University of Leicester

DEPARTMENT OF ECONOMICS

The Behavioral Economics of Crime and Punishment

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

Working Paper No. 10/14
Updated July 2010

The behavioral economics of crime and punishment

Sanjit Dhami* **Ali al-Nowaihi†**

20 July 2010

Abstract

The celebrated *Becker proposition* (BP) states that it is optimal to impose the severest possible punishment (to maintain effective deterrence) at the lowest possible probability of detection (to economize on enforcement costs). However, the BP is not consistent with the evidence. This inconsistency is known as the *Becker paradox*. In fact, the BP is a general result that applies to all low-probability events that lead to 'unbounded loss' of utility. Hence, it is applicable to a wide class of problems in economics. We clarify the BP and its welfare implications under *expected utility*, which remains the favoured framework. We argue that none of the proposed explanations of the Becker paradox is satisfactory. We show that the BP also holds under *rank dependent expected utility* and *cumulative prospect theory*, the two main alternatives to expected utility. We show that *composite prospect theory* (CCP), of al-Nowaihi and Dhami (2010a), can resolve the Becker paradox. Our paper opens the way for incorporating *non-expected utility theories* into the economic analysis of criminal activity.

Keywords: Crime and punishment; non-linear weighting of probabilities; Cumulative prospect theory; Prelec and composite Prelec probability weighting functions; Composite prospect theory; Punishment functions.

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

*Corresponding author. 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), Quoted from Becker (1968, footnote 12).
"*... 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.*" Becker (1968).
"*The absence of sustained and comprehensive economic analysis of legal rules from a perspective informed by insights about actual human behavior makes for a significant contrast with many other fields of economics, where such "behavioral" analysis has become relatively common.*" Jolls, Sunstein, and Thaler (2000).

1. Introduction

In a seminal contribution, Becker (1968) opened the way to a rigorous formal economics analysis of crime. Becker (1968) showed 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 simple and compelling. 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 (monetary and non-monetary), the deterrence effect of the punishment is maintained.

The Becker proposition takes a particularly stark form if the decision maker follows *expected utility theory* (EU) and if we add two assumptions: (1) Risk neutrality or risk aversion. (2) The availability of infinitely severe (monetary and non-monetary) punishments, e.g., capital punishment. With these extra assumptions, the Becker proposition implies that crime would be deterred completely, however small the probability of detection and conviction. Kolm (1973) phrased the Becker proposition as *hang offenders with probability zero.*

1.1. The class of Becker paradoxes

Empirical evidence shows that the Becker proposition does not hold for all individuals. 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, in contradiction to the predictions of 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.

Remark 1 *(Becker-type punishments, Becker paradox) Consider the case where a decision maker's action results in a large loss with a small probability. We call the resulting loss a Becker-type punishments. While Becker-type punishments might deter crime for many, the fact that they do not do so for all is referred to as the Becker paradox.*

Remark 2 *(The class of Becker paradoxes): Our main focus will be on crime and punishment. However, the Becker paradox is quite general. Let A be the set of all acts for which a decision maker faces a small probability of a very large loss* (Becker-type punishment). *Let $\widetilde{A} \subset A$ be the set for which EU theory predicts that the individual will be dissuaded from such acts but observed behavior is to the contrary. Then each act in $\widetilde{A}$ belongs to the class of Becker paradoxes. Examples include not buying insurance against low probability natural hazards, jumping red traffic lights, driving and talking on mobile phones, not wearing seat belts in moving vehicles, etc.; see Section 5.1 for the details.*

1.2. Failure of expected utility theory (EU)

In Section 4 we consider 9 proposed explanations for the Becker paradox, based on adding the following auxiliary assumptions to EU. Risk seeking behavior on the part of offenders, the ability to avoid severe fines by declaring bankruptcy, the need for differential punishments, type-I and type-II errors in conviction, rent seeking behavior in the presence of severe punishments, abhorrence of severe punishments, objectives other than deterrence, including putting the welfare of offenders in the social welfare function, and the psychological traits of offenders. We show, in Section 5, that none of these proposed explanations suffice, singly or jointly.

Consider, for instance, the act of 'running a red traffic light,' a member of the class of Becker paradoxes. In this case, the *low-probability potentially fatal-punishment* (which is a Becker-type punishment) is *self-inflicted*. A little reflection should convince the reader that this rules out most of the proposed explanations given for the Becker paradox, above (see Section 5.1 for the details).

1.3. Some stylized facts on non-linear weighting of probabilities

Given the nature of the Becker paradoxes (see Remark 2), the attitudes to low probability events are crucial. Extensive evidence suggests the following stylized facts.

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]

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

S2. For probabilities close to the endpoints of $[0,1]$:

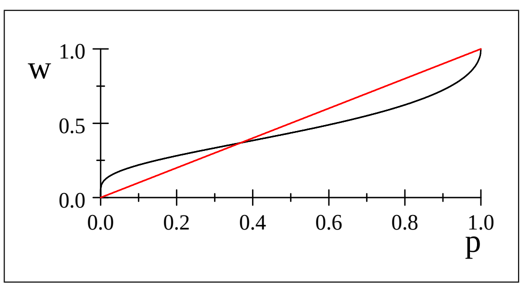
S2a: A fraction $\mu \in [0,1]$ of decision makers (i) ignore events of extremely low probability and/or (ii) treat extremely high probability events as certain.[2]

S2b: A fraction $1-\mu$ places great salience on the size of the outcomes (particularly losses), even if the probabilities are very small or very large.

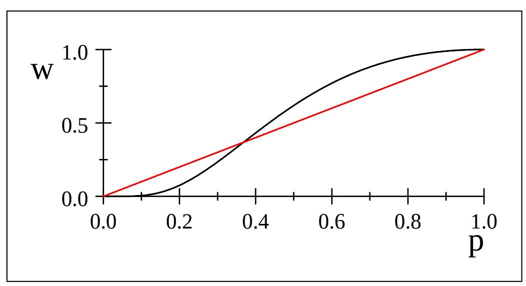
The evidence against EU is now overwhelming.[3] In particular, EU does not take account of S1, S2a or S2b. By contrast, S1 is incorporated into almost all *non-linear weighting models*, including *rank dependent utility* (RDU), *prospect theory* (PT) and *cumulative prospect theory* (CP).

Non-expected utility (non-EU) theories postulate a *probability weighting function*, $w(p) : [0,1] \to [0,1]$, that captures the subjective weight placed by decision makers on the objective probability, p. Such theories, e.g., RDU and CP, account for S1 by incorporating a $w(p)$ function that overweights low probabilities but underweights high probabilities. Such non-EU theories can account for S1 and S2b, but they cannot account for S2a.

A popular probability weighting function is the *Prelec (1998) function*: $w(p) = e^{-\beta(-\ln p)^{\alpha}}$, $\alpha > 0$, $\beta > 0$, which is parsimonious and has an axiomatic foundation. If $\alpha < 1$, then this function is *inverse-S-shaped*. It overweights low probabilities and underweights high probabilities, so it conforms to S1 and S2b but not S2a. All non-EU theories, particularly RDU and CP, assume a weighting function of this form. However, for $\alpha > 1$ the Prelec function is *S-shaped*; it is in conflict with S1 and S2b but respects S2a. The two cases $\alpha < 1$ and $\alpha > 1$ are plotted below for the case $\beta = 1$ and respectively $\alpha = 0.5$ and $\alpha = 2$.



The case $\alpha < 1$ ($\alpha = 0.5$, $\beta = 1$)



The case $\alpha > 1$ ($\alpha = 2$, $\beta = 1$)

Remark 3 : *To distinguish between the two cases $\alpha < 1$ and $\alpha > 1$, we call the former ($\alpha < 1$) the standard Prelec function. This function becomes extremely steep as $p \to 0$ and as $p \to 1$. It infinitely-overweights infinitesimal probabilities in the sense that* $\lim_{p\to 0} \frac{w(p)}{p} = \infty$

[2] The actual fraction μ could be affected by many factors; see the discussion in subsection 5.4. In the context of buying insurance against low probability natural hazards, in one set of experiments (the urn experiments, chapter 7), Kunreuther et al. (1978) find that $\mu = 0.8$.

[3] See, e.g., Kahneman and Tversky (2000), Starmer (2000), and Dhami and al-Nowaihi (2007, 2010b).

and infinitely-underweights near-one probabilities in the sense that $\lim_{p\to 1}\frac{1-w(p)}{1-p}=\infty$*. Most weighting functions in use in RDU and CP have this property. We shall call these the standard probability weighting functions.*

As compared to S1, stylized fact S2 is less well documented but no less important. The evidence for S2 has come from three different kinds of works. These are the *bimodal perception of risks* (see Camerer and Kunreuther, 1989), *prospect theory* (see Kahneman and Tversky, 1979), and *composite cumulative prospect theory or CCP* (see al-Nowaihi and Dhami, 2010a).[4] In conjunction, this evidence is sufficiently important to merit a section on its own. For that reason, we postpone this material to Section 5, below.

1.4. Failure of non-linear weighting theories to explain the Becker paradox

As pointed out in Remark 3, the main non-linear weighting models, e.g., RDU, CP, incorporate S1 and S2b (but not S2a). Hence, they use probability weighting functions that are extremely steep near the origin (as in the case $\alpha<1$ in Remark 3). Decision makers using such models place very high subjective weight, $w(p)$, on the probability, p, of facing the Becker-type punishments. Thus, for the class of Becker paradoxes (see Remark 2) such individuals would always be deterred by Becker-type punishments. Therefore, one would expect that the Becker paradox would be even stronger under RDU and CP. This intuition turns out to be correct, as we prove in Sections 7 and 8. These individuals will always be deterred by capital punishment, will never run red traffic lights, will always buy insurance for low probability natural hazards, will never talk on mobile phones while driving, etc.

1.5. Resolution of the Becker paradoxes

In order to explain the Becker paradoxes one needs a decision theory that simultaneously accounts for S1, S2a, S2b. There are two alternative theories that do so (see Section 5). The first is the Nobel prize winning work of Kahneman and Tversky's (1979) prospect theory (PT). However, PT's treatment of S1, S2a, S2b, while strongly grounded in empirical evidence, and seminal, is informal, heuristic, and can violate monotonicity even when obvious. We discuss these issues in section 5.3.

The second theory that simultaneously accounts for S1, S2a, S2b is al-Nowaihi and Dhami's (2010a) *composite cumulative prospect theory*, CCP. CCP is formal, axiomatically founded, and respects monotonicity (i.e., decision makers will not choose stochastically dominated options). Before explaining CCP, and showing how it is able to explain the Becker paradoxes, we shall introduce the *composite Prelec probability weighting function* (CPF), which is of fundamental importance in CCP.

[4] We shall often abbreviate *composite cumulative prospect theory* to *composite prospect theory.*

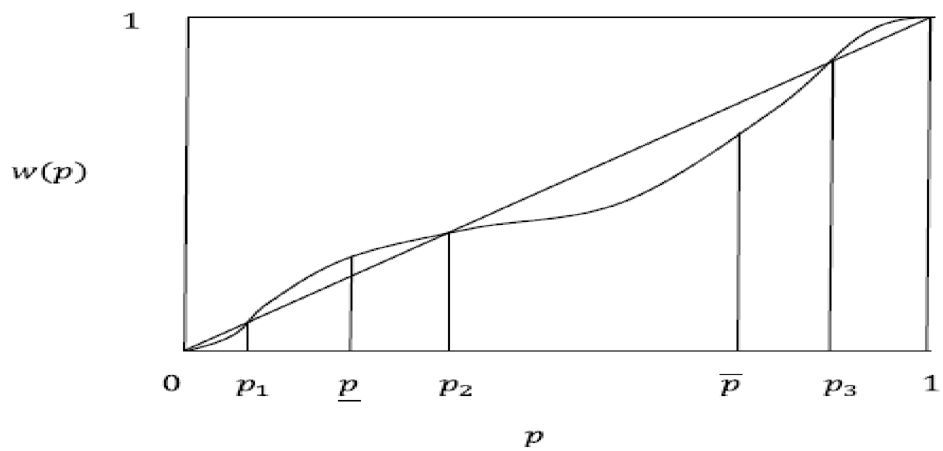


Figure 1.1: A composite Prelec probability weighting function (CPF).

Figure 1.1 gives the general shape of the composite Prelec probability weighting function (CPF). Let p_1, p_2, p_3 be objective probabilities. From Figure 1.1 we see that a CPF overweights probabilities in the range (p_1, p_2) and underweights probabilities in the range (p_2, p_3), thus S1 holds in (p_1, p_3). On the other hand, the CPF underweights probabilities in the range $(0, p_1)$ and overweights probabilities in the range $(p_3, 1)$, so much so that the CPF is almost flat very near 0 and very near 1. This reflects the fact that decision makers who use a CPF ignore events of extremely low probability and treat extremely probable events as certain. Thus S2a holds in $(0, p_1) \cup (p_3, 1)$.

