An extension of the Becker proposition to non-expected utility theory

Sanjit Dhami* **Ali al-Nowaihi†**

7 September 2011

Abstract

In a seminal paper, Becker (1968) showed that the most *efficient* way to deter crime is to impose the severest possible penalty (to maintain adequate deterrence) with the lowest possible probability (to economize on costs of enforcement). We shall call this the *Becker proposition* (BP). The BP is derived under the assumptions of expected utility theory (EU). However, EU is heavily rejected by the evidence. A range of non-expected utility theories have been proposed to explain the evidence. The two leading alternatives to EU are rank dependent utility (RDU) and cumulative prospect theory (CP). The main contributions of this paper are: (1) We formalize the BP in a more satisfactory manner. (2) We show that the BP holds under RDU and CP. (3) We give a formal behavioral approach to crime and punishment that could have applicability to a wide range of problems in the economics of crime.

Keywords: Crime and punishment, non-linear weighting of probabilities, cumulative prospect theory, rank dependent utility, probability weighting functions, punishment functions.

JEL Classification: D03 (Behavioral Economics: Underlying Principles), 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`.

"*If the supply of offenses depended only on $pf-$ offenders were risk neutral $-$ a reduction in p...would lower the loss...The loss would be minimized, therefore, by lowering* p *arbitrarily close to zero and raising* f *sufficiently high so that the produce* pf *would induce the optimal number of offenses.*" Becker (1968, p. 183).

1. Introduction

In a seminal contribution, Becker (1968) opened the way to a rigorous formal economic 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, as illustrated in Becker's quote at the beginning.[1] Kolm (1973) memorably phrased this as *hang offenders with probability zero.*

The Becker proposition (BP) has had a tremendous impact on the economics and law literature and it introduced a formal economic approach to crime.[2] The BP has spawned a great deal of literature that typically relaxes some assumption behind the BP or embeds the BP within a more complex setting in order to see if the BP holds.

1.1. How robust is the BP?

The BP might not hold for a variety of reasons. For instance, criminals could be risk-seeking (Becker, 1968). Infinite fines could be difficult to collect due to bankruptcy constraints (Friedman, 1999; Polinsky and Shavell, 1991). Because of the danger of falsely convicting an innocent person (a Type-I error), society could face huge welfare losses from large fines (Andreoni, 1991; Feess and Wholschlegel, 2009; Polinsky and Shavell, 2000b). High fines might encourage rent-seeking behavior on the part of law-enforcers to collude in acquitting criminals (Friedman, 1999). For reasons of norms and fairness society may

[1] In the quote, p is the probability of detection and conviction and f is the punishment (monetary and non-monetary).

[2] The salience of the Becker proposition in that literature can be gauged from the fact that a Google Scholar search shows that it has been cited 7050 times

not accept the severe punishments required by the BP (Polinsky and Shavell, 2000a). Risk averse criminals whose utility enters the social welfare function would be severely hit by very high fines in the event that they are caught, lowering social welfare (Polinsky and Shavell,1979; Kaplow, 1992). Pathological criminals are not deterred by high fines anyway (Colman, 1995).

The extensions of the Becker proposition noted above have enriched the literature and deepened our understanding of the economics of crime. Like Becker (1968), these extensions use expected utility theory (EU), a decision theory whose predictions, as we discuss below, are routinely refuted by the evidence .

Our first objective is to formalize the BP. Our second objective is to check for the robustness of BP with respect to alternative decision theories that are better supported by the evidence as compared to EU. The third objective of our paper is to outline a formal behavioral approach to crime and punishment that can potentially have a range of applications to related problems.

1.2. Expected utility theory, its refutations and the leading alternatives

A common feature of all the robustness tests of the BP briefly outlined in subsection 1.1 above, is that they assume criminals follow *expected utility theory* (EU). Despite its overwhelming popularity in the literature, EU has been strongly rejected by the evidence gathered over several decades of research. The Allais paradox, in its common ratio and common consequence forms, was an early demonstration of the violation of the independence axiom of EU. Other well documented violations have included preference reversals, violation of framing invariance, violation of the reduction of compound lotteries axiom, violation of the dependence of utility on final wealth rather than wealth relative to a reference point (reference dependence), conflation of risk aversion with loss aversion, attitudes to risk that depend only on the shape of the utility function, unreasonable risk aversion for large stake gambles if there is risk aversion over small stakes, qualitatively and quantitatively incorrect results when pitted in a race with other decision theories, among many others.[3]

EU has two salient features. (1) The objective function is linear in probabilities, and (2) utility is defined over final levels of wealth. For a range of problems, such as the Allais paradox, it was found that the relaxation of linearity in probabilities provided a solution. This lead to the development of several alternatives. According to Machina (2008) the most popular of these alternatives is *rank dependent utility* (RDU), developed by Quiggin (1982). However, RDU retains the second salient feature of EU, namely that utilities are

[3] The literature documenting violations of EU has assumed huge proportions. Comprehensive reviews of the violation of EU can be found in Kahneman and Tversky (2000), Starmer (2000), Dhami and al-Nowaihi (2007, 2010b), Wakker (2010), Crawford and Meng (2011).

defined over final wealth levels. RDU can explain everything that EU can explain because EU is a special case of RDU. But, in addition, RDU can explain a range of phenomena that EU is unable to explain.

Empirical evidence arising from several economic phenomena demonstrates strongly, and robustly, the following facts; see Kahneman and Tversky (1979, 2000), Starmer (2000) and Wakker (2010). Individuals derive utility not from final wealth levels but changes in wealth relative to a reference level (*reference dependence*).[4] This partitions the domain of outcomes into gains (when wealth exceeds the reference point) and the domain of losses (when wealth is lower than the reference wealth). Individuals exhibit *loss aversion*, namely the tendency for losses to bite more than equivalent gains. Furthermore, the utility function is concave in the domain of gains and convex in the domain of losses (diminishing sensitivity). We discuss these features in detail in section 6.

The award of the Nobel Prize to Daniel Kahneman was in recognition for joint work, largely with Amos Tversky.[5] They documented severe violations of EU and provided in its place a successful alternative in the form of *prospect theory* (PT) also known as first generation prospect theory. PT is described in Kahneman and Tversky (1979), which is the second most cited paper in economics since 1970 (see Dellavigna, 2009). In PT, both salient features of EU, namely, linearity in probabilities and utility defined over final wealth, are relaxed. In particular, the second of these two salient features is relaxed by allowing for reference dependence, loss aversion and declining sensitivity. All these features of PT are based on strong experimental evidence. This work was followed by second generation prospect theory, also known as *cumulative prospect theory* (CP) in Tversky and Kahneman (1992), who incorporated insights from Quiggin's work on RDU into PT (see section 6 below). In conjunction, PT and CP are able to address most of the shortcomings of EU briefly mentioned in the opening paragraph of this subsection.

Many economic phenomena require reference dependence, loss aversion and declining sensitivity as essential components. As one might expect, EU and RDU typically do not perform well in the domain of these economic phenomena while PT and CP do far better. The range of such phenomena is very wide and important. In the context of criminal activity, specifically tax evasion, Dhami and al-Nowaihi (2007) show that the predictions of EU are both quantitatively incorrect (by up to 100 fold) and qualitatively incorrect.[6] This raises great difficulty in applying EU to empirical or theoretical work in

[4] In a classic experiment from psychology an individual is asked to dip one hand (hand L) in hot water and another hand (hand R) in cold water. Once the hands become attuned to the respective temperatures, both hands are taken out simultaneously and dipped in lukewarm water. Hand L reports the lukewarm water to be cold while hand R reports the lukewarm water to be hot.

