University of Leicester

DEPARTMENT OF ECONOMICS

HANG 'EM WITH PROBABILITY ZERO: WHY DOES IT NOTWORK?

Sanjit Dhami, University of Leicester, UK

Ali al-Nowaihi, University of Leicester, UK

Working Paper No. 06/14
November 2006

Hang 'em with probability zero: Why does it not work?

Sanjit Dhami* **Ali al-Nowaihi†**

17 November 2006

Abstract

A celebrated result in the economics of crime, which we call the *Becker proposition (BP)*, states that it is optimal to impose the severest possible punishment (to maintain effective deterrence) at the lowest possible probability (to economize on enforcement costs). Several other applications, some unrelated to the economics of crime, arise when an economic agent faces punishments/ rewards with very low probabilities. For instance, insurance against low probability events, principal-agent contracts that impose punitive fines, seat belt usage and the usage of mobile phones among drivers etc. However, the BP, and the other applications mentioned above, are at variance with the evidence. The BP has largely been considered within an expected utility framework (EU). We re-examine the BP under rank dependent expected utility (RDU) and prospect theory (PT). We find that the BP always holds under RDU. However, under plausible scenarios within PT it does not hold, in line with the evidence.

Keywords: Behavioral economics; Illegal activity; Expected utility theory; Rank dependent expected utility; Prospect theory; Prelec and higher order Prelec probability weighting functions.

JEL Classification: D81 (Criteria for Decision Making Under Risk and Uncertainty), K42 (Illegal Behavior and the Enforcement of Law).

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

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

"*Certainty of detection is far more important than severity of punishment.*" Lord Shawness (1965)[1]

"*...a useful theory of criminal behavior can dispense with special theories of anomie, psychological inadequacies, or inheritance of special traits and simply extend the economist's usual analysis of choice.*" Gary Becker (1968)

"*...public finance models that aim for real world relevance ought to take behavioral insights into account.*" Edward McCaffery and Joel Slemrod (2004)

"*Homo economicus is dead but whose Homo behavioralis will replace him?*" Ken Binmore (2004)[2]

1. Introduction

In one of the major contributions to economics, Gary Becker (1968) opened the way to the rigorous analysis of crime using the formal methods of modern economics.[3] A celebrated proposition from Becker (1968) states that the most efficient way to deter a crime is to impose the 'severest possible penalty with the lowest possible probability'. We shall call this the *Becker proposition*. The intuition is compelling and simple. By reducing the probability of detection and conviction, society can economize on costs of enforcement such as policing and trial costs. But by increasing the severity of the punishment, the deterrence effect of the punishment is maintained.

The Becker proposition takes a particularly stark form if we add two assumptions: (1) Risk neutrality or risk aversion on the part of individuals. (2) The availability of infinitely severe punishments such as ruinous fines, slavery, torture, extraction of body parts and capital punishment. With these two extra assumption, 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 *hang offenders with probability zero*.

1.1. The Evidence on the Becker Proposition

Empirical evidence has not been kind to 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. History does not bear out the Becker proposition either. Since the late middle ages, the severity of punishments has been declining while expenditures on

[1] Quoted by Gary Becker (1968, footnote 12).
[2] Cited on the back cover of Samuel Bowles (2004).
[3] See Polinsky and Shavell (2000) for a brief history of the economics of crime and law enforcement.

enforcement has been increasing. So, either policy makers and criminologists have been making a big mistake (quite possible), or the Becker proposition is wrong. We call this the *Becker paradox*.

A number of explanations of the Becker paradox have been attempted. These explanations include risk seeking behavior on the part of offenders, the ability to avoid severe fines by declaring bankruptcy, the need for differential punishments, type-I and type-II errors in conviction, rent seeking behavior, abhorrence of severe punishments, objectives other than deterrence, and the psychological traits of offenders. We analyze these explanations in detail in Section 4.1 below. We then go on to argue (section 4.2) that the work of Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004), on the jumping of red traffic lights, provides near decisive evidence that none of these explanations can provide a satisfactory resolution of the Becker paradox within an expected utility framework.

1.2. Other Possible Applications and Evidence

The idea behind the Becker proposition, namely the response of individuals to low probability punishments/rewards can be found in many other economic contexts. We indicate some of these here.

In principal-agent theory, it is well known that a more efficient contract might be achieved by what Rasmusen (1994) calls a 'boiling in oil contract'. To illustrate, suppose that an agent can undertake two effort levels, a low level (shirking) and a high level. The principal has access to a monitoring technology, whose cost increases with the probability of detection. Suppose the contract specifies the severest possible punishment (boiling in oil) if the agent is caught shirking. Then the high effort level can be induced even with a cheap monitoring technology that has a low probability of detection. However, we do not observe such contracts.

Expected utility predicts that a risk averse decision maker will always buy some positive level of insurance, even when premiums are unfair. What is observed is that many people do not buy any insurance, even when available. Indeed for several types of risk, the government has to legislate the mandatory purchase of insurance. Two particularly striking examples are given in the seminal work of Kunreuther et al. (1978). These are the unpopularity of flood and earthquake insurance, despite government intervention in the forms of high subsidies to overcome transaction costs, reduction of premiums below their actuarially fair rates, provision of reinsurance for firms and the provision of relevant information. To quote from Kunreuther et al. (1978, p248) "This brings us to the key finding of our study. The principal reason for a failure of the market is that most individuals do not use insurance as a means of transferring risk from themselves to others. This behavior is caused by people's refusal to worry about losses whose probability is below some threshold."[4]

Other studies that reach a similar conclusion, reviewed by Kunreuther et al. (1978, section 1.4), cover the decisions to wear seat belts, to obtain breast examinations, to stop smoking, to purchase subsidized crime insurance and to purchase flight insurance. The last of these, however, shows that people purchase too much flight insurance, compared to the prediction of EU.

Yet another example comes from legislation regarding usage of mobile phones in moving vehicles. Such legislation in the UK has been particularly lax up to now. 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. In the UK, however, evidence is to the contrary. There has been a growth in mobile phone usage among drivers. In particular, there is evidence that there has been a relatively greater growth in hand-held mobile usage relative to hands-free phones.[5]

1.3. Alternative frameworks

A possible reaction is to entirely reject the economic approach to the study of law and crime. But we take a less radical approach that maintains an economic analysis, but uses tools from behavioral economics. Behavioral economics seeks to retain the rigor of modern economics, however, it draws its inspiration, basic principles and methodology from cognitive and experimental psychology.

Since von Neumann and Morgenstern (1944), *expected utility theory* (EU) has become the main tool for studying individual decision making in economics. Following Becker (1968), EU has also become the main tool for studying individual decision making in the economics of crime. However, to our minds, an EU framework suffers from two problems. First, there is the large body of refutations of EU; for surveys see, for instance, Kahneman and Tversky (2000) and Starmer (2000). Second, EU gives rise to misleading qualitative and quantitative results when the standard model of crime is applied to tax evasion; see, for instance, Dhami and al-Nowaihi (forthcoming). Furthermore, the evidence on red traffic light jumping given by Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004) provides,

[4] Kunreuther et al. (1978) is a major study, involving a very large data set, using three methodologies i.e. survey data, econometric analysis and experimental evidence. All three methodologies give this same conclusion.

[5] Use of both kinds of mobile phones, hand-held and hands-free, is dangerous. But there is a relatively greater risk when one uses hand-held mobile while driving. A recent report by the 'Royal society for the prevention of accidents' (ROSPA) cites three different kinds of survey evidence (on p.2) that indicates that the usage of mobile phones while driving is respectively, 37%, 39% (for the UK) and 27% (for the US).

at least to our mind, a near decisive refutation of the predictions of EU.

Prospect theory (Kahneman and Tversky, 1979) and its later version, *cumulative prospect theory* (CP, Tversky and Kahneman, 1992), have emerged from behavioral economics as the main rivals to EU. There is a substantial body of evidence in support of CP. CP has been successfully used to explain a range of puzzles in economics, such as the disposition effect, asymmetric price elasticities, elasticities of labour supply that are inconsistent with standard models of labour supply, the excess sensitivity of consumption to income and the equity-premium puzzle in finance; see, for example, Camerer (2000), and, especially, the collection of papers in Kahneman and Tversky (2000). For applications to tax evasion and insurance, see Dhami and al-Nowaihi (forthcoming) and al-Nowaihi and Dhami (2006), respectively.

In EU, carriers of utility are the final levels of wealth (or bundles of goods). By contrast, in CP, carriers of utility are deviations of wealth (or bundles of goods) from a reference point[6]. Another element of CP is the non-linear transformation of probabilities. There is now a large amount of empirical and experimental evidence that supports the hypothesis that decision makers overweight low probabilities and underweight high probabilities. A number of probability weighting functions have been proposed that embody this feature; we discuss these issues in section 5 below.

When non-linear transformation of probabilities is combined with otherwise standard EU, we get *rank dependent expected utility theory* (RDU), which is a less radical departure from EU than CP, see Quiggin (1993).

1.4. Results

The first contribution of our paper is to show that the Becker paradox reemerges in RDU. Just as it does under EU, increasing the punishment under RDU reduces the utility from committing a crime if the offender is caught, but *leaves his utility unchanged if not caught.* This allows the probability of detection and conviction to be reduced, economizing on enforcement costs.

The second contribution of our paper is to show that if we take the reference point for each activity (crime or no crime) to be the expected income from that activity then the Becker paradox can be solved in CP. The intuition is as follows. Increasing the punishment reduces the expected income from crime. This *increases the offender's utility if he commits an offence and is not caught.* To say the same thing in more colorful language, reference

[6] Consider a student whose reference point is a grade B out of three possible grades $A > B > C$ and let the actual grade received by the student be g, where $g = A$, B, C. Then, under EU and RDU the utility of this individual is simply $v(g)$ while under PT it is $v(g - B)$. Such a student is in the *domain of gains* if $g = A$, and in the *domain of losses* if $g = B$. Finally, if the grade is B then the utility under PT is $v(0) = 0$.

dependence allows CP to model the *elation obtained by engaging in a dangerous activity and getting away with it.* Increasing the punishment has two effects: (1) it increases the loss if caught and (2) it *increases the sense of elation obtained from engaging in the dangerous activity but not being caught.* If the probability weighting function does not excessively overweight low probabilities, the effect of the latter can overcome the effect of the former (hence, explaining the Becker paradox).