Remark 4 *: Under CPF,* $\lim_{p \to 0} w(p)/p = 0$ *and* $\lim_{p \to 1} \frac{1-w(p)}{1-p} = 0$ *(compare with Remark 3).*

Remark 5 *(CCP; al-Nowaihi and Dhami, 2010a): Under CCP, a fraction* $1 - \mu$ *of the population (see S2b) uses Tversky and Kahneman's (1992) CP with its standard probability weighting function (see Remark 3). The remaining fraction* μ *(see S2a) uses CP but replaces the standard probability weighting function with the composite Prelec probability weighting function (CPF).*

Remark 6 *(CRDU; al-Nowaihi and Dhami, 2010a) Composite rank dependent utility (CRDU) is defined analogously with rank dependent utility, RDU, replacing CP in Remark 5. We argue, however, that CCP is more satisfactory than CRDU.*

Given Remarks 4, 5, 6, it follows that:

1. In each of CCP and CRDU, a fraction $1 - \mu$ of the individuals follows respectively, CP and RDU and, so, the remarks in subsection 1.4 apply. They infinitely overweight infinitesimal probabilities and, thus, they avoid acts that lead to Becker-type punishments. But then the Becker paradox remains.

2. In each of CCP and CRDU, a fraction μ follows respectively, CP and RDU with one important difference. They use the CPF rather than any of the standard weighting functions and, so, place very low subjective weights on very low probabilities (Remark 4). Hence, one might conjecture that they would *not* be dissuaded from acts that lead to Becker type punishments, thus, resolving the Becker paradox. This intuition is only partially correct because in Becker type punishments, against the 'very low probability' must be set the 'very high level of punishment.' CRDU is unable to explain the Becker paradox while CCP can explain it.

   So why is CCP successful when all other theories fail? The main reason is that in addition to more reasonable behavior for low probabilities (in the region $[0, p_1]$ in Figure 1.1) CCP shares other empirically important features with CP that are absent in CRDU. In particular, the *reference dependence* feature of CCP turns out to be necessary to address the problem. Due to reference points, criminals derive *elation* in the state of the world where they are not caught. The higher the level of punishment, the higher is the elation from escaping it. In conjunction with Remark 4, elation ensures that such decision makers will not be dissuaded by Becker type punishments, hence, explaining the Becker paradox.[5] These individuals may (i) not be dissuaded by capital punishment, (ii) ignore insurance for low probability hazards, (iii) drive and talk on mobile phones, (iv) run red traffic lights, and (v) not take up offers of free breast cancer examinations etc.

1.6. Contributions of our paper

First we show that the Becker proposition is a general result that applies to all low probability events that lead to potentially unbounded loss for the individual. Second, we clarify the necessary and sufficient conditions for the Becker proposition under EU and it's welfare consequences. Third, we show that none of the proposed explanations of the Becker paradox is satisfactory. Fourth, we show that the Becker paradox survives under several main alternatives to EU, specifically, RDU, CP and CRDU. Fifth, we show that, so far, the only decision theory that can explain the general class of Becker paradoxes is composite prospect theory (CCP).

1.7. Organization of the paper

Section 2 formulates a standard economic model of crime. The Becker proposition is considered under EU in Section 3. Potential solutions to the Becker paradox are surveyed

[5] We show that the solution of the Becker paradox under CCP is robust to several considerations on reference points that have been suggested in the recent literature. These include the emphasis on rational expectations (Köszegi and Rabin, 2006) and state-dependent reference points (Schmidt et al., 2008).

in Section 4. Section 5 gives more evidence for stylized fact S2 and shows that none of the explanations proposed in Section 4, either singly or jointly, can explain the Becker paradoxes. Section 6 gives a brief discussion of probability weighting functions which are central to applying non-linear decision weights models (with particular emphasis on the Prelec (1998) function). Section 7 shows that the Becker paradox re-emerges under RDU while Section 8 shows it also re-emerges under CP. Section 9 discusses the composite Prelec function (CPF). Section 10 gives formal definitions of composite prospect theory (CCP) and composite rank dependent utility (CRDU); it also shows that CRDU cannot explain the Becker paradox. Section 11 shows how CCP can resolve the Becker paradox. Section 12 concludes. All proofs are relegated to the Appendix.

2. The model and assumptions

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

$$b = y_1 - y_0 \geq 0. \tag{2.1}$$

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. As in the literature we interpret F as the monetary equivalent of all punishments. We assume that,

$$b \leq F \leq F_{\max} \leq \infty. \tag{2.2}$$

Thus, it is feasible to levy a fine that is at least as great as the benefit from crime, b. If $F_{\max} < \infty$, then a feasible fine is bounded above. If $F_{\max} = \infty$, then there is no upper bound to fines. Given the enforcement parameters p, F the individual makes only one choice: To commit the crime or not.

This minimal framework nests several important settings, for example:

Example 1 *(Theft/robbery): Engaging in theft gives a monetary reward $b \geq 0$. If the thief is caught (with probability $p \geq 0$) the goods, whose value is b, are impounded and, in addition, the offender pays a fine, $f \geq 0$ (or faces other penalties such as imprisonment whose equivalent monetary value is f). Hence, $F = b + f$.*

Example 2 *(Tax evasion): Consider the following widely used model (Allingham and Sandmo, 1972). A taxpayer has taxable incomes $z_1 > 0$ and $z_2 > 0$ 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.*[6] *It follows that $y_0 = (1-t)(z_1+z_2)$. Suppose that the taxpayer chooses to evade income z_2. Hence, $y_1 = (1-t)z_1 + z_2 > y_0$ 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 $f = \delta t z_2$ where $\delta > 0$ is the penalty rate. Hence, $F = (1+\delta)tz_2$.*

Example 3 *(Pollution): Consider a firm that produces a fixed output that is sold for a profit, π. As a by-product, and conditional on the firm's existing technology, the firm creates a level of emissions, E, that is greater than the legal limit, $\overline{E}$. 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 by making changes to existing technology. Hence, $y_0 = \pi - c\left(E - \overline{E}\right)$ and $y_1 = \pi$ so that $b = c\left(E - \overline{E}\right) \geq 0$ is the benefit arising from not lowering emissions to the legal requirement. If caught, the firm is made to pay $b = c\left(E - \overline{E}\right)$ as well as a monetary fine $f \geq 0$. Hence, $F = f + c\left(E - \overline{E}\right)$.*

We have not specified the preferences of the individual, yet. In subsequent sections we shall consider, sequentially, the possibilities that the decision maker has *expected utility* (EU), *rank dependent utility* (RDU), and *cumulative prospect theory* (CP) preferences.

2.1. The social costs of crime and law enforcement

Let $C(p,F) \geq 0$ be the *cost to society of law enforcement*. Also, let $D(p,F) \geq 0$ be the *damage to society caused by crime*. Obviously, these entities depend on p, F.

Remark 7 *(Notation) : We indicate partial differentiation with subscripts. For example, $C_p = \frac{\partial C}{\partial p}$ and $C_{pF} = \frac{\partial^2 C}{\partial p \partial F}$.*

We assume that $C(0,0) = 0$, i.e., in the absence of any law enforcement, costs of such enforcement are zero. We also assume that C and D are differentiable with

$$C_p > 0,\ C_F \geq 0,\ D_p \leq 0,\ D_F \leq 0. \tag{2.3}$$

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). The damages to society can be reduced by increasing p, F, because these act to deter crime.

We note, for future reference, a special case below.

Definition 1 *(Ideal fine): The case $C_F = 0$ 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).*

[6] 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 (2007).

2.2. Society's objective

Let $T(p,F) \geq 0$ be the total cost to society of crime. This is the sum of cost of law enforcement, $C(p,F)$, and the damages to society, $D(p,F)$.

$$T(p,F) = C(p,F) + D(p,F). \tag{2.4}$$

Society aims to choose the substitute instruments p and F so as to minimize $T(p,F)$. Both instruments aim for a common target (minimizing $T(p,F)$) but have different effects on costs and damages. We make the following assumptions on $T(p,F)$,

$$[T_F]_{F=0} < 0, [T_F]_{F=F_{\max}} > 0, T_{FF} > 0, T_{pF} > 0. \tag{2.5}$$

The first condition in (2.5) ensures that total costs can be reduced by raising the fine just above zero, hence, $F = 0$ is not optimal. The second condition ensures that, at $F = F_{\max}$, the costs are increasing in fines, hence, it would be beneficial to reduce fines below $F_{\max}$. Together, they ensure that an optimal fine is an interior point. The third condition ensures that our solution is a global maximum. The fourth condition will ensure that the two instruments, p, F are *strategic substitutes* (see below).

2.3. A principal-agent interpretation of the model

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 $C(p,F)$. The total cost to the principal is, thus, $T(p,F) = C(p,F) + D(p,F)$, where p is the probability of detection and F is the sanction. The analogue of Becker's proposition, in this case, 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.

2.4. Punishment functions

The objective of minimizing $T(p,F)$ with respect to p, F can be broken down into two stages. First, we ask whether, for each p, there is a level of punishment, $F = \varphi(p)$, that minimizes $T(p, \varphi(p))$ given p. If the existence of such an *optimal punishment function* is assured, then we can ask whether there exists a probability, p, that minimizes $T(p, \varphi(p))$. Formal definitions are given below. First, we define a punishment function (optimal or otherwise), then we define an optimal punishment function.

Definition 2 *(Punishment function): By a punishment function we mean a function* $\varphi(p) : [0,1] \to [0, F_{\max}]$ *that assigns to each probability of detection and conviction,* $p \in [0,1]$*, a punishment* $\varphi(p) \in [0, F_{\max}]$.[7]

Note that we allow for the possibility of infinite punishments. In particular, if $F_{\max} = \infty$, then we allow for the possibility that $\varphi(p) = \infty$ for some p.

Remark 8 *(Notation):* $F = \varphi(p)$ *denotes a punishment function.* $F = \overline{F}$ *is the special case where the same fine,* $\overline{F}$*, is levied for all levels of* p.

Definition 3 *(Optimal punishment function): Let* $\varphi : [0,1] \to [0, F_{\max}]$ *be a punishment function. We call* φ *an optimal punishment function if, for all* $p \in [0,1]$*, and for all* $F \in [0, F_{\max}]$, $T(p, \varphi(p)) \leq T(p, F)$.

We now give two results but omit the simple proofs.

Result 1 *(Existence of optimal punishment functions; Dhami and al-Nowaihi, 2010a, Lemma 1): Optimal punishment functions exist.*

Definition 4 *(Cost and fine elasticities):* $\eta_p^C = \frac{p}{C} C_p$ *is the probability elasticity of cost,* $\eta_F^C = \frac{F}{C} C_F$ *is the punishment elasticity of cost and* $\eta_p^F = -\frac{p}{\varphi(p)} \frac{d\varphi}{dp}$ *is the probability elasticity of punishment.*

Result 2 *(Dhami and al-Nowaihi, 2010a, Lemma 2):* $\frac{d}{dp} C(p, \varphi(p)) > 0$ *if, and only if,* $\eta_p^C > \eta_p^F \eta_F^C$ *at* $F = \varphi(p)$.

The condition $\eta_p^C > \eta_p^F \eta_F^C$ is most likely to hold when the costs to society do not increase too rapidly in response to an increase in fines. It will be satisfied for an *ideal fine*, since $\eta_p^C > 0$ and $\eta_F^C = 0$ for an ideal fine (see Definition 1).

2.4.1. Substitutability of the instruments p, F

From Result 1, an optimal punishment function exists. Hence, given p, $F = \varphi(p)$ minimizes $T(p, F)$ with respect to F. From (2.5), the optimal fine $F = \varphi(p)$ lies in the interior of $[0, F_{\max}]$. From the first order condition $[T_F]_{F=\varphi(p)} = 0$, implicit differentiation gives:

$$\frac{d}{dp} [T_F]_{F=\varphi(p)} = [T_{FF} \varphi'(p) + T_{pF}]_{F=\varphi(p)} = 0,$$

$$\Rightarrow \varphi'(p) = -\left[\frac{T_{pF}}{T_{FF}}\right]_{F=\varphi(p)} < 0. \tag{2.6}$$

[7] In fact, $\varphi([0,1]) \subset [b, F_{\max}]$.

In order to determine the sign in (2.6), we have used the strategic substitutability between p, F, from (2.5). Thus, at the optimal solution, for a fixed level of crime, p, F are negatively related, i.e., the optimal punishment function takes the form:

$$\varphi(p) : [0,1] \to [0, F_{\max}], \ \varphi(1) = b, \ \varphi'(p) < 0. \tag{2.7}$$

At $p = 1$, all criminal activity is perfectly detected and, so, deterred. Hence, fines, because, they are costly, should be set to their lowest value, so, $\varphi(1) = b$.