[5] Daniel Kahneman won the 2002 Nobel prize in Economics. Sadly, Amos Tversky died in 1996 and there is no provision to receive the Nobel prize postumously.

[6] The qualitative incorrectness in this case refers to the Yitzhaki puzzle. Under plausible attitudes to

economics. By contrast, the evidence is easily explained by CP. Camerer (2000) documents the superior performance of PT/CP for a range of phenomena that include the disposition effect, asymmetric price elasticities, the excess sensitivity of consumption to income, the equity premium puzzle, elasticities of labour supply and asset pricing, among others.[7]

CP can explain everything that EU and RDU can, but the converse is false, hence, CP is the most satisfactory decision theory among the leading alternatives. Indeed, EU and RDU are special cases of CP. Furthermore, CP is very tractable in actual applications. The continued use of EU in the face of ever mounting refutations, better developed, equally tractable and equally rigorous alternative theories that conform much better with the evidence is one of the great puzzles in the economics profession. We do not assume a knowledge of RDU and CP on the part of the reader in this paper. The relevant concepts are developed during the paper when needed.

1.3. Research problems, results and some implications

The BP has been extended in several directions (see subsection 1.1 above), yet the analysis has steadfastly continued to assume EU in most cases despite EU having faced intense refutations for several decades. We find this puzzling and surprising.[8] A major research aim of our paper is to test if the BP holds under RDU and CP. We find that the BP is robust to a consideration of alternative mainstream decision theories. In particular, we find that the BP holds for RDU as well as for CP. We also provide a clearer derivation of the BP under EU. A related contribution of our paper is that by addressing a basic crime and punishment problem in a non-expected utility framework, it opens up the way for a consideration of other problems and issues using our framework.

Our findings have mixed implication for alternative explanations of the violations of the BP outlined briefly in subsection 1.1 above. On the one hand, our findings strengthen these alternative explanations because if the BP is violated, it is not on account of a refuted theory, namely, EU. We show that the BP continues to hold under RDU and CP,theories which are confirmed much better by the evidence. In this sense, the research program for alternative explanations was fortuitously headed in the right direction. On the other hand, because the alternative explanations themselves continued to use EU, it might be

risk, EU predicts that when the tax rate increases, taxpayers will evade less. In the limit when the tax rate is a 100%, all income is declared. This contradicts the bulk of evidence; see Dhami and al-Nowaihi (2007) for the details.

[7] The interested reader can also consult Kahneman and Tversky (2000), the Nobel lecture by Kahneman (2003) and a review of field studies in Dellavigna (2009).

[8] Neilson and Winter (1997) make an attempt to apply RDU to a problem of crime but their intention was not to see if BP holds under RDU. They were interested in showing that even if individuals violate EU they could still be risk averse. A much more complete treatment of attitudes to risk under RDU, arising as the product of the shapes of the utility function and the probability weighting function can be found in Wakker (2010).

worthwhile to re-examine them in the light of RDU and CP to check if they continue to hold. The latter remark is not purely speculative. In the context of the criminal activity of tax evasion, Dhami and al-Nowaihi (2007, 2010b), show that EU gives results that are qualitatively incorrect and quantitatively incorrect.

1.4. Organization of the paper

Section 2 formulates a standard economic model of crime. The Becker proposition is considered under EU in Section 3. Section 4 gives a brief discussion of probability weighting functions, with particular emphasis on the Prelec (1998) function. Section 5 shows that the Becker proposition holds under RDU and Section 6 shows it also holds under CP. We consider fixed reference points in subsection 6.2, while subsection 6.3 allows for a wider range of reference points as well as heterogeneity among criminals. Section 7 concludes. All proofs are relegated to the Appendix.

2. The model and assumptions

Suppose that an individual receives income, y_0, from being engaged in some legal activity and income, y_1, $y_1 > y_0$, from being engaged in some illegal activity. Hence, the benefit, b, to the perpetrator from the illegal activity is

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

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 much of the existing literature, particularly Becker (1968), we use the simplifying assumption that F is the monetary equivalent of all punishments.[9] We assume that,

$$F \in [0, \infty]. \tag{2.2}$$

In particular, it is feasible to levy a fine at least equal to the benefit from crime, b. This is true of Examples 1 - 3 that we give below, consistent with real world practice and is an assumption that is made in standard models of crime and punishment; see, for instance, Polinsky and Shavell (2007). Given the enforcement parameters p, F, the individual makes only one choice, to commit the crime or not.

This framework nests several important settings. Consider the following examples.

[9] Thus, we are assuming that it is possible to impute a monetary value to all possible kinds of punishments such as incarceration, loss of reputation, naming and shaming etc. While this is not innocuous, it simplifies the analysis substantially.

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*.*[10] *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,* $F = (1+\delta) t z_2$*.*

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. Emissions 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*. Let $D(p, F)$ be the *damage to society caused by crime*. We assume that $C(0, 0) = 0$, i.e., in the absence of any law enforcement, costs of such enforcement are zero. If crime is deterred completely by $p = p_0$ and $F = F_0$ then, we assume, $D(p_0, F_0) = 0$.

We also assume that C and D are continuous functions of p and F with continuous first and second partial derivatives[11], i.e., $C, D \in C^2$. We denote partial derivatives with

[10] 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).

[11] C^n is the class of continuous functions with continuous partial derivatives up to order n.

subscripts, e.g., $C_p = \frac{\partial C}{\partial p}$ and $C_{pF} = \frac{\partial^2 C}{\partial p \partial F}$. We assume that

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

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). 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).*

2.2. Society's objective

Let

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

be the total cost to society of crime. We assume that

$$[T_F]_{F=0} < 0. \tag{2.5}$$

Society aims to choose the instruments p and F so as to minimize $T(p, F)$. The condition in (2.5) ensures that total costs can be reduced by raising the fine just above zero, hence, $F = 0$ is not optimal.

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 $\varphi(p) : [0,1] \rightarrow [0,\infty]$ that assigns to each probability of detection and conviction, $p \in [0,1]$, a punishment $\varphi(p) \in [0,\infty]$.*

Note that we allow for the possibility of infinite punishments, i.e., we allow for the possibility that $\varphi(p) = \infty$ for some p.

Definition 3 *(Optimal punishment function): Let $\varphi : [0,1] \rightarrow [0,\infty]$ be a punishment function. We call φ an optimal punishment function if, for all $p \in [0,1]$ such that $\varphi(p) < \infty$, and for all $F \in [0,\infty)$, $T(p, \varphi(p)) \leq T(p, F)$.*

Proposition 1 *(Existence of optimal punishment functions):*
(a) An optimal punishment function, $\varphi(p) : [0,1] \rightarrow [0,\infty]$, exists.
(b) If $\varphi(p) < \infty$ then $[T_F(p,F)]_{F=\varphi(p)} = 0$ and $[T_{FF}(p,F)]_{F=\varphi(p)} \geq 0$.
(c) If $\varphi(p) < \infty$ and $[T_{FF}(p,F)]_{F=\varphi(p)} > 0$ then $\varphi'(p) = -\left[\frac{T_{pF}}{T_{FF}}\right]_{F=\varphi(p)}$.
(d) If $\varphi(p) < \infty$, $[T_{FF}(p,F)]_{F=\varphi(p)} > 0$ and $[T_{pF}(p,F)]_{F=\varphi(p)} > 0$ then $\varphi'(p) < 0$.