Thus, our paper contributes to a rapidly growing body of literature that shows that much empirical evidence about human behavior, that is very difficult to explain using mainstream economics, can be more easily explained using behavioral economics.

1.5. Organization of the paper

The rest of the paper is organized as follows. Section 2 of our paper formulates a fairly standard but simple economic model of crime in the spirit of Becker (1968). We show how this generic model applies to theft/ robbery, tax evasion, enforcement of pollution and principal-agent problems in general. A proof of the Becker proposition is given (for the case of EU) in section 3. In section 4, we critically survey the main solutions that have been proposed for the Becker paradox. Section 5 is on probability weighting functions, which are critical to the explanation of the Becker paradox in our paper. In sections 6 and 7 we reexamine the Becker Paradox from the perspective of behavioral economics and demonstrate how it can be resolved. Section 8 concludes.

2. A simple model of crime

Suppose that an individual receives income $y_0 \geq 0$ from being engaged in some legal activity but income $y_1 \geq y_0$ from being engaged in some illegal activity. Hence, the benefit, b, from the illegal activity equals $y_1 - y_0 \geq 0$. 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,

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

Given the enforcement parameters p, F the individual makes only one choice: to commit the crime or not. We assume that all individuals are identical. Furthermore, we restrict attention in this paper to those illicit activities where an individual can engage in a *fixed level* of a criminal activity.

This generic situation can describe several important settings in the economics of illicit activity, for example:

Example 1 *(Theft/robbery): If an individual does not engage in theft then $y_0 = 0$. Engaging in theft gives a monetary reward $b = y_1 \geq 0$. If caught (with probability $p \geq 0$)*

the goods in possession of the individual (i.e. b) are impounded and, in addition, the offender pays a fine, f (or faces other penalties such as imprisonment equivalent to this monetary value). Hence $F = b + f$.

Example 2 *(Tax Evasion): Consider the following widely used model (Allingham and Sandmo, 1972; Yitzhaki, 1974). A taxpayer has taxable incomes z_1 and z_2 from two economic activities both of which are taxed at the rate $t > 0$. Income z_1 cannot be evaded (for instance, it could be wage income with the tax withheld at source). However, the individual can choose to evade or declare income z_2.*[7] *It follows that $y_0 = (1 - t)(z_1 + z_2)$. Suppose the taxpayer chooses to evade income z_2. Hence, $y_1 = (1 - t) z_1 + z_2 \geq y_0$ and the benefit from tax evasion is $b = tz_2 \geq 0$. If caught evading, the individual is asked to pay back the tax liabilities owed, $b = tz_2$, and an additional fine $f = \lambda t z_2$ where $\lambda > 0$ is the penalty rate. Hence, $F = (1 + \lambda) t z_2$.*

Example 3 *(Enforcement of Pollution): Consider a firm that produces a fixed output q that is sold in the market at a price p. Let the profit from the product market be π. As a by-product, the firm creates emissions, e, that are greater than the allowable legal limit on emissions, $\overline{e}$. With probability $p \geq 0$ the firm's emissions are audited by the appropriate regulatory authority. Emission can be reduced at a cost of $c > 0$ per unit. Hence, $y_0 = \pi - c(e - \overline{e})$, while $y_1 = \pi$ so that $b = c(e - \overline{e}) \geq 0$ is the benefit arising from not lowering emissions to the legal requirement. If caught, the firm pays $b = c(e - \overline{e})$ as well as a monetary fine $f \geq 0$. Hence, $F = f + c(e - \overline{e})$.*

It is easiest to think of the punishment, F, as a monetary fine (as in the above examples). However, F can also be interpreted as the monetary equivalent of a non-monetary punishment, such as imprisonment.[8]

Let $T(p, F) \geq 0$ be the total cost to society of crime and of law enforcement, so that

$$T(p, F) = C(p, F) + D(p, F),$$

where $C(p, F)$ is the cost to society of law enforcement and $D(p, F)$ is the damage to society caused by crime. Let us assume that society aims to choose p and F so as to minimize $T(p, F)$. A major insight emerging from the economic approach to crime, is that it might be optimal not to eradicate all crime, but to tolerate some level of crime, i.e., the level at which $T(p, F)$ is a minimum.

[7] Examples include income from several kinds of financial assets, domestic work, private tuition, private rent, income from overseas, among many others. In actual practice tax evasion often takes the form of completely hiding certain taxable activities; see Dhami and al-Nowaihi (forthcoming).

[8] In this paper, we do not consider the important issue of how to measure the monetary equivalent of a non-monetary punishment. Nor do we consider the optimal combination of monetary and non-monetary punishments.

Since, in our model, an individual makes only one choice: to commit the crime or not, and since all individuals are identical[9], the choice facing society is either to deter all crime or tolerate all crime. If crime is deterred, then $D = 0$ and $T(p, F) = C(p, F)$. If crime is not deterred, then $D(p, F) = \overline{D} > 0$[10] and $T(p, F) = C(p, F) + \overline{D}$. We shall assume that $C(0, 0) = 0$, thus $T(0, 0) = \overline{D}$. We shall also assume that C is differentiable with

$$\frac{\partial C}{\partial p} > 0; \frac{\partial C}{\partial F} \geq 0.$$

Thus, the cost of law enforcement can be reduced by reducing the probability of detection and conviction, p. In general, an increase in the punishment, F, will increase the cost of law enforcement (for example, increasing the length of prison sentences). However, we note a special case below.

Definition 1 *(Ideal fine): The case* $\frac{\partial C}{\partial F} = 0$ *can be thought of as that of an ideal fine, which has a fixed administrative cost and involves a transfer from the offender to the victim or society (so there is no aggregate loss to society other than the fixed administrative cost).*

Definition 2 *(Punishment function): By a punishment function we mean a differentiable function* $F : (0, 1] \to [0, F_{\max}]$ *that assigns to each probability of detection and conviction,* p, *a punishment* $F(p)$.

Definition 3 *(Cost and fine elasticities):* $C_p = \frac{p}{C}\frac{\partial C}{\partial p}$ *is the probability elasticity of cost,* $C_F = \frac{F}{C}\frac{\partial C}{\partial F}$ *is the punishment elasticity of cost and* $F_p = -\frac{p}{F}\frac{dF}{dp}$ *is the probability elasticity of punishment.*[11]

Lemma 1 *:* $\frac{d}{dp}C(p, F(p)) > 0$ *if, and only if,* $C_p > C_F F_p$.

Proof: $\frac{d}{dp}C(p, F(p)) = \frac{\partial C}{\partial p} + \frac{\partial C}{\partial F}\frac{dF}{dp} = \frac{C}{p}\left[\left(\frac{p}{C}\frac{\partial C}{\partial p}\right) + \left(\frac{F}{C}\frac{\partial C}{\partial F}\right)\left(\frac{p}{F}\frac{dF}{dp}\right)\right] = \frac{C}{p}\left[C_p - C_F F_p\right] > 0 \Leftrightarrow C_p > C_F F_p$. ■

The condition $C_p > C_F F_p$ is most likely to hold when the costs of administering fines is relatively low (compared to imprisonment, say). It will be satisfied for an *ideal fine*, since $C_p > 0$ and $C_F = 0$ for an ideal fine (see definition 1). A particularly interesting punishment function is given in the following definition.

---

[9] Of course, in reality populations are heterogenous. Furthermore, there are many types of crime where an individual might have a choice as to the level of the illegal activity, such as the amount of tax to evade. These can be accommodated within our model but we abstract here from such issues.

[10] Recall our assumption that if individuals choose to be dishonest then they can engage in a *fixed* level of the criminal activity. This fixity of the level ensures that the damage, D, to society is also fixed at some level, $\overline{D}$, should the illegal activity be undertaken.

[11] Rather than the standard practice of using a subscript to indicate a partial derivative, it will be convenient for us to follow a non-standard practice of using a subscript to indicate an elasticity.

Definition 4 *(Hyperbolic punishment function): A hyperbolic punishment function is defined by*

$$F(p) = \frac{b}{p}. \tag{2.1}$$

The name derives from the fact that in p, F space, the hyperbolic punishment function plots as a rectangular hyperbola. We show below that the hyperbolic punishment function will deter a risk neutral or risk averse offender under expected utility theory. Since $F_p = 1$, it will be socially beneficial to reduce p as long as $C_p > C_F$. In particular, this always holds for an ideal fine (since $C_F = 0$ for such a fine).

Finally, we an assumption to guarantee that our model is not vacuous, i.e., that it is worthwhile to prevent crime. We shall assume that

$$C(1, b) < \overline{D}. \tag{2.2}$$

Under the assumptions $\frac{\partial C}{\partial p} > 0$; $\frac{\partial C}{\partial F} \geq 0$, $C(1, b)$ is the maximum cost at the lowest fine. Because, by (2.2), damages to society from the illegal activity exceed this cost, crime prevention is worthwhile.

2.1. A principal-agent interpretation

We can also interpret this model, more generally, as that of a principal-agent relationship. A principal contracts an agent to perform a certain task in exchange for the monetary reward y_0. The agent can either carry out his task honestly, or can improperly exploit the principal's facilities to enhance his income to $y_1 > y_0$. This causes damage, D, to the principal. The principal can introduce a monitoring technology and a system of sanctions at a cost $C(p, F)$ to the principal. The total cost to the principal would then be $T(p, F) = C(p, F) + D(p, F)$, where p is the probability of detection and F is the sanction. The analogue of Becker's proposition is to impose the severest sanction on the agent with the minimum probability of detection, i.e., offer, what Rasmusen (1994) calls, a *boiling in oil* contract.