2.4.2. The hyperbolic punishment function and its rationale

A popular and tractable punishment function, which satisfies all the desired restrictions in (2.7), i.e., $\varphi(1) = b$, $\varphi'(p) < 0$, is the hyperbolic punishment function (HPF).

Definition 5 *(Hyperbolic punishment function, HPF): A HPF is defined by*

$$\varphi(p) = b/p. \tag{2.8}$$

The name derives from the fact that in p, F space, the HPF plots as a rectangular hyperbola. Note that, for (2.8), $\varphi(0) = \infty$. We show below that the HPF will always deter a risk neutral or risk averse offender under expected utility theory (i.e., the Becker proposition holds in this case under the stated assumptions).

There are several important justifications for using the HPF. Using a standard model of crime under the assumption of risk neutrality, Polinsky and Shavell (2007) show that the HPF is an optimal punishment function. Furthermore, Dhami and al-Nowaihi (2010a) prove the following two important results. First, the HPF is optimal for wide, plausible and sensible range of total cost functions. Second, the HPF gives an upper bound on punishments. These are formally stated, below.

Result 3 *(Dhami and al-Nowaihi, 2010a, Proposition 1): The HPF is optimal, whenever the total cost function (see, (2.4)) takes the following form:*

$$T(p,F) = (cF - aF^2) + \frac{2ab}{p}F + \pi(p), \ \ a < 0, c > 0, \pi'(p) > 0.$$

The first term can be thought of as a quadratic approximation to a more general fine function. The second term represents the strategic substitutability between p and F. $\pi(p)$ is any arbitrary increasing function that can be used to fit the data.

Result 4 *(Dhami and al-Nowaihi, 2010a, Proposition 2): Let Φ be the class of optimal punishment functions that arises from the following cost and damage functions:*

$$C(p,F) = \frac{1}{2}mcF^2e^{\omega(p)},$$

$$D(p,F) = -\frac{1}{2}nF^2 e^{\omega(p)},$$

where, $c > 0,\ m > 0,\ \omega'(p) > 0, \omega'(1) < 1, \omega''(p) \geq 0, n > 0,\ \omega'(p) > 0, mc > n, mc^2 > n.$

Then the HPF is an upper bound to punishment functions is Φ*.*

In the cost and damage functions of Result 4, the fines appear as quadratic terms and p appears in an exponential form. Thus, the marginal effect of changes in p are steeper than the corresponding effects of fines. The critical implication of Result 4, is that if the HPF cannot deter crime, then none of the members in Φ can deter crime either.

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

We now consider the Becker proposition under 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. In that case, his payoff from no-crime, U_{NC}, is given by $U_{NC} = u(y_0)$. 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 expected utility from crime, EU_C, is given by $EU_C = pu(y_1 - F) + (1-p)u(y_1)$. The individual does not engage in crime if the *no-crime condition* (NCC) $EU_C \leq U_{NC}$ is satisfied. Thus the no-crime condition is

$$NCC\text{: } pu(y_1 - F) + (1-p)u(y_1) \leq u(y_0). \tag{3.1}$$

Recall from (2.7) that for $p = 1$, $F = \varphi(1) = b$. Thus, since $y_1 - b = y_0$, (3.1) is clearly satisfied for $p = 1$. Let $F = \varphi(p)$ be a differentiable function of p satisfying (2.7). The *NCC* in (3.1) continues to hold, as p is declines from 1, if, and only if

$$\frac{d}{dp}\left[pu(y_1 - \varphi(p)) + (1-p)u(y_1)\right] \geq 0. \tag{3.2}$$

$$\Leftrightarrow u'(y_1 - \varphi(p)) \geq \frac{u(y_1) - u(y_1 - \varphi(p))}{-p\varphi'(p)}. \tag{3.3}$$

For the HPF (2.8), $\frac{-p\varphi'(p)}{\varphi(p)} = 1$, so the *NCC,* (3.3), reduces to

$$\text{NCC for HPF: } u'(y_1 - \varphi(p)) \geq \frac{u(y_1) - u(y_1 - \varphi(p))}{\varphi(p)}. \tag{3.4}$$

If the decision maker is risk averse or risk neutral, so that u is concave, then the *NCC* (3.4) *will* hold for all $p \in (0,1]$. We show in Proposition 1 below that (i) not only does the Becker proposition hold for risk-neutral and risk averse criminals, (ii) 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 HPF $\varphi(p) = \frac{b}{p}$ 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.
(b) If, in addition, $\eta_p^C > \eta_F^C$ (Definition 4), then reducing p reduces the total social cost of crime and law enforcement, $T(p, F)$.

While Proposition 1(a) is well known from Becker (1968), Proposition 1(b) is, as far as we know, a new result.

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 Proposition 1 (a) it would be possible to deter crime, however small the probability of detection and conviction.*

In contrast to these results, Levitt (2004) argues: "... given the rarity with which executions are carried out in this country and the long delays in doing so, a rational criminal should not be deterred by the threat of execution." That might well be true. However, this observation is certainly not consistent with the decision maker following EU, under the conditions of Proposition 1. The probability of capital punishment can be made arbitrarily small but it will certainly deter crime under EU (Becker proposition).

3.1. Risk aversion: The role of necessary and sufficient conditions

We now explore the role of necessity and sufficiency of risk attitudes for the Becker proposition. The conditions in Proposition 1 are sufficient conditions. We first give an example to show that the Becker proposition need not hold under risk-loving behavior.

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 (3.1), 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. Thus, 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 (3.1), that the probability of detection and conviction, $p > 0$, can be made arbitrarily low, by choosing the fixed punishment $F = \overline{F} = y_1$. Hence, the HPF, $F = \varphi(p) = \frac{b}{p}$, is sufficient, but not necessary to deter crime.*

Example 7 *: Consider the following utility function used by Tversky and Kahneman (1992) and supported by a large body of experimental work.*

$$u(y) = \begin{cases} y^{\gamma} & \text{if } y \geq 0 \\ -(-y)^{\gamma} & \text{if } y < 0 \end{cases};\ 0 < \gamma < 1. \tag{3.5}$$

This utility function is (strictly) concave for $y \geq 0$ but strictly convex for $y < 0$. Zero is the reference point in this case. Hence, we have risk seeking behavior in the region $y < 0$. In this case, the NCC, (3.1), holds for any $p \in (0,1]$, if the punishment function is given by $F = \varphi(p) = y_1 + q(y_1, y_0)$, where $q(y_1, y_0) = \left(\frac{y_1^{\gamma} - y_0^{\gamma}}{p}\right)^{\frac{1}{\gamma}}$. However, the elasticity of fine with respect to the probability of detection, η_p^F, satisfies

$$\eta_p^F = \frac{q(y_1, y_0)}{y_1 + q(y_1, y_0)} \to \infty \text{ as } p \to 0.$$

Since the NCC holds and crime is deterred, so $D(p, F) = 0$, therefore, $T(p, F) = C(p, F)$. Hence, it would be socially beneficial to drive the probability of detection and conviction down to zero if ideal fines ($\eta_F^C = 0$) were available or if $\eta_p^C \to \infty$ faster than $\eta_p^F \eta_F^C$ (see Lemma 2). Thus, risk seeking behavior is not sufficient to explain the Becker paradox.

4. The competing explanations of the Becker paradox

The *Becker paradox* arises when observed behavior is contrary to the *Becker proposition.* Considerable effort has gone into explaining the Becker paradox. We first critically discuss these explanations. Section 5.1 then argues that these explanations are insufficient, singly or jointly, to explain the general class of Becker paradoxes (see Remark 2).

1. *Risk seeking behavior*: If decision makers are risk-seekers (compare Examples 4 and 5, above) then the Becker proposition need not hold. This potential explanation is given in Becker (1968). However, this explanation creates great difficulties for other explanations of human behavior that require some form of risk aversion, e.g., insurance, investment, saving, risk management, principal-agent theory and mechanism design. Moreover, risk-seeking behavior is not sufficient, as shown by Example 7.

2. *Bankruptcy issues*: Bankruptcy issues put an upper bound on the level of possible fines.[8] There are several objections to this explanation. First, it takes fines literally

[8]See, for example, Polinsky and Shavell (1991) and Garoupa (2001).

rather than the more general interpretation as the monetary equivalent of punishment. Second, even when interpreted literally, fines 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 the consequences of declaring bankruptcy, e.g., the emergence of the limited company; see 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 (2000b) for a discussion.

4. *Errors in conviction*: To our minds, this is one of the two most persuasive explanations (the other is 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).[9] 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, i.e., increasing p is more costly to society than increasing F; see Polinsky and Shavell (2000b).

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 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 this possibility undermines the assumption that increasing F is less costly to society than increasing p; see, for instance, Friedman (1999). Explanations 4 and 5 seem persuasive but fail for the general class of Becker paradoxes; see section 5.1, below.

6. *Abhorrence of severe punishments*: Society may not, for reasons of norms, fairness etc., accept severe punishments.[10] This provides a strong potential explanation for the Becker paradox, but leaves open the question of an explanation of these norms. However, our focus is on examining the purely economic case for the Becker paradox.

[9] See, for instance, Andreoni (1991) and Feess and Wholschlegel (2009).

[10] For a model in which preferences for fairness ensure that fines are bounded away from their upper bound, see Polinsky and Shavell (2000a).

This explanation, like most others in this section is found wanting when we consider the general class of Becker paradoxes in section 5.1, below.

7. *Objectives other than deterrence*: Punishment might have objectives other than deterrence, e.g., incapacitation and retribution. Incapacitation is easily incorporated 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. *Risk aversion with bounded punishments*: Suppose that criminals are risk averse and differ in their benefits, b, from crime. Consider an increase in fines. On the one hand, the imposition of the fine acts as a deterrent for some individuals (depending on their level of b). But on the other hand, for those who are not deterred, an increase in fines reduces the risk-averse criminals' income, if caught. If the utility of all citizens enters the social welfare function, then fines have opposing effects, leading to an interior solution where these marginal effects balance out.[11]

9. *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 other members in the general class of Becker paradoxes; see section 5.1, below.

5. Evidence for S2 and an assessment of the proposed explanations for the Becker paradox

In subsection 5.1, below, we evaluate the explanations in Section 4 with two further objectives. (i) To provide more evidence for S2, particularly S2a. (ii) To introduce several members of the general class of Becker paradoxes (see Remark 2). We then turn, in subsection 5.2, to the evidence on S2 from the *bimodal perception of risks model.* Kahneman and Tversky's (1979) evaluation of the evidence on S2, and their proposed theoretical solution is discussed in section 5.3. The determinants of μ, the fraction which respects S2a, are discussed in section 5.4.

[11] This expanation is due to Polinsky and Shavell (1979) and Kaplow (1992) but it is subsumed within our point 6, abhorrence of severe punishments, above.

5.1. Why might the explanations in Section 4 not suffice?

We now argue that the explanations in section 4, either singly or in conjunction, cannot explain the *general class of Becker paradoxes* (see Remark 2).

5.1.1. Evidence from jumping red traffic lights

Consider an individual act of running red lights. There is (at least) a *small probability* of an accident. However, the consequences (F in our framework) are *self-inflicted* and potentially have *infinite costs*. Hence, this belongs to the general class of Becker paradoxes.

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 and it is implausible to assume that running red traffic lights are simply 'mistakes'. Bar-Ilan (2000) and Bar-Ilan and Sacerdote (2001, 2004) provide, what is to our minds, near decisive evidence that the explanations in subsection 4 cannot provide a satisfactory explanation of the Becker paradox within an EU framework.

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 he finds that if a slight injury causes a loss greater or equal to 0.9 days, a risk neutral person will be deterred by that risk alone. But, the corresponding numbers for the additional risks of serious and fatal injuries are 13.9 days and 69.4 days respectively (which should deter red traffic light running completely).[12]

Clearly EU combined with risk aversion would find it difficult to explain this evidence. Explanations 2-8 in section 4 do not apply here, because the punishment is *self inflicted*.[13] Explanation 9 is also inadequate, for Bar-Ilan and Sacerdote (2004) report "We find that red-light running decreases sharply in response to an increase in the fine ... Criminals convicted of violent offences or property offences run more red lights on average but have the same elasticity as drivers without a criminal record". This leaves explanation 1, i.e., risk seeking, but we have already noted the problems with it in Section 4.

5.1.2. Driving while talking on hand-held mobile phones

Consider the usage of hand-held mobile phones in moving vehicles. A user of mobile phones faces potentially *infinite punishment* (e.g., loss of one's and/or the family's life) with *low*

[12] 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.

[13] For instance, one cannot argue along the lines of Explanation 6 that there are any norms or fairness considerations involved in jumping red traffic lights.

probability, in the event of an accident.[14] This too belongs to the general class of Becker paradoxes. The Becker proposition applied to this situation suggests that drivers will not use mobile phones while driving. But evidence is to the contrary.

