following proposition. First, define p_1, p_2, p_3 that correspond to the notation used in our proposal for a CPF in Figure 2.3 (see section 2.3).

$$p_1 = e^{-\left(\frac{1}{\beta_0}\right)^{\frac{1}{\alpha_0-1}}},\ p_2 = e^{-\left(\frac{1}{\beta}\right)^{\frac{1}{\alpha-1}}},\ p_3 = e^{-\left(\frac{1}{\beta_1}\right)^{\frac{1}{\alpha_1-1}}} \tag{5.6}$$

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

By Proposition 7, the CPF in (5.4), (5.5) is a probability weighting function in the sense of Definition 1. By Proposition 8, a CPF overweights low probabilities, i.e., those in the range (p_1, p_2), and underweights high probabilities, i.e., those in the range (p_2, p_3). Thus it accounts for stylized fact S1. But, in addition, and unlike all the standard probability weighting functions, it underweights probabilities near zero, i.e., those in the range $(0, p_1)$, and overweights probabilities close to one, i.e., those in the range $(p_3, 1)$ as required in the narrative of Kahneman and Tversky (1979, p. 282-83). Hence, a CPF also accounts for the second stylized fact, S2.

The restrictions $\alpha_0 > 1$ and $\alpha_1 > 1$ in (5.5) ensure that a CPF has the following properties, listed below as Proposition 9, that will help explain human behavior for extremely low probability events; see below.

Proposition 9 : *The CPF (5.4):*
(a) zero-underweights infinitesimal probabilities, i.e., $\lim_{p\to 0} \frac{w(p)}{p} = 0$ *(Definition 3a),*
(b) zero-overweights near-one probabilities, i.e., $\lim_{p\to 1} \frac{1-w(p)}{1-p} = 0$ *(Definition 3b).*

6. Axiomatic derivation of the composite Prelec probability weighting function (CPF).

Recall from remark 5 that al-Nowaihi and Dhami (2006) gave an axiomatic derivation of the Prelec probability weighting function, based on *power invariance*. 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. First, we give a general definition of a CPF.

Definition 9 *(Composite Prelec function, CPF): By the composite Prelec function we mean the function* $w : [0,1] \rightarrow [0,1]$ *given by*

$$w(p) = \begin{cases} 0 & if \quad p = 0 \\ e^{-\beta_i(-\ln p)^{\alpha_i}} & if \quad p_{i-1} < p \leq p_i,\ i = 1, 2, ...n, \end{cases} \tag{6.1}$$

where $\alpha_i > 0$, $\beta_i > 0$, $p_0 = 0$, $p_n = 1$ *and*

$$e^{-\beta_i(-\ln p_i)^{\alpha_i}} = e^{-\beta_{i+1}(-\ln p_i)^{\alpha_{i+1}}},\ i = 1, 2, ...n-1. \tag{6.2}$$

The restriction (6.2) is needed to insure that w is continuous.

Proposition 10 *: The composite Prelec functions (Definition 9) are probability weighting functions in the sense of Definition 1.*

Definition 10 *(Power invariance, al-Nowaihi and Dhami, 2006): A probability weighting function,* w, *satisfies power invariance if:* $\forall p, q \in (0,1)$, $(w(p))^{\mu} = w(q) \Rightarrow \left(w\left(p^{\lambda}\right)\right)^{\mu} = w\left(q^{\lambda}\right)$, $\lambda, \mu \in \{2, 3\}$.

Definition 11 *(Local power invariance): Let* $0 = p_0 < p_1 < ... < p_n = 1$. *A probability weighting function,* w, *satisfies local power invariance if, for* $i = 1, 2, ..., n$, w *is* C^1 *on* (p_{i-1}, p_i) *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* $\left(w\left(p^{\lambda}\right)\right)^{\mu} = w\left(q^{\lambda}\right)$.

Proposition 11 *(al-Nowaihi and Dhami, 2006) The following are equivalent.*
1. The probability weighting function, w, *satisfies power invariance.*
2. $\forall p, q \in (0,1)$, $\forall \lambda, \mu \in (0, \infty)$, $(w(p))^{\mu} = w(q) \Rightarrow \left(w\left(p^{\lambda}\right)\right)^{\mu} = w\left(q^{\lambda}\right)$.
3. There is a function, $\varphi : \mathbb{R}_{++} \rightarrow \mathbb{R}$, *such that,* $\forall p \in (0,1)$, $\forall \lambda \in (0, \infty)$, $w\left(p^{\lambda}\right) = (w(p))^{\varphi(\lambda)}$. *Moreover, for some* $\alpha \in (0, \infty)$, $\varphi(\lambda) = \lambda^{\alpha}$.
4. w *is the Prelec function (Definition 7).*

The interested reader can consult al-Nowaihi and Dhami (2006) for the proof of Proposition 11.

Definition 12 *(Useful notation): 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* Λ_i *as follows.* $\Lambda_1 = [\frac{\ln p_1}{\ln p}, \infty)$, $\Lambda_n = (0, \frac{\ln p_{n-1}}{\ln p}]$, $\Lambda_i = \left[\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right]$, $i = 2, 3, ..., n-1$.

Remark 10 *: The problem with the notation* $\left[\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right]$ *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 12 to avoid this, by excluding the points* 0 *and* 1.

Lemma 1 *: Let p_i, P_i and Λ_i be as in Definition 12. Then,*

$$\text{Let } p \in (p_{i-1}, p_i)\text{. Then } p^{\lambda} \in (p_{i-1}, p_i) \Leftrightarrow \lambda \in \left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right), \tag{6.3}$$

$$\text{furthermore, } 0 < \frac{\ln p_i}{\ln p} < 1 < \frac{\ln p_{i-1}}{\ln p}. \tag{6.4}$$

$$\text{Let } p \in P_i\text{. Then } p^{\lambda} \in P_i \Leftrightarrow \lambda \in \Lambda_i. \tag{6.5}$$

Proposition 12 *(CPF representation): The following are equivalent.*
(a) The probability weighting function, w, satisfies local power invariance.
(b) There are functions, $\varphi_i : \Lambda_i \rightarrow \mathbb{R}_{++}$, such that φ_i is C^1 on $\left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right)$, $i = 1, 2, ..., n$, where $0 = p_0 < p_1 < ... < p_n = 1$, and, $\forall p \in P_i$, $\forall \lambda \in \Lambda_i$, $w\left(p^{\lambda}\right) = (w(p))^{\varphi_i(\lambda)}$. Moreover, for each $i = 1, 2, ..., n$, $\exists\ \alpha_i \in (0, \infty)$, $\varphi_i(\lambda) = \lambda^{\alpha_i}$.
(c) w is a composite Prelec function (Definition 9).

7. Composite Cumulative Prospect Theory (CCP)

Composite prospect theory (CCP) requires standard cumulative prospect theory (CP) to use the axiomatically founded cumulative Prelec probability weighting function CPF. As noted in section 5, the CPF is consistent with S1 and S2 in contrast to the standard probability weighting functions in CP which are inconsistent with S2. Since the other components of CP and CCP are identical, and CP is well known, we restrict ourselves to a brief formal statement of CCP. Consider a lottery in the form

$$L = (y_{-m}, p_{-m}; y_{-m+1}, p_{-m+1}; ...; y_{-1}, p_{-1}; y_0, p_0; y_1, p_1; y_2, p_2; ...; y_n, p_n),$$

where $y_{-m} < ... < y_{-1} < y_0 < y_1 < ... < y_n$ are the outcomes or the *final* positions of aggregate wealth and $p_{-m}, ..., p_{-1}, p_0, p_1, ...p_n$ are the corresponding objective 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 for wealth, y_0.[25]

Definition 13 *(Lotteries in incremental form) Let $x_i = y_i - y_0, i = -m, -m+1, ..., n$ be the increment in wealth relative to y_0 and $x_{-m} < ... < x_0 = 0 < ... < x_n$. Let the restriction on probabilities be $\Sigma_{i=-m}^{n} p_i = 1$, $p_i \geq 0$, $i = -m, -m+1, ..., n$. Then, a lottery is presented in incremental form if it is represented as:*

$$L = (x_{-m}, p_{-m}; ...; x_{-1}, p_{-1}; x_0, p_0; x_1, p_1; ...; x_n, p_n), \tag{7.1}$$

[25] y_0 could be initial wealth, status-quo wealth, average wealth, desired wealth, rational expectations of future wealth etc. depending on the context.

Definition 14 *(Set of Lotteries): Denote by $\mathcal{L}_P$ the set of all lotteries of the form given in (7.1) subject to the restrictions in definition 13.*

Definition 15 *(Domains of losses and gains): The decision maker is said to be in the domain of gains if $x_i > 0$ and in the domain of losses if $x_i < 0$. The reference point x_0 lies neither in the domain of gains nor in the domain of losses.*

7.1. The value function under CCP

The utility function under CCP is defined over the set $\mathcal{L}_P$.