3. The Becker proposition under expected utility theory (EU)

In this section, we consider Becker's proposition in the framework he originally formulated it, namely, expected utility theory (EU). Consider an individual with continuously differentiable and strictly increasing utility of income, u. If he does not engage in crime, his income is y_0. His payoff, U_{NC}, is given by

$$U_{NC} = u(y_0). \tag{3.1}$$

On the other hand, if the individual engages in crime, his income is $y_1 \geq y_0$ if not caught but $y_1 - F \leq y_1$ if caught. Since he is caught with probability p, his payoff, EU_C, under EU is given by

$$EU_C = pu(y_1 - F) + (1-p)\,u(y_1). \tag{3.2}$$

The individual does not find it worthwhile to engage in crime if the following *no-crime condition* (NCC) is satisfied i.e.

$$EU_C \leq U_{NC}. \tag{3.3}$$

Substituting (3.2), (3.1) in (3.3), the NCC becomes

$$pu(y_1 - F) + (1-p)\,u(y_1) \leq u(y_0). \tag{3.4}$$

Since $y_1 - b = y_0$, (3.4) is clearly satisfied for $p = 1$ and $F = b$. Let F be a differentiable function of p satisfying

$$F : [0,1] \to [0, F_{\max}]\,,\ F(1) = b,\ F'(p) < 0. \tag{3.5}$$

The NCC (3.4) continues to hold, as p is reduced from 1, if, and only if, the following is the case,

$$\frac{d}{dp}\left[pu(y_1 - F(p)) + (1-p)\,u(y_1)\right] \geq 0, \tag{3.6}$$

which will be the case if, and only if,

$$u'(y_1 - F(p)) \geq \frac{u(y_1) - u(y_1 - F(p))}{-pF'(p)}. \tag{3.7}$$

For the case of the hyperbolic punishment function (2.1), the NCC (3.7) reduces to

$$u'(y_1 - F(p)) \geq \frac{u(y_1) - u(y_1 - F(p))}{F(p)}. \tag{3.8}$$

If the decision maker is risk averse or risk neutral, so that u is concave, then the NCC (3.8) *will* hold for all $p \in (0,1]$. In Figure 3.1 we show, in p, u space, the utility function in bold (the curve oe) for a risk averse individual. The NCC is illustrated by the fact that ac is greater than bc.

We show in Proposition 1 below that not only does the Becker proposition hold for risk-neutral and risk averse criminals, its implementation is socially desirable because it reduces the total cost of crime $T(p, F)$ for society.

Proposition 1 : *Under EU:*
(a) If the individual is risk neutral or risk averse, so that u is concave, then the hyperbolic punishment function $F(p) = \frac{b}{p}$ will deter crime. It follows that given any probability of

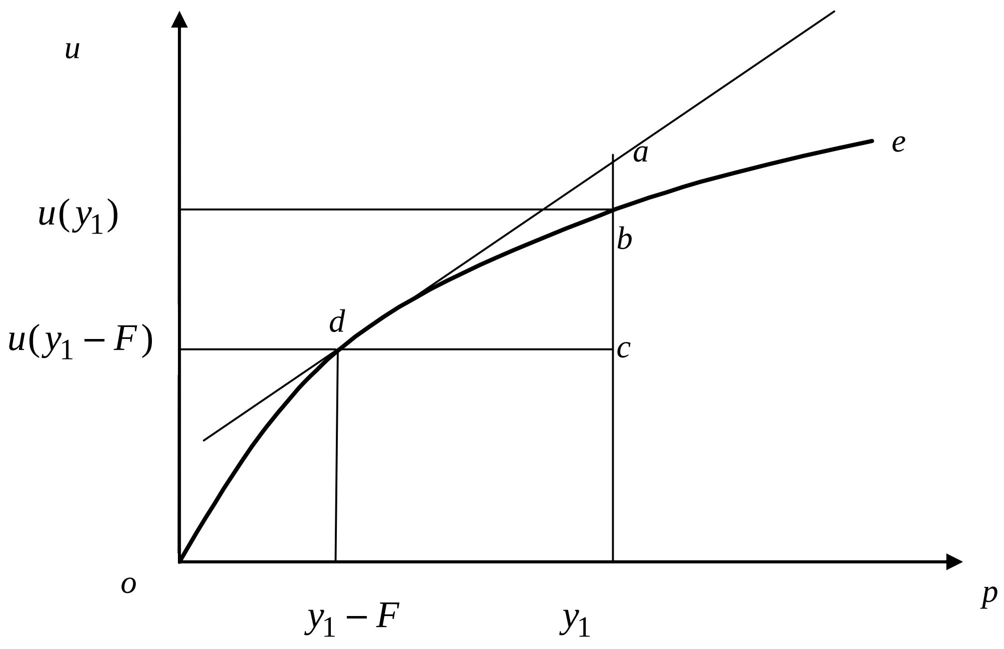

Figure 3.1: **The *NCC* holds for risk-averse and risk-neutral criminals**

detection and conviction, $p > 0$, no matter how small, then crime can be deterred by a sufficiently large punishment.
(b) If, in addition, $C_p > C_F$ (Definition 3), then reducing p reduces the total social cost of crime and law enforcement, $T(p, F)$.

Proof: If $F(p) = \frac{b}{p}$ then the *NCC* (3.8) holds for concave u and, hence, (a) follows. If $F(p) = \frac{b}{p}$, then (from Definition 3) $F_p = -\frac{p}{F}\frac{dF}{dp} = 1$. Since $C_p > C_F$ it follows, from Lemma 1, that $\frac{d}{dp}C(p, F(p)) > 0$. Since crime is deterred, $D(p, F(p)) = 0$. Hence, $T(p, F(p)) = C(p, F(p))$. Hence, $\frac{d}{dp}T(p, F(p)) > 0$. This establishes part (b). ■

As illustrations, we give two examples from Bar-Ilan (2000):

Example 4 : *Consider the utility function $u(y) = -e^{-y}$. Note that $u'(y) = e^{-y} > 0$, $u''(y) = -e^{-y} < 0$. From the second inequality, we see that this utility function exhibits risk averse behavior. Hence from part (a) of Proposition 1, it would be possible to deter crime, however small the probability of detection and conviction.*

Example 5 : *Consider the utility function $u(y) = e^{y}$. Note that $u'(y) = e^{y} > 0$, $u''(y) = e^{y} > 0$. From the second inequality, we see that this utility function exhibits risk seeking behavior. Hence, Proposition 1 (a) does not apply. In fact, substituting $u(y) = e^{y}$ in the NCC (3.4), and allowing infinitely large fines, gives that crime is deterred if, and* *only if,* $p > p_{\min} = 1 - \frac{u(y_0)}{u(y_1)} > 0$. *Hence, even if infinite punishments were available, it would be possible to deter crime only if the probability of conviction was above a certain minimum. Hence, the Becker proposition need not hold in the case of risk seeking behavior.*

The following two examples show that the conditions of Proposition 1(a), although sufficient, are not necessary.

Example 6 : *Consider the utility function* $u(y) = \ln y$. *Then, it follows from the NCC (3.4), that the probability of detection and conviction,* $p > 0$, *can be made arbitrarily low, by choosing the punishment* $F = y_1$. *Hence, the hyperbolic punishment function* $F = \frac{b}{p}$ *is sufficient, but not necessary.*

Example 7 : *Consider the following utility function used by Tversky and Kahneman (1992).*

$$u(y) = y^{\gamma},\ y \geq 0;\ u(y) = -(-y)^{\gamma},\ y < 0;\ 0 < \gamma < 1 \tag{3.9}$$

This utility function is (strictly) concave for $y \geq 0$ *but strictly convex for* $y < 0$. *Hence, we have risk seeking behavior in the region* $y < 0$. *However, the NCC (3.4) holds for any* $p \in (0,1]$, *if the punishment is given by* $F(p) = y_1 + \left(\frac{y_1^{\gamma} - y_0^{\gamma}}{p}\right)^{\frac{1}{\gamma}}$. *However,* $F_p = \frac{\left(\frac{y_1^{\gamma} - y_0^{\gamma}}{p}\right)^{\frac{1}{\gamma}}}{y_1 + \left(\frac{y_1^{\gamma} - y_0^{\gamma}}{p}\right)^{\frac{1}{\gamma}}} \to \infty$ *as* $p \to 0$. *Hence, it would be socially beneficial to drive the probability of detection and conviction down to zero if ideal fines* ($C_F = 0$) *were available or if* $C_p \to \infty$ *faster than* $C_F F_p$. *Hence, risk seeking behavior is not sufficient to explain the Becker paradox.*

4. The Becker paradox

The consensus of opinion is that, if the Becker proposition is correct, then the prevailing system of punishments is far from optimal, with punishments too low. Moreover, the historic trend has been towards more lenient punishments. Since Becker (1968), a considerable effort has gone into explaining the 'Becker paradox'. Several explanations have been proposed. These are briefly and critically discussed in section 4.1 below. Section 4.2 then argues that these explanations are not sufficient on their own.

4.1. Explanations

1. *Risk seeking behavior*: If decision makers are not risk-averse but risk-seekers (compare Examples 4 and 5, above) then the Becker proposition need not go through. Hence, this could be a potential explanation, and is the one given by Becker (1968),

for why we do not observe it in the real world. The main problem with this explanation is that it creates great difficulties for almost all other areas of mainstream economics in explaining behavior under uncertainty. Examples include insurance, investment, saving, risk management, principal-agent theory and mechanism design. A crucial aspect of all of these is the assumption of risk averse behavior (or the different degrees of risk aversion among the decision makers). Moreover, risk-seeking behavior is not sufficient, as shown by Example 7.