Various survey evidence in the UK indicates that up to 40 percent of individuals drive and talk on mobile phones; see, for example, the RSPA (2005). Pöystia et al. (2005) report that two thirds of Finnish drivers and 85% of American drivers use their phone while driving, which increases the risk of an accident by two to six fold.[15]

5.1.3. Purchase of insurance for natural hazards

Consider individual choice to buy non-mandatory insurance against *low probability but high-cost natural hazards* (e.g., earthquake, flood and hurricane damage in areas prone to these hazards). This is yet another member of the general class of Becker paradoxes. The Becker proposition would suggest that if decision makers use expected utility then they should buy at least some insurance, even if it is actuarially unfair. This applies with even greater bite as the probability becomes smaller and the size of the loss becomes larger.[16]

By contrast, based on extensive field and lab evidence, Kunreuther et al. (1978) found that there is a probability below which the take-up of insurance drops dramatically. These results were robust to a very large number of controls, perturbations and other relevant factors; see al-Nowaihi and Dhami (2010b) for the details. 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 [as in S2a]." This evidence contradicts the Becker proposition. Furthermore, none of the explanations in section 4 provide a satisfactory resolution.[17]

[14] Extensive evidence suggests that the perceived probability of an accident might be even lower than the actual probability because drivers are *overconfident* of their driving abilities. Taylor and Brown (1998) suggest that upto 90 percent of car accidents might be caused by overconfidence. The overconfidence finding is pervasive in behavioral economics and arises from many diverse contexts. In each case, the individual's perceived probability of a loss is lower than the actual probability, which further strengthens our argument for the general class of Becker paradoxes.

[15] Hands-free equipment, although now obligatory in many countries, seems not to offer essential safety advantages. This suggests that it is the mental distraction that is dangerous, rather than the physical act of holding a mobile phone.

[16] For formal proofs of these claims, see al-Nowaihi and Dhami (2010b).

[17] For the literature that extends the original work of Kunreuther et al (1978), see al-Nowaihi and Dhami (2010b). Recent evidence suggests that some individuals overinsure for modest risks; see Sydnor (2010). Under composite prospect theory (CCP) this can be accomodated because the evidence indicates that a fraction $1-\mu$ of the individuals respect S2(b) (these conform to the evidence in Sydnor, 2010), while the remaining fraction, μ, respects S2(a) (these conform to the evidence in Kunreuther et al, 1978).

5.1.4. Breast cancer examination

Consider the decision to have a breast cancer examination before the recent spread of awareness a few decades ago. Breast cancer has a *low unconditional probability* but *potentially infinite private cost* (Becker-type punishment).[18] Hence, this decision qualifies it to be a member of the general class of Becker paradoxes. The unconditional probability of breast cancer was perceived to be even lower prior to the recent ris in awareness. Even as evidence accumulated about the dangers of breast cancer, women took up the offer of breast cancer examination, only sparingly.[19] This clearly runs counter to the predictions of the Becker proposition which would suggest that all women should have over-subscribed to the non-mandatory examination. The suggested explanations of the Becker paradox in section 4 do not seem to offer a satisfactory solution.

5.2. The bimodal perception of risk

There is strong evidence of a *bimodal perception of risks*, see Camerer and Kunreuther (1989) and Schade et al. (2001) that directly supports S2. A fraction μ of individuals do not pay attention to losses whose probability falls below a certain threshold (as in S2a). For the remaining fraction $1-\mu$ of individuals, the size of the loss is relatively more salient despite the low probability (as in S2b).

McClelland et al. (1993) find strong evidence of a bimodal perception of risks for insurance for low probability events. The evidence indicates that decision makers have a threshold below which they underestimate risk and above which they overestimate it. Furthermore, individuals have different thresholds. Across any population of individuals, for any given probability one would then observe a bimodal perception of risks; see Viscusi (1998). Kunreuther et al. (1988) argue that the bimodal response to low probability events is observed in most field studies. This line of work, that we have only briefly reviewed, provides strong evidence for S2.

5.3. The prospect theory approach to modelling S1, S2a, S2b

Kahneman and Tversky's (1979) Noble prize winning work on *prospect theory*, PT, is the only theoretical attempt at incorporating S1, S2. PT makes a distinction between an *editing* and an *evaluation* phase. From our perspective, the most important aspect of the editing phase takes place when decision makers decide which improbable events to treat

[18] We know now from recent research in genetics that the conditional probability, conditional on a close female relative having had the disease is much higher.

[19] In the US, this changed after the greatly publicized events of the mastectomies of Betty Ford and Happy Rockefeller; see Kunreuther et al. (1978, p. xiii, p. 13-14).

as impossible and which probable events to treat as certain.[20] Under PT, decision makers evaluate the value, V, of a lottery $L = (x_1, p_1; x_2, p_2; ...; x_n, p_n)$ as

$$V(L) = \sum_{i=1}^{n} \pi_i(p_i)u(x_i), \tag{5.1}$$

where $u(x_i)$ is the utility of x_i and $\pi(p_i)$ is the associated decision weight. Kahneman and Tversky (1979) drew $\pi(p)$ as in Figure 5.1, which is *undefined* at both ends, reflecting the vexed issue of how decision makers behave over these ranges of probabilities.

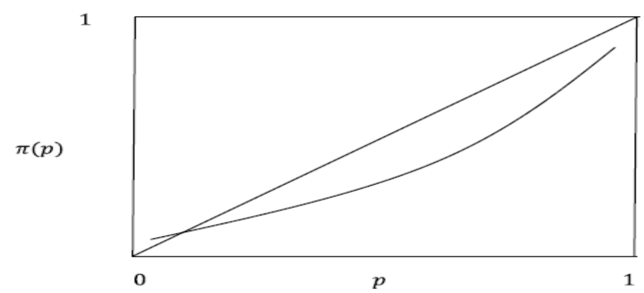


Figure 5.1: Ignorance at the endpoints. Source: Kahneman and Tversky (1979, p. 283)

Kahneman and Tversky (1979, pp.282-283) wrote the following to summarize the evidence for S2, which is worth reading carefully and in full. "*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...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.*"

In Kahneman and Tversky's words, low probability events are either *ignored* (S2a) or *overweighted* (S2b). After the prospects are 'psychologically cleaned' in this *editing phase*, the decision maker then applies (5.1) in the *evaluation phase*.

Remark 9 *(Problems with PT): Under PT, $\pi(p)$ uses point transformations of probabilities which can violate monotonicity*[21]. *Furthermore, $\pi(p)$ is not defined at the endpoints.*

[20] Kahneman and Tversky (1979) also identify other heuristics in the editing phase, e.g., the *isolation effect*. This allows decision makers to cancel 'nearly common' components of two prospects before evaluating them.

[21] See Quiggin (1982) and Starmer (2000).

Hence, one cannot apply cumulative transformations of probability suggested by Quiggin (1982, 1993) in rank dependent utility (RDU) to deal with the monotonicity problem. Quiggen's machinery now lies at the heart of most modern non-linear weights theories. Also, while PT can simultaneously incorporate S1, S2a, S2b, it does so in an informal manner that is often difficult to apply.

Remark 10 *(Cumulative prospect theory, CP): In response to the problems in Remark 9, Tversky and Kahneman (1992) incorporated Quiggin's (1982, 1993) insights by modifying PT to cumulative prospect theory (CP). To do so, they needed a decision weights function, π, that is defined for the entire probability domain $[0, 1]$ (unlike the case shown in Figure 5.1). This meant eliminating the psychologically rich editing phase, which by incorporating S2a created the gaps at the end-points. However, (and in conjunction with the standard probability weighting functions under CP; see Remark 3) this implies that under CP, S2a cannot be taken into account.*

Remark 11 *(Composite cumulative prospect theory, CCP): Remarks 9 and 10 provide the motivation behind al-Nowaihi and Dhami's (2010a) composite cumulative prospect theory (CCP). Like PT, CCP incorporates S1, S2a, S2b but, unlike PT, it does so in a formal manner that is axiomatically founded and consistent with the evidence. Also like CP, CCP does not violate monotonicity.*

5.4. Determinants of μ, the fraction which respects S2a

The bimodal perception of risks framework, PT, and the general class of Becker paradoxes do not specify the respective fractions, μ and $1 - \mu$ of decision makers, i.e., those who follow respectively S2a and S2b. These fractions are likely to depend on the context and the problem. We believe that the salience of low-probability large-outcomes, particularly losses (and so the size of μ) can be influenced by the media, family, friends, and public policy. For instance, as discussed in section 5.1.4, the 'low-take up of free breast cancer examination' rose phenomenally in the US after the greatly publicized mastectomies of Betty Ford and Happy Rockefeller in the media. Vivid public warnings of the fatal consequences of running red traffic lights and speeding (as in the UK) can have a similar effect.

The size of μ can also be influenced by other factors such as emotions, experience, time available to make a decision, bounded rationality, framing, incentive effects and so on. We take an agnostic position as to the source of μ. For our purposes we shall simply take μ as given. For $\mu = 0$ our theory reduces to CP. However, the evidence strongly suggests that (1) $\mu > 0$, and (2) the size of μ is significant in many relevant contexts.

6. Probability weighting functions (PWF)

In all non-linear weights models (e.g. RDU and CP), the individual uses a *probability weighting function* (PWF), denoted by $w(p)$, that transforms objective probabilities, p.

Definition 6 *: A PWF is a strictly increasing function* $w : [0,1] \overset{onto}{\longrightarrow} [0,1]$.

Result 5 *(al-Nowaihi and Dhami, 2010a, Proposition 1): A PWF, $w(p)$, 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.*

Since the Becker proposition hinges critically on the behavior of decision makers as $p \rightarrow 0$, we now offer some definitions that establish the relevant terminology.

Definition 7 *: For $\gamma > 0$, $w(p)$ infinitely-overweights infinitesimal probabilities if* $\lim\limits_{p\rightarrow 0} \frac{w(p)}{p^{\gamma}} = \infty$*. This is the sense in which a PWF is extremely steep as $p \rightarrow 0$.*

Definition 8 *(Standard probability weighting functions): We shall call the entire class of probability weighting functions that satisfy Definition 7, for empirically relevant values of γ, the class of standard probability weighting functions.*

Some examples of standard probability weighting functions are Tversky and Kahneman (1992), Gonzalez and Wu (1999), Lattimore, Baker and Witte (1992) and Prelec (1998) (for $\alpha < 1$; see subsection 6.1 below). These functions satisfy S1, S2b but, crucially, not S2a. To account for S2a, it will be necessary for a PWF to satisfy the condition of Definition 9, below, for empirically relevant values of γ (recall section 1.5 in the introduction).

Definition 9 *: For $\gamma > 0$, $w(p)$ zero-underweights infinitesimal probabilities if* $\lim\limits_{p\rightarrow 0} \frac{w(p)}{p^{\gamma}} = 0$.

6.1. Prelec's probability weighting function

The Prelec (1998) PWF is parsimonious, consistent with S1, S2b, and has an axiomatic foundation. For these reasons we choose the Prelec function, but our observations apply to all standard probability weighting functions used in RDU and CP.

Definition 10 *(Prelec, 1998): By the Prelec function we mean the PWF* $w : [0,1] \overset{onto}{\longrightarrow} [0,1]$ *given by $w(0) = 0$, and*

$$w(p) = e^{-\beta(-\ln p)^{\alpha}},\ 0 < p \leq 1,\ \alpha > 0,\ \beta > 0. \tag{6.1}$$

We now make a distinction between the *Prelec function* and the *standard Prelec function*; Prelec (1998) prefers the later for reasons we specify below.

Definition 11 *(Standard Prelec function): By the standard Prelec PWF we mean the Prelec function, defined in Definition 10, but with* $0 < \alpha < 1$.

A simple proof that we omit gives the following result.

Result 6 *(al-Nowaihi and Dhami, 2010a, Proposition 3): The Prelec function (Definition 10) is a PWF in the sense of Definition 6.*

For $\alpha < 1$ the Prelec function is strictly concave for low probabilities but strictly convex for high probabilities, i.e., it is *inverse-S shaped* as in $w(p) = e^{-(-\ln p)^{\frac{1}{2}}}$ $(\alpha = \frac{1}{2},\ \beta = 1)$, sketched in subsection 1.3, above. Conversely, for $\alpha > 1$ the Prelec function is strictly convex for low probabilities but strictly concave for high probabilities, i.e., it is *S shaped*, as in $w(p) = e^{-(-\ln p)^2}$ $(\alpha = 2,\ \beta = 1)$, sketched in subsection 1.3, above. Clearly, for $\alpha < 1$ the Prelec function is consistent with S1 but not S2a. Conversely, for $\alpha > 1$ the Prelec function is consistent with S2a but not S1.

