UNIVERSIDAD DE LA REPÚBLICA

FACULTAD DE CIENCIAS SOCIALES

DEPARTAMENTO DE ECONOMÍA

Tesis de Doctorado en Economía

Essays on behavioral decision making

Martín Egozcue

Tutor: Jorge Ponce

2013

Profesores Guía: Dr. Jorge Ponce y Dr. Ričardas Zitikis

Fecha:

Autor: Martín Egozcue

Calificación:

Dr. Elvio Accinelli | Dr. Juan Dubra | Dr. Gerardo Licandro

Acknowledge

I owe a large debt of gratitude to my supervisors Jorge Ponce and Ričardas Zitikis for their constant support, guidance, patience and encouragement throughout my doctoral study during these years.

My deepest thanks also to Luis Fuentes García, a third uncredited supervisor.

My gratitude also goes to all the staff of the Department of Economics at Facultad de Ciencias Sociales, UdelaR. Specially, my thanks to the Heads of the Doctoral program: Rossana Patrón, Ruben Tansini, Patricia Triunfo and Marcel Vaillant. The department gave me full support to complete this endeavor.

All the chapters benefits from helpful comments by Elvio Accinelli, Dante Amengual, Ana Balsa, Marcelo Caffera, Carlos Casacuberta, Juan Dubra, Alvaro Forteza, Gerardo Licando, Konstantinos Katsikopolous, Ignacio Munyo, Danilo Trupkin, Hugh Schwartz, Eduardo Siandra, Michael Smithson and participants at Jornadas Anuales de Economía del Banco Central, Seminars at Universidad de Montevideo, Seminar at Central Bank of Uruguay, and Seminars at the FCS, UdelaR.

I gratefully acknowledge financial support from Agencia Nacional de Investigación e Innovación (ANII).

Finally, my thanks to my family for their constant love and support.

Resumen

Este trabajo de tesis consiste en cuatro ensayos sobre la teoría de la toma de decisiones. El primer ensayo analiza las preferencias por las ganancias derivadas de la diversificación que tienen los agentes económicos de acuerdo la teoría del arrepentimiento. El segundo ensayo presenta nuevas desigualdades de la covarianza de funciones no monótonas de una variable aleatoria. Se muestran dos aplicaciones de estos nuevos resultados. El tercer ensayo propone una extensión de la heurística del reconocimiento. En esta nueva propuesta se distinguen tres niveles de reconocimiento. Se compara el poder predictivo de esta heurística para vectores con tres y diez objetos. Por último, el cuarto ensayo analiza los principios de la contabilidad mental cuando los tomadores de decisiones deben integrar o segregar tres o más experiencias, resultados, etc.

Palabras clave: Teoría del arrepentimiento, desigualdades de la covarianza, heurística del reconocimiento, contabilidad mental.

Abstract

This thesis consists in four essays on behavioral decision making. The first essay analyses the preferences for diversification of decision makers according to regret theory. The second essay presents some new covariance inequalities of non-monotonic functions of a random variable. I also show two applications of these new results. The third essay proposes an extension of the recognition heuristic. I also compare the predictive power of this heuristic for recognition vectors with three and ten objects. Finally, the fourth essay analyses the mental accounting principles when decision makers must decide to integrate or segregate three or more experiences, outcomes, etc.

Keywords: Regret theory, covariance inequalities, recognition heuristic, mental accounting.

Contents

[1]Jointly with Sebastien Massoni, Wing Keung Wong and Ričardas Zitikis.

General introduction

This thesis aims to contribute to the theoretical literature on behavioral decision making by proposing answers to the following questions:

(i) When does a decision maker gains from diversification? More precisely, how should an investor choose among different assets? Should he invest in a single asset or should he invest in many different assets? Could the feeling of regret affect this decision?

(ii) What are the optimal hedging policies of an enterprise? How decision changes when are considered variation in profits instead of their level?

(iii) Suppose you are driving on a highway and your car is running out of gas. You see in the highway signal the names of the two next neighborhoods, one of which owns a very bad reputation. You have never heard about the second one. How would you decide which exit to take? and

(iv) Suppose that your doctor schedule a minor surgery for you on the same day of an exciting concert you would like to attend, what would you do? Would you postpone attending the concert? Or would you do both ?

Von Neumann and Morgenstern expected utility theory provides economics with a powerful model to analyze decision making under uncertainty. There is, however, an agreement of two major lines of critique of expected utility theory. First, despite the wide acceptance of this theory many empirical studies reveal the inconsistency of its predictions with current human behavior. Hence, new models propose alternatives to expected utility theory (Starmer, 2000; Sugden, 2004). These new models use different types of utility functions and/or relax the linearity on probabilities. Second, many studies suggest that laypeople, experts and professional decision makers do not necessarily decide according to mathematical models. That is, in contrast to the use of complex mathematical models, this approach proposes that decision makers use shortcuts and rules of thumb (usually referred as heuristics) to decide (Gigerenzer & Selten, 2001).

These new lines of research not only allow to provide answers to old questions, but they also raise new ones. For instance: Are human being rational? Do they have bounded rationality? How these facts influence their decisions? How do feelings like regret and rejoice influence their choices? Do humans always use complex models to make decisions or sometimes they decide according to simple models, such as naive rules of thumb?

This thesis contributes to these two lines of research. More precisely, I shall analyze some economic problems, such as the above mentioned questions, from these perspectives. The first two chapters focus on the study of the first group of criticism of expected utility theory. The remaining two chapters analyze human behavior from the perspective of the second stream of critiques.

Chapter 1 is based in Egozcue (2012). In this chapter, I study the gains from diversification within regret theory. This is an important issue that applies to many real economic problems such as portfolio selection, remuneration schemes and international trade, among others. This problem has been analyzed within expected utility theory. However, few studies have approached these problems with regret theory. The aim of this chapter is to contribute to the literature by: (i) providing conditions under which a regret averse decision maker will diversify between two risky options; (ii) showing the differences between the optimal choices of regret averse and risk averse individuals; (iii) analyzing the conditions under which the results for two risky options can be generalized to many number of alternatives; and finally (iv) proposing two applications of the main results to existing models of decision making under uncertainty.

In several analytical problems of decision making under uncertainty, it is necessary to study the sign of a covariance that involves marginal utilities. Chebyschev's integral inequality is an important tool that helps to elucidate the sign of this covariance. Its application, however, requires the functions be monotonic. For instance, alternative theories to expected utility theory, often assume non-monotonic marginal utilities.

In **Chapter 2**, I derive some new covariance inequalities for utility functions with non-monotonic marginal utilities. In particular, I establish the conditions to determine the sign of the covariance for utility functions that start concave and then turn convex with an inflection point at the origin. I also derive conditions for those that are concave for positive values and convex for negative values. I apply these covariances inequalities to two problems in economics. First, I study some properties of the indifference curve in the mean-variance space for Prospect Theory and for Markowitz utility functions. Second, I analyze the asset's hedging policies of an enterprise that behaves as predicted by Prospect Theory.

In **Chapter 3,** I propose a generalization of the recognition heuristic model originally introduced by Goldstein and Gigerenzer (1999, 2002). The recognition heuristic surged to explain why some people could respond correctly to questions on some topics that a priori they do not know? In this chapter, instead of considering only two levels of recognition, I propose a three levels recognition model. I derive explicit formulas for all the parameters of the model. This allows me to study the expected accuracy rate of the three levels recognition heuristic and compare it with the performance of the two levels model. Besides, I characterize the conditions under which the recognition heuristic expected accuracy rate is equal to 50%. Finally, I discuss whether less information could lead to higher accuracy rates in the three levels of recognition model.

Finally, **Chapter 4** is based on a joint project with Sebastien Massoni, Wing Keung Wong and Ričardas Zitikis forthcoming in IMA Journal of Management Mathematics with the title: Integration-segregation decisions under general value functions: "Create your own bundle – choose 1, 2, or all 3!". In this work, we study whether to keep products segregated (e.g., unbundled) or integrate some or all of them (e.g., bundle). This problem has been of big interest in areas such as portfolio theory, risk capital allocations, taxation and marketing. Our findings show that the celebrated Thaler's principles of mental accounting hold as originally postulated when the values of all exposure units are positive (i.e., all are gains) or all negative (i.e., all are losses). In the case of exposure units with mixed-sign values, decision rules are much more complex and rely on cataloging the

Bell-number of cases, which grow very fast depending on the number of exposure units. So, in this chapter we provide detailed rules for the integration and segregation decisions in the case up to three exposure units, and partial rules for the arbitrary number of units. Also, we show various possible applications of mental accounting in different areas such as: product bundling, legislation and taxation, among others.

Chapter 1

Regret and diversification

1.1 Introduction

The diversification problem is about allocating an individual's initial wealth between risky prospects (random variables). If the optimal allocation includes many prospects, there may be gains from diversification.