2. *Bankruptcy issues*: The possibility of declaring bankruptcy puts an upper bound on the level of fines that can be usefully imposed. There are several objections to this explanation. First, it takes fines too literally, rather than the more general interpretation as the monetary equivalent of punishment. Second, even when fines are interpreted literally, they can be backed up by other punishments such as imprisonment (which is currently the case) or penal slavery (which used to be the case) for those (and their descendants) unwilling or unable to pay the fine. Third, the historic trend has been to limit, rather than enhance, the consequences of declaring bankruptcy. Witness, for example, the emergence of the limited company. See, for instance, Friedman (1999).

3. *Differential punishments*: The argument for a system of differential punishments is unassailable. However, it does not explain why the whole portfolio of punishments cannot be made more severe while maintaining differentiation. For example, we could combine imprisonment and capital punishment with various degrees of torture. In fact, the historic trend is to make prisons (and capital punishment, where it still remains) more humane. See Polinsky and Shavell (2000).

4. *Errors in conviction*: To our minds, this is one of the two most cogent explanations for the Becker paradox (the other being rent seeking behavior). The penal system may fail to convict an offender (a type I error), or might falsely convict an innocent person (a type II error). The possibility of falsely convicting an innocent person causes a loss to society. Unboundedly severe punishments then cause potentially unbounded losses to society. This destroys one of the fundamental assumptions of the economic model of crime (namely, that increasing p is more costly to society than increasing F). Furthermore, more severe punishments encourage defendants (or society on their behalf) to spend more on detection and trial, which is socially wasteful. Again, this undermines the basic assumption that an increase in F is less costly to society than an increase in p. See Polinsky and Shavell (2000).

5. *Rent seeking behavior*: The possibility of a false conviction and the availability of out of court settlements, encourages malicious accusations. This temptation increases

with increasing F, thus undermining the basic assumption that increasing F is less costly to society than increasing p. The possibility of failing to convict an offender encourages payments by offenders to lawyers to defend them or (even worse) to pay police (and other monitoring authorities) to 'turn a blind eye'. Again the possibility of these undermines the assumption that increasing F is less costly to society than increasing p; see, for instance, Friedman (1999). To our minds, explanations 4 and 5 should certainly be part of a full explanation. But they cannot be the full story, in light of the evidence discussed in section 4.2, below.

6. *Abhorrence of severe punishments*: If we accept this explanation (and there is much truth in it), and if we want to retain an economic explanation of crime, then we have to explain why it is beneficial for individuals in society to adopt such an attitude.

7. *Objectives other than deterrence*: It is usual to attribute objectives to punishment other than deterrence. These include incapacitation and retribution. It is straightforward to incorporate incapacitation into an economic model of crime (because it has measurable monetary benefits and costs). It may be possible to give an evolutionary-economic explanation to the emergence of the desire by individuals for retribution. Such a desire would clearly help law enforcement, so would be beneficial to society and, hence, to its members.

8. Pathological traits of offenders. Colman (1995) shows how the persistence of "criminal types" (the most notorious being psychopaths) can be part of an evolutionary stable Nash equilibrium. These individuals are predisposed to commit crime irrespective of the enforcement parameters p, F. Although this explains why the Becker proposition fails when applied to the most heinous crimes, it does not explain why the Becker proposition fails, for instance, for ordinary principal-agent problems where, according to the proposition, principals should impose the severest possible punishment on an agent with the lowest possible probability (boiling in oil contracts).

4.2. Why do these explanations not suffice? Evidence from jumping red traffic lights

Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004) provide, what is to our minds, near decisive evidence that none of the explanations in subsection 4.1 above, singly or jointly, can provide a satisfactory explanation of the Becker paradox within an EU framework. Bar-Ilan and Sacerdote (2001, 2004) 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. Clearly, this is an activity of economic significance. 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 "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. The corresponding numbers for the additional risks of serious and fatal injuries are 13.9 days and 69.4 days respectively". 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.

Clearly expected utility theory combined with risk aversion would find it very difficult to explain this evidence. Explanations 2-7 in section 4.1, are not applicable here, because the punishment is self inflicted. Explanation 8 is also inadequate, for Bar-Ilan and Sacerdote (2004) report "We find that red-light running decreases sharply in response to an increase in the fine... Criminals convicted of violent offences or property offences run more red lights on average but have the same elasticity as drivers without a criminal record".

This leaves explanation 1, i.e., that offenders are risk seeking. Unfortunately, Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004) do not report the attitude of offenders to insurance. It is clear that offenders do have car-insurance, for Bar-Ilan and Sacerdote (2004) report that "in California a single traffic ticket would raise the insurance premiums of the average driver by approximately \$160 per year for 3 years". However, it is not reported whether this insurance is compulsory or entered into voluntarily. 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.

4.3. Other offenses: Driving and talking on car mobile phones

The work of Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004), on jumping red traffic lights, receives corroboration from other sources. For example, 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. In the UK, however, evidence is to the contrary. 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).

4.4. Alternative frameworks

In light of this evidence, we, therefore, turn to frameworks other than EU. In section 6 we consider *rank dependent expected utility* (RDU) and in section 7 we consider *cumulative* *prospect theory* (CP). For each of these alternative frameworks, non-linear transformations of objective probabilities is a fundamental element. We explain this in more detail, next.

5. Probability weighting functions

Empirical evidence strongly suggests that decision makers overweight small probabilities and underweight large probabilities; see, for example, the collection of papers in Kahneman and Tversky (2000). Here we discuss two broad classes of probability weighting functions. The first, which include those of Tversky and Kahneman (1992) and Prelec (1998), we call *standard probability weighting functions.* The second, we call *new probability weighting functions.* In section 7, below, we show that they lead to very different behavior under CP.

Definition 5 *: By a probability weighting function we mean a strictly increasing function* $w : [0,1] \overset{onto}{\longrightarrow} [0,1]$.

Remark 1 *: Note that a probability weighting function,* w*, has a unique inverse,* $w^{-1} : [0,1] \overset{onto}{\longrightarrow} [0,1]$ *and that* w^{-1} *is strictly increasing. Furthermore, it follows that* w *and* w^{-1} *are continuous and must satisfy* $w(0) = w^{-1}(0) = 0$ *and* $w(1) = w^{-1}(1) = 1$.

5.1. Standard probability weighting functions

5.1.1. Prelec's probability weighting function

The Prelec (1998) probability weighting function has the attraction that it is parsimonious, is consistent with much of the available empirical evidence and has an axiomatic foundation[12]. Therefore, we chose it as our starting point.

Definition 6 *: (Prelec, 1998). By the Prelec function we mean the probability weighting function* $w : [0,1] \overset{onto}{\longrightarrow} [0,1]$ *given by*

$$w(0) = 0, \tag{5.1}$$

$$w(p) = e^{-\beta(-\ln p)^{\alpha}}, \ 0 < p \leq 1, \ 0 < \alpha < 1, \ \beta > 0. \tag{5.2}$$

Note that $\alpha = \beta = 1$ gives the the identity transformation $w(p) = p$, which is the case used by EU. For $\alpha = 0.35$, $\beta = 1$ we get $w(p) = e^{-(-\ln p)^{0.35}}$, which we plot below. This represents a decision maker who overweights low probability and underweights high probabilities.

[12] Prelec (1998) gives a derivation based on 'compound invariance', Luce (2001) gives a derivation based on 'reduction invariance' and al-Nowaihi and Dhami (forthcoming) give a derivation based on 'power invariance'. Since the Prelec function satisfies all three, 'compound invariance', 'reduction invariance' and 'power invariance' are all equivalent.

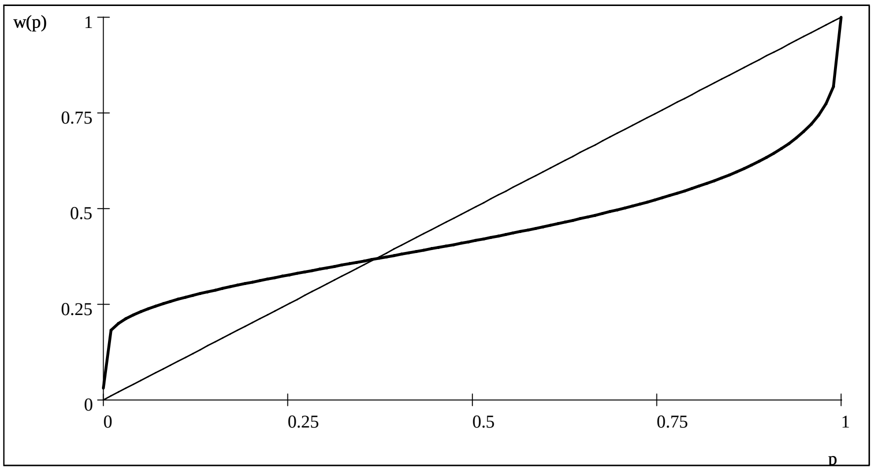

Figure 5.1: **A Prelec function**

For $\beta > 1$, we have $e^{-\beta(-\ln p)^{\alpha}} < e^{-(-\ln p)^{\alpha}}$. This represents a decision maker who, in addition to overweighting low probabilities and underweighting high probabilities, also systematically underweights all probabilities (relative to the case $\beta = 1$). So, we may describe such a decision maker as 'optimistic' in the sense that he systematically underweights the probability of being caught, relative to the case $\beta = 1$. Similarly we may describe a decision maker with $\beta < 1$ as 'pessimistic'. We record this in the following definition.