Recall Definitions 10 and 11. Prelec's (1998) own preference is for the standard Prelec function in Definition 11. According to Prelec (1998, p.505), the infinite limit in Definition 7 capture the qualitative change as we move from improbability to impossibility. However, the standard Prelec function contradicts stylized fact S2a, i.e., the observed behavior that people ignore events of very low probability; see, e.g., Kahneman and Tversky (1979). A related and useful result, whose proof we omit, is the following.

Result 7 *(al-Nowaihi and Dhami, 2010a, Proposition 7):*
(a) If $\alpha < 1$ *then, for all* $\gamma > 0$*, the Prelec function (Definition 10) infinitely overweights infinitesimal probabilities (Definition 7), i.e.,* $\lim_{p \to 0} \frac{w(p)}{p^{\gamma}} = \infty$.
(b), If $\alpha > 1$ *then, for all* $\gamma > 0$*, the Prelec function zero-underweights infinitesimal probabilities (Definition 9) i.e.,* $\lim_{p \to 0} \frac{w(p)}{p^{\gamma}} = 0$.

7. The Becker paradox under rank dependent utility

Rank dependent utility (RDU) is a very conservative extension of expected utility (EU). The objective function under RDU is succinctly described in Definition 12, below.

Definition 12 *(Quiggin 1982) Consider the lottery* $(x_1, x_2, ..., x_n;\ p_1, p_2, ..., p_n)$ *that pays* x_i *with probability* p_i*, where* $x_1 < x_2 < ... < x_n$*. For RDU, the decision weights,* π_j*, are defined by* $\pi_i = w\left(\Sigma_{j=i}^{n}\ p_j\right) - w\left(\Sigma_{j=i+1}^{n}\ p_j\right)$*, where* $w(.)$ *is a standard probability weighting function (see definition 8).*[22] *The decision maker's RDU is*

$$U(x_1, x_2, ..., x_n;\ p_1, p_2, ..., p_n) = \Sigma_{j=1}^{n} \pi_j u(x_j)\,. \tag{7.1}$$

[22] It can be shown that these cumulative transformations of probability ensure that stochastically dominated choices are not made by the decision maker; see Quiggin (1982, 1993).

Machina (2008) describes RDU as the most popular alternative to EU. Since RDU uses *standard probability weighting functions* (Definition 8) it explain S1 but not S2a.

Consider the model of crime in Section 2 and let the decision maker uses RDU. Using Definition 12, the payoff from no-crime is $U_{NC} = u(y_0)$, while that from crime is

$$EU_C = [1 - w(1-p)]\, u(y_1 - F) + w(1-p)\, u(y_1). \tag{7.2}$$

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

$$[1 - w(1-p)]\, u(y_1 - F) + w(1-p)\, u(y_1) \leq u(y_0). \tag{7.3}$$

After some simple algebra, the NCC (7.3) becomes

$$p \geq 1 - w^{-1}\left[\frac{1 - \frac{u(y_0)}{u(y_1 - F)}}{1 - \frac{u(y_1)}{u(y_1 - F)}}\right]. \tag{7.4}$$

Since $F_{\max}$ is the maximum fine, the corresponding utility is $u_{\min} = u(y_1 - F_{\max})$. Let

$$p_{\min} = 1 - w^{-1}\left[\frac{1 - \frac{u(y_0)}{u_{\min}}}{1 - \frac{u(y_1)}{u_{\min}}}\right]. \tag{7.5}$$

Then, from (7.4), (7.5), the NCC becomes

$$p \geq p_{\min}. \tag{7.6}$$

Proposition 2 : *Under RDU, 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.*

Proposition 3, below, the main result of this section, holds for any weighting function $w(p)$, standard (Definition 8) or otherwise.

Proposition 3 : *Under RDU and for any PWF $w(p)$,*
(a) If the utility function, u, is concave, then, given any $p > 0$, no matter how small, crime is deterred by choosing the punishment function $\varphi(p) = \frac{b}{1-w(1-p)}$.
(b) If, in addition, $\eta_p^C > \frac{pw'(1-p)}{1-w(1-p)}\eta_F^C$ (Definition 4 and Result 2), then reducing p reduces the total social cost of crime and law enforcement, $T(p, F)$.

So, under RDU, the Becker proposition survives and the Becker paradox remains.

8. Cumulative prospect theory (CP) and fixed reference points

We first outline cumulative prospect theory (CP). Consider a lottery of the form

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

where $y_{-m} < ... < y_0 < ... < y_n$ are the outcomes and $p_{-m}, ..., p_n$ are the corresponding probabilities, such that $\sum_{i=-m}^{n} p_i = 1$ and $p_i \geq 0$. In CP, decision makers derive utility from wealth relative to a reference point for wealth, y_0.[23]

Definition 13 *(Lotteries in incremental form or 'prospects') Let $x_i = y_i - y_0, i = -m, ..., n$ be the increment in wealth relative to y_0 when the outcome is y_i and $x_{-m} < ... < x_0 = 0 < ... < x_n$. Then, a lottery in incremental form (or a prospect) is:*

$$L = (x_{-m}, ..., x_{-1}, x_0, x_1, ..., x_n; p_{-m}; ..., p_{-1}, p_0, p_1; ..., p_n). \tag{8.1}$$

Denote by $\mathcal{L}_P$ the set of all *prospects* of the form given in (8.1).

Remark 12 *: An outcome is in the domain of gains if $x_i > 0$ and in the domain of losses if $x_i < 0$.*

Definition 14 *(Tversky and Kahneman, 1979). A utility function, $v(x)$, under CP 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).[24]
4. $-v(-x) > v(x)$ for $x > 0$ (loss aversion, i.e., losses bite more than equivalent gains).

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

$$v(x) = \begin{cases} x^{\gamma} & \text{if } x \geq 0 \\ -\theta(-x)^{\rho} & \text{if } x < 0 \end{cases} \tag{8.2}$$

where γ, θ, ρ are constants. The coefficients of the power function satisfy $0 < \gamma < 1$, $0 < \rho < 1$. $\theta > 1$ is known as the *coefficient of loss aversion*.[25]

[23] y_0 could be initial wealth, status-quo wealth, average wealth, desired wealth, rational expectations of future wealth etc. depending on the context. See Kahneman and Tversky (2000), Köszegi and Rabin (2006), and Schmidt et al. (2008).

[24] Concavity in the domain of gains and convexity in the domain of losses does not mean, however, that the decision maker is risk averse in the domian of gains and risk seeking in the domian of losses. The reason is that attitudes to risk are also influenced by the shape of the probability weighting function. See, for instance, the *four-fold classification of risk* outlined in Kahneman and Tversky (2000).

[25] 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 the value function in (8.2). al-Nowaihi et al. (2008) give a formal proof, as well as some other results (e.g. that γ is necessarily identical to ρ). Tversky and Kahneman (1992) estimated that $\gamma \simeq \rho \simeq 0.88$ and $\theta \simeq 2.25$.

Definition 15 *: For CP, the decision weights, π_i, are defined as follows. In the domain of gains, $\pi_i = w\left(\Sigma_{j=i}^{n} p_j\right) - w\left(\Sigma_{j=i+1}^{n} p_j\right)$ $i = 1, ..., n$ while in the domain of losses, $\pi_j = w\left(\Sigma_{i=-m}^{j} p_i\right) - w\left(\Sigma_{i=-m}^{j-1} p_i\right)$, $j = -m, ..., -1$.[26] As in RDU, $w(.)$ is a standard probability weighting function (see Definition 8)*

A decision maker using CP maximizes the following *value function* defined over $\mathcal{L}_P$,

$$V(L) = \Sigma_{i=-m}^{n} \pi_i v(x_i); \; L \in \mathcal{L}_P. \tag{8.3}$$

Remark 13 *: Since CP (like RDU) uses standard probability weighting functions (e.g. the standard Prelec function with $\alpha < 1$), thus, using Result 7a, $\lim_{p \to 0} \frac{w(p)}{p^\gamma} = \infty$. It follows that, like RDU, CP can address S1 and S2b but not S2a which is necessary to resolve the general class of Becker paradoxes.*

Let the reference incomes for crime and no-crime, be respectively, y_c and y_{nc}, which are assumed to be fixed in this section.[27] Then, using reference dependence, the payoff from not committing crime is

$$V_{NC} = v(y_0 - y_{nc}). \tag{8.4}$$

From Remark 12, the outcomes under CP are split into the domain of gains and loss. We make the sensible assumption that the decision maker who commits a crime is in the domain of gains if 'not caught' and in the domain of losses if 'caught'. Thus, if 'caught' (with probability p), the outcome, $y_1 - F$, is in the domain of losses (i.e., $y_1 - F - y_c < 0$). If 'not caught' (with probability $1 - p$), the outcome, y_1, is in the domain of gains (i.e., $y_1 - y_c > 0$). Thus, we have one outcome each in the domain of losses and gains.[28] Using Definition 15, the respective decision weights are $w(p)$ and $w(1-p)$.[29] Then, under CP, the individual's payoff from committing a crime is given by

$$V_C = w(p) v(y_1 - F - y_c) + w(1-p) v(y_1 - y_c). \tag{8.5}$$

[26] 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).

[27] In the original version of CP, in Tversky and Kahneman (1992), the reference point is fixed. In their third generation prospect theory, denoted by PT^3, Schmidt et al (2008) suggest using state-dependent reference points. However, the model by Schmidt et al (2008) relies on reference dependent subjective expected utility of Sugden (2003) and so PT^3 is linear in probabilities (as is, of course, EU). Thus, PT^3 cannot address stylized facts S1 and S2. For that reason we do not follow PT^3, but we exploit it's insight of using state dependent reference points.

[28] In the context of tax evasion, Example 2, Dhami and al-Nowaihi (2007, 2010b), show that the *legal after-tax income* is the unique reference point such that for all levels of declared income, the taxpayer is in the domain of loss if caught and in the domain of gains if not caught.

[29] In the special case of one outcome each in the domain of gains and losses, point transformations of probabilities coincide with cumulative transformations.

Definition 16 *(Elation): We shall refer to* $v(y_1 - y_c)$ *as the elation from committing a crime and getting away with it.*

The 'no crime condition' (*NCC)* is $V_C \leq V_{NC}$, which implies that

$$w(p)\, v(y_1 - F - y_c) + w(1-p)\, v(y_1 - y_c) \leq v(y_0 - y_{nc})\,. \tag{8.6}$$

The *NCC* depends on the two reference points, y_{nc} and y_c, assumed fixed in this section. Section 11 considers alternative specifications of reference incomes.[30] Recall from Remark 8 that punishment could be fixed, $F = \overline{F}$, or variable, $F = \varphi(p)$. We consider the former case here; the latter case is considered in section 11 below.

Proposition 4 *: Assume fixed reference points* y_{nc}, y_c, *fixed punishments* $F = \overline{F} \in [0, \infty)$, *and a general utility function* v *that is unbounded below. Then, under CP, crime can be deterred with arbitrarily low* p, *i.e., the Becker proposition holds.*

Under the conditions of Proposition 4, the Becker paradox emerges under CP.

9. The composite Prelec function (CPF)

From subsection 6.1, the standard probability weighting functions either violate S1 or S2a. This subsection outlines the *composite Prelec function* (CPF) of al-Nowaihi and Dhami (2010a) which simultaneously accounts for S1, S2a. In conjunction with the feature of reference points in CP, it will enable us to address the Becker paradox.[31]

The CPF in Figure 1.1 is composed of three segments of the Prelec function, respectively defined over $(0, \underline{p}]$, $(\underline{p}, \overline{p}]$, $(\overline{p}, 1]$. Using the notation in Figure 1.1, let

$$\underline{p} = e^{-\left(\frac{\beta}{\beta_0}\right)^{\frac{1}{\alpha_0 - \alpha}}},\ \overline{p} = e^{-\left(\frac{\beta}{\beta_1}\right)^{\frac{1}{\alpha_1 - \alpha}}}. \tag{9.1}$$

Definition 17 *(Composite Prelec weighting function, CPF): By the CPF we mean the probability weighting function* $w : [0,1] \to [0,1]$ *given by*

$$w(p) = \begin{cases} 0 & if \quad p = 0 \\ e^{-\beta_0(-\ln p)^{\alpha_0}} & if \quad 0 < p \leq \underline{p} \\ e^{-\beta(-\ln p)^{\alpha}} & if \quad \underline{p} < p \leq \overline{p} \\ e^{-\beta_1(-\ln p)^{\alpha_1}} & if \quad \overline{p} < p \leq 1 \end{cases} \tag{9.2}$$

where $\underline{p}$ *and* $\overline{p}$ *are given by (9.1) and*

$$0 < \alpha < 1,\ \beta > 0;\ \alpha_0 > 1,\ \beta_0 > 0;\ \alpha_1 > 1,\ \beta_1 > 0, \beta_0 < 1/\beta^{\frac{\alpha_0 - 1}{1-\alpha}},\ \beta_1 > 1/\beta^{\frac{\alpha_1 - 1}{1-\alpha}}. \tag{9.3}$$

[30] These include the state-dependent reference point specification of Schmidt et al (2008) as well as the state-independent-rational-expectations specification of Köszegi and Rabin (2006).