Usually, this analysis uses expected utility theory, where risk aversion represents decision maker's behavior. For instance, Samuelson (1967) proves that if two random variables (i.e., risky prospects) are independent and identically distributed then there are always gains from diversification. Moreover, he shows that assigning equal shares of the initial wealth in each asset is the optimal choice. Brumelle (1974) further shows that assuming negative correlation between two random variables is neither necessary nor sufficient to assure gains from diversification. In fact, when the two random variables are identically distributed and have finite mean, then diversification is optimal (Hadar & Russell, 1971, 1974; Tesfatsion, 1976). Nevertheless, this problem has been extended by relaxing the independent and identically distributed assumption, making the analysis with multiple random variables and so forth (see, for example, Gollier, 2004; Hadar, Russell & Seo, 1977; Landsberger & Meilijson, 1990; Kira & Ziemba, 1980; Ma, 2010; Pellerey & Semeraro, 2005; Wright, 1987).

There are, however, many experiments showing limited predictive accuracy of expected utility theory. So, these pieces of evidence questioned whether expected utility theory is a good model of economic behavior, and many alternative models were proposed.[1] One of these alternative theories is regret theory. Originally suggested by Savage (1951), this theory assumes that decision makers may include in their decision process the feelings of regret and rejoice.[2] The seminal papers by Bell (1982), Fishburn (1982) and Loomes and Sugden (1982, 1987) present a formal analysis of regret theory. Sugden (1993) gives an axiomatic approach, while Quiggin (1994) extends the analysis to multiple choices.

There is an extensive body of research that has found empirical support for regret theory (Loomes & Sugden, 1982; Loomes, Starmer & Sugden, 1992). Since then, it has been increasingly used as an alternative model for the expected utility theory and it has been applied to different disciplines such as economics, finance and psychology (see, e.g., Braun & Muermann, 2004; Muermann, Mitchell & Volkman, 2006; Mulaudzi, Petersen & Schoeman, 2008; Solnik, 2008; Wong , 2011).

In this chapter, I shall assume that the decision maker takes into account both risk and regret, instead of considering risk only. Namely, the decision maker is regret averse. This means that: (i) he experiences regret of having allocated a small portion of his wealth in a prospect that yields a higher payoff ex-post; and (ii) he experiences regret of having allocated a large portion of his wealth in a prospect that turns out to have a lower payoff ex-post. Therefore, the disutility of regret is crucial when decision maker should select initial shares of their wealth at the beginning of the period and cannot be modified afterwards.

This chapter contributes with the following. First, I study whether a regret averse decision maker prefers to diversify between two risky prospects rather than to specialize

[1] Harless and Camerer (1994) show that expected utility theory accuracy rate does better than other theories when the gambles (lotteries) in a pair have the same support, and does poorly when they have different support. However, in their work they did not specifically compare expected utility theory with regret theory.

[2] Baron (2000) illustrates these feelings with a simple example. For instance, regret is experienced if we decide to carry an umbrella and find that it does not rain or if we decide not to carry an umbrella and find that it does rain. On the other hand, rejoice is experienced if we carry an umbrella and it rains, or if we do not carry an umbrella and it finally does not rain.

and allocate all the initial wealth in one prospect. I also show conditions to generalize the results to multiple random variables. Second, I analyze whether regret averse and risk averse decision makers coincide or differ in their optimal choices . Last, my results might be used to extend a series of existing models of decision making under uncertainty. In fact, I explain how my main findings can be used in a variety of applications. For instance, first, I apply the results to the portfolio selection problem. Here, I generalize the framework of a risky asset and a risk-free asset model studied by Muermann et al. (2006) and Mulaudzi et al. (2008), but now considering two risky assets. Second, I apply the main findings to study the salesman remuneration scheme (Hildreth & Tesfatsion, 1977) for a regret averse agent.

The rest of the chapter continues as follows. In Section 2, I explain the characteristics of the regret utility function. In Section 3 and 4, I develop the main results of the chapter. In Section 5, I present the mentioned applications, including some illustrative numerical examples. I finish the chapter with concluding remarks.

1.2 A utility function with regret

In this section, I present the regret utility function. We will see that this utility function not only considers risk, but regret as well. As I have mentioned earlier, in regret theory individuals compare between what is received after choosing one option with what might have been received, under the same states of nature, if they had chosen differently. In other words, they compare the chosen outcome with the foregone outcomes.

Loomes and Sugden (1982) propose a utility function represented as follows

$$u(x,y) = v(x) + \varphi\left[v(x) - v(y)\right], \tag{2.1}$$

where v is a conventional Bernoulli's utility function and φ is an increasing function that reflects the valuation of the regret-rejoice feelings. The first argument of this utility x is the chosen alternative, while the second argument y is the foregone alternative. The function φ serves to anticipate and incorporate in the decision making process the regret or rejoice that the individual would experience as a result of having chosen x and not y. After the state of nature occurs we have the following cases: (i) if $x > y$, then the decision maker would experience the pleasure (rejoice) of having made the correct choice; (ii) if $x < y$, then the decision maker would experience regret of not having chosen the best alternative.

Many different regret utility functions, apart from (2.1), have been proposed (see, e.g., Paroush & Venezia, 1979; Braun & Muermann, 2004). For instance, Braun and Muermann (2004) propose the following two attribute additive utility function u given by the formula

$$v(W) - \theta\varphi\left[R\right], \tag{2.2}$$

where $W > 0$ is the final wealth, and R is a measure of regret. Herein, v is a standard Bernoulli utility function with $v' > 0$ and $v'' < 0$. The function φ is coined the regret function. It is continuous and differentiable in its domain, with $\varphi(0) = 0$, $\varphi' > 0$ and $\varphi'' > 0$. Laciana and Webber (2008) propose a regret function that satisfies certain properties that helps to explain the preference patterns described in Allais' paradox (Allais, 1953; Allais & Hagen, 1979). Specifically, their proposal is defined as follows

$$\varphi(x) = \beta^x - 1 \text{ where } \beta > 1.$$

The parameter $\theta \geq 0$ in (2.2) measures the weight of the regret attribute with respect to the first risk aversion attribute. Naturally, if $\theta = 0$ then the utility function (2.2) simplifies and becomes $u\,(W) = v(W)$, which is the traditional utility function of a risk-averse decision maker.

Notice that all assumptions determine that $u_W > 0$ and $u_R < 0$, which means that decision makers like more wealth, but dislike more regret. Besides, they also imply that $u_{WW} < 0$ and $u_{RR} < 0$, reflecting risk aversion and regret aversion respectively. Utility (2.2) considers only regret and does not consider rejoice. Nevertheless, there are many studies that have found that anticipating rejoice has little influenced in the decision making (cf.e.g., Beattie, Baron, Hershey, & Spranca, 1994 and references therein).

From now on, I call regret averse to those decision makers possessing a utility function as defined in (2.2) with $\theta > 0$, and risk averse to those having a utility function (2.2) with $\theta = 0$.

Remark 1.2.1 *Utility function (2.2) has some differences with utility (2.1). First, x and y in (2.1) are replaced by W and R respectively. Second, as I have noticed earlier, it focuses solely on regret, an approach similar in spirit to Savage's regret minimax criterion and to Sarver (2008) utility function representation. Third, utility function (2.2) is mathematically more tractable than utility (2.1). This is one of the reasons that explains its frequent use in recent research (see, e.g., Muermann et al., 2006; Mulaudzi et al., 2008; Wong, 2011). Finally, utility function (2.2) is consistent with experimental evidence that supports regret theory (Bleichrodt, Cillo & Diecidue, 2010; Laciana & Weber, 2008).*

In the next sections, I develop the main results of the chapter. I divide the analysis in two cases. First, I establish some results for two random variables. Second, I characterize the conditions under which the results for two random variables can be extended to multiple random variables.

1.3 Diversification and regret with two random variables

As it is common in the literature on diversification, I restrict myself to the case when the final wealth is the convex combination of two random variables (see, for instance, Hadar, Russell & Seo, 1977). Thus, without loss of generality, the gains from diversification consists in studying the following mathematical problem.