Definition 7 : *Consider the Prelec weighting function given in (5.2). A decision maker is said to be optimistic if $\beta > 1$ and pessimistic if $\beta < 1$.*

Proposition 2 : *For Prelec's function (Definition 6) and for $\gamma > 0$, $\lim_{p\to 0} \frac{w(p)}{p^{\gamma}} = \infty$.*

Proof: Since $p^{-\gamma} = e^{-\gamma \ln p}$, (5.2) 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 $\lim_{p\to 0} \frac{w(p)}{p^{\gamma}} = \infty$. ■

5.1.2. Other standard probability weighting functions

Tversky and Kahneman (1992) proposed the following probability weighting function

$$w(p) = \frac{p^{\sigma}}{\left[p^{\sigma} + (1-p)^{\sigma}\right]^{\frac{1}{\sigma}}}, 0.5 \leq \sigma < 1, \tag{5.3}$$

where the lower bound on σ comes from Rieger and Wang (2006). Experimental evidence from Tversky and Kahneman (1992) suggests that $\gamma \cong 0.88$, where γ is a parameter of the utility function given in (3.9) and that furthermore, $\sigma < \gamma$. Hence, this function also has the property that $\lim_{p \to 0} \frac{w(p)}{p^{\gamma}} = \infty$. It can be shown that other probability weighting functions that have been proposed, for example, Gonzalez and Wu (1999) and Lattimore, Baker and Witte (1992), also have this feature. For this reason we refer to these, as well as the Prelec function, as *the standard probability weighting functions.* Thus we have,

Proposition 3 *For the standard probability weighting functions* $\lim_{p \to 0} \frac{w(p)}{p^{\gamma}} = \infty$.

5.2. New probability weighting functions

Whilst the standard probability weighting functions appear to fit well the evidence in the middle ranges of probability, they are less successful in their predictions when the probability of an event is at one of the two extremes. Blavatsky (2004) and Rieger and Wang (2006) have shown that the St. Petersburg paradox reemerges under CP with the standard probability weighting functions. Al-Nowaihi and Dhami (2006) show that for low probability events, the standard probability weighting functions predict over-insurance, even when insurance is actuarially unfair. We consider below some of the new probability weighting functions that have been proposed and are able to address these problems.

5.2.1. The probability weighting function of Rieger and Wang

Rieger and Wang (2006) show that the probability weighting function:

$$w(p) = p + \frac{3(1-b)}{1-a+a^2}\left[ap - (1+a)p^2 + p^3\right],\ a \in \left(\frac{2}{9}, 1\right), b \in (0,1) \tag{5.4}$$

solves the St. Petersburg paradox by generating a finite expected utility under cumulative prospect theory.[13]

Proposition 4 : *For* $0 < \gamma < 1$ *the Rieger-Wang probability weighting function (5.4) satisfies:*

$$\lim_{p \to 0} \frac{w(p)}{p^{\gamma}} = 0.$$

Proof: Obvious from (5.4). ■

5.2.2. Higher order Prelec probability weighting functions

According to Prelec (1998, p505), the infinite limits of propositions 2 and 3 capture the qualitative change as we move from impossibility to probability. Al-Nowaihi and Dhami (2006) show that this leads people to fully insure against all losses of sufficiently low

[13]The lower bound on a comes from al-Nowaihi and Dhami (2006).

probability, even with actuarially unfair premiums and fixed costs to insurance. This is contrary to observation; see, for example, Kunreuther et al. (1978).

The results of propositions 2 and 3 also contradict the observed behavior that people ignore events of very low probability and code very high probability events as certain. Following Kahneman and Tversky (1979), we could rely on an initial editing phase, where the decision maker chooses which improbable events to treat as impossible and which probable events to treat as certain. While we are persuaded by this choice heuristic, as yet there is no general theory of the editing phase.

Al-Nowaihi and Dhami (2006) propose a class of probability weighting functions that combine the editing phase of Kahneman and Tversky (1979) with the probability weighting phase. These probability weighting functions have the property that $\lim_{p\to 0} \frac{w(p)}{p^\gamma} = 0$ and give a better explanation of insurance (and also solve the St. Petersburg paradox under CP). Al-Nowaihi and Dhami call these *higher order Prelec probability weighting functions* because they are generalizations of Prelec's function. We discuss these weighting functions next.

Lemma 2 : *(Prelec, 1998, p507, footnote). Prelec's function (Definition 6) can be written as*

$$w(0) = 0, w(1) = 1, \tag{5.5}$$

$$-\ln(-\ln w) = (-\ln \beta) + \alpha(-\ln(-\ln p)), 0 < p < 1. \tag{5.6}$$

Lemma 2 motivates the following development. Assume that $-\ln(-\ln w)$ can be expanded as a power series in $-\ln(-\ln p)$, i.e.,

$$-\ln(-\ln w) = \sum_{k=0}^{\infty} a_k (-\ln(-\ln p))^k, 0 < p < 1, \tag{5.7}$$

or, equivalently,

$$w(p) = \exp\left(-\exp\left(-\sum_{k=0}^{\infty} a_k (-\ln(-\ln p))^k\right)\right), 0 < p < 1. \tag{5.8}$$

Definition 8 : *By a higher order Prelec probability weighting function, we mean a probability weighting function given by (5.5) and (5.8). If $a_n \neq 0$ but $a_k = 0$ for all $k > n$, then we say $w(p)$ is a Prelec probability weighting function of order n. In particular, the original Prelec probability weighting function is of order 1.*

We give an example of an infinite order Prelec function. Take $a_0 = -\ln\beta$, $\beta > 0$, $a_1 = \alpha > 0$ and, for $k \geq 1$, $a_{2k} = 0$, $a_{2k+1} \geq 0$, we get that (5.5) and (5.7) define a strictly increasing function, provided the series is convergent. An easy way to guarantee convergence is to take $a_{2k+1} = \frac{a}{(2k+1)!}$. Then, for any $p \in (0,1)$, the series in (5.7) converges (absolutely and uniformly) to $-\ln\beta + \frac{1}{2}\alpha\left(e^{-\ln(-\ln p)} - e^{\ln(-\ln p)}\right) = -\ln\beta + \alpha\sinh\left(-\ln\left(-\ln p\right)\right)$.[14] To get the 'right shape', we take $a \in (0,1)$. For probabilities in the middle range, $\sinh\left(-\ln\left(-\ln p\right)\right) \simeq -\ln\left(-\ln p\right)$, hence this function is a good approximation to the (first order) Prelec function for such probabilities.

Definition 9 *: By the hyperbolic Prelec function (HP), we mean the probability weighting function defined by*

$$w\left(0\right) = 0, w\left(1\right) = 1, \tag{5.9}$$

$$-\ln\left(-\ln w\right) = -\ln\beta + \alpha\sinh\left(-\ln\left(-\ln p\right)\right), 0 < p < 1, 0 < \alpha < 1, \beta > 0. \tag{5.10}$$

Note that (5.9) and (5.10) define a two-parameter family of functions. Hence, a hyperbolic Prelec function is just as parsimonious as the (1st order) Prelec function (Definition 6). Figure 5.2 is a graph of (5.9) and (5.10) for $\beta = 1$ and $\alpha = \frac{1}{2}$. Its shape is similar to that of a Prelec weighting function. However, the essential difference from the latter is shown in Figure 5.3; the hyperbolic Prelec function underweights (not overweights) extremely low probabilities. This is proved formally in proposition 5 below.

Proposition 5 *: Let $w\left(p\right)$ be the hyperbolic Prelec probability weighting function. Then*
(a) $w : [0,1] \to [0,1]$ is continuous and strictly increasing; w is C^{∞} on $(0,1)$,
(b) $\lim\limits_{p\to 0}\frac{w(p)}{p^{\gamma}} = 0$.

Proof: Part (a) is obvious from the properties of $\ln p$ and $\sinh x$. To prove part (b), let

$$\ln\frac{w}{p^{\gamma}} = e^{x}\left[\gamma - \beta e^{\alpha x\left[\frac{\sinh x}{x} - 1\right]}\right], \text{ where } x = \ln\left(-\ln p\right). \tag{5.11}$$

Let $p \to 0$. Then $x \to \infty$ and $\frac{\sinh x}{x} \to \infty$. Thus, $e^{x} \to \infty$ and $\gamma - \beta e^{\alpha x\left[\frac{\sinh x}{x} - 1\right]} \to -\infty$. Hence, $\ln\frac{w}{p^{\gamma}} \to -\infty$ and, thus, $\frac{w}{p^{\gamma}} \to 0$. ■

A similar proposition holds for Prelec probability weighting functions of order $n > 1$. See Proposition 3 of al-Nowaihi and Dhami (2006).

[14] The *hyperbolic sin* function is defined as

$$\sinh x = \frac{1}{2}\left(e^{x} - e^{-x}\right) = \Sigma_{k=0}^{\infty}\frac{x^{2k+1}}{(2k+1)!}.$$