Definition 16 *(Tversky and Kahneman, 1979). A utility function, $v(x)$, is a continuous, strictly increasing, mapping $v : \mathbb{R} \to \mathbb{R}$ 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:

$$v(x) = \begin{cases} x^{\gamma} & \text{if } x \geq 0 \\ -\lambda(-x)^{\theta} & \text{if } x < 0 \end{cases} \tag{7.2}$$

where γ, θ, λ 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*. Tversky and Kahneman (1992) assert (but do not prove) that the axiom of *preference homogeneity* ($(x, p) \sim y \Rightarrow (kx, p) \sim ky$) generates this value function. al-Nowaihi et al. (2008) give a formal proof, as well as some other results (e.g. hat γ is necessarily identical to θ). Tversky and Kahneman (1992) estimated that $\gamma \simeq \theta \simeq 0.88$ and $\lambda \simeq 2.25$. The reader can check the properties listed in definition 16 for the utility function, (7.2), drawn in figure 7.1 for the case:

$$\begin{cases} v(x) = \begin{matrix} \sqrt{x} & \text{if } x \geq 0 \\ -2.5\sqrt{-x} & \text{if } x < 0 \end{matrix} \end{cases} \tag{7.3}$$

7.2. Construction of decision weights under CCP

Let $w(p)$ be the CPF in Definition 8. We could have different weighting functions for the domain of gains and losses, respectively, $w^{+}(p)$ and $w^{-}(p)$. However, we make the empirically founded assumption that $w^{+}(p) = w^{-}(p)$; see Prelec (1998).

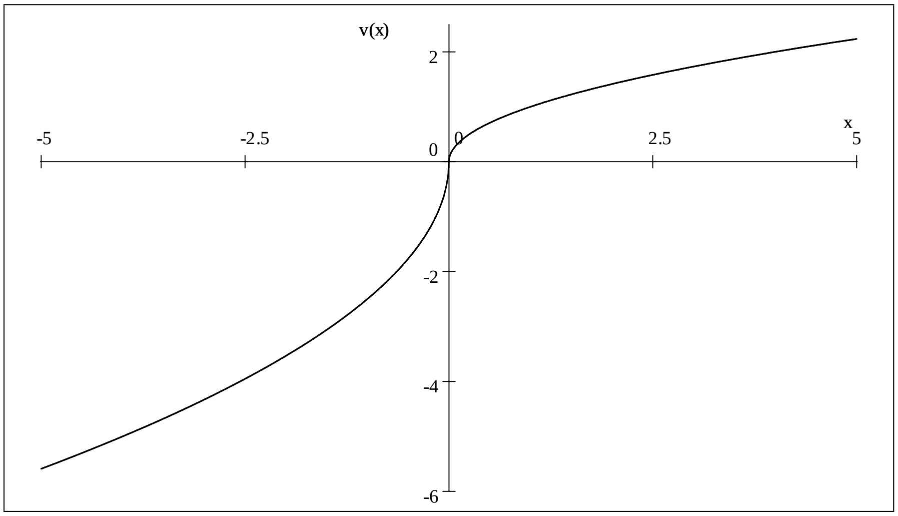


Figure 7.1: The utility function under CCP for the case in (7.3)

Definition 17 *(Tversky and Kahneman, 1992). For CCP, the decision weights, π_i, are defined as follows:*

Domain of Gains	*Domain of Losses*
$\pi_n = w(p_n)$	$\pi_{-m} = w(p_{-m})$
$\pi_{n-1} = w(p_{n-1} + p_n) - w(p_n) \ldots$	$\pi_{-m+1} = w(p_{-m} + p_{-m+1}) - w(p_{-m}) \ldots$
$\pi_i = w\left(\Sigma_{j=i}^{n} p_j\right) - w\left(\Sigma_{j=i+1}^{n} p_j\right) \ldots$	$\pi_j = w\left(\Sigma_{i=-m}^{j} p_i\right) - w\left(\Sigma_{i=-m}^{j-1} p_i\right) \ldots$
$\pi_1 = w\left(\Sigma_{j=1}^{n} p_j\right) - w\left(\Sigma_{j=2}^{n} p_j\right)$	$\pi_{-1} = w\left(\Sigma_{i=-m}^{-1} p_i\right) - w\left(\Sigma_{i=-m}^{-2} p_i\right)$

7.3. The objective function under prospect theory

As in EU, a decision maker using CCP maximizes a well defined objective function, called the value function, which we now define.

Definition 18 *(The value function under CCP) The value of the lottery, $\mathcal{L}_P$, to the decision maker is given by*

$$V(\mathcal{L}_P) = \Sigma_{i=-m}^{n} \pi_i v(x_i) \tag{7.4}$$

Note that the decision weights across the domain of gains and losses *do not* necessarily add up to 1.[26] To see this, from definition 17, we get that

$$\sum_{j=-m}^{n} \pi_j = w\left(\Sigma_{j=1}^{n} p_j\right) + w\left(\Sigma_{i=-m}^{-1} p_i\right) \neq 1 \tag{7.5}$$

[26] This contrasts with the case of RDU, in which there is no conception of different domains of gains and losses and the decision weights add up to one.

If all outcomes were in the domain of gains then we get $\sum_{j=1}^{n}\pi_j = w\left(\Sigma_{j=1}^{n}\, p_j\right) = 1$ because $\Sigma_{j=1}^{n}\, p_j = 1$ and $w(1) = 1$ (as in RDU). If all outcomes were in the domain of losses then similarly $\sum_{j=-m}^{-1}\pi_j = w\left(\Sigma_{j=-m}^{-1}\, p_j\right) = 1$ because $\Sigma_{j=-m}^{-1}\, p_j = 1$ and $w(1) = 1$ (as in RDU). For the general case when there are some outcomes in the domain of gains and others in loss then, since $v(0) = 0$, the decision weight on the reference outcome, π_0, can be chosen arbitrarily. We have found it technically convenient to define $\pi_0 = 1 - \Sigma_{i=-m}^{-1}\pi_i - \Sigma_{i=1}^{n}\pi_i$, so that $\Sigma_{i=-m}^{n}\pi_i = 1$.

8. Some Applications of CCP

We give three applications of CCP in this section. The first application is the take-up of insurance for low probability events (see section 3.1). Using one of Kunreuther's (1978) experiments, it shows, numerically, that the data is consistent with CCP but not CP (or RDU). The second application shows how CCP can address the Becker paradox (see section 3.2) but CP (and RDU) cannot. The third application shows how the St. Petersburg paradox that re-emerges under CP can be solved under CCP.

8.1. 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, so $r = p$. 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. $y_0 = W$.

Expressed in incremental form, the lottery faced by the decision maker under CCP is $(-L + C - rC, p; -rC, 1-p)$. Thus, in both states of the world the outcomes are in the domain of loss. In terms of our exposition of CCP in section 7, we have a lottery of the form: $(x_1, p; x_2, 1-p)$, i.e., loose x_1 with probability p or lose x_2 with probability $1-p$, $x_1 < x_2$, $0 \leq p \leq 1$. In this case, using the construction of decision weights in definition 17, and the utility function $v(x)$ in (7.2), the value function in (7.4) implies

$$V(x_1, p; x_2, 1-p) = w(p)\, v(x_1) + [1 - w(p)]\, v(x_2), \tag{8.1}$$

Now consider a decision maker whose behavior is described by CCP and who faces the insurance problem, given above. For such a decision maker, the value function from the act of purchasing some level of coverage $C \in [0, L]$ is given by

$$V_I(C) = w(p)\, v(-L + C - rC) + [1 - w(p)]\, v(-rC). \tag{8.2}$$

Since $V_I(C)$ is a continuous function on the non-empty compact interval $[0, L]$, an optimal level of coverage, C^*, exists. For full insurance, $C = L$, (8.2) gives:

$$V_I(L) = v(-rL). \tag{8.3}$$

On the other hand, if the decision maker does not buy any insurance coverage (i.e. $C = 0$), the value function is (recall that $v(0) = 0$):

$$V_{NI} = w(p)\, v(-L). \tag{8.4}$$

For the decision maker to buy full coverage, under the actuarially-fair condition, the following participation constraint must be satisfied:

$$V_{NI} \leq V_I(C^*). \tag{8.5}$$

We now provide an example that is motivated by Kunreuther's (1978) empirical results.