A decision maker needs to determine α, with $\alpha \in [0, 1]$, such that maximizes the following expected utility function:

$$H(\alpha) := \mathbf{E}\left[u\left(\eta(\alpha)\right)\right], \tag{3.1}$$

where u is defined as follows,

$$u(\eta(\alpha)) = v(\eta(\alpha)) - \theta\varphi\left(v(\eta^{\max}) - v(\eta(\alpha))\right). \tag{3.2}$$

with $\eta(\alpha) = Y + \alpha\Delta$ and $\Delta = X - Y$, where X and Y are two non-degenerate random variables.[3]

The term $\eta^{\max}$ is the ex post optimal final value if the decision maker had chosen the optimal choice for each state of the world. Note that $\eta^{\max}$ is a random variable and is independent of α. In more detail, for two random variables X and Y, $\eta^{\max}$ is defined as follows:

$$\eta^{\max} = \max\{X, Y\} = \begin{cases} X & \text{if } Y \leq X, \\ Y & \text{if } Y \geq X. \end{cases} \tag{3.3}$$

In fact, the reason for considering (3.3) can be interpreted as follows:

(i) If X is larger than Y then the best choice the decision maker would have chosen is $\alpha^* = 1$;

(ii) On the other hand, if Y is larger than X the decision maker would have wanted to choose only Y, thus $\alpha^* = 0$.

There are gains from diversification when the optimal solution of the maximization of (3.1), denoted by α^*, is in $(0, 1)$. Otherwise, specialization is optimal (i.e., $\alpha^* = 0$ or $\alpha^* = 1$). (In some parts, I shall use the notation α_θ to denote the optimal α for a given θ).

To solve the aforementioned problem, we need to study the two order conditions. The first derivative of (3.1) with respect to α is :

$$\begin{aligned} H'(\alpha) &= \frac{\delta \mathbf{E}\left[u(\eta(\alpha))\right]}{\delta\alpha} \\ &= \mathbf{E}\left[\Delta v'(\eta(\alpha))\right] + \theta\mathbf{E}\left[\Delta v'(\eta(\alpha))\varphi'\left(v(\eta^{\max}) - v(\eta(\alpha))\right)\right], \end{aligned} \tag{3.4}$$

[3]Throughout this chapter, I shall assume that all the expectations exist.

and the second derivative is equal to

$$
\begin{aligned}
H''(\alpha) &= \frac{\delta^2 \mathbf{E}\left[u\left(\eta(\alpha)\right)\right]}{\delta\alpha^2} \\
&= \mathbf{E}\left[\Delta^2 v''(\eta(\alpha))\right] - \theta\mathbf{E}\left[\Delta^2 v'(\eta(\alpha))^2 \varphi''\left(v(\eta^{\max}) - v(\eta(\alpha))\right]\right. \\
&\quad + \theta\mathbf{E}\left[\Delta^2 v''(\eta(\alpha))\varphi'\left(v(\eta^{\max}) - v(\eta(\alpha))\right]\right. .
\end{aligned}
$$

Note, that $H''(\alpha) \leq 0$ is guaranteed, since: (i) $\theta \geq 0$; (ii) v is strictly concave; and (iii) φ is strictly convex. Therefore, there is a global solution of (3.1). However, I cannot assure that the solution of (3.1) is an interior optimum (i.e., diversification is preferred) or there is a corner solution (i.e., specialization is preferred). Since H is a concave function, $\alpha = 0$ is optimal if and only if $H'(\alpha)|_{\alpha=0} \leq 0$. Similarly, $\alpha = 1$ is optimal if and only if $H'(\alpha)|_{\alpha=1} \geq 0$. Nevertheless, I will later discuss the conditions that assure the existence of an interior optimum.

We know that risk-averse decision makers prefer to choose a mixture of equal shares of independent and identically distributed random variables (Samuelson, 1967). Will a regret-averse decision maker also choose this mixture? The following proposition provides an answer of it.

Proposition 1.3.1 *Both regret averse and risk averse decision makers will allocate theirs initial wealth equally (i.e. $\alpha^* = 1/2$) among the risky choices provided the two stochastic variables X and Y are independent and identically distributed.*

Proof. For convenience, I shall prove this proposition for the continuous case. Let $f(x)f(y)$ be the joint density function of X and Y. Let

$$\eta_\alpha = \alpha X + (1-\alpha)Y. \tag{3.5}$$

Then its distribution function is equal to

$$F_{\eta_\alpha(z)} = \iint\limits_{\alpha x + (1-\alpha)y \leq z} f(x)f(y)dxdy = \int_{-\infty}^{\infty} f(y)\left[\int_{-\infty}^{\frac{z-(1-\alpha)y}{\alpha}} f(x)dx\right]dy.$$

Hence the density function of (3.5) is the derivative of the distribution function with respect to z, which is equal to

$$F'_{\eta_\alpha(z)} = f_{\eta_\alpha}(z) = \int_{-\infty}^{\infty} \frac{1}{\alpha} f\left(\frac{z-(1-\alpha)y}{\alpha}\right) f(y)dy.$$

Now, consider the random variable

$$\eta_{1-\alpha} = (1-\alpha)X + \alpha Y. \tag{3.6}$$

In the same manner, we find that its density function (denoted by $f_{\eta_{1-\alpha}}(z)$) is equal to

$$f_{\eta_{1-\alpha}}(z) = \int \frac{1}{1-\alpha} f\left(\frac{z-\alpha x}{1-\alpha}\right) f(x)dx. \tag{3.7}$$

Changing variables in (3.7) ($t = \frac{z-\alpha x}{1-\alpha}$, with $dt = -\frac{\alpha}{1-\alpha}dx$) yields

$$\begin{aligned} f_{\eta_{1-\alpha}}(z) &= \int \frac{1}{\alpha} f(\frac{z-(1-\alpha)t}{\alpha}) f(t)\, dt \\ &= \int \frac{1}{\alpha} f(\frac{z-(1-\alpha)x}{\alpha}) f(x)\, dx. \end{aligned}$$

Therefore, we conclude that (3.5) and (3.6) have the same density function. So that,

$$H(\alpha) = \mathbf{E}\left[u\left(\alpha X + (1-\alpha)Y\right)\right] = \mathbf{E}\left[u\left((1-\alpha)X + \alpha Y\right)\right] = H(1-\alpha). \tag{3.8}$$

Differentiating (3.8) with respect to α, we obtain

$$H'(\alpha) = -H'(1-\alpha) \tag{3.9}$$

Evaluating (3.9) at $\alpha = 1/2$ we have that

$$H'(\alpha)|_{\alpha=1/2} = -H'(1-\alpha)|_{\alpha=1/2},$$

hence

$$2H'(\alpha)|_{\alpha=1/2} = 0.$$

Therefore we conclude that $H'(\alpha)|_{\alpha=1/2} = 0$. The second order condition holds by the concavity of $H(\alpha)$. ■

The conclusion above proves that a regret-averse decision maker and a risk-averse decision maker would coincide in their allocation weights in the special case when the stochastic variables are independent and identically distributed.

Next, I move forward to consider the case when the two random variables X and Y are independent, but not necessarily identically distributed. But, first, to establish this and further results, I need the following lemma that studies the sign of the first derivative of (3.1) evaluated at $\alpha = 0$ and $\alpha = 1$.

Lemma 1.3.1 *Let X and Y be two random variables. Consider the function (3.1), with $\theta > 0$ then*

$$H'(\alpha)|_{\alpha=0} > (1 + \theta\varphi'(0)) \left[\mathbf{Cov}[\Delta, v'(Y)] + \mathbf{E}[\Delta]\, \mathbf{E}[v'(Y)]\right] \tag{3.10}$$

and

$$H'(\alpha)|_{\alpha=1} < (1 + \theta\varphi'(0)) \left[\mathbf{Cov}[\Delta, v'(X)] + \mathbf{E}[\Delta]\, \mathbf{E}[v'(X)]\right]. \tag{3.11}$$

Proof. Using equation (3.4) evaluated at $\alpha = 0$, we have:

$$\begin{aligned}
&H'(\alpha)|_{\alpha=0} \\
&= \mathbf{E}[\Delta v'(Y)] + \theta\mathbf{E}[\Delta v'(Y)\varphi'(v(\eta^{\max}) - v(Y))] \\
&= \mathbf{E}[\Delta v'(Y)] + \theta\mathbf{E}[\Delta v'(Y)\varphi'(v(Y) - v(Y))) \cdot \mathbf{1}_{X<Y}] \\
&+ \theta\mathbf{E}[\Delta v'(Y)\varphi'(v(X) - v(Y))) \cdot \mathbf{1}_{X\geq Y}] \\
&= \mathbf{E}[\Delta v'(Y)] + \theta\mathbf{E}[\Delta v'(Y)\varphi'(0) \cdot \mathbf{1}_{X<Y}] \\
&+ \theta\mathbf{E}[\Delta v'(Y)\varphi'(v(X) - v(Y))) \cdot \mathbf{1}_{X\geq Y}],
\end{aligned}$$

where $\mathbf{1}_{X<Y}$ is the indicator function which is equal to 1 if $X < Y$, and is equal to 0 otherwise (similar definition applies to $\mathbf{1}_{X\geq Y}$).