Hence,

$$\frac{\sinh x}{x} = 1 + \Sigma_{k=1}^{\infty}\frac{x^{2k}}{(2k+1)!},$$

and, hence,

$$\lim_{x\to\infty}\frac{\sinh x}{x} = \infty.$$

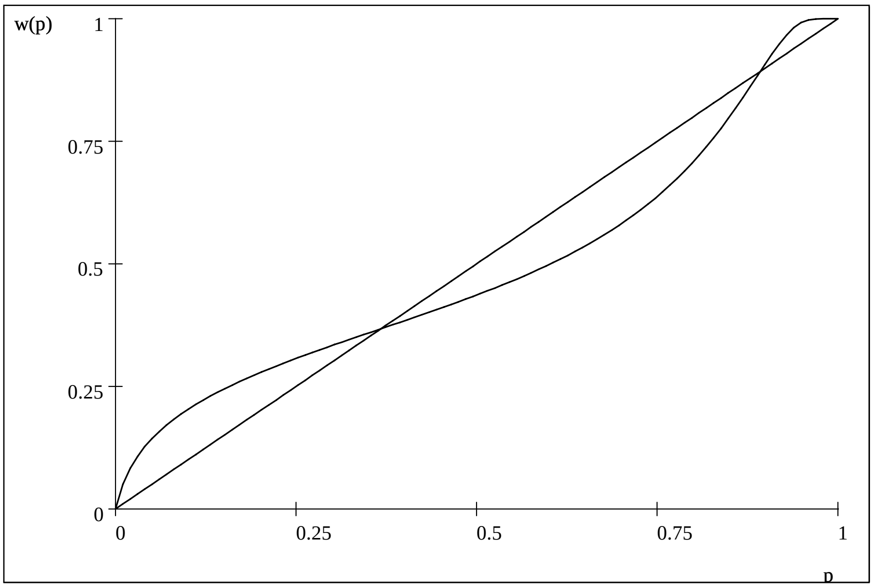

Figure 5.2: **A Hyperbolic Prelec Function**

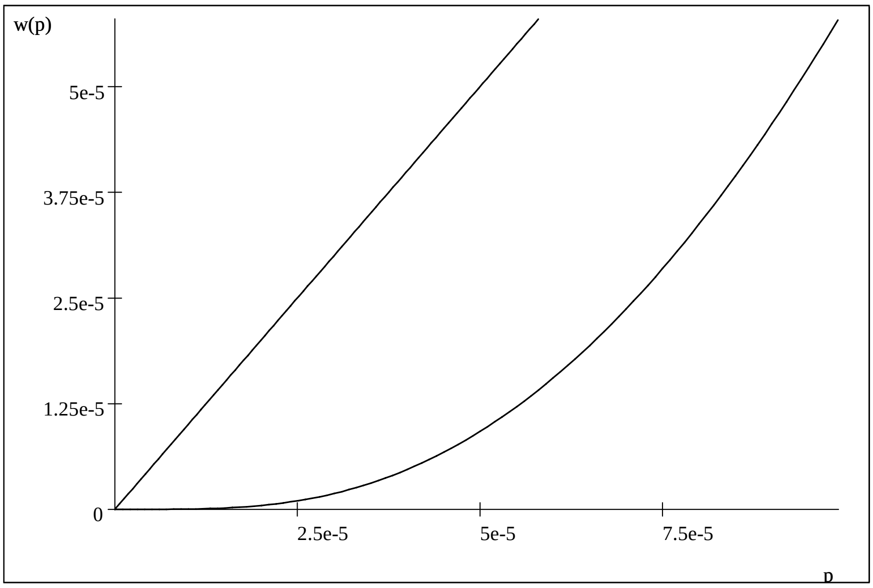

Figure 5.3: **A Hyperbolic Prelec Function For Low Probabilities**

In section 7 we compare the behavior of the decision maker when she uses, respectively, a probability weighting function of order 1 (which is just the standard Prelec function) and of order greater than one. We show that her behavior differs significantly between the two cases.

Propositions 4 and 5(b) distinguish the new probability weighting functions from the standard probability weighting functions and, as we shall demonstrate below, are critical to understanding differences in the results between the two sets of models.

6. Rank dependent expected utility theory (RDU)

We now model the behavior of an individual using RDU. Appendix-A outlines the basic concepts in RDU that we shall require in this paper. The payoff from not committing the criminal activity under RDU is $U_{NC} = u(y_0)$. Under RDU, the individual's payoff from the criminal activity is

$$EU_C = \pi u(y_1 - F) + (1 - \pi) u(y_1), \tag{6.1}$$

where

$$\pi = w(p). \tag{6.2}$$

Hence, the no-crime condition (NCC), $EU_C \leq U_{NC}$, gives

$$\pi u(y_1 - F) + (1 - \pi) u(y_1) \leq u(y_0) \tag{6.3}$$

After some simple algebra, the NCC (6.3) becomes

$$p \geq w^{-1}\left[\frac{1 - \frac{u(y_0)}{u(y_1)}}{1 - \frac{u(y_1 - F)}{u(y_1)}}\right]. \tag{6.4}$$

As before, let $F_{\max}$ be the maximum fine that can be imposed and let the corresponding level of utility be

$$u_{\min} = u(y_1 - F_{\max}), \tag{6.5}$$

and let

$$p_{\min} = w^{-1}\left[\frac{1 - \frac{u(y_0)}{u(y_1)}}{1 - \frac{u_{\min}}{u(y_1)}}\right]. \tag{6.6}$$

Then, from (6.4)-(6.6), the NCC becomes

$$p \geq p_{\min}. \tag{6.7}$$

Proposition 6 : *Under rank dependent expected utility theory, if the utility function is unbounded below, then for any probability of punishment $p > 0$, no matter how small, crime can be deterred by a sufficiently severe punishment, F.*

Proof: Follows from (6.5), (6.6) and (6.7). ■

Proposition 7 : *Under rank dependent expected utility theory and for the probability weighting function* $w(p)$*:*
(a) If the utility function, u*, is concave, then, given any probability of detection and conviction,* $p > 0$*, no matter how small, crime can be deterred by choosing the punishment* $F(p) = \frac{b}{w(p)}$*.*
(b) If, in addition, $C_p > \frac{w'(p)}{w(p)} C_F$ *(Definition 3 and Lemma 1), then reducing* p *reduces the total social cost of crime and law enforcement,* $T(p, F)$*.*

Proof: Similar to the proof of Proposition 1, except for the following points: (a) $F(p) = \frac{b}{w(p)}$ (instead of $F(p) = \frac{b}{p}$) and (b) $F_p = \frac{pw'(p)}{w(p)}$ (instead of $F_p = 1$). ■

7. Prospect theory (PT)

We now turn to the modelling of crime under cumulative prospect theory (CP). Appendix-A outlines the basic concepts in CP that we shall require in this paper. 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

$$V_{NC} = v(y_0 - y_r). \tag{7.1}$$

Assume that if an individual commits a crime, and is caught, then the outcome is in the domain of losses; while if he commits a crime and is not caught, then the outcome is in the domain of gains. Then we have only one outcome in the domain of losses and one outcome in the domain of gains. So the decision weights are simply $w^-(p)$ and $w^+(1-p)$.[15] Then the individual's payoff from committing a crime is

$$V_C = w^-(p)\, v(y_1 - F - y_R) + w^+(1-p)\, v(y_1 - y_R). \tag{7.2}$$

Definition 10 *(Elation): We shall refer to* $v(y_1 - y_R)$ *as the elation from committing a crime and getting away with it.*

The *NCC* is

$$V_C \leq V_{NC}. \tag{7.3}$$

Substituting from (7.1) and (7.2) into (7.3), the *NCC* becomes

$$w^-(p)\, v(y_1 - F - y_R) + w^+(1-p)\, v(y_1 - y_R) \leq v(y_0 - y_r). \tag{7.4}$$

[15] In this case, the cumulative prospect theory of Tversky and Kahneman (1992) reduces to the original prospect theory of Kahneman and Tversky (1979).

7.1. Fixed reference points

The *NCC* in (7.4) depends on the two reference points, y_r and y_R. We now explore the implications of a prospect theory approach to crime with alternative specifications of the reference incomes. In this section we assume that the reference points are fixed. In section 7.3 the reference point for each activity will be the expected income from that activity. In appendix-B below the expected income from crime is the reference point for both activities.

Proposition 8 *: Assume that the reference points, y_r and y_R, are fixed. If Unboundedly severe punishments are available and the value function is unbounded below, then, under prospect theory, crime can be deterred with arbitrarily low probabilities of detection and conviction.*

Proof: Let the probability of detection and conviction be $p > 0$. Then the *NCC* (7.4) can be satisfied by taking the punishment, F, to be sufficiently large. ■

Hence, the Becker paradox reemerges under prospect theory if (1) the reference points are fixed, (2) punishments of unlimited severity are available, (3) the value function is unbounded below and (4) it is economically worthwhile to deter crime.

7.2. The value function is a power function

To facilitate further development, we adopt the power form of the value function, which has become almost standard in applications.

Definition 11 *(Tversky and Kahneman, 1992, the analogue of CRRA in EU) The value function satisfies preference homogeneity if $(x, p) \sim y \Rightarrow (kx, p) \sim ky$, where $x, y, p, k \in \mathbf{R}$, $p \in [0, 1]$, $k > 0$ and $\sim$ is the indifference relation.*

Proposition 9 *: Under preference homogeneity the value function takes the simple form[16]:*
$v(x) = x^{\gamma}, x \geq 0,$
$v(x) = -\theta(-x)^{\gamma}, x < 0,$
where γ, θ are constants satisfying $0 < \gamma < 1$, $\theta > 1$.

Experimental evidence from Tversky and Kahneman (1992) suggests that $\gamma \cong 0.88$, $\theta \cong 2.25$.

We shall adopt the value function

$$\begin{aligned} v(x) &= x^{\gamma},\, x \geq 0, \\ v(x) &= -\theta(-x)^{\gamma},\, x < 0, \\ \text{where } \theta &> 1,\, 0 < \gamma < 1. \end{aligned} \tag{7.5}$$

[16] Tversky and Kahneman (1992) state, without proof, that preference homogeneity is necessary and sufficient for the value function to take the form $v(x) = x^{\gamma_+}, x \geq 0$; $v(x) = -\theta(-x)^{\gamma_-}, x < 0$. Loss aversion, however, implies that $\gamma_+ = \gamma_-$. See al-Nowaihi, Bradley and Dhami (2006).