Example 2 : *Suppose that a decision maker faces a loss, L, of \$200,000 with probability $p = 0.001$. Under CP, the decision maker uses a standard probability weighting function, say, the Prelec function, $w(p) = e^{-\beta(-\ln p)^{\alpha}}$ with $\beta = 1$ and $\alpha = 0.50$. Using experimental values suggested by Kahneman and Tversky (1979), the utility function (7.2) is*

$$v(x) = \begin{cases} x^{0.88} & \text{if} \quad x \geq 0 \\ -2.25(-x)^{0.88} & \text{if} \quad x < 0 \end{cases}.$$

For CCP, we take the CPF in (5.1) from Kunreuther's urn experiments. For $p = 0.001$, for CCP, (5.1) gives $w(0.001) = e^{-0.612\,66(-\ln 0.001)^2}$. We now see if it is optimal for a decision maker under, respectively, CP and CCP, to fully insure, i.e., $C = L$. Using (8.5), we need to check, in each case, the following condition that ensures full insurance (we have used the actuarially fair condition $r = p$):

$$w(p)\, v(-L) \leq v(-pL) \tag{8.6}$$

(a) Decision maker uses CP: In this case, (8.6) requires that

$$e^{-(-\ln 0.001)^{0.50}} \left(-2.25\left(2 \times 10^5\right)^{0.88}\right) \leq \left(-2.25\left(0.001 \times 2 \times 10^5\right)^{0.88}\right)$$
$$\Leftrightarrow -7510.2 \leq -238.28,$$

which is true. Hence, a decision maker who uses CP will fully insure. However, as noted in section 5.1.1, Kunreuther's (1978) data shows that only 20% of the decision makers insure.
(b) Decision maker uses CCP: In this case, (8.6) requires that

$$e^{-0.612\,66(-\ln 0.001)^2} \left(-2.25\left(2 \times 10^5\right)^{0.88}\right) \leq \left(-2.25\left(0.001 \times 2 \times 10^5\right)^{0.88}\right)$$
$$\Leftrightarrow -2.093 \times 10^{-8} \leq -238.28,$$

which is not true. Hence, a decision maker using CCP will not insure, which is in conformity with Kunreuther's (1978) data.

Example 3 *: Now continue to use the set-up of Example 2. However, let the probability of a loss be* $p = 0.25$ *(instead of* $p = 0.001$*). Kunreuther's (1978) data shows that 80% of the experimental subjects took up insurance in this case. For CCP, as in Example 2, we take the CPF in (5.1) from Kunreuther's urn experiments. For* $p = 0.25$ *the Prelec and CCP functions coincide and* $w(p) = e^{-\beta(-\ln p)^{\alpha}}$*. Thus, in each case the full insurance condition* $w(p)\,v(-L) \leq v(-pL)$ *in (8.6) is given by*

$$e^{-(-\ln 0.25)^{0.50}} \left(-2.25\left(2 \times 10^5\right)^{0.88}\right) \leq \left(-2.25\left(0.25 \times 2 \times 10^5\right)^{0.88}\right)$$

$$\Leftrightarrow -32044 \leq -30710,$$

which is true. Hence, for losses whose probability is bounded well away from the end-points, the predictions of, both, CP and CCP are in conformity with the evidence.

In conjunction, Examples 2 and 3 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 reality, of course, the insurance problem is much more complicated.[27] In particular, there could be considerations of fixed costs of insurance, and insurance premiums might not be actuarially fair. A more satisfactory treatment of the insurance problem is beyond the scope of this paper. However, al-Nowaihi and Dhami (2010i) give a full treatment of the insurance problem that includes fixed costs and actuarially unfair premiums. It turns out that the central insight of Examples 2 and 3 survives in these more complicated settings. Decision makers who use CCP exhibit much more realistic insurance behavior as compared to decision makers who use CP or RDU.

We note here the intuition, more generally, for why CP fails to address the insurance problem but CCP is able to address it. For $\alpha < 1$, the Prelec function infinitely-overweights infinitesimal probabilities (Definitions 2a and 7 and Proposition 6(ai)). In this case, and as illustrated in Figure 2.2, the probability weighting function is very steep near 0 (becoming infinitely steep in the limit as $p \to 0$) and considerably overweights probabilities close to 0. Now return to (8.1). Recall that $x_1 < x_2 < 0$ in this case. Hence, x_1 is the difference between wealth if the loss occurs (with probability, p) and reference wealth. On the other hand, x_2 is the difference between wealth if the loss does not occur (with probability $1-p$) and reference wealth. In the latter case, the only costs incurred are the minor ones of the premium payment. As $p \to 0$, $w(p)$ increasingly overweights p. This increases the relative salience of the term, $w(p)\,v(x_1)$ but reduces the relative salience of the term,

[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).

$[1 - w(p)]\, u(x_1)$. This makes insurance even more attractive under CP than under EU (which was already too high, given the evidence).

The converse occurs under CCP. Here, the probability weighting function zero underweights infinitesimal probabilities (Definition 3a). This is the case for the Prelec function with $\alpha > 1$ (Definition 7 and Proposition 6(bii)). In this case, and as illustrated in Figure 4.1, and section 5, the probability weighting function is very shallow near 0 (and the slope, in fact, approaches zero) and considerably underweights probabilities close to 0. Now return to (8.1). As $p \to 0$, $w(p)$ underweights p. This reduces the relative salience of the term, $w(p)\, v(x_1)$ but increases the relative salience of the term, $[1 - w]\, u(x_2)$. This makes insurance against very low probability events unattractive under CCP, in conformity with the evidence.

8.2. Is it optimal to hang offenders with probability zero?

We now examine the Becker proposition that we outlined in section 3.2. 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,

$$b \leq F \leq F_{\max} \leq \infty. \tag{8.7}$$

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, $F_{\max}$. Given the enforcement parameters (p, F) the individual makes only one choice: to commit the crime or not.

We consider a *hyperbolic punishment function* which encapsulates in a simple manner, the idea that p, F are substitutes in deterrence,[28]

$$F = \varphi(p) = b/p. \tag{8.8}$$

Notice that in (8.8), fines vary continuously with p.

8.2.1. Illustration of the Becker proposition under EU

We first show that Becker's proposition holds under EU. Consider an individual with continuously differentiable and strictly increasing utility of income, u. So, from the activity

[28] Dhami and al-Nowaihi (2010h) show that the hyperbolic punishment function is optimal for a wide class of *cost of deterrance* and *damage from crime* functions. Furthermore, they show that it provides an upper bound on punishments for a large and sensible class of cost and damage functions. Thus, if the hyperbolic punishment function is not able to support the Becker proposition, then the optimal punishment functions with a weaker level of punishment cannot support the proposition either.

‘no-crime’, his payoff, U_{NC}, is $U_{NC} = u(y_0)$. His payoff from the activity ‘crime’, EU_C, is given by

$$EU_C = pu(y_1 - F) + (1-p)\,u(y_1)\,. \tag{8.9}$$

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

$$pu(y_1 - F) + (1-p)\,u(y_1) \leq u(y_0)\,. \tag{8.10}$$

Since $y_1 - b = y_0$, (8.10) is clearly satisfied for $p = 1$ and $F = b$. The NCC (8.10) continues to hold, as p is reduced from 1, if, and only if, the following is the case,

$$\frac{d}{dp}\left[pu(y_1 - \varphi(p)) + (1-p)\,u(y_1)\right] \geq 0, \tag{8.11}$$

Since $F = \varphi(p) = b/p$, (8.11) can be written as:

$$u'(y_1 - \varphi(p)) \geq \frac{u(y_1) - u(y_1 - \varphi(p))}{\varphi(p)}. \tag{8.12}$$

If the decision maker is risk averse or risk neutral, so that u is concave, then the NCC (8.12) *will* hold for all $p \in (0,1]$.