Since $\Delta = 0$ when $\mathbf{1}_{X=Y}$, we can write the following identity

$$\mathbf{E}\left[\Delta v'(Y)\varphi'(v(X)-v(Y))\cdot \mathbf{1}_{X\geq Y}\right] = \mathbf{E}\left[\Delta v'(Y)\varphi'(v(X)-v(Y))\cdot \mathbf{1}_{X>Y}\right].$$

Since we have assumed that: (i) $\theta > 0$, (ii) v and φ' are strictly increasing functions, then

$$\theta\mathbf{E}\left[\Delta v'(Y)\varphi'(v(X)-v(Y))\cdot\mathbf{1}_{X>Y}\right] > \theta\mathbf{E}\left[\Delta v'(Y)\varphi'(0)\cdot\mathbf{1}_{X>Y}\right].$$

Therefore, we obtain

$$\begin{aligned}
H'(\alpha)|_{\alpha=0} &> \mathbf{E}\left[\Delta v'(Y)\right] + \theta\mathbf{E}\left[\Delta v'(Y)\varphi'(0)\cdot \mathbf{1}_{X<Y}\right] \\
&\quad + \theta\mathbf{E}\left[(\Delta v'(Y)\varphi'(0)\cdot \mathbf{1}_{X>Y}\right] \\
&= \mathbf{E}\left[\Delta v'(Y)\right] + \theta\mathbf{E}\left[\Delta v'(Y)\varphi'(0)\right] \\
&= \left[1+\theta\varphi'(0)\right]\mathbf{E}\left[\Delta v'(Y)\right] \\
&= \left[1+\theta\varphi'(0)\right]\Bigg[\mathbf{Cov}[\Delta, v'(Y)] + \mathbf{E}\left[\Delta\right]\mathbf{E}\left[v'(Y)\right]\Bigg].
\end{aligned}$$

This completes the proof of the first part.

The proof of the second inequality follows the same argument, but now using equation (3.4) evaluated at $\alpha = 1$. ■

We are now in a position to relax the assumptions in Proposition 1.3.1. I exploit the results of Lemma 1.3.1 to establish the gains from diversification considering two independent random variables. I emphasize that these random variables are not necessarily identically distributed.

Proposition 1.3.2 *Suppose the random variables X and Y are independent. Then for a regret averse decision maker we have:*

- *if* $\mathbf{E}[X] - \mathbf{E}[Y] \geq \frac{\mathbf{Cov}[Y,v'(Y)]}{\mathbf{E}[v'(Y)]}$ *then* $\alpha^* > 0$, *and*
- *if* $\mathbf{E}[X] - \mathbf{E}[Y] \leq -\frac{\mathbf{Cov}[X,v'(X)]}{\mathbf{E}[v'(X)]}$ *then* $\alpha^* < 1$.

Proof. Now, I prove the first case. Since H is strictly concave function, we need to check the sign of (3.4) evaluated at $\alpha = 0$. If it is positive, the decision maker would prefer to hold some amount of X. Now, using inequality (3.10) of Lemma 1.3.1

$$H'(\alpha)|_{\alpha=0} > (1+\theta\varphi'(0))\left[\mathbf{Cov}[\Delta, v'(Y)] + \mathbf{E}\left[\Delta\right]\mathbf{E}\left[v'(Y)\right]\right] \geq 0$$

Since X and Y are independent then $\mathbf{Cov}[X, v'(Y)] = 0$ and:

$$\mathbf{Cov}[\Delta, v'(Y)] = \mathbf{Cov}[X - Y, v'(Y)] = -\mathbf{Cov}[Y, v'(Y)].$$

So that if

$$\mathbf{E}\left[X\right] - \mathbf{E}\left[Y\right] \geq \frac{\mathbf{Cov}[Y, v'(Y)]}{\mathbf{E}\left[v'(Y)\right]},$$

then $\alpha^* > 0$.

For the second part, in the same manner, we need to study the sign of Eq. (3.4) evaluated at $\alpha = 1, H'(\alpha)|_{\alpha=1}$. If it is negative, the decision maker would prefer to allocate some portion of his wealth in Y. Now, using inequality (3.11) of Lemma 1.3.1

$$H'(\alpha)|_{\alpha=1} < (1+\theta\varphi'(0))\left[\mathbf{Cov}[\Delta, v'(X)] + \mathbf{E}\left[\Delta\right]\mathbf{E}\left[v'(X)\right]\right] \leq 0.$$

Again, since X and Y are independent then $\mathbf{Cov}[\Delta, v'(X)] = \mathbf{Cov}[X, v'(X)]$. So that, if

$$\mathbf{E}\left[X\right] - \mathbf{E}\left[Y\right] \leq -\frac{\mathbf{Cov}[X, v'(X)]}{\mathbf{E}\left[v'(X)\right]},$$

then $\alpha^* < 1$. ■

The implications of Proposition 1.3.2 deserve some comments.

Remark 1.3.1 *Notice that by the concavity of v then $\frac{\mathbf{Cov}[Y, v'(Y)]}{\mathbf{E}[v'(Y)]} \leq 0$.*[4] *Therefore, if $\mathbf{E}\left[X\right] \geq \mathbf{E}\left[Y\right]$ then the decision maker will allocate some of his wealth in X. The interesting case is when $\mathbf{E}\left[X\right] \leq \mathbf{E}\left[Y\right]$ and condition $\mathbf{E}\left[X\right] - \mathbf{E}\left[Y\right] \geq \frac{\mathbf{Cov}[Y, v'(Y)]}{\mathbf{E}[v'(Y)]}$ may still hold. This implies that as long $\mathbf{E}\left[X\right] - \mathbf{E}\left[Y\right]$ is not sufficiently negative, then the decision maker will still allocate some amount of his wealth in X.*

[4] *This covariance inequality is known as the covariance rule (Gollier, 2004). We refer to Lehmann (1966), Gurland (1967) and Egozcue et.al.(2009, 2010) for the proof and further inequalities of the covariance.*

Remark 1.3.2 *Notice that for a regret averse decision maker Proposition 1.3.2 does not always imply that diversification is optimal. In fact, specialization could be the best choice. Because in the first case the optimal α can rise to 1, if $\mathbf{E}[\Delta]$ is sufficiently large. Whereas in the second case, the optimal α can fall to 0, if $\mathbf{E}[\Delta]$ is sufficiently low.*

To give a sufficient condition for diversification for regret averse decision makers, we need to combine both inequalities (3.10) and (3.11). This shall be done with the following observation. Since H is strictly concave, a necessary and sufficient condition for diversification to be optimal is that $H'(\alpha)|_{\alpha=0} > 0$ and $H'(\alpha)|_{\alpha=1} < 0$ hold at the same time. The following corollary uses the facts of this observation.

Corollary 1.3.1 *Assume two independent stochastic variables X and Y such that*

$$\frac{\mathbf{Cov}[Y, v'(Y)]}{\mathbf{E}[v'(Y)]} \leq \mathbf{E}[X] - \mathbf{E}[Y] \leq \frac{\mathbf{Cov}[-X, v'(X)]}{\mathbf{E}[v'(X)]}, \tag{3.12}$$

then a regret-averse decision maker would prefer diversification.

I skip the proof of this corollary, since it can be proved invoking Proposition 1.3.2.

Remark 1.3.3 *Notice that since $\mathbf{Cov}[Y, v'(Y)]$ is non-positive and v is increasing, then the lower bound of (3.12) is non-positive. Using similar arguments, the upper bound of (3.12) is non-negative. Therefore, it is obvious that inequalities in (3.12) hold for independent random variables with the same mean. Nevertheless, we will later see that assuming only that X and Y have the same mean is not a sufficient condition to assure preferences for diversification.*

The next examples illustrate the above results.