7.3. The reference point for each activity is the expected income from that activity

It would seem more plausible to assume that the reference point for each activity is the expected income arising from that activity, i.e., the reference incomes from the honest and the criminal activity are respectively given by:

$$y_r = y_0; \; y_R = y_1 - pF. \tag{7.6}$$

Hence, as before, $y_1 - y_R = pF$ and $y_1 - F - y_R = -(1-p)F$. However, $y_0 - y_r = 0$. Hence, the *NCC* (7.4) becomes

$$w^-(p)\, v(-(1-p)F) + w^+(1-p)\, v(pF) \leq 0. \tag{7.7}$$

For the power function form (7.5) the *NCC* (7.7) becomes:

$$-\theta (1-p)^\gamma F^\gamma w^-(p) + p^\gamma F^\gamma w^+(1-p) \leq 0. \tag{7.8}$$

For $F > 0$, this simplifies to

$$\frac{w^-(p)}{p^\gamma} \geq \frac{w^+(1-p)}{\theta (1-p)^\gamma}. \tag{7.9}$$

Proposition 10 : *As the probability of detection approaches zero, a prospect theory decision maker facing a strictly positive punishment, i.e.* $F > 0$*, who satisfies preference homogeneity and whose reference points are given by (7.6), does not engage in the criminal activity if the probability weighting function satisfies the condition*

$$\lim_{p \to 0} \frac{w^-(p)}{p^\gamma} > \frac{1}{\theta}. \tag{7.10}$$

On the other hand, the same individual engages in crime if

$$\lim_{p \to 0} \frac{w^-(p)}{p^\gamma} < \frac{1}{\theta}. \tag{7.11}$$

Proof: If (7.10) holds, then the *NCC* (7.9) will hold with strict inequality in some non-empty interval $(0, p_1)$. Hence, no crime will occur if $p \in (0, p_1)$. If (7.11) holds, then the converse of the *NCC* (7.9) holds with strict inequality in some non-empty interval $(0, p_2)$. Hence, for punishment to deter in this case, we must have $p > p_2$. ■

Proposition 11 *If the reference point for each activity, crime and no crime, is taken to be the expected income from that activity, then the Becker paradox reemerges under prospect theory if we use any of the "standard" probability weighting functions.*

Proof: For any of the standard Probability weighting functions, $\lim_{p\to 0} \frac{w^-(p)}{p^\gamma} = \infty$ (Proposition 3). From Proposition 10, it follows that, for some non-empty interval $(0, p_1)$, no crime will occur if $p \in (0, p_1)$. Hence, $D(p, F) = 0$ for $p \in (0, p_1)$. Now note that the only condition on F is that $F > 0$ (in particular F may be less than b). Since $\frac{\partial}{\partial p} C(p, F) > 0$, $\frac{\partial}{\partial F} C(p, F) \geq 0$ and $C(0, 0) = 0$ (and C is continuous), it follows that $C(p, F)$ can be made as close to 0 as we like by taking p and F sufficiently small. Hence, the total social cost of crime, $T(p, F)$, can be made as close to 0 as we like by choosing p and F to be positive but sufficiently small. ■

We now state the key proposition in the paper, Proposition 12 below.

Proposition 12 *If the reference point for each activity, crime and no-crime, is taken to be the expected income from that activity, then the Becker paradox is solved if we use the identity probability weighting function, the Rieger-Wang probability weighting function or a higher order Prelec probability weighting function.*

Proof: Since $\lim_{p\to 0} \frac{w^-(p)}{p^\gamma} = \lim_{p\to 0} \frac{p}{p^\gamma} = 0$ for the identity probability weighting function, $\lim_{p\to 0} \frac{w^-(p)}{p^\gamma} = 0$ for the Rieger-Wang function (Proposition 4), the hyperbolic Prelec function (Proposition 5) and for Prelec function of order $n > 1$ (al-Nowaihi and Dhami (2006), Proposition 3c)) it follows, from Proposition 10, that, for some non-empty interval $(0, p_2)$, no level of punishment, F, no matter how large, will deter crime, if $p \in (0, p_2)$. ■

7.4. Discussion: Becker's proposition under CP

We have found that Becker's paradox reemerges under prospect theory in each of the following cases:

1. The reference points are fixed (independent of p and F) and the value function is unbounded below.

2. The value function is the power function and the reference point for both activities is the expected income from crime (this result is shown in appendix-B).

3. The value function is the power function, the reference point for each activity is the expected income from that activity and the probability weighting function is any of the standard probability weighting functions.

On the other hand, we have found that Becker's proposition does not hold under prospect theory, and hence Becker's paradox is resolved, if the following hold jointly:

1. The value function is of the power form.

2. The reference point for each activity, crime and no crime, is the expected income from that activity.

3. The probability weighting function is the identity function, the Rieger-Wang function or a Prelec function of order greater than 1.

The intuition behind these results is as follows. If the reference points are fixed, then so also are the gains from not engaging in crime or engaging in crime but not being caught. Thus, crime can be deterred by imposing sufficiently high punishment if caught. On the other hand, if the reference point for criminal activity is the expected income from that activity, then increasing the punishment, not only increases the loss (or pain) from being caught, but also increases the gain (or elation) from committing the crime and not being caught. This increase in gain can compensate for the increase in loss, if there is no similar increase in gain from not committing the crime, and if the low probability (of being caught) is not excessively overweighted.

8. Conclusions

The Becker proposition summarized eloquently in Kolm's (1973) phrase "hang offenders with probability zero" has a long and distinguished tradition in the crime and punishment literature. However it is often noted that it does not hold empirically. This, we called the Becker paradox. A convenient fix proposed in the literature is to close the model with an arbitrary upper bound on the amount of fines that can be imposed. This is, of course, reasonable only if punishments were restricted to purely monetary terms. However, if one allows for the use of non-monetary punishments, then the Becker paradox reemerges.

A sizeable literature addresses this paradox in an expected utility (EU) framework. We argued, in section 4, that it is very difficult to explain the evidence on the basis of EU. For these reasons we reexamined the Becker paradox from the perspective of behavioral economics.

We found that the Becker paradox reemerges under rank dependent expected utility theory. However, we are able to explain the Becker paradox within prospect theory by adopting (1) a power function specification for the value function, (2) a probability function that does not excessively overweight low probabilities and (3) choosing the reference point for each activity (crime or no crime) to be the expected income under that activity. The intuition behind this result is as follows. Increasing the punishment has two effects: (1) it increases the loss if caught and (2) it *increases the sense of elation obtained from engaging in the dangerous activity and not being caught.* If the probability weighting function does not excessively overweight low probabilities, the effect of the latter can overcome the effect of the former.

9. Appendix-A

9.1. A note on Rank Dependent Expected Utility (RDU)

Rank dependent expected utility (RDU) was discovered by a number of independent researchers beginning with Quiggin (1979). The major contribution of Quiggin is that he, for the first time, provided a coherent theory of behavior with non-linear weighting of probabilities. He did this by proposing that it is not individual probabilities that should be transformed (which is what all earlier researchers had done) but cumulative probabilities. We shall now explain.

Consider the lottery $(x_1, x_2, ..., x_n;\ p_1, p_2, ..., p_n)$ that pays x_i with probability p_i, where $x_1 \leq x_2 \leq ... \leq x_n$. Let w be a probability weighting function. For RDU, the *decision weights*, π_j, are defined by $\pi_j = w\left(\Sigma_{i=1}^{j} p_i\right) - w\left(\Sigma_{i=1}^{j-1} p_i\right)$. The decision maker's *rank dependent expected utility* is given by

$$U(x_1, x_2, ..., x_n;\ p_1, p_2, ..., p_n) = \Sigma_{j=1}^{n} \pi_j u(x_j). \tag{9.1}$$

The success of RDU stems from the following facts. (1) All objects of interest such as random variables, various concepts of stochastic dominance, etc., are defined in terms of *cumulative* probability distributions, not individual probabilities. (2) $w(p)$ is *not* a probability measure unless w is the identity transformation. (3) On the other hand, if $F(x)$ is a cumulative distribution function, then so is[17] $\Phi(x) = w(F(x))$. (4) The point transformation approach cannot be extended to continuous distributions because, necessarily, the probability, p, of each point is zero and, hence, the transformed 'probabilities', $w(p)$, are all zero. It follows that all concepts of stochastic dominance in EU carry over to RDU. In particular, under RDU, a decision maker will never choose a first order stochastically dominated prospect. In fact, we can view RDU as simply EU applied to a transform of the cumulative probability distribution. This is formalized by Quiggin's *correspondence principle* (Quiggin, 1993). This is clearly of tremendous utility, since the whole machinery of risk analysis in EU can be transferred to RDU. See Quiggin (1993) for further details.

9.2. A note on Cumulative Prospect Theory (CP)

In prospect theory, carriers of utility are not final bundles of goods (as under *EU* and *RDU*) but deviations of these from a reference point. In prospect theory, the utility function is often called a *value function*, which is defined below. Our definition of a value function is as in Kahneman and Tversky (1979).

[17] If $F(x)$ is a probability distribution, then it is non-decreasing, right-continuous, $F(-\infty) = 0$ and $F(\infty) = 1$. It follows from the properties of w that Φ also has these properties and, hence, is also a probability distribution. Alternatively, we could define $\Psi(x) = 1 - w(1 - F(x))$. The apparently more complex Ψ is sometimes slightly more convenient.

Definition 12 *: For simplicity of exposition, assume one good, 'wealth', y. The individual has reference wealth given by y_r. Let $x = y - y_r$. A value function is a mapping $v : \mathbf{R} \rightarrow \mathbf{R}$ that satisfies:*
1. $v(x)$ is continuous,
2. $v(x)$ is strictly increasing,
3. $v(0) = 0$ (reference dependence),
4. $v(x)$ is concave for $x \geq 0$ and $v(x)$ is convex for $x \leq 0$ (declining sensitivity),
5. $|v(-x)| > v(x)$, $x > 0$ (loss aversion).

These points are illustrated in the following figure.