Proposition 13 *(Becker Proposition): Under EU if the individual is risk neutral or risk averse, so that u is concave, 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.*

8.2.2. The Becker paradox under CP and CCP

Recall from section 3.2 that the empirical evidence is not supportive of the Becker proposition (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 the payoff from not committing crime is

$$V_{NC} = v(y_0 - y_r)\,. \tag{8.13}$$

Assume that if an individual commits a crime which gives income y_1, and is caught with probability p, then 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,[29] with respective decision weights, $w^{-}(p)$ and $w^{+}(1-p)$.[30] Then, under CP, the individual's payoff from committing a crime is given by the value function

$$V_C = w^{-}(p)\, v\,(y_1 - F - y_R) + w^{+}(1-p)\, v(y_1 - y_R). \tag{8.14}$$

Definition 19 *(Elation): We shall refer to $v(y_1 - y_R)$ as the elation from committing a crime and getting away with it.*

Substituting (8.13), (8.14) into the 'no crime condition' (*NCC*), $V_C \leq V_{NC}$,

$$w\,(p)\, v\,(y_1 - F - y_R) + w\,(1-p)\, v(y_1 - y_R) \leq v\,(y_0 - y_r)\,. \tag{8.15}$$

The *NCC* in (8.15) 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.[31] Our model is consistent with a perfect foresight rational expectations path. Hence, the rational expectation of income from an activity is the expected income from that activity. Thus,

$$y_r = y_0;\ y_R = y_1 - p\varphi\,(p)\,. \tag{8.16}$$

Since the carrier of utility under CP and CCP is income relative to the reference point, we get $y_1 - y_R = p\varphi\,(p)$, $y_1 - \varphi\,(p) - y_R = -\,(1-p)\,\varphi\,(p)$, $y_0 - y_r = 0$. Hence, and recalling that $v(0) = 0$, the *NCC* (8.15) becomes

$$w^{-}(p)\, v\,(-\,(1-p)\,\varphi\,(p)) + w^{+}(1-p)\, v(p\varphi\,(p)) \leq 0. \tag{8.17}$$

For the power function form of utility (7.2) with $\theta = \gamma$ (which is consistent with the empirical evidence), the *NCC* (8.17) becomes:

$$-\lambda\,(1-p)^{\gamma}\,\varphi\,(p)^{\gamma}\, w^{-}(p) + p^{\gamma}\varphi\,(p)^{\gamma}\, w^{+}(1-p) \leq 0. \tag{8.18}$$

For $\varphi\,(p) > 0$, this simplifies to

$$\frac{w^{-}(p)}{p^{\gamma}} \geq \frac{w^{+}(1-p)}{\lambda\,(1-p)^{\gamma}}. \tag{8.19}$$

[29] This accords with basic intuition: If a criminal gets away with the crime, he/she is in the domain of gains, otherwise, if caught, the criminal is in the domain of losses.

[30] In this case, CP reduces to PT. We have superscripted the decision weights to indicate if the outcome is in the domain of gains or losses. For the Prelec weighting function, however, the weighting function is identical in the domain of gains and losses, so $w^{-}(p) = w^{+}(p) = w(p)$.

[31] See, for instance, Koszegi and Rabin (2006).

Proposition 14 *: As the probability of detection approaches zero, a decision maker using CP who (i) faces a strictly positive punishment, i.e.,* $\varphi(p) > 0$*, (ii) satisfies preference homogeneity*[32] *and (iii) whose reference points are given by (8.16),*
(a) does not engage in the criminal activity if the probability weighting function satisfies the condition

$$\lim_{p\to 0}\frac{w^{-}(p)}{p^{\gamma}} > \frac{1}{\lambda}. \tag{8.20}$$

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

$$\lim_{p\to 0}\frac{w^{-}(p)}{p^{\gamma}} < \frac{1}{\lambda}. \tag{8.21}$$

Lemma 2 *: For Prelec's function, and for* $\gamma > 0$*,* $\lim_{p\to 0}\frac{w(p)}{p} = \infty$ *and* $\lim_{p\to 0}\frac{w(p)}{p^{\gamma}} = \infty$*.*

In view of Proposition 14(a), Lemma 2, the Becker paradox reemerges under CP for any "standard" probability weighting function, for which $\lim_{p\to 0}\frac{w(p)}{p^{\gamma}} = \infty$. However, under CCP, using the CPF in Definition 8 we get that $\lim_{p\to 0}\frac{w^{-}(p)}{p^{\gamma}} = 0$. The Becker paradox can then be explained under CCP by directly using Proposition 14(b).

We have considered a simplified model of crime that can admit several possible modifications. For instance, issues of heterogeneity among the population with respect to their reference points could be important. It is also desirable to explicitly model the *costs of deterrence* and *damages from crime* functions explicitly, in the tradition of Becker (1968). These extensions are beyond the scope of the current paper. The interested reader can consult Dhami and al-Nowaihi (2010h) for these extension, but the results for the elementary model, that we presented above, continue to hold.

8.3. The St. Petersburg paradox

Blavatskyy (2005) and Rieger and Wang (2006) have shown that the St. Petersburg paradox reemerges under CP. They prove that, even with a strictly concave utility function, the Bernoulli lottery will have an infinite expected utility under CP. Their analysis shows that this is due to the fact that all the standard probability weighting functions (see Definition 4) *infinitely* overweight infinitesimal probabilities. To solve this problem, Rieger and Wang (2006) propose a probability weighting which (in our terminology) *finitely* overweights infinitesimal probabilities.

[32] This ensures that the value function under CP and CCP has the power form in (7.2); see al-Nowaihi et al. (2008).

Definition 20 *: The Rieger and Wang (2006) probability weighting function is given by*[33]

$$w(p) = p + \frac{3(1-b)}{1-a+a^2}\left[ap - (1+a)p^2 + p^3\right],\ a \in \left(\frac{2}{9}, 1\right),\ b \in (0,1),\ 0 \leq p \leq 1. \quad (8.22)$$

The following proposition can be proved using a simple proof that we omit.

Proposition 15 *: The Rieger and Wang function (8.22) finitely-overweights infinitesimal probabilities, i.e.,* $\lim_{p\to 0}\frac{w(p)}{p} \in (1,\infty)$*; see Definition 4. Depending on the values of the parameters* a *and* b*, the Rieger and Wang function (8.22) finitely underweights near-one probabilities, or positively overweights near-one probabilities, i.e.,* $\lim_{p\to 1}\frac{1-w(p)}{1-p} \in (1,\infty)$ *or* $\lim_{p\to 1}\frac{1-w(p)}{1-p} \in (0,1)$*, respectively; see Definition 4b and Definition 5b.*

Rieger and Wang then go on to show that the probability weighting function (8.22) solves the St. Petersburg paradox under CP by generating a finite expected utility.[34] It does so, because, although it overweights low probabilities, it only *finitely* overweights them. However, because it *overweights* infinitesimal probabilities it, like the other probability weighting functions, is unable to explain the insurance paradox and, also like the other probability weighting functions, it does not capture the fact that decision makers code very small probability events as impossible.

In contrast to the Rieger and Wang function, CCP because it zero underweights near zero probabilities, can also explain the St. Petersburg paradox (a simple proof, that we omit, will demonstrate this fact). However, in applications (for instance, insurance implications) the CPF (and, hence, CCP) is a significant improvement. Furthermore, there are no known axiomatic foundations of the Rieger and Wang function, unlike the CPF which is axiomatically founded.

9. Why is CCP possibly the best available decision theory?

It is widely accepted that EU is refuted by a large body of literature.[35] For this reason, we shall focus here only on the salient mainstream alternatives to EU. As the reader will appreciate, our observations will apply to a much larger set of theories.

One could combine the Composite Prelec weighting functions (CPF) in our paper with Quiggen's (1982) RDU to propose *composite rank dependent utility* (CRDU). However, we believe that *composite prospect theory* (CCP) is a much better alternative than CRDU.

[33] Rieger and Wang state, incorrectly, that $a \in (0,1)$. For sufficiently low a and $p \simeq \frac{1}{3}$, $w(p)$, as given by (8.22), is *decreasing* in p. The lower bound of $\frac{2}{9}$ on a is sufficient, but not necessary, for $w(p)$ to be strictly increasing.

[34] We omit the proof; the interested reader can directly consult their paper.

[35] See, for instance, Kahneman and Tversky (2000) and Starmer (2000) for surveys.

CRDU could be thought of as a special case of CCP with a reference point of zero and absence of the domain of gains and losses. However, this rules out the psychologically powerful notion of loss aversion and reference dependence, which have strong explanatory power in a variety of contexts.[36] Our strong hunch is that this will considerably weaken the power of theory. Reference dependence and loss aversion are robust findings from experimental and field data, that have enormous explanatory potential and are too important to be ignored. For that reason, we have not focussed on CRDU in this paper.

Regret theory tries to incorporate the ideas of *regret*, and its counterpart, *rejoice*, into decision theory.[37] It postulates the existence of a regret function over binary comparisons of outcomes and then choose an action that has the lowest regret, using objective weighting of probabilities. There is no notion of loss aversion and reference dependence; problems that it shares with RDU. Furthermore, regret theory is not able to account for the two important stylized facts S1 and S2 outlined in section **??** in the introduction. Hence, it would, for instance, find it hard to solve the 3 applications that we discuss in section 8.

Similar comments apply to the *case based decision theory* of Gilboa and Schmeidler (2001) although it would seem to have greater applicability in situations of uncertainty rather than risk. However, in one important approach to uncertainty, Tversky and Koehler (1994), Rottenstreich and Tversky (1997) outline 'support theory'.[38] In essence, they propose axioms which allow, in the first stage, for a situation of uncertainty to be transformed into a situation of risk. Then, in the second stage, decision makers simply use PT or CP (and by implication, CCP) to choose among risky prospects. This, once again, allows for a central role for PT, CP, and CCP, even in situations of uncertainty.