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.*

Lemma 1 *: $\frac{d}{dp} C(p, \varphi(p)) > 0$ 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).

2.5. The hyperbolic punishment function

A popular and tractable subset of $\varphi(p)$ is the hyperbolic punishment function (HPF), $H(p)$.

Definition 5 *: A hyperbolic punishment function, HPF is defined by*

$$H(p) = c/p. \tag{2.6}$$

where c is a positive constant.

The name derives from the fact that in p, F space, the HPF plots as a rectangular hyperbola. Note that, for (2.6), $\varphi(0) = \infty$. The justification for the hyperbolic punishment function is considered in al-Nowaihi and Dhami (2011a). The HPF has been widely used in the law and economics field.[12]

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

We now consider the Becker proposition (BP) under EU. One of the objectives of this section is to give a more satisfactory formal treatment of the BP.[13] 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 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

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

Proposition 2 *(Becker, 1968): Under EU, if the utility function is unbounded below, so that $u(y_1 - F) \to -\infty$ as $F \to \infty$, then, for any probability of punishment $p > 0$, no matter how small, crime can be deterred by a sufficiently severe punishment, F.*

We show in Proposition 3 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 3 *(Becker, 1968)*[14]*: 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, $T(p, F)$.

---

[12] The following quote from Polinsky and Shavell (2007, footnote 16) testifies to the importance of the HPF: "[The HPF] or its equivalent, was put forward by Bentham (1789, p.173), was emphasized by Becker (1968), and has been noted by many others since then."

[13] In Becker (1968) the first order conditions are given with equality. When the optimum lies on the boundary this poses obvious problems. An intuitive argument is then given in Becker (1968) to justify a boundary solution. We provide the formal underpinning to Becker's informal argument.

[14] While Proposition 3(a) is well known from Becker (1968), Proposition 3(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 3 (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 as an empirical finding. However, this empirical observation is certainly not consistent with the decision maker following EU, under the conditions of Proposition 3. The probability of capital punishment can be made arbitrarily small but it will certainly deter crime under EU (Becker proposition).

4. Probability weighting functions (PWF)

All mainstream decision theory models that relax the "linearity in probabilities" assumption of EU, e.g., rank dependent utility (RDU) and cumulative prospect theory (CP), have one thing in common. Namely, that the decision maker uses a *probability weighting function* (PWF), denoted by $w(p)$, that transforms probabilities, p.

Definition 6 : *A probability weighting function (PWF) is a strictly increasing function* $w : [0,1] \xrightarrow{onto} [0,1]$.

A simple proof, that we omit, can be used to derive the following Proposition.

Proposition 4 : *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 $p \to 0$, we now offer some definitions that establish the relevant terminology.

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

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 well known examples of standard probability weighting functions are Tversky and Kahneman (1992), Gonzalez and Wu (1999), Lattimore, Baker and Witte (1992) and Prelec (1998).

Example 5 *(Prelec's probability weighting function): The Prelec (1998) probability weighting function, $w : [0,1] \overset{onto}{\longrightarrow} [0,1]$, is one of the most popular and satisfactory probability weighting functions. It is parsimonious, consistent with the stylized facts and has an axiomatic foundation.*[15] *It is given by $w(0) = 0$ and*

$$w(p) = e^{-\beta(-\ln p)^{\alpha}},\ 0 < p \leq 1,\ 0 < \alpha < 1,\ \beta > 0. \tag{4.1}$$

The Prelec function is strictly concave for low probabilities but strictly convex for high probabilities, i.e., it is inverse-S shaped as in $w(p) = e^{-(-\ln p)^{\frac{1}{2}}}$ ($\alpha = 0.5$, $\beta = 1$), sketched in Figure 4.1 as the bold curve. In Figure 4.1, the point of inflection is on the 45^o line. This arises because $\beta = 1$.[16] *Furthermore, it can be easily shown that in the special case $\alpha = \beta = 1$, we get $w(p) = p$, as under expected utility theory.*

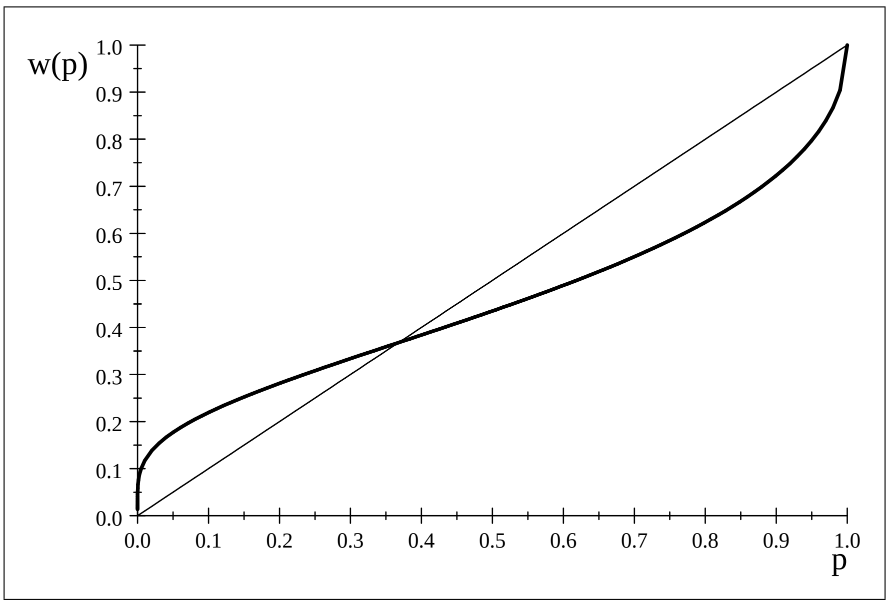

Figure 4.1: The Prelec function $w(p) = e^{-(-\ln p)^{\frac{1}{2}}}$.

Proposition 5 *: (a) The Prelec function (Example 5) is a probability weighting function in the sense of Definition 6.*
(b) For all $\gamma > 0$, the Prelec function (Example 5) infinitely overweights infinitesimal probabilities (Definition 7), i.e., $\lim_{p \to 0} \frac{w(p)}{p^{\gamma}} = \infty$. Thus, the Prelec function is a standard probability weighting function (Definition 8).

[15] For the axiomatic foundations, see Prelec (1998) and al-Nowaihi and Dhami (2006).

[16] For $\beta < 1$ the inflection point is above the 45^o line while for $\beta > 1$ the inflection point is below the 45^o line. The interested reader can consult al-Nowaihi and Dhami (2011b) for further details on these and other related issues on the Prelec function.