Example 1.3.1 *Let u defined as in (3.2) with $v(x) = \sqrt{x}$, $\theta = 2$ and $\varphi(x) = e^x - 1$. Consider X and Y two independent random variables, with the following probability mass* *function*

$$f(x,y)=\begin{cases} 0.25 & \textit{if } x=50, y=50 \\ 0.25 & \textit{if } x=a, y=50 \\ 0.25 & \textit{if } x=50, y=80 \\ 0.25 & \textit{if } x=a, y=80 \\ 0 & \textit{otherwise.} \end{cases}$$

Obviously, both random variables are independent. Let

$$\begin{aligned} H(\alpha) &= 0.25\left[\sqrt{50}-2\left[\exp\left\{\sqrt{50}-\sqrt{50\alpha+50(1-\alpha)}\right\}-1\right]\right] \\ &+0.25\left[\sqrt{a\alpha+50(1-\alpha)}-2\left[\exp\left\{\sqrt{\max\{a,50\}}-\sqrt{a\alpha+50(1-\alpha)}\right\}-1\right]\right] \\ &+0.25\left[\sqrt{50\alpha+80(1-\alpha)}-2\left[\exp\left\{\sqrt{80}-\sqrt{50\alpha+80(1-\alpha)}\right\}-1\right]\right] \\ &+0.25\left[\sqrt{a\alpha+80(1-\alpha)}-2\left[\exp\left\{\sqrt{\max\{a,80\}}-\sqrt{a\alpha+80(1-\alpha)}\right\}-1\right]\right]. \end{aligned}$$

Let $a = 100$ *which implies that* $E[X] = 75$ *and* $E[Y] = 65$ *we are under the assumptions of Proposition 1.3.2. Numerical solution of this equation shows there is a maximum at* $\alpha_2^* = 0.73$, *which can be seen in Figure 1.*

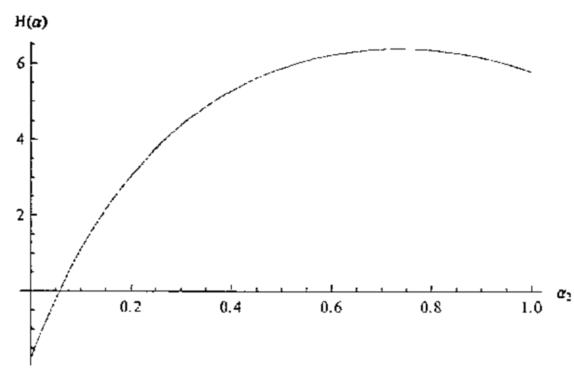

Figure 1 Function $H(\alpha)$ *when* $a = 100$

The interesting case is when $\mathbf{E}[X] \leq E[Y]$ *and diversification is still optimal. Now, suppose that* $a = 78$ *then* $\mathbf{E}[X] - E[Y] = -1$, *and diversification is preferred. Actually,* *the maximum of $H(\alpha)$ is achieved at $\alpha^* = 0.46$. We display $H(\alpha)$ in Figure 2.*

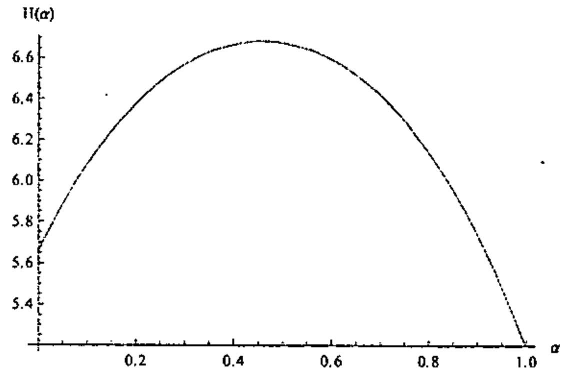

Figure 2 Function $H(\alpha)$ when $a = 78$

So far, I have relaxed the identically distributed assumption, now I shall consider the case when the random variables are stochastically dependent. To relax the independence assumption, I first need to define the concept of stochastic dependence. The intuition of positive dependence between two random variables X and Y implies that larger values of a random variable X are accompanied by larger values of Y. While negative dependence means that larger values of one variable tend to accompany small values of the other variable. However, this basic dependence notion has been improved and more sophisticated definitions of stochastic dependence were developed.

For instance, a well known measure of dependence is defined by Lehmann (1966), which I recall in the next definition.

Definition 1.3.1 *Two random variables X and Y are positive (negative) quadrant dependent if*

$$\mathbf{P}(X \leq x, Y \leq y) \geq (\leq)\ \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y) \ \ \textit{for all } x, y \in R.$$

As it is well known, Lehman's dependence measure implies other weaker notions of dependence. Esary, Proschan and Walkup (1967) introduce the idea of associated random variables and its relation with quadrant dependence and they derived the following inequalities.

Theorem 1.3.1 *Let α and β be two real functions. If X and Y are positive (negative) quadrant dependent then:*

1. *if α and β are increasing (or both decreasing) then $\mathbf{Cov}[\alpha(X), \beta(Y)] \geq (\leq)\ 0$, and*
2. *if one function is increasing and the other decreasing then $\mathbf{Cov}[\alpha(X), \beta(Y)] \leq (\geq)$ 0.*

In the next Proposition, I prove that when the two stochastic variables are negative quadrant dependent with the same mean, it is sufficient to assure preferences for diversification for a regret averse decision maker.

Proposition 1.3.3 *Let X and Y be two random variables that are negative quadrant dependent and have the same mean then a regret averse individual would prefer diversification (i.e., $0 < \alpha^*_{\theta>0} < 1$).*

Proof. We need to show that $H'(\alpha)|_{\alpha=0}$ is positive and $H'(\alpha)|_{\alpha=1}$ is negative. Since $f(x) = x$ is an increasing function and $v'(y)$, by the concavity assumption, is decreasing, by Theorem 1.3.1 negative quadrant dependence implies that $\mathbf{Cov}[X, v'(Y)]$ is non-negative. Hence, using inequality (3.10) of Lemma 1.3.1 and the assumption that $\mathbf{E}[\Delta] = 0$, we conclude that

$$H'(\alpha)|_{\alpha=0} > (1 + \theta\varphi'(0))\,\mathbf{Cov}[\Delta, v'(Y)] \geq 0.$$

By the assumption of negative quadrant dependence and since $f(y) = -y$ is a decreasing function and knowing the assumption that $v''(x) < 0$, thus invoking Theorem 1.3.1 implies that $\mathbf{Cov}[-Y, v'(X)]$ is non positive. Likewise, using inequality (3.11) of Lemma 1.3.1, it follows that

$$H'(\alpha)|_{\alpha=1} < (1 + \theta\varphi'(0))\,\mathbf{Cov}[\Delta, v'(X)] \leq 0.$$

This completes the proof. ■

Notice that in Proposition 1.3.3, gains from diversification might not be optimal for two stochastic variables with the same mean. The assumption of negative quadrant dependence is crucial to achieve this statement. In the next example, I consider two random variables with the same mean and I show that specialization is optimal, contradicting the natural intuition that equality in means implies preference for diversification.

Example 1.3.2 *Let u defined as in (3.2) with $v(x) = \sqrt{x}$ and $\varphi(x) = e^x - 1$. Consider X and Y two random variables with the following probability mass function*

$$f(x,y) = \begin{cases} 0.05 & \text{if } x = 100, y = 81 \\ 0.95 & \text{if } x = 200, y = 201 \\ 0 & \text{otherwise.} \end{cases}$$

The random variables are neither independent nor identically distributed. They have the same mean equal to 195. One can easily check that both random variables are positive quadrant dependent. Now,

$$H(\alpha) = 0.05\left[\sqrt{100\alpha + 81(1-\alpha)} - \theta\left[\exp\left\{\sqrt{\max\{100,81\}} - \sqrt{100\alpha + 81(1-\alpha)}\right\} - 1\right]\right]$$
$$+ 0.95\left[\sqrt{200\alpha + 201(1-\alpha)} - \theta\left[\exp\left\{\sqrt{\max\{200,201\}} - \sqrt{200\alpha + 201(1-\alpha)}\right\} - 1\right]\right].$$

Let $\theta = 2$, as it can be seen in Figure 3 below, $H(\alpha)$ is increasing for all $\alpha \in [0,1]$.

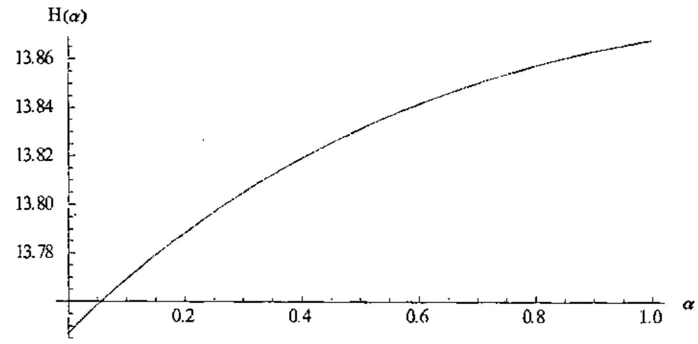

Figure 3 Function $H(\alpha)$ when $\theta = 2$

Thus, the maximum of $H(\alpha)$ is attained at $\alpha = 1$.

So far, I have studied when regret averse behavior is similar or differ to the risk averse behavior. Next, I shall analyze whether the regret averse investor prefers more or less diversification than its risk averse investor counterpart. In addition, how does the parameter, θ influence the optimal choice, α^*? That is, how is the comparative statics between α^* and θ? To accomplish this answer I present the next proposition.

Proposition 1.3.4 *Let X and Y be two random variables.*

- *If* $\mathbf{E}[Y] - \mathbf{E}[X] \geq \frac{\mathbf{Cov}[\Delta, v'(\eta(\alpha^*))]}{\mathbf{E}[v'(\eta(\alpha^*))]}$ *then* $\frac{d\alpha^*}{d\theta} \geq 0$.
- *If* $\mathbf{E}[Y] - \mathbf{E}[X] \leq \frac{\mathbf{Cov}[\Delta, v'(\eta(\alpha^*))]}{\mathbf{E}[v'(\eta(\alpha^*))]}$ *then* $\frac{d\alpha^*}{d\theta} \leq 0$.