These observations lead us to assert that, of the available alternatives at the moment, CCP is possibly the best decision theory, particularly under risk.

10. Conclusions

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

[36] See, for instance, the applications discussed in Camerer (2000), Kahneman and Tversky (2000) and Barberis and Thaler (2003).

[37] Important early ideas that formed the basis of subsequent developments in this area are Bell (1982), Loomes and Sugden (1982) and Fishburn (1982).

[38] Existing theories of uncertainty have been recently strongly criticised by al-Najjar and Weinstein (2009). See the same issue of *Economics and Philosophy* for replies by emiment theorists working in the area and the rejoinder by al-Najjar and Weinstein. However, support theory is not subjected to these critiques.

to treat as certain. But this was not rigorously formalized. Furthermore, PT allowed for the choice of stochastically dominated options, which was not well received.

In their update to PT, 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. 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 (ii) ignore events of extremely low probability and treat extremely high probability events as certain. We also provide three applications of CCP to outstanding puzzles in economics: insurance behavior towards low probability events, St. Petersburg paradox, and the inefficacy of low-probability capital punishment.

We believe that CCP offers, at the moment, the best choice, among the alternatives.

11. Appendices

11.1. Appendix-A: Proofs of the results

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

We shall use the following simple lemma.

Lemma 3 *: Let $w(p)$ be a probability weighting function (Definition 1). Then:*
(a) If $w(p)$ is differentiable in a neighborhood of $p=0$, then $\lim_{p\to 0}\frac{w(p)}{p}=\lim_{p\to 0}w'(p)$.
(b) If $w(p)$ is differentiable in a neighborhood of $p=1$, then $\lim_{p\to 1}\frac{1-w(p)}{1-p}=\lim_{p\to 1}w'(p)$.

Proof of Lemma 3: (a) Let $p\to 0$. Since w is continuous (Proposition 1c), $w(p)\to w(0)=0$ (Proposition 1a). By L'Hospital's rule, $\frac{w(p)}{p}\to\frac{dw(p)/dp}{dp/dp}=w'(p)$.
(b) Similarly, if $p\to 1$, then $w(p)\to w(1)=1$. By L'Hospital's rule, $\frac{1-w(p)}{1-p}\to\frac{d[1-w(p)]/dp}{d(1-p)/dp}=w'(p)$. ■.

Proof of Proposition 2: $\frac{w(p)}{p}=\frac{1}{[p^{\gamma}+(1-p)^{\gamma}]^{\frac{1}{\gamma}}p^{1-\gamma}}\to\infty$, as $p\to 0$.

$\frac{1-w(p)}{1-p}\to w'(p)=\left\{\frac{\gamma}{p}+\frac{1}{p^{\gamma}+(1-p)^{\gamma}}\left[\frac{1}{(1-p)^{1-\gamma}}-\frac{1}{p^{1-\gamma}}\right]\right\}w(p)\to\infty$, as $p\to 1$. ■.

Proof of Proposition 3: Straightforward from Definition 7. ■.

Proof of Proposition 4: Obvious from Definition 7. ■.

The following lemmas will be useful.

Lemma 4 *: For $\alpha \neq 1$, the Prelec function (Definition 7) has exactly three fixed points, at 0, $p^* = e^{-\left(\frac{1}{\beta}\right)^{\frac{1}{\alpha-1}}}$ and 1. In particular, for $\beta = 1$, $p^* = e^{-1}$.*

Proof of Lemma 4: From Propositions 1a and 3 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 $e^{-\beta(-\ln p^*)^\alpha} = p^* \Leftrightarrow p^* = e^{-\left(\frac{1}{\beta}\right)^{\frac{1}{\alpha-1}}}$. In particular, $\beta = 1$ gives $p^* = e^{-1}$. ■.

Lemma 5 *: Let $w(p)$ be the Prelec function (Definition 7) and let*

$$f(p) = \alpha\beta(-\ln p)^\alpha + \ln p + 1 - \alpha,\ p \in (0,1), \tag{11.1}$$

then

$$f'(p) = \frac{1}{p}\left[1 - \alpha^2\beta(-\ln p)^{\alpha-1}\right],\ p \in (0,1), \tag{11.2}$$

$$w''(p) = \frac{w'(p)}{p(-\ln p)} f(p),\ p \in (0,1), \tag{11.3}$$

$$w''(p) \lesseqgtr 0 \Leftrightarrow f(p) \lesseqgtr 0,\ p \in (0,1). \tag{11.4}$$

Proof of Lemma 5: Differentiate (11.1) to get (11.2). Differentiate (4.3) twice and use (11.1) to get (11.3). $-\ln p > 0$, since $p \in (0,1)$. $w'(p) > 0$ follows from Definitions (1) and (7) and Proposition (3). (11.4) then follows from (11.3). ■.

Lemma 6 *: Let $w(p)$ be the Prelec function (Definition 7). Suppose $\alpha \neq 1$. Then*
(a) $w''(\widetilde{p}) = 0$ for some $\widetilde{p} \in (0,1)$ and, for any such $\widetilde{p}$:

$$\text{(i) For } \alpha < 1: \ p < \widetilde{p} \Rightarrow w''(p) < 0,\ p > \widetilde{p} \Rightarrow w''(p) > 0 \tag{11.5}$$

$$\text{(i) For } \alpha > 1: p < \widetilde{p} \Rightarrow w''(p) > 0,\ p > \widetilde{p} \Rightarrow w''(p) < 0 \tag{11.6}$$

(b) The Prelec function has a unique inflexion point, $\widetilde{p} \in (0,1)$, and is characterized by $f(\widetilde{p}) = 0$, where $f(p)$ is defined in (11.1) i.e., $\alpha\beta(-\ln \widetilde{p})^\alpha + \ln \widetilde{p} + 1 - \alpha = 0$.
(c) $\beta = 1 \Rightarrow \widetilde{p} = e^{-1}$.
(d) $\frac{\partial \widetilde{p}}{\partial \beta} = \frac{\alpha \widetilde{p}(-\ln \widetilde{p})^{1+\alpha}}{(\alpha-1)(\alpha - \ln \widetilde{p})}$.
(e) $\frac{\partial[w(\widetilde{p}) - \widetilde{p}]}{\partial \beta} = \frac{\widetilde{p}(-\ln \widetilde{p})^{1+\alpha}}{(\alpha-1)(\alpha - \ln \widetilde{p})}\left(e^{\frac{1-\alpha}{\alpha}}(\widetilde{p})^{\frac{1-\alpha}{\alpha}} - \alpha\right)$.
(f) $\widetilde{p} \lesseqgtr w(\widetilde{p}) \Leftrightarrow \beta \lesseqgtr 1$.

Proof of Lemma 6: (a) Suppose $\alpha < 1$. From (11.1) we see that

$$\lim_{p \to 1} f(p) = 1 - \alpha > 0; \lim_{p \to 0} f(p) = \left[\frac{\alpha\beta}{(-\ln p)^{1-\alpha}} + 1\right] \ln p + 1 - \alpha = -\infty$$

Since f is continuous, it follows that $f(\widetilde{p}) = 0$, for some $\widetilde{p} \in (0,1)$. From (11.4), it follows that $w''(\widetilde{p}) = 0$. Since $\alpha < 1$, (11.1) gives $\alpha\beta(-\ln\widetilde{p})^{\alpha} + \ln\widetilde{p} < 0$ and, hence,

$$\widetilde{p} < e^{-(\alpha\beta)^{\frac{1}{1-\alpha}}}. \tag{11.7}$$

Consider the case, $p < \widetilde{p}$. From (11.7) it follows that $p < e^{-(\alpha\beta)^{\frac{1}{1-\alpha}}}$ and, hence, $1 - \frac{\alpha^2\beta}{(-\ln p)^{1-\alpha}} > 1 - \alpha > 0$. Thus, from (11.2), $f'(p) > 0$. Since $f(\widetilde{p}) = 0$, it follows that $f(p) < 0$. Hence, from (11.4), it follows that $w''(p) < 0$. This establishes the first part of (**??**). The derivation of the second part of (**??**) is similar. The case $\alpha > 1$ is similar.

(b) follows from (a) and (11.1), (11.4).

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

(d) Differentiating the identity $f(\widetilde{p}) = 0$ with respect to β gives $\frac{\partial\widetilde{p}}{\partial\beta} = \frac{\alpha\widetilde{p}(-\ln\widetilde{p})^{\alpha}}{\alpha^2\beta(-\ln\widetilde{p})^{\alpha-1}-1}$, then using $f(\widetilde{p}) = 0$ gives $\frac{\partial\widetilde{p}}{\partial\beta} = \frac{\alpha\widetilde{p}(-\ln\widetilde{p})^{1+\alpha}}{(\alpha-1)(\alpha-\ln\widetilde{p})}$.