The bulk of the empirical evidence points to individuals having an inverse-S shaped weighting function (as in the bold curve in Figure 4.1). Extensive empirical evidence for this stylized fact is described in Wakker (2010). From Proposition 5, and as illustrated in Figure 4.1, the Prelec function is consistent with this stylized fact.[17] According to Prelec (1998, p.505), the infinite limit in Definition 7, and Proposition 5(b), captures the qualitative change as we move from improbability to impossibility.

5. The Becker paradox under rank dependent expected utility

Expected utility has two main features. It is (i) linear in probabilities, and (ii) utility is defined over final wealth levels. The refutations of EU have lead to the relaxation of both these features. The most common relaxation has been to relax linearity in probabilities. This was motivated initially by the inability of EU to explain various versions of the Allais paradox arising from a violation of the independence axiom. The main alternative to EU in decision theory that arose from work in this direction is *rank dependent expected utility* (RDU). Machina (2008) describes RDU as the most popular alternative to EU.

The objective function under RDU is described in Definition 9, below. We follow it up with an explanation.

Definition 9 *(Quiggin 1982) Consider the lottery $(x_1, p_1; x_1, p_1; ...; x_n, p_n)$ that pays some monetary amount, x_i, with probability, p_i, where $x_1 \leq x_2 \leq ... \leq x_n$. For RDU, the decision weights, π_j, are defined by*

$$\pi_i = w\left(\Sigma_{j=i}^{n} p_j\right) - w\left(\Sigma_{j=i+1}^{n} p_j\right), \tag{5.1}$$

where w is a standard probability weighting function (see Definition 8). The decision maker's rank dependent utility, RDU, is

$$RDU\left(x_1, p_1; x_1, p_1; ...; x_n, p_n\right) = \Sigma_{j=1}^{n} \pi_j u\left(x_j\right). \tag{5.2}$$

RDU offers a huge improvement over EU in terms of predictive success. The main difference between (5.2) and expected utility is that RDU replaces probabilities with decision weights. With this modification to EU, RDU can explain everything that EU can, but the converse is false (e.g., RDU can explain the Allais paradox while EU cannot).[18] The construction of decision weights in (5.1) shows the sense in which cumulative transformations

[17] The interested reader can consult al-Nowaihi and Dhami (2011b) for more details.

[18] More formally, to see that EU is a special case of RDU, take the probability weighting function to be $w(p) = p$. Then, from (5.1), we get $\pi_i = p_i$. From (5.2) we then get $RDU\left(x_1, p_1; x_1, p_1; ...; x_n, p_n\right) = \Sigma_{j=1}^{n} p_j u\left(x_j\right)$, which is the EU objective function.

of probabilities are made under RDU. This ensures that stochastically dominated choices are never made by the decision maker; see Quiggin (1982).[19]

The decision weights in (5.1) may look complicated but they are actually very intuitive. It can be shown that if, for instance, the probability weighting function is convex (concave) throughout then a decision maker who uses (5.1) places much greater (smaller) decision weight on smaller outcomes as compared to larger outcomes. Such a decision maker appears to an outside observer as pessimistic (optimistic).[20] Thus, while (5.1) may appear to be cognitively challenging, it encapsulates simple psychological principles.

Almost all versions of RDU use *standard probability weighting functions* (Definition 8 and Proposition 5b).

Consider the model of crime in Section 2 and assume that the decision maker uses RDU. Using Definition 9, the payoff from no-crime, the non-risky activity, is $U_{NC} = u(y_0)$, while that from crime is

$$EU_C = [1 - w(1-p)]\, u(y_1 - F) + w(1-p)\, u(y_1). \tag{5.3}$$

Hence, the no-crime condition (NCC), $RDU_C \leq U_{NC}$, holds if, and only if,

$$[1 - w(1-p)]\, u(y_1 - F) + w(1-p)\, u(y_1) \leq u(y_0). \tag{5.4}$$

Proposition 6 *: (a) Under RDU, if the utility function is unbounded below, so that $u(y_1 - F) \to -\infty$ as $F \to \infty$, then, for any probability of punishment, $p > 0$, no matter how small, crime can be deterred by a sufficiently severe punishment, F. In other words, the Becker proposition holds.*

Proposition 7 *: 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 $\varphi(p) = \frac{b}{1-w(1-p)}$, and
(b) If, in addition, $\eta_p^C > \frac{pw'(1-p)}{1-w(1-p)}\eta_F^C$ (Definition 4 and Lemma 1), then reducing p reduces the total social cost of crime and law enforcement, $T(p, F)$.

Lemma 2 *: The NCC (5.4) can be rewritten as:*

$$\Gamma(p, F) \leq u(y_0), \tag{5.5}$$

where

$$\Gamma(p, F) = [1 - w(1-p)]\, u(y_1 - F) + w(1-p)\, u(y_1). \tag{5.6}$$

$\Gamma(p, F)$ is continuous in p and F for all p, F with partial derivatives $\Gamma_p(p, F) < 0$, for all F and p, and $\Gamma_F(p, F) \leq 0$, for all F and p, where the latter inequality is strict for $p > 0$.

[19] This solved an important problem with Kahneman and Tversky's (1979) prospect theory. Indeed it was the incorporation of this insight that was the major impetus to Tversky and Kahneman's (1992) second generation prospect theory, which is known as cumulative prospect theory (CP).

[20] See Wakker (2010) for the details.

We now give some comparative static effects of the policy instruments, p, F, for the case of RDU in the Proposition below.

Proposition 8 *: Suppose that the utility function is unbounded below, so that $u(y_1 - F) \to -\infty$ as $F \to \infty$. Then:*
(a) For every $F > b$, there exists some $p = p_c \in (0,1)$ such that for all $p \geq p_c$ decision makers choose the legal activity and for all $p < p_c$ they choose the illegal activity.
(b) For $p > 0$, there exists some $F = F_c$ such that for all $F \geq F_c$ decision makers choose the legal activity and for all $F < F_c$ they choose the illegal activity.

6. The Becker proposition under cumulative prospect theory (CP)

6.1. An introduction to CP

Recall from the introduction to section 5 that RDU relaxes the assumption of linearity in probabilities. However, like EU, RDU continues to assume that decision makers derive utility from final levels of wealth. By contrast, a range of empirical evidence that has a rich history in psychology and has been robustly confirmed in thousands of experiments shows that decision makers behave in the following manner.[21] (i) They derive utility from changes in wealth relative to some reference point and not final wealth levels (*reference dependence*). This partitions the domain of outcomes into *the domain of gains and the domain of losses* relative to the reference point. (ii) Losses bite more than equivalent gains (*loss aversion*). (iii) The utility function is concave in the domain of gains but convex in the domain of losses (*declining sensitivity*). Hence, in each domain there is diminished sensitivity to a move away from the reference point.

We now outline cumulative prospect theory (CP), the second generation of prospect theory due to Tversky and Kahneman (1992). CP is rigorous, has axiomatic foundations, is consistent with much of the evidence and has better explanatory power relative to EU and RDU.[22] Consider a lottery of the form

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

where $y_{-m} \leq ... \leq y_0 \leq ... \leq y_n$ are the outcomes, possibly wealth levels, and $p_{-m}, ..., p_n$ are the corresponding probabilities, such that $\sum_{i=-m}^{n} p_i = 1$ and $p_i \geq 0$. The reference point for wealth is denoted by y_0.[23] Thus, we have m outcomes in the domain of losses

[21] See Kahneman and Tversky (1979), Starmer (2000), Wakker (2010), al-Nowaihi and Dhami (2011b).