Proof. We only prove the first case, the other case can be proved similarly. Taking the total differential of the first order condition $H'(\alpha)$ with respect to α^* and θ yields

$$\frac{d\alpha^*}{d\theta} = -\frac{\mathbf{E}\left[(\Delta)\, v'\varphi'\right]}{\mathbf{E}\left[\Delta^2 v''\right] + \theta \mathbf{E}\left[\Delta^2\left(v''\varphi' - (v')^2\varphi''\right)\right]}.$$

Since (i) v is increasing and concave; and (ii) φ is increasing and convex, then the denominator is negative. Thus,

$$sign\left\{\frac{d\alpha^*}{d\theta}\right\} = sign\left\{\mathbf{E}\left[\Delta v'(\eta(\alpha))\varphi'\left[v(\eta^{\max}) - v(\eta(\alpha^*))\right]\right]\right\}.$$

Now, by the first order condition, we have that

$$\mathbf{E}\left[\Delta v'(\eta(\alpha^*)]\right) + \theta \mathbf{E}\left[\Delta v'(\eta(\alpha^*))\varphi'\left[v(\eta^{\max}) - v(\eta(\alpha^*))\right]\right] = 0.$$

Hence,

$$sign\left\{\mathbf{E}\left[\Delta v'(\eta(\alpha^*))\varphi'\left[v(\eta^{\max}) - v(\eta(\alpha^*))\right]\right]\right\} = -sign\left\{\mathbf{E}\left[\Delta v'(\eta(\alpha^*))\right]\right\}.$$

Thus,

$$sign\left\{\frac{d\alpha^*}{d\theta}\right\} = -sign\left\{\mathbf{E}\left[\Delta v'(\eta(\alpha^*))\right]\right\}.$$

The conclusion follows upon observing that $sign\left\{\mathbf{E}\left[\Delta v'(\eta(\alpha^*))\right]\right\} \leq 0$ is equivalent to

$$\mathbf{E}\left[\Delta\right] \leq -\frac{\mathbf{Cov}\left[\Delta, v'(\eta(\alpha^*))\right]}{\mathbf{E}\left[v'(\eta(\alpha^*))\right]},$$

and the statement follows. ∎

These results allow us to compare the optimal choices of risk averse and regret averse decision makers, as I do in the following corollary.

Corollary 1.3.2 *Let X and Y be two random variables. We have that*

- *if* $\mathbf{E}\left[Y\right] - \mathbf{E}\left[X\right] \geq \frac{\mathbf{Cov}\left[\Delta, v'(\eta(\alpha^*))\right]}{\mathbf{E}\left[v'(\eta(\alpha^*))\right]}$, *then* $\alpha^*_{\theta>0} \geq \alpha^*_{\theta=0} = 0$, *and*
- *if* $\mathbf{E}\left[Y\right] - \mathbf{E}\left[X\right] \leq \frac{\mathbf{Cov}\left[\Delta, v'(\eta(\alpha^*))\right]}{\mathbf{E}\left[v'(\eta(\alpha^*))\right]}$, *then* $\alpha^*_{\theta>0} \leq \alpha^*_{\theta=0} = 1$.

Proof. Notice that if $\mathbf{E}\left[\Delta\right] \leq -\frac{\mathbf{Cov}\left[\Delta, v'(\eta(\alpha^*))\right]}{\mathbf{E}\left[v'(\eta(\alpha^*))\right]}$ then $H'(\alpha)|_{\theta=0} \leq 0$, and thus $\alpha^*_{\theta=0} = 0$. Hence, by Proposition 1.3.4 we conclude that

$$\alpha^*_{\theta>0} \geq \alpha^*_{\theta=0} = 0.$$

I skip the proof of the second part since it can be proved in the same manner. ∎

This result characterizes the behavior of the optimal choice as the regret term θ changes and compares it with the optimal choice of a risk averse individual. In other words, Corollary 1.3.2 establishes the conditions under which regret averse decision makers prefer to move more towards diversification than risk averse counterparts.

So far, I have made the analysis of two random variables and find the conditions under which both risk averse and regret averse coincide and differ in their optimal choices. In the next section, I study the gains from diversification considering many random variables.

1.4 Diversification and regret with multiple random variables

The purpose of this section is to state a generalization of Proposition 1.3.1 with more than two random variables. The analysis of the portfolio problem with more than two random variables is, in general, a complex task. The utility function (3.2) can be extended to multiple random variables. In this case, we consider n random variables denoted by $X_1, X_2, ..., X_n$. It is natural to define $\eta^{\max}$ of (3.1) as follows

$$\eta^{\max} = \max\{X_1, X_2, ..., X_n\}.$$

Consequently, the optimization problem defined in (3.1) transforms to the following new objective function

$$g(\alpha_1, \alpha_2, \ldots, \alpha_n) = \mathbf{E}\left[u\left(\sum_{i=1}^{n}\alpha_i X_i\right)\right], \tag{4.1}$$

with $\mathbb{T} = \{\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbf{R}^n | \alpha_i \geq 0, \sum_{i=1}^{n}\alpha_i = 1\}$ as the choice set.

Our objective is to find $\alpha \in \mathbf{R}^n$ that maximizes (4.1). Since $\mathbb{T}$ is a convex set, and since g is a concave function, the critical point of (4.1) is the global optimum (Simon & Blume, 1994). Therefore, first, I shall prove that g is a concave function.

Proposition 1.4.1 *Let g be as defined in (4.1). Then g is concave.*

Proof. Since $\mathbb{T}$ is a convex set, then we only need to prove that the Hessian of g is semidefinite negative (Simon & Blume, 1994 p. 513). The Hessian of the objective function at any point is

$$\mathbb{H} = \begin{bmatrix} \mathbf{E}\left[X_1^2 u''(.)\right] & \mathbf{E}\left[X_1 X_2 u''(.)\right] & \ldots & \mathbf{E}\left[X_1 X_n u''(.)\right] \\ \mathbf{E}\left[X_2 X_1 u''(.)\right] & \mathbf{E}\left[X_2^2 u''(.)\right] & \ldots & \mathbf{E}\left[X_2 X_n u''(.)\right] \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{E}\left[X_n X_1 u''(.)\right] & \ldots & \ldots & \mathbf{E}\left[X_n^2 u''(.)\right] \end{bmatrix}.$$

Note that the quadratic form associated with $\mathbb{H}$ is

$$\begin{aligned}\mathbb{Q}(y_1, y_2, ..., y_n) &= \sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \mathbf{E}\left[X_i X_j u''(.)\right] \\ &= \mathbf{E}\left[(y_1X_1 + y_2X_2 + \ldots + y_nX_n)^2 u''(.)\right],\end{aligned}$$

where $(y_1, y_2, ..., y_n) \in \mathbf{R}^n$. Recall that by (2.2) the utility function u is given by

$$u(x) = v(x) - \theta\varphi\left[v(\eta^{\max}) - v(x)\right].$$

Since $\theta \geq 0$, v is concave and φ is convex then the utility function u is[5] concave. Thus, $(y_1X_1 + y_2X_2 + \ldots + y_nX_n)^2 u''(.) \leq 0$, which implies that $\mathbb{Q}$ is negative semidefinite for all α_i. This proves that g is concave. ■

To get consistent results with many variables, we need to restrict the analysis to certain types of random variables. In the following proposition, I relax the independence and identically distributed condition studied in Proposition 1.3.1 and consider, instead, exchangeable random variables. Rigorously, this means that, for every permutation π,

$$\left(X_{\pi(1)}, X_{\pi(2)}, ..., X_{\pi(n)}\right) \stackrel{d}{=} (X_1, X_2, ..., X_n),$$

where $\stackrel{d}{=}$ stands for the equality in distribution. Hence, for example, independent and identically distributed random variables are exchangeable, but the opposite is not necessarily true.

Proposition 1.4.2 *Let g defined as in (4.1). Suppose that $X_1, X_2, ..., X_n$ are exchangeable random variables then*

$$g\left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right) \geq g(\alpha_1, \alpha_2, ..., \alpha_n),$$

[5] The concavity follows from the equation

$$u''(x) = v''(x) - \theta\left(v'(x)\right)^2 \varphi''(\cdot) + \theta\varphi'(\cdot)\, v''(x)$$

and the properties of θ, v and φ.

for all $(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{T}$.

Proof. The proof is done by maximizing the Lagrangian

$$L(\alpha_1, \alpha_2, ..., \alpha_n) = g(\alpha_1, \alpha_2, \ldots, \alpha_n) + \lambda \left[1 - \sum_{i=1}^{n} \alpha_i\right] \tag{4.2}$$

with respect to α_i.