(e) Differentiate $w(\widetilde{p}) - \widetilde{p} = e^{-\beta(-\ln\widetilde{p})^{\alpha}} - \widetilde{p}$ with respect to β, and use (d) and $f(\widetilde{p}) = 0$, to get $\frac{\partial[w(\widetilde{p})-\widetilde{p}]}{\partial\beta} = \frac{\widetilde{p}(-\ln\widetilde{p})^{1+\alpha}}{(\alpha-1)(\alpha-\ln\widetilde{p})}\left(e^{\frac{1-\alpha}{\alpha}}(\widetilde{p})^{\frac{1-\alpha}{\alpha}} - \alpha\right)$.

(f) Assume $\alpha < 1$. For $\beta = 1$, $\widetilde{p} = e^{-1}$ and, hence, $e^{\frac{1-\alpha}{\alpha}}(\widetilde{p})^{\frac{1-\alpha}{\alpha}} - \alpha = e^{\frac{1-\alpha}{\alpha}}(e^{-1})^{\frac{1-\alpha}{\alpha}} - \alpha = 1 - \alpha > 0$. Since $\frac{\partial\widetilde{p}}{\partial\beta} < 0$ for $\alpha < 1$, it follows that $e^{\frac{1-\alpha}{\alpha}}(\widetilde{p})^{\frac{1-\alpha}{\alpha}} - \alpha > 0$ for $\beta \leq 1$. Hence, $\frac{\partial[w(\widetilde{p})-\widetilde{p}]}{\partial\beta} < 0$ for $\beta \leq 1$. We have $w(\widetilde{p}) - \widetilde{p} = w(e^{-1}) - e^{-1} = w(p^*) - p^* = 0$ for $\beta = 1$ (recall part c and Lemma 4). Hence, $w(\widetilde{p}) > \widetilde{p}$ for $\beta < 1$. The case $\beta \geq 0$ is similar. The case $\alpha > 1$ is similar. ■.

Proof of Proposition 5: (a) is established by Lemma 4. (b) is established by Lemma 6b. (c) follows from Lemma 6a(i). (d) follows from Lemma 6a(ii). (e), (f) and (g) follow from Lemma 6f ■.

Proof of Corollary 1: Immediate from Proposition 5. ■.

Proof of Proposition 6: From (4.3) we get $\ln\frac{w(p)}{p} = \ln w(p) - \ln p = -\beta(-\ln p)^{\alpha} - \ln p = (-\ln p)^{\alpha}\left((-\ln p)^{1-\alpha} - \beta\right)$. Hence, if $\alpha < 1$, then $\lim_{p\to 0}\ln\frac{w(p)}{p} = \infty$ and, hence, $\lim_{p\to 0}\frac{w(p)}{p} = \infty$. This establishes (ai). On the other hand, if $\alpha > 1$, then $\lim_{p\to 0}\ln\frac{w(p)}{p} = -\infty$ and, hence, $\lim_{p\to 0}\frac{w(p)}{p} = 0$. This establishes (bi). From (4.3) we get $w'(p) = \frac{\alpha\beta}{p}(-\ln p)^{\alpha-1}w(p)$. If $\alpha < 1$, then $\lim_{p\to 1}w'(p) = \infty$. Part (aii) then follows from Lemma 3b. If $\alpha > 1$, then $\lim_{p\to 1}w'(p) = 0$. Part (bii) then follows from Lemma 3b. ■.

Proof of Proposition 7: Straightforward from Definition 8. ■.

Proof of Proposition 8: Follows by direct calculation from (5.4) and (5.5). ■.

Proof of Proposition 9: Part (a) follows from part (bi) of Proposition 6, since $\alpha_0 > 1$. Part (b) follows from part (bii) of Proposition 6, since $\alpha_1 > 1$. ■.

Proof of Proposition 13: If $\varphi(p) = \frac{b}{p}$ then the *NCC* (8.12) holds for concave u and, hence, the result follows.■

Proof of Proposition 10: Straightforward from Definitions 1 and 9. ■.

Proof of Lemma 1: Straightforward from Definition 12. ■.

Proof of Proposition 12: *(a)⇒(b)*. Suppose the probability weighting function, w, satisfies local power invariance.

Let

$$f(x,\lambda) = w\left(\left(w^{-1}\left(e^{-x}\right)\right)^{\lambda}\right),\ x,\lambda \in \mathbb{R}_{++}, \tag{11.8}$$

and

$$\varphi(\lambda) = -\ln f(1,\lambda) = -\ln w\left(\left(w^{-1}\left(e^{-1}\right)\right)^{\lambda}\right),\ \lambda \in \mathbb{R}_{++}. \tag{11.9}$$

Clearly,

$$\varphi \text{ maps } \mathbb{R}_{++} \text{ into } \mathbb{R}_{++}. \tag{11.10}$$

Since $w^{-1}\left(e^{-1}\right) \in (0,1)$, it follows that $\left(w^{-1}\left(e^{-1}\right)\right)^{\lambda}$ is a strictly decreasing function of λ, and so are $w\left(\left(w^{-1}\left(e^{-1}\right)\right)^{\lambda}\right)$ and $\ln w\left(\left(w^{-1}\left(e^{-1}\right)\right)^{\lambda}\right)$. Hence, from (11.9),

$$\varphi \text{ is a strictly increasing function of } \lambda. \tag{11.11}$$

From (11.8) we get

$$f(-\mu \ln w(p), \lambda) = w\left(\left(w^{-1}\left((w(p))^{\mu}\right)\right)^{\lambda}\right),\ p \in (0,1),\ \lambda,\mu \in \mathbb{R}_{++}. \tag{11.12}$$

Let $0 = p_0 < p_1 < ... < p_n = 1$.

Since w is C^1 on (p_{i-1}, p_i) it follows, from (11.9) and (6.3), that

$$\varphi \text{ is } C^1 \text{ on } \left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right). \tag{11.13}$$

Let

$$p,q \in (p_{i-1}, p_i),\ (w(p))^{\mu} = w(q),\ p^{\lambda}, q^{\lambda} \in (p_{i-1}, p_i). \tag{11.14}$$

From (**??**) we get

$$q = w^{-1}\left((w(p))^{\mu}\right). \tag{11.15}$$

From (11.14) and local power invariance, we get

$$\left(w\left(p^{\lambda}\right)\right)^{\mu} = w\left(q^{\lambda}\right). \tag{11.16}$$

Substituting for q from (11.15) into (11.16), we get

$$\left(w\left(p^{\lambda}\right)\right)^{\mu} = w\left(\left(w^{-1}\left((w(p))^{\mu}\right)\right)^{\lambda}\right),\ p, p^{\lambda} \in (p_{i-1}, p_i). \tag{11.17}$$

From (11.17) and (11.12) we get

$$f(-\mu \ln w(p), \lambda) = \left(w\left(p^{\lambda}\right)\right)^{\mu},\ p, p^{\lambda} \in (p_{i-1}, p_i). \tag{11.18}$$

In particular, for $\mu = 1$, (11.18) gives

$$f\left(-\ln w\left(p\right),\lambda\right) = w\left(p^{\lambda}\right),\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right). \tag{11.19}$$

From (11.19) we get

$$\left(f\left(-\ln w\left(p\right),\lambda\right)\right)^{\mu} = \left(w\left(p^{\lambda}\right)\right)^{\mu},\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right). \tag{11.20}$$

From (11.18) and (11.20) we get

$$f\left(-\mu\ln w\left(p\right),\lambda\right) = \left(f\left(-\ln w\left(p\right),\lambda\right)\right)^{\mu},\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right). \tag{11.21}$$

Put

$$z = -\ln w\left(p\right). \tag{11.22}$$

From (11.21) and (11.22) we get

$$f\left(\mu z\right) = \left(f\left(z,\lambda\right)\right)^{\mu},\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right). \tag{11.23}$$

From (11.9) and (11.23) we get

$$f\left(\mu\right) = \left(f\left(1,\lambda\right)\right)^{\mu} = e^{-\mu\varphi(\lambda)},\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right), \tag{11.24}$$

and, hence,

$$f\left(-\ln w\left(p\right),\lambda\right) = \left(w\left(p\right)\right)^{\varphi(\lambda)},\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right). \tag{11.25}$$

From (11.19) and (11.25) we get

$$w\left(p^{\lambda}\right) = \left(w\left(p\right)\right)^{\varphi(\lambda)},\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right), \tag{11.26}$$

from which we get,

$$\varphi\left(\lambda\right) = \frac{\ln w\left(p^{\lambda}\right)}{\ln w\left(p\right)},\ p, p^{\lambda} \in \left(p_{i-1}, p_i\right). \tag{11.27}$$