[31] This combination can also potentially address the other problems outlined in section 5.1. See, for instance, al-Nowaihi and Dhami (2010a,b).

Result 8 *(al-Nowaihi and Dhami, 2010a, Proposition 8): The CPF (Definition 17) is a PWF in the sense of Definition 6.*

The restrictions $\alpha > 0$, $\beta > 0$, $\beta_0 > 0$ and $\beta_1 > 0$, in (9.3), are required by the axiomatic derivations of the Prelec function (see Prelec (1998), Luce (2001) and al-Nowaihi and Dhami (2006)). The restrictions $\beta_0 < 1/\beta^{\frac{\alpha_0-1}{1-\alpha}}$ and $\beta_1 > 1/\beta^{\frac{\alpha_1-1}{1-\alpha}}$ guarantee that the interval $\left(\underline{p}, \overline{p}\right)$ is not empty. The interval limits are chosen so that the CPF in (9.2) 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 Figure 1.1 in the introduction.

$$p_1 = e^{-\left(\frac{1}{\beta_0}\right)^{\frac{1}{\alpha_0-1}}}, \; p_2 = e^{-\left(\frac{1}{\beta}\right)^{\frac{1}{\alpha-1}}}, \; p_3 = e^{-\left(\frac{1}{\beta_1}\right)^{\frac{1}{\alpha_1-1}}}. \tag{9.4}$$

Result 9 *(al-Nowaihi and Dhami, 2010a, Proposition 9): (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 Result 8, the CPF in (9.2), (9.3) is a PWF in the sense of Definition 6. It will be helpful to bear in mind Figure 1.1 at this stage. By Result 9, 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) in subsection 5.3. Hence, a CPF also accounts for S2a.

The restrictions $\alpha_0 > 1$ and $\alpha_1 > 1$ in (9.3) ensure that a CPF has the following properties, that help explain human behavior for extremely low probability events.

Result 10 *(al-Nowaihi and Dhami, 2010a, Proposition 10c): Since for the first segment of the CPF, $\alpha_0 > 1$, hence, using Result 7(b) we get that for all $\gamma > 0$, a CPF, zero-underweights infinitesimal probabilities, i.e., $\lim_{p \to 0} \frac{w(p)}{p^\gamma} = 0$.*

al-Nowaihi and Dhami (2010a) provide the axiomatic foundations for the CPF, which the interested reader can pursue. They relies on a modification of their axiom of *power invariance* that they used in al-Nowaihi and Dhami (2006) to axiomatically derive the Prelec probability weighting function. Their new axiom is called *local power invariance*.[32]

[32] Local power invariance is defined as follows. 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 $\left(w\left(p^\lambda\right)\right)^\mu = w\left(q^\lambda\right)$. The al-Nowaihi and Dhami (2010a) axioms can also be directly imposed on lotteries.

9.1. An example of a composite Prelec function

Kunreuther et al. (1978, ch.7) report that the take-up of actuarially fair insurance begins to decline if the probability of the loss goes below 0.05. al-Nowaihi and Dhami (2010b) use the following CPF to fit this data, which is plotted in Figure 9.1.

$$w(p) = \begin{cases} e^{-0.19286(-\ln p)^2}, & i.e.,\ \alpha = 2, \beta = 0.19286 \quad if \quad 0 < p < 0.05 \\ e^{-(-\ln p)^{\frac{1}{2}}}, & i.e.,\ \alpha = 0.5, \beta = 1 \quad if \quad 0.05 \leq p \leq 0.95 \\ e^{-86.081(-\ln p)^2}, & i.e.,\ \alpha = 2, \beta = 86.081 \quad if \quad 0.95 < p \leq 1 \end{cases} \tag{9.5}$$

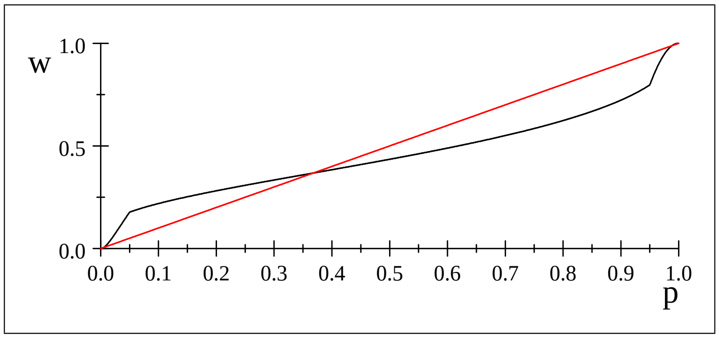


Figure 9.1: The composite Prelec function.

Comparing (9.2) with (9.5) we see that $\underline{p} = 0.05$ and $\overline{p} = 0.95$. For $0 \leq p < 0.05$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_0(-\ln p)^{\alpha_0}}$, with $\alpha_0 = 2$, $\beta_0 = 0.192\,86$. β_0 is chosen to make $w(p)$ continuous at $p = 0.05$. For $0.05 \leq p \leq 0.95$, the CPF is identical to the inverse-S shaped Prelec function of Figure 1.3 ($\alpha = 0.5$, $\beta = 1$). For $0.95 < p \leq 1$, the CPF is identical to the S-shaped Prelec function, $e^{-\beta_1(-\ln p)^{\alpha_1}}$, with $\alpha_1 = 2$, $\beta_1 = 86.081$. β_1 is chosen to make $w(p)$ continuous at $p = 0.95$.

Remark 14 *(Fixed points): This CPF has five fixed points, $p^* = w(p^*)$. These are 0, 0.0055993, e^{-1}, 0.98845 and 1. The CPF is strictly concave for $0.05 < p < e^{-1}$ and $0.95 < p < 1$ and strictly convex for $e^{-1} < p < 0.95$ and $0 < p < 0.05$.*

Remark 15 *(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 $e^{-1} = 0.367\,88 < p < 0.98845$. This accounts for stylized fact S1. Behavior near $p = 0$, and near $p = 1$, is not obvious from Figure 9.1. So, Figures 9.2 and 9.3, below, respectively magnify the regions near 0 and near 1. From Figure 9.2, we see that (9.5) underweights very low probabilities, in the range $0 < p < 0.0055993$. For p close to zero, we see that this PWF is nearly flat, thus, capturing Arrow's astute observation in the foreword to Kunreuther et al. (1978) "...it does appear from the data that the sensitivity*

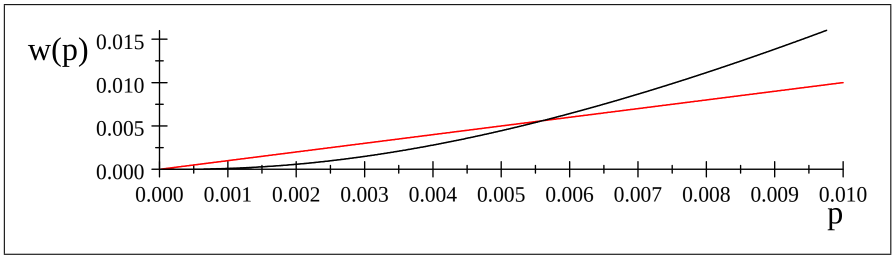


Figure 9.2: Behaviour of Figure 9.1 near 0.

goes down too rapidly as the probability decreases." From Figure 9.3, we see that (9.5) overweights very high probabilities, in the range $0.98845 < p < 1$. *For* p *close to one, we see that this PWF is nearly flat.*

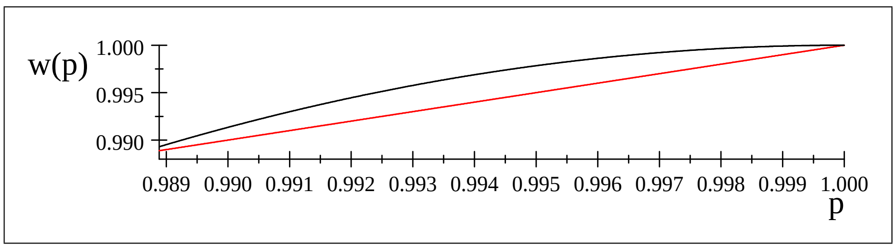


Figure 9.3: Behaviour of Figure 9.1 near 1.

10. Composite cumulative prospect theory (CCP) and composite rank dependent theory (CRDU)

We gave informal definitions of *composite prospect theory* (CCP) and *composite rank dependent utility* (RDU) in Remarks 5 and 6, which we now restate more formally.

Definition 18 *(Composite prospect theory (CCP), al-Nowaihi and Dhami, 2010a): Under CCP, a fraction* $1-\mu$ *of the population uses cumulative prospect theory, CP (see section 8), with its standard PWF, so* $\lim_{p\to 0}\frac{w(p)}{p^{\gamma}}=\infty$ *(see Remark 13). This fraction conforms to S1 for non-extreme probabilities and to S2b for extreme probabilities. The remaining fraction* μ *uses CP but replaces the standard PWF with the composite Prelec probability weighting function, CPF, defined in Definition 17 for which* $\lim_{p\to 0}\frac{w(p)}{p^{\gamma}}=0$ *(see Result 10). This fraction conforms to S1 for non-extreme probabilities and to S2a for extreme probabilities.*

Definition 19 *(CRDU; al-Nowaihi and Dhami, 2010a) In CRDU, RDU replaces CP in Definition 18.*

Other than the probability weighting function, CP and CCP share all other elements. The same holds true of RDU and CRDU. CCP can explain everything that CP can, which in turn can explain everything that RDU can, which in turn can explain everything that EU can. However, the reverse is not true. Hence, CCP is the most satisfactory theory under risk in economics.[33]

10.1. The Becker paradox under CRDU

The focus of our paper is not on CRDU. Hence, we state the following remark before focussing on CP/CCP in subsequent sections.

Remark 16 : *Notice that Proposition 3 under RDU holds for any PWF $w(p)$. In particular, it holds for the standard weighting functions such as the Prelec function, which are used in RDU, but also the composite Prelec weighting function used in CRDU. Thus, Proposition 3 also holds for composite rank dependent utility theory (CRDU). Therefore, the Becker paradox survives under all versions of rank dependent utility.*

11. A resolution of the Becker paradox using composite prospect theory

The results in Section 8 show that CP cannot explain the Becker paradox in the presence of fixed fines $F = \overline{F}$, which is a strong assumption. In actual practice, one observes a punishment function $F = \varphi(p)$, with some minimal desirable properties, such as substitutability of the instruments, p, F, given in (2.7). Section 8 also did not incorporate recent research on reference points. We address these issues for CP and CCP in this section. We make three assumptions, A1-A3, in this section.

Assumption A1 (Hyperbolic Punishment function, HPF): We have already considered, in section 2.4.2, the rationale for the HPF, $F = \varphi(p) = b/p$ (see Definition 5).[34]

Assumption A2 (Heterogeneity in reference points): As in the formulation of Schmidt et al. (2008), we allow for the reference incomes from the two activities, crime and no-crime, to differ. To enable a formulation that nests several interesting cases, we take the reference income from crime, y_c, as the *expected income from crime*, i.e.,

$$y_c = y_1 - p\varphi(p), \tag{11.1}$$

[33] See al-Nowaihi and Dhami (2010a) for these claims. CCP can also incorporate the insight from *third generation prospect theory* of Schmidt et al. (2008) about state dependent reference points. One may also refer to CCP as *fourth generation prospect theory.*

[34] The HPF is optimal for a wide class of cost and damage functions. Even when the HPF is not necessarily optimal, it provides an upper bound on punishments for a large and sensible class of cost and damage functions. Thus, if the HPF is unable to support the Becker proposition, then the optimal punishment functions, if different from the HPF, cannot support the Becker proposition either.

while the reference income from no-crime, y_{nc}, is specified as

$$y_{nc} = \lambda y_c = \lambda \left(y_1 - p\varphi \left(p \right) \right), \lambda \geq 0. \tag{11.2}$$

The *variable* λ is distributed across the population with some distribution $g(\lambda)$ on $\left[0, \overline{\lambda}\right]$, $\overline{\lambda} > 1$. This is the only source of individual heterogeneity in the model. For the HPF, $\varphi\left(p\right) = b/p$, and, so, using (2.1) we get that

$$p\varphi\left(p\right) = b \equiv y_1 - y_0. \tag{11.3}$$

Using (11.3), the reference incomes, y_c, y_{nc}, in (11.1) and (11.2) are:

$$y_c = y_1 - p\varphi\left(p\right) = y_0;\ y_{nc} = \lambda y_c = \lambda y_0. \tag{11.4}$$

The level of income from no-crime relative to the reference income from no-crime is

$$y_0 - y_{nc} = y_0 - \lambda y_0 = y_0(1 - \lambda). \tag{11.5}$$

Depending on the value of λ, we get three important cases.