The first order condition of (4.2) is equal to

$$\mathbf{E}\left[X_i u'\left(\sum_{i=1}^{n} \alpha_i X_i\right)\right] - \lambda = 0 \text{ for all } i = 1, 2, ..., n,$$

which is also equivalent to

$$\mathbf{E}\left[X_1 u'\left(\sum_{i=1}^{n} \alpha_i X_i\right)\right] = \mathbf{E}\left[X_2 u'\left(\sum_{i=1}^{n} \alpha_i X_i\right)\right] = ... = \mathbf{E}\left[X_n u'\left(\sum_{i=1}^{n} \alpha_i X_i\right)\right]. \tag{4.3}$$

Since $X_1, X_2, ..., X_n$ are exchangeable then a solution of (4.3) is achieved when $\alpha_i = 1/n$ for all $i = 1, 2, ..., n$. The reason for this can be inferred by the same argument as in the proof of Proposition 1.3.1.

Moreover, since g is a concave function by Proposition 1.4.1, then the solution of (4.3) is a global optimum. This completes the proof of this Proposition. ■

Proposition 1.4.2 implies that when facing exchangeable random variables, regret averse investors prefer to choose equal shares of their initial wealth in every prospect rather than any other linear combinations of the prospects. Notice that the same optimal choice holds true for risk averse decision makers.

1.5 Applications

In this section, I show two applications of the main results of this chapter. First, I study the standard portfolio allocation problem within regret theory. Second, I study the optimal remuneration scheme of an agent that considers risk and regret.

1.5.1 The standard portfolio problem

A main issue in portfolio theory is to study whether diversification is optimal. As noted earlier, there are many studies that deal with this problem considering risk averse decision makers. For instance, Arrow (1971), Brumelle (1974), Hadar and Russell (1971, 1974), Hildreth (1974), Pratt (1964), Ross (1981), Samuelson (1967) and Tesfatsion (1976), among others, analyze the convenience of diversification within the classic von Neuman and Morgenstern (1947) expected utility theory. Dekel (1989), Egozcue, Fuentes García, Wong and Zitikis (2011), among others, study the portfolio diversification problem without expected utility theory. However, the study of the standard portfolio problem within regret theory is not very large.

Muermann et al. (2006) and Mulaudzi et al. (2008) study the standard portfolio problem with a risky asset and a risk-free asset. These works establish conditions for preferences for diversification when the choices consist of one-safe asset and one-risky asset. Both works use a regret utility function similar to equation (2.2). They show that if the expected return of the risky asset is equal to the risk free asset return then a regret averse investor prefers to invest some amount of the initial wealth in the risky asset. Just in case, to prevent regret, it is optimal to purchase some amount of the risky asset. This will avoid the feeling of regret if the realized return is larger than the risk free asset. However, a risk-averse decision maker would invest the entire amount of the initial wealth in the risk-free asset. On the other hand, when there is a net premium, a regret-averse decision makers would invest some amount of the initial wealth in the risk-free asset. Contrary, for a large risk premium, a risk-averse decision maker would specialized, investing all the initial wealth in the risky asset (Arrow, 1971; Dalal, 1983). Since, by definition, the variance of the random return of a risk-free asset is zero, we note that this result differs from that in Samuelson (1967), who considers non-degenerated random variables.

Zeelenberg, Beattie, van der Pligt and de Vries (1996) run some experiments where the participants must choose between a risky choice and a safe choice. Their results show that regret agents could promote risk aversion or risk seeking, contrary to the usual claim that regret implies risk aversion. Therefore, their findings contradict the claim that the anticipation of regret only implies risk aversion. Nevertheless, their study does not involve making a formal analysis of the portfolio choice with two risky options.

The framework with a risk-free asset has several limitations. First, several experiments have shown what is called the asymmetric feedback effect (Zeelenber, 1999; Zeelenberg et al., 1996; and Zeelebnerg and Beattie, 1997). That is, the outcome of a certain option is known in advance, thus by choosing the risky choice you will always know the foregone choice. This does not happen when the chosen alternative is the certain option. For this reason, the risk free option might bias the participants' choices reducing the regret influence in their decisions. Second, one can argue that there is not a risk-free asset. For instance, usually US Bonds are associated as a risk free asset. However, since inflation is random US Bonds real rate of return is also random. These two limitations justify the portfolio analysis with two stochastic returns.

The model is as follows. Assume a decision maker must determine the weights of an initial wealth W_0 ($W_0 > 0$), to be invested in two assets $\mathbb{A}_1$ and $\mathbb{A}_2$ with random net returns $\mathbb{R}_1$ and $\mathbb{R}_2$. Therefore the final wealth, W, is a function of $\alpha \in [0, 1]$, and can be expressed as follows

$$W(\alpha) = W_0 [1 + \alpha\mathbb{R}_1 + (1 - \alpha)\mathbb{R}_2].$$

Therefore, the regret averse decision maker's optimization problem is to maximize

$$T(\alpha) = \mathbf{E}[u(W(\alpha))] = \mathbf{E}[u(W_0 [1 + \alpha\mathbb{R}_1 + (1 - \alpha)\mathbb{R}_2])], \tag{5.1}$$

where u defined as in (3.2).[6]

Nevertheless, my main results can be applied in a similar context and study also the risk-free asset case. In particular, we can apply Proposition (1.3.2) to get the following result.

Proposition 1.5.1 *Let $\mathbb{R}_1$ and $\mathbb{R}_2$ and be random independent returns. Suppose $\theta > 0$.*

[6] We do not consider short sales. That is, we do not allow α to be larger than one or less than zero.

1. *If* $\mathbf{E}[\mathbb{R}_1] - \mathbf{E}[\mathbb{R}_2] \geq \frac{\mathbf{Cov}[\mathbb{R}_2, v'(W_0(1+\mathbb{R}_2))]}{\mathbf{E}[v'(W_0(1++\mathbb{R}_2))]}$, *then* $\alpha^* > 0$.

2. *If* $\mathbf{E}[\mathbb{R}_1] - \mathbf{E}[\mathbb{R}_2] \leq -\frac{\mathbf{Cov}[\mathbb{R}_1, v'(W_0(1+\mathbb{R}_1))]}{\mathbf{E}[v'(W_0(1+\mathbb{R}_1))]}$, *then* $\alpha^* < 1$.

This result deserves some comments, and a connection with results already known in the literature.

Remark 1.5.1 *First, Proposition 1.5.1 it is a generalization of the main results in Muermann et al. (2006) and Mulaudzi et al. (2008). Notice that if* $\mathbb{R}_2$ *is a degenerate random variable, then* $\mathbf{Cov}[\mathbb{R}_2, v'(W_0(1+\mathbb{R}_2))] = 0$, *and the model collapses to the risk-free and risky asset model. Second, Proposition 1.5.1 characterizes the conditions under which regret averse will invest in both assets. In the first case, investing all the initial wealth in asset* $\mathbb{A}_2$ *is suboptimal, while in the second case investing all the initial wealth in asset* $\mathbb{A}_1$ *is suboptimal. In fact, the first part of the proposition says that if the difference between the expected return of* $\mathbb{A}_1$ *and the expected return of* $\mathbb{A}_2$ *is large enough, then the regret averse decision maker would invest some amount of its initial wealth in asset* $\mathbb{A}_1$. *Notice that since* $\mathbf{Cov}[\mathbb{R}_2, v'(W_0(1+\mathbb{R}_2))] \leq 0$ *then the condition* $\mathbf{E}[\mathbb{R}_1] > \mathbf{E}[\mathbb{R}_2]$ *alone is not enough for a regret averse decision maker to invest some amount in asset* $\mathbb{A}_1$. *In other words, the feel of not having chosen the asset with largest mean return is not sufficient to assure that decision maker would choose it.*

1.5.2 Mixed remuneration scheme

In this subsection, I show a second application of the main findings of this chapter. In this case, I apply the results to determine a salesman remuneration scheme. This problem has been studied among others by Basu, Lal, Srinivasan and Staelin (1985), Farley (1964) and Hildreth and Tesfatsion (1977).

The model set up is as follows. Suppose an agent (e.g., salesman) must decide on a remuneration scheme. To simplify the analysis, I am not assuming any salesman costs, but I assume that the salesman's remuneration depends only on the total branches' sale and not in the agent individual sales. Suppose $\mathbb{S}_1$ and $\mathbb{S}_2$ are the random sales of two branches of a certain company. The salesman receives a fix percentage $\lambda \in (0,1)$ from one branch sales or from both. Thus, the salesman problem is to find the optimal branches' sale weights of his compensation plan. The agent's income I is equal to

$$I(\alpha) = \alpha\lambda\mathbb{S}_1 + (1-\alpha)\lambda\mathbb{S}_2,$$

where $\alpha \in [0,1]$ is the weight of the branches' sales. Therefore the agent's problem is to maximize

$$\max_{\alpha} \mathbf{E}\left[u(I(\alpha))\right],$$

where u defined in (3.2).