Let $p, p^{\lambda}, p^{\mu}, p^{\lambda\mu} \in \left(p_{i-1}, p_i\right)$. From (11.26) and (11.27) we get $\varphi\left(\lambda\mu\right) = \frac{\ln w\left(p^{\lambda\mu}\right)}{\ln w(p)} = \frac{\ln w\left(\left(p^{\mu}\right)^{\lambda}\right)}{\ln w(p)} = \frac{\ln\left[\left(w(p^{\mu})\right)^{\varphi(\lambda)}\right]}{\ln w(p)} = \frac{\varphi(\lambda)\ln(w(p^{\mu}))}{\ln w(p)} = \frac{\varphi(\lambda)\ln\left[\left(w(p)\right)^{\varphi(\mu)}\right]}{\ln w(p)} = \frac{\varphi(\lambda)\varphi(\mu)\ln w(p)}{\ln w(p)} = \varphi\left(\lambda\right)\varphi\left(\mu\right)$, i.e.,

$$\varphi\left(\lambda\mu\right) = \varphi\left(\lambda\right)\varphi\left(\mu\right),\ p, p^{\lambda}, p^{\mu}, p^{\lambda\mu} \in \left(p_{i-1}, p_i\right). \tag{11.28}$$

From (6.3), (6.4), (11.13) and (11.28) we have: φ is C^1 on $\left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right)$, $0 < \frac{\ln p_i}{\ln p} < 1 < \frac{\ln p_{i-1}}{\ln p}$, $\forall\lambda,\mu \in \left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right)$, s.t. $\lambda\mu \in \left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right)$, $\varphi\left(\lambda\mu\right) = \varphi\left(\lambda\right)\varphi\left(\mu\right)$. Hence, by Theorem 4 (see Appendix 2, below), we have,

$$\exists\alpha_i \in \mathbb{R}, \forall\lambda \in \left(\frac{\ln p_i}{\ln p}, \frac{\ln p_{i-1}}{\ln p}\right),\ \varphi\left(\lambda\right) = \lambda^{\alpha_i}. \tag{11.29}$$

But, by (11.11), φ is a strictly increasing function of λ. Hence,

$$\alpha_i > 0. \tag{11.30}$$

Let P_i and Λ_i be as in Definition12. Let $p \in P_i$. Define $\varphi_i : \Lambda_i \to \mathbb{R}_{++}$ by $\varphi_i(\lambda) = \lambda^{\alpha_i}$. Then, clearly, φ_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 $\forall p \in P_i$, $\forall \lambda \in \Lambda_i$, $w\left(p^\lambda\right) = (w(p))^{\varphi_i(\lambda)}$. 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 fist establish the result for P_i, then we use induction, and the continuity conditions (6.2), to extend the result to P_{i+1}, P_{i+2}, ... , P_n and P_{i-1}, P_{i-2}, ... , P_1. Let $\beta_i = -\ln w\left(e^{-1}\right)$. Then $w\left(e^{-1}\right) = e^{-\beta_i}$. Let $p \in P_i$. Let $\lambda = -\ln p$. Then $p = e^{-\lambda}$. Hence $w(p) = w\left(e^{-\lambda}\right) = w\left(\left(e^{-1}\right)^\lambda\right) = \left(w\left(e^{-1}\right)\right)^{\varphi_i(\lambda)} = \left(e^{-\beta_i}\right)^{\lambda^{\alpha_i}} = e^{-\beta_i \lambda^{\alpha_i}} = e^{-\beta_i(-\ln p)^{\alpha_i}}$. Thus we have shown

$$w(p) = e^{-\beta_i(-\ln p)^{\alpha_i}}, \; p \in P_i. \tag{11.31}$$

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

$$w(p) = e^{-\beta_{i+1}(-\ln p)^{\alpha_{i+1}}}, \; p \in P_{i+1}. \tag{11.32}$$

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

$$w(p) = e^{-\beta_{i-1}(-\ln p)^{\alpha_{i-1}}}, \; p \in P_{i-1}. \tag{11.33}$$

Continuing the above process, we get

$$w(p) = e^{-\beta_i(-\ln p)^{\alpha_i}}, \; p \in P_i, \; i = 1, 2, ..., n, \tag{11.34}$$

which establishes part (c).

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

Proof of Proposition 14: First, note that $\lim_{p \to 0} \frac{w^+(1-p)}{(1-p)^\gamma} = 1$ because $w(1) = 1$. If (8.20) holds, then the *NCC* (8.19) will hold with strict inequality in some non-empty interval $(0, p_1)$.[39] Hence, no crime will occur if $p \in (0, p_1)$. If (8.21) holds, then the converse of the *NCC* (8.19) holds with strict inequality in some non-empty interval $(0, p_2)$. Hence, for punishment to deter in this case, we must have $p > p_2$. ■

[39] 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 $1-p \to 1$, $w(1) = 1$.

Proof of Lemma 2: Since $p^{-\gamma} = e^{-\gamma \ln p}$, (4.3) gives $\frac{w(p)}{p^{\gamma}} = e^{(-\ln p)\left(\gamma - \frac{\beta}{(-\ln p)^{1-\alpha}}\right)}$. Note that $\lim_{p \to 0} (-\ln p) = \infty$. Since $0 < \alpha < 1$, we get $\lim_{p \to 0} \ln (-\ln p)^{1-\alpha} = \infty$. Hence, since $\gamma > 0$, we get $\lim_{p \to 0} \frac{w(p)}{p^{\gamma}} = \infty$. This must also hold for $\gamma = 1$, so $\lim_{p \to 0} \frac{w(p)}{p} = \infty$, i.e., infinite overweighting of zero probabilities. ■

Proof of Proposition 15: $w(p) = 1 + \frac{3(1-b)}{1-a+a^2} \left[a - (1+a)p + p^2\right] \to 1 + \frac{3a(1-b)}{1-a+a^2} \in (1, \infty)$, as $p \to 0$. $\frac{1-w(p)}{1-p} \to w'(p) = \frac{3(1-b)}{1-a+a^2} \left[2p - (1+a)\right] \to \frac{3(1-b)}{1-a+a^2} (1-a) = \frac{3(1-b)}{1+\frac{a^2}{1-a}} \in (0, 2.8209)$, as $p \to 1$, since $a \in \left(\frac{2}{9}, 1\right)$, $b \in (0,1)$. ■.

11.2. Appendix-B: Cauchy's algebraic functional equations.

We start with Cauchy's first algebraic functional equation, with its classic proof. Our notation is standard. In particular: $\mathbb{R}$: reals, $\mathbb{R}_{+}$: non-negative reals, $\mathbb{R}_{++}$: positive reals and C^1: class of continuous functions with continuous first derivatives.

11.3. Cauchy's first algebraic functional equation

Theorem 1 *: If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous and satisfies $\forall x \in \mathbb{R}$, $\forall y \in \mathbb{R}$, $f(x+y) = f(x) + f(y)$, then $\exists c \in \mathbb{R}$, $\forall x \in \mathbb{R}$, $f(x) = cx$.*

Proof. By mathematical induction it follows that, $\forall n \in \mathbb{N}$, $\forall x_1, x_2, ..., x_n \in \mathbb{R}$, $f(\Sigma_{i=1}^{n} x_i) = \Sigma_{i=1}^{n} f(x_i)$. In particular, $\forall n \in \mathbb{N}, \forall x \in \mathbb{R}, f(nx) = nf(x)$. Let $n \in \mathbb{N}$, $x \in \mathbb{R}$. Let $y = \frac{1}{n}x$. Then $x = ny$. Hence, $f(x) = f(ny) = nf(y) = nf\left(\frac{1}{n}x\right)$. Thus, $\forall n \in \mathbb{N}$, $\forall x \in \mathbb{R}$, $f\left(\frac{1}{n}x\right) = \frac{1}{n}f(x)$. And, so, $\forall m, n \in \mathbb{N}$, $\forall y \in \mathbb{R}$, $f\left(\frac{m}{n}y\right) = \frac{1}{n}f(my) = \frac{m}{n}f(y)$. From the continuity of f it follows that $\forall x, y \in \mathbb{R}$, $f(xy) = xf(y)$. In particular, for $y = 1$, we get $\forall x \in \mathbb{R}$. $f(x) = xf(1)$. Letting $c = f(1)$, we get $\forall x \in \mathbb{R}$. $f(x) = cx$. ■.