(i) *Socially responsible individuals* ($0 \leq \lambda < 1$): In this case $\lambda < 1$, so from (11.5) $y_0(1-\lambda) > 0$. Such an individual feels positively rewarded on account of his honesty.

(ii) *Regretful individuals* ($1 < \lambda \leq \overline{\lambda}$): In this case $1 < \lambda$. (11.5) then implies that $y_0(1 - \lambda) < 0$. Such an individual experiences *regret* from not committing the crime and having to forego the higher income from crime.

(iii) *Individuals with reference income equal to the rational expectation of income* ($\lambda = 1$): The recent literature has suggested that the reference point could be the rational expectation of income, which in this perfect foresight model equals the expected income level.[35] The expected income from no-crime is y_0 while the expected income from crime is $y_1 - p\varphi\left(p\right)$, which using (11.3) also equals y_0. Clearly this case corresponds to a value of $\lambda = 1$ in (11.4) and we get $y_c = y_{nc} = y_0$.

Assumption A3: (Power form of utility): We use the utility function in (8.2). As already noted, this is consistent with the evidence and has axiomatic foundations.

Remark 17 *: Using (11.3), $y_0/\left(y_1 - p\varphi\left(p\right)\right) = 1$. Thus $\lambda \leq 1 \Leftrightarrow \lambda < y_0/\left(y_1 - p\varphi\left(p\right)\right)$ and $\lambda > 1 \Leftrightarrow \lambda \geq y_0/\left(y_1 - p\varphi\left(p\right)\right)$.*

[35] See, for instance, Koszegi and Rabin (2006), Crawford and Meng (2008)

We now compute the *income relative to the reference point* (the source of utility under CP and CCP) in each state for each activity. Using (11.2), for no-crime, this is

$$y_0 - y_{nc} = y_0 - \lambda\left(y_1 - p\varphi\left(p\right)\right) = y_1\left(1-\lambda\right) - b + \lambda p\varphi\left(p\right). \tag{11.6}$$

From (11.5) we know that $y_0 - y_{nc} = y_0(1-\lambda)$, but the seemingly more cumbersome form in (11.6) will turn out to be more convenient. Now suppose that the individual chooses the criminal activity instead. Using $y_c = y_1 - p\varphi\left(p\right)$ from (11.4), in the two states, caught and not-caught, income relative to the reference point is:

$$\begin{cases} y_1 - \varphi\left(p\right) - y_c = -(1-p)\varphi\left(p\right), & \text{if caught.} \\ y_1 - y_c = p\varphi\left(p\right), & \text{if not caught.} \end{cases} \tag{11.7}$$

From (11.7), the income from the criminal activity relative to the reference point if 'not caught' equals the expected fine $p\varphi\left(p\right)$ that the individual has avoided paying. This is the *elation* from the criminal activity, provided one gets away with it (see Definition 16). Substituting (11.6), (11.7) in the *NCC* (8.6) we get

$$w\left(p\right)v\left(-(1-p)\varphi\left(p\right)\right) + w\left(1-p\right)v(p\varphi\left(p\right)) \leq v\left(y_1\left(1-\lambda\right) - b + \lambda p\varphi\left(p\right)\right). \tag{11.8}$$

Using assumption A3 (the power form of utility), the NCC in (11.8) becomes:

$$-\theta w\left(p\right)\left(1-p\right)^{\gamma}\varphi\left(p\right)^{\gamma} + p^{\gamma}\varphi\left(p\right)^{\gamma}w\left(1-p\right) \leq \varphi\left(p\right)^{\gamma}\left(\frac{y_1\left(1-\lambda\right)-b}{\varphi\left(p\right)} + \lambda p\right)^{\gamma}. \tag{11.9}$$

We now use this condition to analyze the three cases: $0 \leq \lambda < 1$; $\lambda = 1; 1 < \lambda \leq \overline{\lambda}$.

11.1. Socially responsible individuals ($0 \leq \lambda < 1$)

Divide through by $p^{\gamma}\left(1-p\right)^{\gamma}\varphi\left(p\right)^{\gamma}$ to rewrite the NCC in (11.9) as:

$$-\theta\frac{w\left(p\right)}{p^{\gamma}} + \frac{w\left(1-p\right)}{\left(1-p\right)^{\gamma}} \leq \left(\frac{y_1\left(1-\lambda\right)-b}{p\left(1-p\right)\varphi\left(p\right)} + \frac{\lambda}{1-p}\right)^{\gamma} \tag{11.10}$$

Using Remark 17, since $\lambda < y_0/\left(y_1 - p\varphi\left(p\right)\right)$, so $\lambda < y_0/y_1$. Thus, an increase in fines, $F = \varphi\left(p\right)$, by increasing the RHS of the NCC, deters crime. For the hyperbolic fine function $\varphi\left(p\right) = b/p$ (assumption A2) we can write (11.10) as:

$$-\theta\frac{w\left(p\right)}{p^{\gamma}} + \frac{w\left(1-p\right)}{\left(1-p\right)^{\gamma}} \leq \left(\frac{y_1\left(1-\lambda\right)}{b(1-p)} - \frac{1-\lambda}{1-p}\right)^{\gamma} \tag{11.11}$$

Now take limits on both sides as $p \to 0$ to get

$$-\theta\lim_{p\to 0}\frac{w\left(p\right)}{p^{\gamma}} + 1 \leq \left(1-\lambda\right)^{\gamma}\left(\frac{y_1}{b} - 1\right)^{\gamma} \tag{11.12}$$

We now consider separately the cases of cumulative prospect theory (CP)[36] and composite prospect theory (CCP).

[36] Recall that section 8 only considered CP in the special case of fixed reference points and fixed fines.

Proposition 5 *(CP): Socially responsible individuals* ($0 \leq \lambda < 1$)*, who use CP are completely dissuaded from crime, i.e., the Becker proposition holds.*

Proposition 5 is obvious from (11.12). Under CP, $\lim_{p\to 0} \frac{w(p)}{p^{\gamma}} = \infty$ (see Remark 13), so the NCC in (11.12) holds. A decision maker who uses CP, overweights low probabilities to such an extent that the salience of the punishment overwhelms all other considerations, such as the relative benefit from the crime. Such a decision maker is completely dissuaded from crime by Becker-type punishments and, hence, the Becker paradox remains.

Proposition 6 *(CCP): Consider socially responsible individuals* ($0 \leq \lambda < 1$) *who follow CCP. A fraction* $1 - \mu$ *of the these decision makers follow CP, hence, they behave as in Proposition 5. The remaining fraction,* μ*, uses the composite Prelec function, CPF. For these individuals, the Becker proposition fails when income from crime is at least twice as large as the income from no-crime (i.e.,* $y_1 \geq 2y_0$*).*

From Proposition 6, a fraction $1-\mu$ of socially responsible individuals are deterred from crime by Becker-type punishments. However, the remaining fraction, μ, is not deterred if the income from crime is at least twice as great as the income from no-crime. For these individuals, $\lim_{p\to 0} \frac{w(p)}{p^{\gamma}} = 0$ (see Definition 18) and, so, unlike a CP decision maker, they place relatively low salience on the possibility of the punishment. Furthermore, they derive *elation* from getting away with the act of crime (see Definition 16 and the discussion following (11.7)).

Because a socially responsible individual derives gratification from honesty, a higher inducement in criminal benefits is needed to commit the crime that requires $y_1 \geq 2y_0$. In conjunction, for the two types of decision makers under CCP, with respective fractions μ, $1 - \mu$ we can explain why some individuals are deterred and others are not deterred in the face of Becker-type punishments. Since the intuition is similar to remaining cases, we shall be brief.

We derive even stronger results in the remaining two subsections for *regretful individuals* and *individuals with rational expectations*. For such individuals, Proposition 6 holds unconditionally, i.e., we require no restrictions on benefits from crime except that $b > 0$.

11.2. Regretful individuals ($1 < \lambda \leq \overline{\lambda}$)

In this case, $y_0 - y_{nc} = y_0(1 - \lambda) < 0$, thus, the NCC in (11.8) becomes:

$$-\theta w\left(p\right)\left(1-p\right)^{\gamma}\varphi\left(p\right)^{\gamma}+p^{\gamma}\varphi\left(p\right)^{\gamma}w\left(1-p\right)\leq-\theta\varphi\left(p\right)^{\gamma}\left(\frac{b-y_{1}\left(1-\lambda\right)}{F}-\lambda p\right)^{\gamma}. \quad (11.13)$$

Proposition 7 *(CP): Consider regretful individuals* ($1 < \lambda \leq \overline{\lambda}$) *who follow CP: For any of the standard probability weighting functions (for which* $\lim_{p\to 0} \frac{w(p)}{p^{\gamma}} = \infty$*) under CP, the NCC is satisfied and so the individual does not commit the crime*

Proposition 8 *(CCP): Now suppose that the decision maker uses CCP. A fraction* $1-\mu$ *of individuals use CP, so Proposition 7 applies to them. The Becker proposition fails for the remaining fraction,* μ*, which uses the composite Prelec function, that has the property* $\lim_{p\to 0}\frac{w(p)}{p^{\gamma}}=0$ *(see Definition 18). Thus, the Becker paradox is resolved.*

11.3. The case of rational expectations $(\lambda = 1)$.

For $\lambda = 1$, the reference point for each activity is the expected income arising from that activity. Thus, from Assumption A3(iii) $y_c = y_{nc} = y_0$. Thus, $y_0 - y_{nc} = 0$, while, from (11.7), $y_1 - y_c = p\varphi(p)$ and $y_1 - \varphi(p) - y_c = -(1-p)\varphi(p)$. Hence, the *NCC* (8.6) is

$$w(p)\, v(-(1-p)\varphi(p)) + w(1-p)\, v(p\varphi(p)) \leq 0. \tag{11.14}$$

Using assumption A3 (the power form of utility), the *NCC* (11.14) becomes:

$$-\theta(1-p)^{\gamma}\varphi(p)^{\gamma} w(p) + p^{\gamma}\varphi(p)^{\gamma} w(1-p) \leq 0. \tag{11.15}$$

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

$$\frac{w(p)}{p^{\gamma}} \geq \frac{w(1-p)}{\theta(1-p)^{\gamma}}. \tag{11.16}$$

Proposition 9 *(CP): Consider a decision maker who uses CP and the reference income is the rational expectation of income. Then the Becker paradox reemerges under CP and fines have no deterrent effect.*

The result in Proposition 9 runs counter to one's intuition for the case of individuals with rational expectations. First, fines have no deterrent effect (unlike the cases of *socially responsible* and *regretful individuals*), which runs counter to the evidence; see, e.g., Levitt (2004). Second, extremely small probabilities of detection eliminate crime, and so it is also optimal to set vanishingly small fines. This is also not borne out by the evidence. Hence, the combination of standard weighting functions and rational expectations on the part of criminals would seem not to be supported by the evidence. Proposition 10 considers the case when decision makers use CCP, and the results conform much better to the evidence.

Proposition 10 *: Suppose that the decision maker uses CCP and the reference income is the rational expectations of income. The result in Proposition 9 applies to a fraction* $1-\mu$*, of individuals. For the remaining fraction,* μ*, the Becker paradox is solved.*

In each of these alternative specifications of the reference point, CCP is able to resolve the Becker paradox but EU, RDU, CP cannot.

12. Conclusions

The *Becker proposition*, summarized eloquently in Kolm's (1973) phrase "hang offenders with probability zero", is a cornerstone in the 'economics and law' literature and has provided the basis for much further development of the field. However it is often noted that it does not hold empirically for all individuals. This, we called the Becker paradox.

A sizeable literature addresses the Becker paradox in an *expected utility* (EU) framework. We argue that it is very difficult to explain the evidence on the basis of EU. Hence, we re-examine the Becker paradox from the perspective of alternative mainstream non-linear decision theories. We show that the Becker paradox reemerges under *rank dependent expected utility* (RDU), which, Machina (2008) describes as the most popular alternative to the EU model. The Becker paradox also reemerges under cumulative prospect theory (CP), the Noble prize winning work of Tversky and Kahneman (1992).

At the heart of the Becker paradox, "hang offenders with probability zero", is the behavior of individuals for low probability events. We argue that the only successful resolution of the Becker paradox (and we consider about 10 different alternatives) is the one based on al-Nowaihi and Dhami's (2010a) *composite cumulative prospect theory* (CCP). CCP combines prospect theory (Kahneman and Tversky, 1979) and cumulative prospect theory (Tversky and Kahneman, 1992). The main insight, emerging from strong evidence, is that while people overweight small probabilities and underweight large ones, *they also ignore extremely small probabilities and treat extremely large probabilities with near certainty.*

The explanation of the Becker paradox under CCP is robust to several plausible assumptions about the reference points of potential criminals. These allow us to explicitly build-in heterogeneity among potential criminals into our analysis. We allow for rational expectations, regret and social responsibility while considering heterogeneity issues. In each case, so long as the decision maker is rational (as opposed to being pathological) CCP allows the Becker paradox to be explained. The Becker paradox has been a particularly elusive problem to explain, and the success of CCP provides independent verification of that theory.