[22] Readers interested in the axiomatic developments can follow up the references in Wakker (2010). The empirical success of CP is highlighted and described in several places; see for instance Kahneman and Tvesky (2000), Starmer (2000) and Wakker (2010).

[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).

and n outcomes in the domain of gains for a total of $m+n+1$ outcomes.

In CP, decision makers derive utility from wealth relative to a reference point for wealth, y_0. In order to capture this fact, we define lotteries in incremental form (also sometimes known as "prospects"; hence, the name "prospect theory").

Definition 10 *(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* $x_{-m} \leq ... \leq x_0 = 0 \leq ... \leq x_n$*. Then, a lottery in incremental form (or a prospect) is:*

$$L = (x_{-m}, p_{-m}; x_{-m+1}, p_{-m+1}; ...; x_{-1}, p_{-1}; x_0, p_0; x_1, p_1; ...; x_n, p_n) . \tag{6.1}$$

Denote by $\mathcal{L}_P$ the set of all *prospects* of the form given in (6.1).

Remark 1 *: An outcome is in the domain of gains if* $x_i \geq 0$ *and in the domain of losses if* $x_i \leq 0$*. Thus, the reference outcome,* $x_0 = 0$ *is in both domains.*

Definition 11 *(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, e.g., a loss of \$1 is more painful than the gain of \$1 is pleasurable).*

Tversky and Kahneman (1992) propose the following utility function which has different shapes in the domain of gains and loss.

$$v(x) = \begin{cases} x^\gamma & \text{if} \quad x \geq 0 \\ -\theta(-x)^\rho & \text{if} \quad x < 0 \end{cases}, \tag{6.2}$$

where γ, θ, ρ are constants. The coefficients of the power function satisfy $0 < \gamma < 1$, $0 < \rho < 1$. $\theta \geq 1$ is known as the *coefficient of loss aversion*. The utility function in (6.2) fits well the data in Kahneman and Tversky (1979). Tversky and Kahneman (1992) estimated that $\gamma \simeq \rho \simeq 0.88$ and $\theta \simeq 2.25$. This utility function also has an axiomatic foundation.[25]

[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 domain of gains and risk seeking in the domain 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). See Wakker (2010) for a formal textbook treatment.

[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 (6.2). al-Nowaihi et al. (2008) give a formal proof, as well as some other results (e.g. that γ is necessarily identical to ρ).

We now show how the decision weights are constructed under CP. This is the analogue of the construction of decision weights under RDU (see Definition 9). The main difference from RDU, in this respect, is that under CP, one computes decision weights separately in the domain of gains and losses. In principle, one could use a probability weighting function, w^{+}, for the domain of gains and a different weighting function, w^{-}, for the domain of losses. However, the empirical evidence indicates that these functions are very similar so, as in Prelec (1998), we set $w^{+} = w^{-} = w$.[26]

Definition 12 *: 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,

$$\pi_j = w\left(\Sigma_{i=-m}^{j}\ p_i\right) - w\left(\Sigma_{i=-m}^{j-1}\ p_i\right),\ j = -m, ..., -1.$$

As in RDU, w is a standard probability weighting function (see Definition 8).

Definition 13 *(Value function): A decision maker using CP maximizes the following value function defined over $\mathcal{L}_P$,*

$$V(L) = \Sigma_{i=-m}^{n} \pi_i v(x_i),\ L \in \mathcal{L}_P. \tag{6.3}$$

where the utility function, v, is defined in Definition 11 and the decision weights are defined in Definition 12.

6.2. Fixed reference points and fixed punishment levels

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

$$V_{NC} = v(y_0 - y_{nc}). \tag{6.4}$$

[26] Abdellaoui (2000) and Abdellaoui et al. (2005) find that there is no significant difference in the curvature of the weighting function for gains and losses. For the Prelec function (see Example 5), we know that the parameter α controls the curvature. However, the elevation (which in the Prelec function is controlled by the parameter β) can be different in the domain of gains and losses. Empirically, however, it appears that $\beta = 1$ or is very close to 1.

[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 the inverse S-shape of the probability weighting function. For that reason we do not follow PT^3, but we exploit it's insight of using state dependent reference points in section 6.3.

Recall that the outcomes under CP are split into the domain of gains and losses.[28] We make the plausible 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., $y_1 - F - y_c \leq 0$). If not caught (with probability $1-p$), the outcome, y_1, is in the domain of gains (i.e., $y_1 - y_c \geq 0$). Thus, we have one outcome each in the domain of losses and gains.[29] Using Definition 12, the respective decision weights are $w(p)$ and $w(1-p)$.[30] Then, under CP, the individual's payoff from committing a crime is given by

$$V_C = w(p)\, v(y_1 - F - y_c) + w(1-p)\, v(y_1 - y_c). \tag{6.5}$$

The no crime condition (*NCC*) in the case of CP is $V_C \leq V_{NC}$, which is equivalent to

$$w(p) v(y_1 - F - y_c) + w(1-p)\, v(y_1 - y_c) \leq v(y_0 - y_{nc}). \tag{6.6}$$

From (6.6), ceteris-paribus, an increase in fines, F, reduces the LHS and so makes it more likely for the NCC to be satisfied. Hence, as one would expect, an increase in fines reduces crime.

The *NCC* depends on the two reference points, y_{nc} and y_c, assumed fixed in this section. Section 6.3 considers alternative specifications of reference incomes drawing on the work of Schmidt et al. (2008) and Köszegi and Rabin (2006).

Proposition 9 : *Assume CP with fixed reference points, y_{nc} and y_c, and a utility function v that is unbounded below, so that $v(y_1 - F - y_c) \to -\infty$ as $F \to \infty$, ((6.2) for example). Then, given a probability of detection, $p > 0$, no matter how small crime can be deterred with a sufficiently high punishment, F. Thus, the Becker proposition holds under CP.*

Under the conditions of Proposition 9, the Becker proposition holds under CP. The reader might wonder if this result is due to the special assumption of fixed reference points. In subsection 6.3, below, we show that the Becker proposition is robust enough to survive a relaxation of this restriction.

6.3. Alternative specifications of reference points under CP

The result (Proposition 9) in subsection 6.2, above, assumed fixed reference points. We now relax this assumption. We make four assumptions, A1-A4, in this subsection.

[28] It is possible to have all outcomes in the domain of gains or all outcomes in the domain of losses but this is not the case for our problem.

[29] 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.

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

1. *Assumption A1 (Hyperbolic Punishment function, HPF)*: The punishment function is the hyperbolic punishment function, $F = \varphi(p) = b/p$ (see Definition 5), where $b = y_1 - y_0$ is the benefit from crime to the perpetrator.

2. *Assumption A2 (Power form of utility)*: We use the utility function in (6.2) with $\gamma = \rho$ (as already noted, this is consistent with the evidence and has axiomatic foundations).

3. *Assumption A3 (Standard probability weighting functions)*: The weighting function $w(p)$ is a standard weighting function (see Definition 8) and so $\lim_{p \to 0} \frac{w(p)}{p^\gamma} = \infty$ (e.g., the Prelec (1998) function, Example 5), which is consistent with the empirically observed inverse-S shape of a probability weighting function.