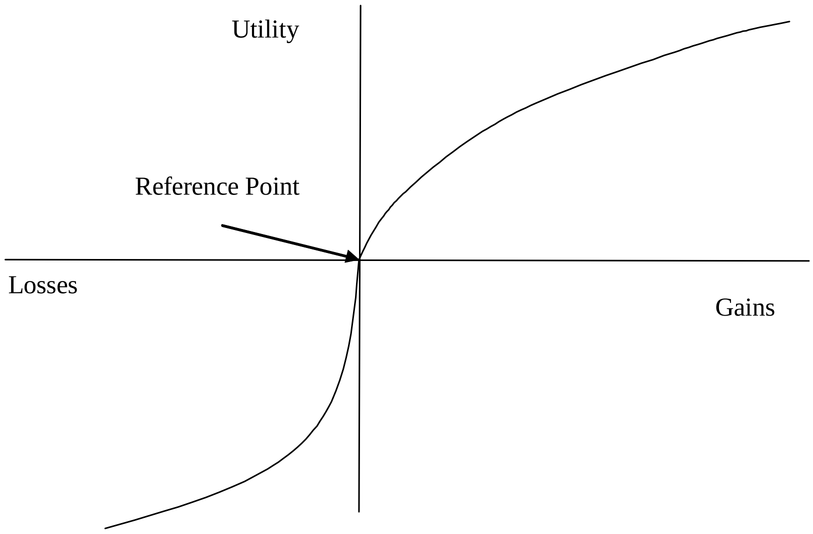

Figure 9.1: **Preferences Under Prospect Theory**

Consider the lottery

$$(x_{-m}, x_{-m+1}, ..., x_{-1}, x_0, x_1, x_2, ..., x_n \ ; p_{-m}, p_{-m+1}, ..., p_{-1}, p_0, \ p_1, p_2, ..., p_n)$$

that pays x_i with probability p_i, where

$$x_{-m} \leq x_{-m+1} \leq ... \leq x_{-1} \leq x_0 = 0 \leq x_1 \leq x_2 \leq ... \leq x_n.$$

To get the decision weights for losses, apply the usual RDU calculations, starting at the negative extreme and use a probability weighting function, w^-. The *decision weights for losses*, π_j, are defined by $\pi_j = w^- \left(\Sigma_{i=-m}^{j} p_i\right) - w^- \left(\Sigma_{i=-m}^{j-1} p_i\right)$; $j = -m, -m+1, ..., -1$. For gains, start at the positive extreme, and use a probability weighting function, w^+. The *decision weights for gains*, π_j, are defined by $\pi_j = w^+ \left(\Sigma_{k=j}^{n} p_k\right) - w^+ \left(\Sigma_{k=j+1}^{n} p_i\right)$;

$j = 1, 2, ..., n$.[18] The value of this prospect is then given by

$$V = \Sigma_{i=-m}^{n} \pi_i v(x_i),$$

which is the objective function that a decision maker using prospect theory maximizes.

In EU, attitude to risk is entirely determined by the shape of the utility function. The decision maker is, respectively, risk averse, risk neutral or risk seeking according to the utility function being concave, linear or convex. In CP the situation is more complex. Attitude to risk is the result of the interaction of the shape of the value function, loss aversion and non-linear weighting of probabilities. This allows CP to explain much behavior under risk that is difficult to explain with EU or RDU. For example, the widely observed phenomenon that people simultaneously gamble and insure can be explained by CP as follows. Overweighting of low probabilities makes a high reward with low probability attractive, despite the concavity of the value function for gains. Overweighting of low probabilities also makes insurance against a large loss with a low probability attractive, despite the convexity of the value function for losses. For further examples, see Tversky and Kahneman (1992), Benartzi and Thaler (1995) and al-Nowaihi and Dhami (2006).

10. Appendix-B

10.1. Both reference points are expected income from crime

Assume now that the reference point for both activities is the expected income arising from crime, i.e.,

$$y_r = y_R = y_1 - pF. \tag{10.1}$$

Then

$$y_0 - y_r = pF - b,\ y_1 - y_R = pF,\ y_1 - F - y_R = -(1-p)F. \tag{10.2}$$

The *NCC* (7.4) becomes

$$w^-(p)\, v(-(1-p)F) + w^+(1-p)\, v(pF) \leq v(pF - b). \tag{10.3}$$

For the power function form (7.5), and for $pF \geq b$, the *NCC* (10.3) becomes:

$$-\theta(1-p)^\gamma F^\gamma w^-(p) + p^\gamma F^\gamma w^+(1-p) \leq (pF-b)^\gamma,\ \text{for } pF \geq b. \tag{10.4}$$

For $F > 0$, this simplifies to

$$p^\gamma w^+(1-p) - \theta(1-p)^\gamma w^-(p) \leq \left(p - \frac{b}{F}\right)^\gamma,\ \text{for } F > 0 \text{ and } pF \geq b. \tag{10.5}$$

[18] To quote from Prelec (1998, last line of Appendix A): "CPT reduces to RDU if $w^-(p) = 1 - w^+(1-p)$. Empirically, however, one observes $w^+(p) = w^-(p)$." Our results do not depend on this. Note that the decision weights do not, necessarily, add up to 1. Since $v(0) = 0$, 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 $\Sigma_{i=-m}^{n}\pi_i = 1$.

We now construct a punishment function as follows:

$$\begin{aligned} p^{\gamma}w^{+}(1-p) &< \theta(1-p)^{\gamma}w^{-}(p) \Rightarrow F(p) = \frac{b}{p}, \\ p^{\gamma}w^{+}(1-p) &\geq \theta(1-p)^{\gamma}w^{-}(p) \Rightarrow F(p) = \frac{b}{p-\left[p^{\gamma}w^{+}(1-p)-\theta(1-p)^{\gamma}w^{-}(p)\right]^{\frac{1}{\gamma}}}. \end{aligned} \tag{10.6}$$

Note that the right hand side of (10.6) is well defined since $p^{\gamma}w^{+}(1-p)-\theta(1-p)^{\gamma}w^{-}(p) \geq 0$ and $p > \left[p^{\gamma}w^{+}(1-p)-\theta(1-p)^{\gamma}w^{-}(p)\right]^{\frac{1}{\gamma}}$.

Proposition 13 : *Assume that both reference points are expected income from crime and that the value function is the power function. Then, under prospect theory, and for the probability weighting function $w(p)$, given any probability of detection and conviction, $p > 0$, no matter how small, crime can be deterred by choosing the punishment, $F(p)$, given by (10.6).*

Proof: First check that $F(p)$, as given by (10.6), satisfies $pF \geq b$. Then verify that the *NCC* (10.5) is satisfied. ■

Thus, the Becker paradox reemerges under prospect theory if (1) both reference points are the expected income from crime, (2) punishments of unlimited severity are available, (3) the value function is a power function and (4) it is economically worthwhile to deter crime.

References

[1] al-Nowaihi, Ali, Ian Bradley and Sanjit Dhami (2006) 'Preference homogeneity and the power function in prospect theory', *mimeo. University of Leicester.*

[2] Allingham, M.G., Sandmo A., 1972. Income tax evasion: A theoretical analysis. Journal of Public Economics 1, 323-38.

[3] al-Nowaihi, Ali and Sanjit Dhami (forthcoming) "A Simple Derivation of the Prelec Probability Weighting Function" *Journal of Mathematical Psychology.*

[4] al-Nowaihi, Ali and Sanjit Dhami (2006) "Insurance and Probability Weighting Functions", *University of Leicester Discussion Paper* No. 09/2006.

[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 (2001) "The response of criminals and noncriminals to fines" *Journal of Law and Economics*, Vol. XLVII, 1-17.

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

[9] Benartzi S, Thaler R. H., 1995. Myopic loss aversion and the equity premium puzzle. Quarterly Journal of Economics 110, 73-92.

[10] Blavatskyy, P.R., 2004. Back to the St. Petersburg paradox? CERGE-EI working paper series No. 227.

[11] Bowles, Samuel (2004) *Microeconomics: Behavior, Institutions, and Evolution* (Princeton University Press: New York).

[12] Camerer C., 2000. Prospect theory in the wild: Evidence from the field. In: Kahneman D, Tversky A., Choices, values and frames, Cambridge University Press: Cambridge.

[13] Colman, A. M. (1995). Prisoner's Dilemma, Chicken, and mixed-strategy evolutionary equilibria. Behavioral and Brain Sciences, 18, 550-551.

[14] Dhami, Sanjit and Ali al-Nowaihi (forthcoming) Why Do People Pay Taxes? Prospect Theory Versus Expected Utility Theory, *Journal of Economic Behavior and Organization.*

[15] Friedman, David (1999) "Why Not Hang Them All: The Virtues of Inefficient Punishment," *Journal of Political Economy*, 107, S259-S269.

[16] Gonzalez R. and Wu G. (1999). On the shape of the probability weighting function. *Cognitive Psychology* 38. 129-166.

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

[18] Kahneman D. and Tversky A. (2000) Choices, values and frames (Cambridge University Press: Cambridge).

[19] Kolm, Serge-Christophe (1973) "A Note on Optimum Tax Evasion," *Journal of Public Economics*, 2, 265-270.

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

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

[22] Luce R. D. (2001). Reduction invariance and Prelec's weighting functions. *Journal of Mathematical Psychology* 45. 167-179.

[23] McCaffery, Edward J. and Slemrod, Joel B. (2004) "Toward an Agenda for Behavioral Public Finance" USC Law and Economics Research Paper No. 04-25; and USC CLEO Research Paper No. C04-22.

[24] Polinsky, Mitchell and Steven Shavell (2000) "The Economic Theory of Public Enforcement of Law" *Journal of Economic Literature,* Vol. XXXVIII, 45-76.

[25] Prelec D. (1998) "The probability weighting function," *Econometrica* 60. 497-528.

[26] Quiggin, J. (1979) "The theory of the multiproduct firm under uncertainty" Australian National University, B.Ec. Honors thesis.

[27] Quiggin, J. (1993). *Generalized Expected Utility.* Kluwer.

[28] 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.

[29] Rasmusen, Eric, 1994. *Games and Information. An Introduction to Game Theory* (2nd edition). Cambridge: Blackwell.

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

[31] 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.

[32] Starmer C., 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. Journal of Economic Literature 38, 332-82.

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

[34] von Neumann, J. and Morgenstern, O.(1944) *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, New Jersey (second edition, 1947, third edition, 1953).

[35] Yitzhaki S., 1974. A note on income tax evasion: A theoretical analysis. Journal of Public Economics 3, 201-202.