Acknowledgements: We are grateful to useful comments from seminar participants at the 2010 XIV FUR conference in Newcastle and comments from the audience at University of Leicester. Dr. Sanjit Dhami would acknowledge a one semester study leave during which this research was partly carried out.

13. Appendix: Proofs of the Propositions

Proof of Proposition 1: If $\varphi(p) = \frac{b}{p}$ then the *NCC* (3.4) holds for concave u and, hence, (a) follows. If $\varphi(p) = \frac{b}{p}$, then (from Definition 4) $\eta_p^F = -\frac{p}{F}\frac{d\varphi}{dp} = 1$. Since $\eta_p^C > \eta_F^C$ it follows, from Lemma 2, that $\frac{d}{dp}C(p, \varphi(p)) > 0$. Since crime is deterred, $D(p, \varphi(p)) = 0$. Hence, $T(p, \varphi(p)) = C(p, \varphi(p))$ which implies that $\frac{d}{dp}T(p, \varphi(p)) > 0$. Since the objective is to minimize $T(p, \varphi(p))$, this establishes part (b). ■.

Proof of Proposition 2: Follows from the definition of $u_{\min}$, (7.5) and (7.6). Since $u_{\min}$ can be made as small as possible by raising fines. In particular, if $u_{\min} \to -\infty$ as $F_{\max} \to -\infty$, then $p_{\min} = 1 - w^{-1}(1) = 0$. Hence, the result follows. ■.

Proof of Proposition 3: Similar to the proof of Proposition 1, except for the following points: (a) $\varphi(p) = \frac{b}{1-w(1-p)}$ (instead of $\varphi(p) = \frac{b}{p}$) and (b) $\eta_p^F = \frac{pw'(1-p)}{1-w(1-p)}$ (instead of $\eta_p^F = 1$). ■

Proof of Proposition 4: Let the probability of detection and conviction be $p > 0$. Then the *NCC* (8.6) can be satisfied by taking the punishment, $F = \overline{F}$, to be sufficiently large. ■

Proof of Proposition 5: This follows directly by substituting $\lim_{p\to 0}\frac{w(p)}{p^\gamma} = \infty$ (which holds for under CP) in (11.12). ■

Proof of Proposition 6: Now suppose that the decision maker follows CCP. It is clear that the fraction $1-\mu$ of individuals follow CP and so the proof of Proposition 5 applies to them. For individuals who belong to the remaining fraction μ (see Definition 18), $\lim_{p\to 0}\frac{w(p)}{p^\gamma} = 0$. Then as $p \to 0$, from (11.12), we get that the NCC holds if

$$1 \leq (1-\lambda)\left(\frac{y_1}{b} - 1\right) \Leftrightarrow \lambda \leq 1 - \frac{b}{y_0} \equiv 2 - \frac{y_1}{y_0},$$

where $\frac{b}{y_0}$ is the benefit of the crime as a fraction of the no-crime income. If follows, conversely that the NCC is not satisfied and so hanging offenders with probability zero does not work if

$$\lambda > 1 - \frac{b}{y_0} \equiv 2 - \frac{y_1}{y_0}.$$

This completes the proof. ■

Proof of Proposition 7: Divide (11.13) through by $-p^\gamma(1-p)^\gamma\varphi(p)^\gamma$ to rewrite the NCC in (11.9) as:

$$\theta\frac{w(p)}{p^\gamma} - \frac{w(1-p)}{(1-p)^\gamma} \geq \theta\left(\frac{b - y_1(1-\lambda)}{p(1-p)\varphi(p)} - \frac{\lambda}{1-p}\right)^\gamma. \tag{13.1}$$

Using remark 17, we know that $\lambda > y_0/(y_1 - p\varphi(p))$, but this does not necessarily imply that $\lambda > y_0/y_1$. If $\lambda > y_0/y_1$ then an increase in fines, $F = \varphi(p)$, by reducing the RHS of the NCC, reduces crime. Note that the RHS of (13.1) is positive and so the inequality is well defined. For the hyperbolic fine function $\varphi(p) = b/p$ we can write (13.1) as:

$$\theta \frac{w(p)}{p^\gamma} - \frac{w(1-p)}{(1-p)^\gamma} \geq \theta \left(\frac{b - y_1(1-\lambda)}{b(1-p)} - \frac{\lambda}{1-p} \right)^\gamma .$$

Now take limits on both sides as $p \to 0$ to get

$$\theta \lim_{p \to 0} \frac{w(p)}{p^\gamma} - 1 \geq \theta \left(\frac{b - y_1(1-\lambda)}{b} - \lambda \right)^\gamma . \tag{13.2}$$

The result now follows directly by substituting $\lim_{p\to 0} \frac{w(p)}{p^\gamma} = \infty$ in (13.2).

Proof of Proposition 8: Now suppose that decision makers follow CCP. A fraction $1-\mu$ uses CP and so the proof of Proposition 7 applies to them. The remaining fraction, μ, uses the composite Prelec function and so $\lim_{p\to 0} \frac{w(p)}{p^\gamma} = 0$ (see Definition 18). For these individuals, the NCC in (13.2) implies that

$$-1 \geq \theta \left(\frac{b - y_1(1-\lambda)}{b} - \lambda \right)^\gamma .$$

The RHS is positive while the LHS is negative, hence, the NCC is violated. Indeed the decision maker will not be dissuaded from crime, i.e., the Becker proposition does not hold. And so, the Becker paradox is resolved for these individuals. ■

Proof of Proposition 9: Using (11.16), as the probability of detection approaches zero, the decision maker does not engage in the criminal activity if $\lim\limits_{p\to 0} \frac{w(p)}{p^\gamma} > \frac{1}{\theta}$ but engages if $\lim\limits_{p\to 0} \frac{w(p)}{p^\gamma} < \frac{1}{\theta}$. None of these conditions involve fines, hence, fines have no deterrent effect. When $\lim\limits_{p\to 0} \frac{w(p)}{p^\gamma} > \frac{1}{\theta}$, the *NCC* (11.16) will hold with strict inequality in some non-empty interval $(0, p_1)$.[37] Hence, no crime will occur if $p \in (0, p_1)$. Conversely, if $\lim\limits_{p\to 0} \frac{w(p)}{p^\gamma} < \frac{1}{\theta}$, then the *NCC* (11.16) is violated in some non-empty interval $(0, p_2)$. Hence, for punishment to deter in this case, we must have $p > p_2$. For any of the standard Probability weighting functions in CP, $\lim\limits_{p\to 0} \frac{w(p)}{p^\gamma} = \infty$ (see Remark 13) and so the Becker paradox survives for all individuals. ■

Proof of Proposition 10: Now suppose that decision makers follow CCP. A fraction $1-\mu$ uses CP and so the proof of Proposition 9 applies to them. For the remaining fraction, μ, since $\lim\limits_{p\to 0} \frac{w(p)}{p^\gamma} = 0$, it follows, from Proposition 9, that, for some non-empty interval $(0, p_2)$, no level of punishment, F, no matter how large, will deter crime, if $p \in (0, p_2)$. ■

Acknowledgements: We are very grateful for insightful comments made at the presentation of this paper at the 2010 FUR XIV conference in Newcastle and seminars at the University of Leicester and ISET in Georgia. Dr. Sanjit Dhami would also like to acknowledge a one semester study leave from the University of Leicester.

[37] 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 $1 - p \to 1$, $w(1) = 1$.

References

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

[2] al-Nowaihi, A., Bradley, I. and Dhami, S., 2008. The Utility Function Under Prospect Theory. *Economics Letters*. 99, 337-339.

[3] al-Nowaihi, A., and Dhami, S., 2006. A Simple Derivation of the Prelec Probability Weighting Function. *Journal of Mathematical Psychology*. 50, 521-524.

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

[5] al-Nowaihi, A., and Dhami, S., 2010b. The behavioral economics of the demand for insurance. *University of Leicester Discussion Paper*.

[6] Andreoni, J., 1991. Reasonable doubt and the optimal magnitude of fines: Should the penalty fit the crime? *RAND Journal of Economics*. 22, 385-95.

[7] Bar Ilan, A., 2000. The Response to Large and Small Penalties in a Natural Experiment. Department of Economics, University of Haifa, 31905 Haifa, Israel.

[8] Bar Ilan, A., and Sacerdote, B., 2001. The response to fines and probability of detection in a series of experiments. *National Bureau of Economic Research Working Paper* 8638.

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

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

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

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

[13] Crawford, V. P., and Meng, J., 2008. New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. mimeo. University of California, San Diego.

[14] Dhami, S., and al-Nowaihi, A., 2010a. Hyperbolic Punishment Functions. *University of Leicester Discussion Paper Series.*

[15] Dhami, S., and al-Nowaihi, A., 2010b. Optimal income taxation in the presence of tax evasion: Expected utility versus prospect theory. *Journal of Economic Behavior and Organization.* 75, 313-337.

[16] Dhami, S., al-Nowaihi, A., 2007. Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behavior and Organization.* 64, 171–192.

[17] Feess, E., and Wohlschlegel, A., 2009. Why higher punishments may reduce deterrence. *Economics Letters.* 104, 69-71.

[18] Friedman, D., 1999. Why Not Hang Them All: The Virtues of Inefficient Punishment. *Journal of Political Economy.* 107, S259-S269.

[19] Garoupa, N., 2001. Optimal magnitude and probability of fines. *European Economic Review.* 45, 1765-1771.

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

[21] Jolls, C., Sunstein, C. R., and Thaler, R., 2000. A Behavioral Approach to Law and Economics. reprinted in Cass R. Sunstein, ed., *Behavioral Law and Economics*, Cambridge University Press, Cambridge.

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

[23] Kahneman, D., and Tversky, A., 2000. *Choices, values and frames.* Cambridge University Press, Cambridge.

[24] Kaplow, L., 1992. The optimal probability and magnitude of fines for acts that definitely are undesirable. *International Review of Law and Economics*, 12, 3-11.

[25] Kolm, S-C., 1973. A Note on Optimum Tax Evasion. *Journal of Public Economics.* 2, 265-270.

[26] Koszegi, B., and Matthew, R., 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics.* 121, 1133-1165.

[27] Kunreuther, H.R., Desvousges, W. H., and Slovic, P., 1988. Nevada's predicament: public perceptions of risk from the proposed nuclear waste repository. Environment, 30, 16-20.

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

[29] 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.

[30] 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.

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

[32] 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.

[33] McClelland, G. H., Schulze, W. D., and Coursey, D. L., 1993. Insurance for low-probability hazards: A bimodal response to unlikely events. Journal of Risk and Uncertainty. 7, 95-116.

[34] Polinsky, M., and Shavell, S., 2007. The theory of public enforcement of law. Chapter 6 in Polinsky, M., and Shavell, S., (eds.) *Handbook of Law and Economics*, Vol. 1, Elsevier.

[35] Polinsky, M., and Shavell, S., 2000a. The fairness of sanctions: Some implications for optimal enforcement policy. *American Law and Economics Review*. 2, 223-237.

[36] Polinsky, M., and Shavell, S., 2000b. The Economic Theory of Public Enforcement of Law. *Journal of Economic Literature*,. 38, 45-76.

[37] Polinsky, M., and Shavell, S., 1991. A note on optimal fines when the wealth varies among individuals. *American Economic Review*. 81, 618-621.

[38] Polinsky, M., and Shavell, S., 1979. The optimal trade-off between the probability and magnitude of fines. *American Economic Review*. 69, 880-891.

[39] 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.

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

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

[42] Quiggin, J., 1993. *Generalized Expected Utility*. Kluwer-Nijhoff, Amsterdam.

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

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

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

[46] Schmidt, U., Starmer, C., and Sugden, R., 2008. Third-generation prospect theory. Journal of Risk and Uncertainty. 36, 203-223

[47] 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.

[48] Sugden, R., 2003. Reference-dependent subjective expected utility. Journal of Economic Theory. 111, 172-191.

[49] Sydnor, J., 2010. (Overinsuring) modest risks. Forthcoming in the American Economic Journal: Applied Economics.

[50] *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.

[51] Taylor, S. E., and Brown, J. D., 1998. Illusion and Well-Being: A Social Psychological Perspective on Mental Health. Psychological Bulletin. 103, 193-210.

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

[53] Viscusi, K. W., 1998. Rational Risk Policy: The 1996 Arne Ryde Memorial Lectures. Oxford University Press, Oxford.