4. *Assumption A4 (Heterogeneity in reference points)*: To enable a formulation that nests several interesting cases, we describe the reference incomes from crime and no-crime as follows. Take the reference income from crime, y_c, as the *expected income from crime*, i.e.,
$$y_c = y_1 - p\varphi(p). \tag{6.7}$$
The reference income from no-crime, y_{nc}, is specified as some fraction, $\lambda \geq 0$, of the reference income from crime, i.e.,
$$y_{nc} = \lambda y_c = \lambda (y_1 - p\varphi(p)), \lambda \geq 0, \tag{6.8}$$
where λ (to be interpreted below) is distributed across the population with support, $[0, \overline{\lambda}]$, $\overline{\lambda} > 1$. This is the only source of individual heterogeneity in the model.

We now analyze our model of crime when the decision maker uses CP and Assumptions 1-4 hold.

Using Assumption 1, for the HPF, $\varphi(p) = b/p$. Using (2.1) we get that
$$p\varphi(p) = b = y_1 - y_0. \tag{6.9}$$
Using (6.9), the reference incomes, y_c, y_{nc}, in (6.7) and (6.8) are
$$y_c = y_1 - p\varphi(p) = y_0,\ y_{nc} = \lambda y_c = \lambda y_0. \tag{6.10}$$
Thus,
$$y_0 - y_{nc} = y_0 - \lambda y_0 = (1 - \lambda) y_0. \tag{6.11}$$
Depending on the value of λ, we get three important cases.

(i) *Socially responsible individuals* ($0 \leq \lambda < 1$): In this case, from (6.11) we get $(1-\lambda) y_0 > 0$, i.e., income relative to the reference point from not committing the crime is positive. Such an individual feels positively rewarded on account of his honesty.

(ii) *Regretful individuals* ($1 < \lambda \leq \overline{\lambda}$): In this case (6.11) implies that $(1-\lambda) y_0 < 0$. 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.[31] By direct calculation, the expected income from no-crime is y_0 while the expected income from crime is $y_1 - p\varphi(p)$. From (6.10), this corresponds to a value of $\lambda = 1$. Thus, for the case of rational expectations of income we get $y_c = y_{nc} = y_0$.

Proposition 10 *: Suppose that the decision maker follows CP and Assumptions 1-4 hold. Then the Becker proposition holds for all three kinds of decision makers, namely, socially responsible, regretful and those who have rational expectations.*

We have argued that CP is the most satisfactory decision theory among the mainstream alternatives. It can explain everything that EU (or its variants such as generalized expected utility) or RDU can explain but the converse is false. Proposition 10 demonstrates the resilience of the Becker proposition even when the decision maker follows CP and a sufficiently wide range of reference point behaviors is allowed for.

6.4. Loss aversion and the extent of crime

An important consideration in models of crime has been the effect of risk aversion on the decision to commit a crime. In RDU and CP, however, one cannot infer the degree of risk aversion simply from the magnitude of the parameter γ in the utility function in (6.2). The reason is that under EU, which is characterized by linearity in probabilities, risk attitudes are captured entirely by the shape of the utility function. For instance, under EU, concavity of utility implies and is implied by risk aversion.

By contrast, when linearity in probabilities is relaxed, as in RDU and CP (see Definitions 9 and 12), then risk aversion is determined *jointly* by the shape of the probability weighting function and the shape of the utility function. Indeed in CP it can be shown that

[31] See, for instance, Koszegi and Rabin (2006), Crawford and Meng (2011).

there is a rich four-fold pattern of attitudes to risk that is supported by the evidence.[32]

Novemsky and Kahneman (2005) show that there is "no risk aversion over and above loss aversion". In other words, it is quite possible that what empirical researchers are picking out as risk aversion is really caused by loss aversion instead.[33] Therefore, it is worth analyzing the effect of loss aversion on crime in the model that we set out above.

The individuals in our model make one of two possible decisions: To engage in the legal activity or the illegal activity. From Proposition 10, under the conditions of the Becker proposition ($p \to 0, F \to \infty$), and Assumptions A1-A4, all individuals decide to opt for the legal activity. Hence, an increase in loss aversion by making the consequences of the illegal activity even worse, has no impact on individual decisions. In other words, as is typically the case in mechanism design problems, we get a corner solution that entirely eliminates crime and so many of the comparative static effects are simply zero. Therefore, in order to isolate the impact of loss aversion we make the following assumption.

5. *Assumption A5 (Punishments bounded above)*: We assume that for whatever reasons, say, those mentioned in subsection 1.1 in the introduction, $F_{\max} < \infty$.[34] Hence, some individuals are not dissuaded from crime.

Let us then look at the impact of the loss aversion parameter, θ, on the three kinds of individuals when assumptions A1, A4, A5 hold (we do not require assumptions A2, A3 for the results in this subsection).

1. Socially responsible individuals ($0 \leq \lambda < 1$): Rewrite the NCC (8.7) as

$$\lambda \leq \lambda_S, \tag{6.12}$$

where

$$\lambda_S = 1 - g\left(\theta, \gamma, p, b, y_0\right), \tag{6.13}$$

$$g\left(\theta, \gamma, p, b, y_0\right) = \frac{b}{y_0}\left[-\theta\left(\frac{1-p}{p}\right)^{\gamma} w(p) + w(1-p)\right]^{1/\gamma}. \tag{6.14}$$

If $g > 0$ then all individuals with $0 \leq \lambda \leq \min\{0, \lambda_S\}$ will not commit the crime, whereas all socially responsible individuals with $\lambda > \min\{0, \lambda_S\}$ will commit the crime. It is straightforward to check that $\frac{\partial g}{\partial \theta} > 0$. Hence, an increase in loss aversion,

[32] Under CP, the four-fold pattern is as follows; see Tversky and Kahneman (1992) and Kahneman and Tversky (2000, p. 56). In the domain of gains, individuals are risk loving for small probabilities and risk averse for larger probabilities. In the domain of losses individuals are risk loving for large probabilities and risk averse for small probabilities.

[33] See also, for instance, Kahneman and Tversky (2000), Kahneman (2003), Marquis and Homer (1996) and Rizzo and Zeckhauser (2004).

[34] Thus, the Becker prescription ($p \to 0, F \to \infty$) does not hold.

by increasing λ_S, reduces the extent of crime in society. The intuition is that an increase in loss aversion makes the existing level of punishment even more onerous, hence, reducing the incentive to engage in the illegal activity. If $g < 0$ then all socially responsible people commit the crime anyway and an increase in θ has no effect on the extent of crime for such people.

2. Regretful individuals ($1 < \lambda \leq \overline{\lambda}$): Rewrite the NCC (8.9) as

$$\lambda \leq \lambda_R, \tag{6.15}$$

where

$$\lambda_R = 1 + h(\theta, \gamma, p, b, y_0), \tag{6.16}$$

$$h(\theta, \gamma, p, b, y_0) = \frac{b}{y_0}\left[\left(\frac{1-p}{p}\right)^{\gamma} w(p) - \frac{1}{\theta} w(1-p)\right]^{1/\gamma}.$$

When $h > 0$, all regretful individuals for whom $1 < \lambda \leq \min\{\lambda_R, \overline{\lambda}\}$ will not commit the crime while all those with $\lambda > \lambda_R$ will commit the crime. It is simple to check that $\frac{\partial h}{\partial \theta} > 0$. Hence, λ_R is increasing in θ. Therefore, as in in the case of socially responsible individuals, an increase in loss aversion, by increasing λ_R, reduces the extent of crime. When $h < 0$, all regretful individuals will commit the crime and so an increase in loss aversion will have no effect on the extent of crime.