Remark 11 *: Note that the rational number, $\frac{m}{n}$, can be arbitrarily large. Hence, for the proof to go through, we do need $f : \mathbb{R} \longrightarrow \mathbb{R}$ (or $f : \mathbb{R}_{+} \longrightarrow \mathbb{R}_{+}$ or $f : \mathbb{R}_{++} \longrightarrow \mathbb{R}_{++}$). In particular, this proof in not valid for the case $f : (a,b) \longrightarrow \mathbb{R}$ when (a,b) is a bounded interval; which is what we need. This is why we need a local form of this theorem.*

11.4. Local forms of Cauchy's algebraic functional equations.

The third Cauchy equation arises naturally in the course of our proof of Proposition *12*. However, we need a local form of it. That is, a form restricted to a bounded real interval around zero, rather than the whole real line (recall Remark11). We achieve this by replacing the assumption of continuity with the stronger assumption of differentiability. As in the classical approach, we give the proof for the first equation. We then transform the third equation to the second which, in turn, we transform to the first. There is a fourth Cauchy algebraic functional equation, but we need not be concerned with it here.

Theorem 2 *: Let $f:(a,b)\longrightarrow\mathbb{R}$ be C^1, $a<0<b$, $\forall x,y\in(a,b)$, s.t. $x+y\in(a,b)$, $f(x+y)=f(x)+f(y)$. Then $\exists c\in\mathbb{R}$, $\forall x\in(a,b)$, $f(x)=cx$.*

Proof. $f(0)=f(0+0)=f(0)+f(0)$. Hence, $f(0)=0$.
$\frac{f(x+\delta x)-f(x)}{\delta x}=\frac{f(x)+f(\delta x)-f(x)}{\delta x}=\frac{f(\delta x)}{\delta x}$. Hence, $f'(x)=\lim_{\delta x\to 0}\frac{f(x+\delta x)-f(x)}{\delta x}=\lim_{\delta x\to 0}\frac{f(\delta x)}{\delta x}$, which is independent of x, thus, $\exists c\in\mathbb{R}$, $f'(x)=c$ and so $f(x)=cx+C$. Hence, $0=f(0)=C$. It follows that $\exists c\in\mathbb{R}$, $f(x)=cx$. $\blacksquare$.

Theorem 3 *: Let $g:(A,B)\longrightarrow\mathbb{R}$ be C^1, $0<A<1<B$, $\forall X,Y\in(A,B)$, s.t. $XY\in(A,B)$, $g(XY)=g(X)+g(Y)$. Then $\exists c\in\mathbb{R},\forall X\in(A,B),g(X)=c\ln X$.*

Proof. Let $a_i=\ln A$, $b_i=\ln B$. Then $a_i<0<b_i$, $x\in(a_i,b_i)\Leftrightarrow e^x\in(A,B)$ and $x,y,x+y\in(a_i,b_i)\Rightarrow e^x,e^y,e^xe^y=e^{x+y}\in(A,B)$.

Define $f:(a_i,b_i)\longrightarrow\mathbb{R}$ by $\forall x\in(a_i,b_i)$, $f(x)=g(e^x)$. Since $e^x\in(A,B)$, f is well defined. Since g is C^1, f is also C^1 and for $x,y,x+y\in(a_i,b_i)$, $f(x+y)=g(e^{x+y})=g(e^xe^y)=g(e^x)+g(e^y)=f(x)+f(y)$. Hence, from Theorem 1, $\exists c\in\mathbb{R}$, $\forall x\in(a_i,b_i)$, $f(x)=cx$. Hence, $\forall x\in(a_i,b_i)$, $g(e^x)=cx$. Let $X\in(A,B)$, $x=\ln X$, then $x\in(a_i,b_i)$, $g(X)=g(e^x)=cx=c\ln X$, i.e.,$\forall X\in(A,B)$, $g(X)=c\ln X$. $\blacksquare$.

Theorem 4 *: Let $G:(A,B)\longrightarrow\mathbb{R}_{++}$ is C^1, $0<A<1<B$, $\forall X,Y\in(A,B)$, 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. Define $g:(A,B)\longrightarrow\mathbb{R}$ by $\forall X\in(A,B)$, $g(X)=\ln G(X)$. Since $G(X)>0$, g is well defined. Since G is C^1, g is also C^1 and for $X,Y\in(A,B)$, s.t. $XY\in(A,B)$, $g(XY)=\ln G(XY)=\ln(G(X)G(Y))=\ln G(X)+\ln G(Y)=g(X)+g(Y)$. Hence, from Theorem 3, $\exists c\in\mathbb{R}$, $\forall X\in(A,B)$, $g(X)=c\ln X$. Hence, $\forall X\in(A,B)$, $\ln G(X)=c\ln X=\ln X^c$. Hence, $\forall X\in(A,B)$, $G(X)=X^c$. $\blacksquare$.

References

[1] Al-Najjar, N. I. and J. Weinstein (2009): The Ambiguity Aversion Literature: A Critical Assessment," Economics & Philosophy. 25: 249-84.

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

[3] al-Nowaihi, A., Dhami, S., 2006. A simple derivation of Prelec's probability weighting function. The Journal of Mathematical Psychology 50 (6), 521-524.

[4] al-Nowaihi, A., Dhami, S. and Bradley, I., (2008) 'The Utility Function Under Prospect Theory', Economics Letters 99, p.337–339.

[5] Bar Ilan, Avner (2000) "The Response to Large and Small Penalties in a Natural Experiment", Department of Economics, University of Haifa, 31905 Haifa, Israel.

[6] Bar Ilan, Avner and Bruce Sacerdote (2001) "The response to fines and probability of detection in a series of experiments" National Bureau of Economic Research Working Paper 8638.

[7] Bar Ilan, Avner and Bruce Sacerdote (2004) "The response of criminals and noncriminals to fines" Journal of Law and Economics, Vol. XLVII, 1-17.

[8] Barberis, N. and Thaler, R., (2003). "A survey of behavioral finance", Handbook of the Economics of Finance, in: G.M. Constantinides & M. Harris & R. M. Stulz (ed.), Handbook of the Economics of Finance, edition 1, volume 1, chapter 18, pages 1053-1128. Elsevier.

[9] Becker, Gary (1968) "Crime and Punishment: an Economic Approach, " Journal of Political Economy, 76, 169-217.

[10] Bell, David. 1982. "Regret in Decision Making under Uncertainty," Operations Research. 20, pp. 961–81.

[11] Bernoulli, D., 1738. Exposition of a New Theory of the Measurement of Risk. English translation in Econometrica 22, 1954, 23-36.

[12] Blavatskyy, P.R., 2005. Back to the St. Petersburg paradox? Management Science, 51, 677-678.

[13] Camerer, Colin F. (2000) Prospect theory in the wild: Evidence from the field. In D. Kahneman and A. Tversky (eds.), Choices, Values, and Frames, Cambridge University Press; Cambridge.

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

[15] Deaton, A.S., Muellbauer, J., 1980. Economics and Consumer Behaviour. Cambridge University Press.

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

[17] Fishburn, Peter C. 1982. "Nontransitive Measurable Utility", Journal of Mathematical Psychology. 26, pp. 31–67.

[18] Gilboa, I. and Schmeidler, D. (2001) A theory of case based decisions. Cambridge University Press: Cambridge.

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

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

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

[22] Koszegi, B. and M. Rabin (2006) "A Model of Reference-Dependent Preferences", Quarterly Journal of Economics, 121: 1133-1166.

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

[24] Lattimore, J. R., Baker, J. K., Witte, A. D.,1992. The influence of probability on risky choice: A parametric investigation. Journal of Economic Behavior and Organization 17, 377-400.

[25] Levitt, S. D. (2004) "Understanding why crime fell in the 1990's: Four factors that explain the decline and six that do not." Journal of Economic Perspectives, 18: 163-190.

[26] Loomes, Graham and Robert Sugden. 1982. "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty," Economic Journal. 92, pp. 805–24.

[27] Luce, R. D., 2001. Reduction invariance and Prelec's weighting functions. Journal of Mathematical Psychology 45, 167-179.

[28] Machina, M. J. (2008) Non-expected utility theory. in Durlaf, S. N. and Blume, L. E. (eds.) The New Palgrave Dictionary of Economics, 2nd Edition, Macmillan (Basingstoke and New York).

[29] Polinsky, Mitchell and Steven Shavell (2007) "The theory of public enforcement of law." Chapter 6 in Polinsky, Mitchell and Steven Shavell (eds.) Handbook of Law and Economics, Vol. 1, Elsevier.

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

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

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

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

[34] Radelet, M. L. and R. L. Ackers, (1996) "Deterrence and the Death Penalty: The Views of the Experts." Journal of Criminal Law and Criminology 87 : 1-16.

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

[36] Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. Psychological Review, 104, 406–415.

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

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

[39] Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. Psychological Review, 101, 547-567.

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

[41] Weinstein, N.D. (1980). Unrealistic optimism about future life events. Journal of Personality and Social Psychology, 39, 806-820.

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