3. Rational individuals ($\lambda = 1$): In this case, the no crime condition (NCC) is

$$\theta \geq \theta^*,$$

where

$$\theta^* = \frac{w(1-p)}{w(p)}\left(\frac{p}{1-p}\right)^{\gamma}.$$

Then all rational individuals will commit the crime if their loss aversion is low enough in the sense that $\theta < \theta^*$ and not commit the crime otherwise.

7. 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.

A sizeable literature addresses the Becker paradox in an *expected utility* (EU) framework and considers several worthwhile extensions. We argue that it is very difficult to reconcile EU with the evidence, a well known fact, but, to our mind, not given enough prominence in the law and economics literature. Hence, we re-examine the Becker paradox from the perspective of alternative mainstream non-linear decision theories that fit the evidence much better than EU. These theories are as tractable as EU and equally rigorous. Two main alternatives that we consider are *rank dependent expected utility* (RDU), which, Machina (2008) describes as the most popular alternative to the EU model and cumulative prospect theory (CP), the Nobel prize winning work of Tversky and Kahneman (1992). The Becker proposition reemerges under RDU and CP.

When put under the lens of CP, in a standard model of crime, the Becker proposition is robust to several plausible assumptions about the reference points of potential criminals. We allow for heterogeneity among potential criminals in terms of their reference points, which allows us to consider behavior that is consistent with rational expectations, regret and social responsibility. In each case, so long as the decision maker is rational (as opposed to being pathological) the Becker proposition holds under CP.

In addition to demonstrating that the Becker proposition holds under RDU and CP, our paper serves two alternative purpose as well. First, we give a more satisfactory formal account of BP. Second, we demonstrate how a standard model of crime can be analyzed under these alternative decision theories. Our analysis gives readers a feel for the possible problems that could be analyzed under alternative decision theories and demonstrates that the analysis under these alternative theories is rigorous and tractable. This should give further impetus to the emerging literature on Behavioral Law and Economics.

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

8. Appendix

Proof of Proposition 1: For the proofs of (a) and (b) see al-Nowaihi and Dhami (2011, Proposition 1), (c) then follows using the identity

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

Finally, (d) is an immediate consequence of (c). ■

Proof of Lemma 1:

$$\frac{d}{dp}C(p,\varphi(p)) = C_p + \left[C_F\frac{d\varphi}{dp}\right]_{F=\varphi(p)} = \left[\frac{C}{p}\left(\frac{p}{C}C_p + \frac{F}{C}C_F\frac{p}{F}\frac{d\varphi}{dp}\right)\right]_{F=\varphi(p)}$$
$$= \left[\frac{C}{p}\left(\eta_p^C - \eta_p^F\eta_F^C\right)\right]_{F=\varphi(p)} > 0 \Leftrightarrow \eta_p^C > \eta_p^F\eta_F^C \text{ at } F=\varphi(p). \ ■$$

Proof of Proposition 2: Immediate from the NCC (3.1). ■

Proof of Proposition 3: (a) The NCC (3.1) is clearly satisfied for $p = 1$ and $F \geq b = y_1 - y_0$. Let $F = \varphi(p)$ be a differentiable punishment function. The *NCC* in (3.1) continues to hold, as p declines from 1, if, and only if

$$\frac{d}{dp}\left[pu\left(y_1 - \varphi(p)\right) + (1-p)\,u(y_1)\right] \geq 0, \tag{8.1}$$

which simplifies to

$$-p\varphi'(p)\,u'\left(y_1 - \varphi(p)\right) \geq u(y_1) - u\left(y_1 - \varphi(p)\right). \tag{8.2}$$

For the HPF (2.6), $-p\varphi'(p) = \varphi(p)$, so the *NCC,* (8.2), becomes

$$\text{NCC for HPF: } u'\left(y_1 - \varphi(p)\right) \geq \frac{u(y_1) - u\left(y_1 - \varphi(p)\right)}{\varphi(p)}. \tag{8.3}$$

Since the decision maker is risk averse or risk neutral, u is concave. It then follows that the *NCC* (8.3) *will* hold for all $p \in (0,1]$. This proves (a).

(b) Since $\varphi(p) = \frac{b}{p}$, then (from Definition 4) $\eta_p^F = -\frac{p}{F}\frac{d\varphi}{dp} = 1$. By assumption $\eta_p^C > \eta_F^C$, thus, it follows, from Lemma 1, that $\frac{d}{dp}C(p,\varphi(p)) > 0$. Since crime is deterred, $D(p,\varphi(p)) = 0$ (see subsection 2.1 for this assumption). 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 5: We omit the simple proof of part (a).

(b) Since $p^{-\gamma} = e^{-\gamma \ln p}$, (4.1) gives

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

Note that $\lim_{p\to 0}(-\ln p) = \infty$. Since $0 < \alpha < 1$, we get $\lim_{p\to 0}(-\ln p)^{1-\alpha} = \infty$ and, hence, $\lim_{p\to 0}\frac{\beta}{(-\ln p)^{1-\alpha}} = 0$. Thus, since $\gamma > 0$, we get $\lim_{p\to 0}\frac{w(p)}{p^{\gamma}} = \lim_{p\to 0} e^{-\gamma \ln p} = \lim_{p\to 0} p^{-\gamma} = \infty$. ■

Proof of Proposition 6: Immediate from the NCC (5.4). ■

Proof of Proposition 7: Similar to the proof of Proposition 3, 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 Lemma 2: Differentiating (5.6) partially with respect to p gives $\Gamma_p(p,F) = -w'(1-p)\left[u(y_1) - u(y_1 - F)\right] < 0$ for all F and p. Differentiating with respect to F gives $\Gamma_F(p,F) = -\left[1 - w(1-p)\right]u'(y_1 - F) \leq 0$, for all F and p, with $\Gamma_F(p,F) < 0$ for $p > 0$. ■

Proof of Proposition 8: We make essential use of Lemma 2.

We first prove (a). Let $F > b$. At $p = 0$, $\Gamma(0,F) = u(y_1) > u(y_0)$ so the individual commits the crime. At $p = 1$, $\Gamma(1,F) = u(y_1 - F) < u(y_0)$ so the individual does not commit the crime. Since $\Gamma(p, F)$ is continuous, there exists some p_c at which $\Gamma(p_c, F) = u(y_0)$. Since $\Gamma_p < 0$, for all $p \geq p_c$, $\Gamma(p_c, F) \leq u(y_0)$ and so individuals do not commit the crime. For all $p < p_c$, $\Gamma(p_c, F) > u(y_0)$ and so individuals commit the crime.

Part (b) can be proved in an analogous manner. Let $p > 0$. At $F = 0$, $\Gamma(p, 0) = u(y_1) > u(y_0)$, hence, all individuals commit the crime. Because $u(y_1 - F) \to -\infty < u(y_0)$, as $F \to \infty$, no individual commits the crime for sufficiently large F. Since (1) Γ is continuous, (2) $\Gamma(p, 0) > u(y_0)$, and (3) $\Gamma(p, F) < u(y_0)$ for sufficiently large F, there exists some critical value F_c at which $\Gamma(p, F_c) = u(y_0)$. Since $\Gamma_F < 0$ for $F \geq F_c$, $\Gamma(p, F_c) \leq u(y_0)$ and so individuals do not commit the crime. For $F < F_c$, $\Gamma(p, F_c) > u(y_0)$ and so individuals commit the crime. ■

Proof of Proposition 9: Immediate from the NCC (6.6). ■

Proof of Proposition 10: We have computed income relative to the reference point for the legal activity in (6.11). For our model of crime, using (6.10) and Assumption 1, we now compute the *income relative to the reference point* in each state from the criminal activity.

$$\begin{cases} y_1 - \varphi(p) - y_c = b - \varphi(p) = -b\frac{1-p}{p}, & \text{if caught,} \\ y_1 - y_c = b, & \text{if not caught.} \end{cases} \tag{8.4}$$

From (8.4), if the criminal is not caught then income relative to the reference point is the benefit, b, from the crime. If caught, this benefit is reduced by the amount of the fine, $\varphi(p)$. Substituting (6.11), (8.4) in the *NCC* (6.6) we get

$$w(p)\, v\left(-b\frac{1-p}{p}\right) + w(1-p)\, v(b) \leq v(y_0(1-\lambda)). \tag{8.5}$$

We now use (8.5) to analyze the three cases, $0 \leq \lambda < 1$, $1 < \lambda \leq \overline{\lambda}$, and $\lambda = 1$.

1. Socially responsible individuals ($0 \leq \lambda < 1$): In this case, $y_0(1-\lambda) > 0$. Hence, using Assumption 2 (the power form of utility), the NCC in (8.5) becomes

   $$-\theta w(p)\, b^{\gamma} \frac{(1-p)^{\gamma}}{p^{\gamma}} + w(1-p)\, b^{\gamma} \leq y_0^{\gamma}(1-\lambda)^{\gamma}, \tag{8.6}$$

   which on simplifying gives

   $$-\theta \frac{w(p)}{p^{\gamma}} + \frac{w(1-p)}{(1-p)^{\gamma}} \leq \left(\frac{y_0(1-\lambda)}{b(1-p)}\right)^{\gamma}. \tag{8.7}$$

   To see if the Becker proposition holds, we take limits on both sides as $p \to 0$ (for the HPF this automatically implies $F = \infty$ as required by the Becker proposition). From Assumption A3, $\lim\limits_{p\to 0} \frac{w(p)}{p^{\gamma}} = \infty$. From the definition of a probability weighting function, $w(1) = 1$ so $\lim_{p \to 0} \left(\frac{w(1-p)}{1-p}\right)^{\gamma} = 1$. Finally, $\lim_{p \to 0} \left(\frac{y_0(1-\lambda)}{b(1-p)}\right)^{\gamma} = \left(\frac{y_0(1-\lambda)}{b}\right)^{\gamma}$. Hence, as $p \to 0$, the inequality in (8.7) becomes

$$-\infty + 1 < \left(\frac{y_0(1-\lambda)}{b}\right)^{\gamma}, \tag{8.8}$$

which is true. Hence, the Becker proposition holds for socially responsible individuals.

2. Regretful individuals ($1 < \lambda \leq \overline{\lambda}$): In this case, $y_0(1-\lambda) < 0$. Analogous to (8.7), the NCC in (8.5) can be written as

$$-\theta \frac{w(p)}{p^{\gamma}} + \frac{w(1-p)}{(1-p)^{\gamma}} \leq -\theta \left(\frac{y_0(\lambda - 1)}{b(1-p)}\right)^{\gamma}. \tag{8.9}$$

Taking the limits as $p \to 0$ in (8.9) we get that

$$-\infty + 1 < -\theta \left(\frac{y_0(\lambda - 1)}{b}\right), \tag{8.10}$$

which is true. Hence, the Becker proposition also holds for regretful individuals.

3. Individuals with rational expectations ($\lambda = 1$). In this case $y_0(1-\lambda) = 0$. Recalling that $v(0) = 0$ (see Definition 11), the analogue of the NCC in (8.8), (8.10) is

$$-\infty + 1 < 0,$$

which is true. Hence, the Becker proposition also holds for individuals with rational expectations. ■

References

[1] Abdellaoui, M., 2000. Parameter-free elicitation of utility and probability weighting functions. Management Science, 46: 1497–1512.

[2] Abdellaoui, M., Vossmann, F., and Weber, M., 2005. Choice-Based Elicitation and Decomposition of Decision Weights for Gains and Losses Under Uncertainty. Management Science 51: 1384-1399.

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

[4] al-Nowaihi, A. and Dhami, S., 2011a. Hyperbolic Punishment Functions. *University of Leicester Discussion Paper Series* (forthcoming in the *Review of Law and Economics, BE Press Journals*).

[5] al-Nowaihi, A., and Dhami, S., 2011b. Probability weighting functions. in James J. Cochran, Louis Anthony Cox, Jr., Pinar Keskinocak, Jeffrey P. Kharoufeh, J. Cole Smith (eds.) Wiley encyclopedia of operations research and management science, John Wiley and Sons. ISBN: 978-0-470-40063-0.

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

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

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

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

[10] Camerer, C. (2000) Prospect Theory in the Wild: Evidence From the Field, in Kahneman, D., and Tversky, A., (eds.) 2000. *Choices, values and frames*. Cambridge University Press, Cambridge.

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

[12] Crawford, V. P. and Meng, J., 2011. "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependence Preferences with Rational-Expectations Targets for Hours and Income". The American Economic Review, 101: 1912-1932(21)

[13] Dellavigna, S. (2009) Psychology and Economics: Evidence from The Field, Journal of Economic Literature, 47: 315-372.

[14] 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.

[15] 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.

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

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

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

[19] Kahneman, D. (2003). "Maps of Bounded Rationality: Psychology for Behavioral Economics," American Economic Review, American Economic Association, 93: 1449-1475.

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

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

[22] 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.

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

[24] Koszegi, B., and Rabin, M. 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics.* 121, 1133-1165.

[25] 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.

[26] 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.

[27] 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.

[28] Marquis, M.S. Holmer, M. R. (1996) Alternative models of choice under uncertainty and demand for health insurance, The Review of Economics and Statistics, 78: 421–427

[29] Neilson, W. S. and Winter, H. (1997) On criminals' risk attitudes, Economics Letters, 55: 97-102.

[30] Novemsky, N. and D. Kahneman. 2005. The boundaries of loss aversion. Journal of Marketing Research. 42. 119-128.

[31] 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.

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

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

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

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

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

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

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

[39] Rizzo, J.A. and Zeckhauser, R. J. (2004) Reference incomes, loss aversion, and physician behavior, Review of Economics and Statistics. 85: 909–922

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

[41] 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.

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

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

[44] Wakker, Peter P. (2010), "Prospect Theory for Risk and Ambiguity." Cambridge University Press, Cambridge.