I first assume the case that the branches' sales are independent. Then using the result in Proposition 1.3.2 we deduce the following salesman behavior.

Proposition 1.5.2 *Assume that $\mathbb{S}_1$ and $\mathbb{S}_2$ are independent.*

1. *If $\mathbf{E}[\mathbb{S}_1] - \mathbf{E}[\mathbb{S}_2] \geq \frac{\mathbf{Cov}[\mathbb{S}_2, v'(\mathbb{S}_2)]}{\mathbf{E}[v'(\mathbb{S}_2)]}$, then $\alpha^* > 0$.*
2. *If $\mathbf{E}[\mathbb{S}_1] - \mathbf{E}[\mathbb{S}_2] \leq -\frac{\mathbf{Cov}[\mathbb{S}_1, v'(\mathbb{S}_1)]}{\mathbf{E}[v'(\mathbb{S}_1)]}$, then $\alpha^* < 1$.*
3. *If $\mathbf{E}[\mathbb{S}_1] = \mathbf{E}[\mathbb{S}_2]$, then $0 < \alpha^* < 1$*

This result characterizes the condition under which the salesman will prefer to have a compensation plan that includes the overall company sales. As expected, the weight depends on the difference between $\mathbf{E}[\mathbb{S}_1]$ and $\mathbf{E}[\mathbb{S}_2]$. The remaining results can be applied similarly. For instance, diversification is optimal if the branches' sales are either: (i) exchangeable or (ii) negative quadrant dependent having the same mean.

We can give an illustration of this application with the following example.

Example 1.5.1 *Let u defined as in (3.2) with $v(x) = \sqrt{x}, \theta = 2$ and $\varphi(x) = \exp\{x\} - 1$. Assume the sales of $\mathbb{S}_1$ and $\mathbb{S}_2$ follow the bivariate exponential distribution, which has a joint distribution function*

$$F(x,y) = 1 - e^{-x} - e^{-y} - e + \left(e^{x} + e^{y} - 1\right)^{-1}, \ x, y \geq 0.$$

These random variables are positive quadrant dependent, but are exchangeable (Balakrishnan & Lai, 2009, p. 123). The agent objective function is then equal to

$$
\begin{aligned}
H(\alpha) = & \int_0^{+\infty}\int_0^{+\infty} \sqrt{\alpha x + (1-\alpha)} \frac{2e^{x+y}}{(e^x + e^y - 1)^3} dxdy \\
& - \theta \int_0^{+\infty}\int_0^{+\infty} \left(\exp\left\{\sqrt{\max\{x,y\}} - \sqrt{\alpha x + (1-\alpha)}\right\} - 1\right) \frac{2e^{x+y}}{(e^x + e^y - 1)^3} dxdy.
\end{aligned}
$$

Since $\mathbb{S}_1$ and $\mathbb{S}_2$ are exchangeable then by Proposition 1.4.2 the maximum is attained at $\alpha^* = 0.5$ as it can be seen in Figure 4.

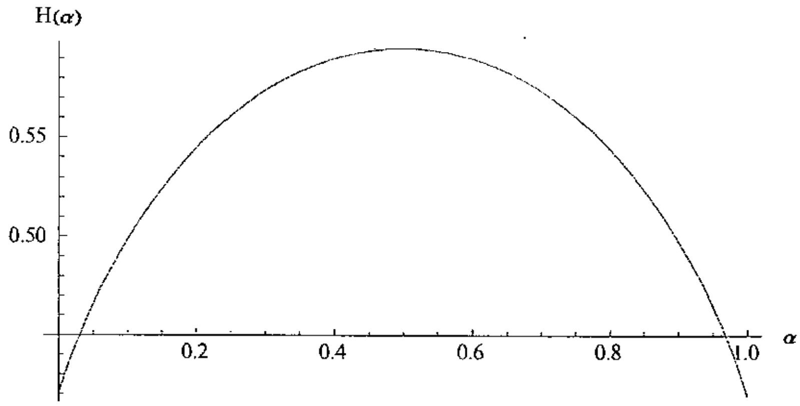

Figure 4 Graph of $H(\alpha)$ considering two exchangeable random variables

This example ends the application section. Although, I have limited the applications to two simple cases, the previous analysis can be applied to similar models of choice under uncertainty.

1.6 Concluding remarks

In this chapter, I study the preferences for diversification of a regret averse decision maker, instead of one with the traditional risk averse behavior. First, I study this optimal allocation problem considering two stochastic variables, instead of considering one risky prospect and one risk-free prospect as in the existing literature. For instance, I prove that if the random variables are independent and identically distributed then complete diversification is optimal. Moreover, when the random variables are independent, then their means play a crucial role to determine preferences for diversification. I also provide conditions to study the diversification behavior of the regret averse decision maker that faces many random variables. In this case, I show that complete diversification is optimal if the random variables are exchangeable. Second, I compare the diversfication preferences behavior of regret averse individuals and that of risk averse counterparts. I do so studying the dynamic relationship between preferences for diversification and regret. I provide the conditions when both behaviors coincide and when they may differ. Finally, I illustrate the practical use of my main findings to two applications: the portfolio selection and the optimal salesman remuneration scheme.

This work can be further improved in several directions. First, a more complete analysis for multiple random variables without the exchangeability assumption is desirable. Second, it would be interesting to use utility function that considers, also, the feeling of rejoice. These remain as tasks for future research.

Chapter 2

Covariance inequalities for non-monotonic functions: theory and applications

2.1 Introduction

Many problems of choice under uncertainty involve studying the sign of a covariance. In particular, many times it is necessary to determine the sign of the covariance of two real functions α and β of a random variable X:

$$\mathbf{Cov}[\alpha(X), \beta(X)]. \tag{1.1}$$

The sign of (1.1) is deduced with the following argument: If these two functions are increasing (or both decreasing) the sign of this covariance is non-negative, while if one function is increasing and the other is decreasing the sign is non-positive (cf., e.g., Gurland 1967; Lehmann, 1966; McEntire, 1984; Schmidt, 2003). This argument relies on Chebyshev's integral inequality (cf., e.g., Mitrinovic & Vasic, 1970; Simonovits, 1995).

We shall see there are some economic problems where α or β is a marginal utility function. For instance, suppose that u is an increasing and concave utility function, then setting $\alpha(x) = x$ and $\beta(x) = u'(x)$, equation (1.1) can be written as follows

$$\mathbf{Cov}[X, u'(X)]. \tag{1.2}$$

In this particular case, as x is an increasing function and u' is a decreasing function (by the concavity of u), the sign of (1.2) is deduced to be non-positive. This particular case of (1.1) has been used in many papers in economics. For instance, Sandmo (1971) studies the sign of covariance (1.2) to characterize the conditions under which a competitive firm, that faces an uncertain price, would produce more or less than under certainty. Similarly, Batra and Russell (1974) use this tool to analyze the effect of international price uncertainty over the social welfare of a small country with two goods. While Mossin (1968) uses this covariance sign to show that full insurance is optimal at an actuarial fair price, while partial insurance is optimal if the premium includes a positive loading.

Nevertheless, Chebyshev's integral inequality crucially depends on the monotonicity behavior of both functions. Sometimes this assumption does not hold. For instance, Wagener (2006) studies the sign of an expression similar to (1.2), that involves a non monotonic function, that helps to derive some results of comparative statics under uncertainty. Besides, other types of utility functions, apart from the traditional one with risk averse behavior, have non-monotonic marginal utilities. For instance, prospect theory proposes a utility function that is S-shaped, which means the marginal utility is non-monotonic. On the other hand, Markowitz (1952) proposes a utility function that, in its simplest case, is reverse S-shaped (RS-shaped), implying that the marginal utility is also non-monotonic. Therefore, Cheyshev's integral inequality doesn't work for marginal utilities of a S-shaped or RS-shaped utility functions.

This chapter contributes in the following. First, I derive some new covariance inequalities for non-monotonic functions that covers the cases when the marginal utilities could be non-monotonic. In particular, I study the sign of covariance (1.2) for prospect theory and for Markowitz utility functions. Second, I apply these new results to two problems in economics. In the first application, I study the shape of the mean variance indifference curves for S-shaped and RS-shaped utility functions. In this application, I shall address the question of whether the monotonicity of the indifference curve in the (μ, σ) still holds for these types of utility functions. Finally, I establish the optimal strategies for hedging asset price risk within prospect theory. Specifically, I examine the optimal strategy for a enterprise that behaves according to prospect theory.

The chapter continues as follows. In the next section, I give a brief view of non-monotonic marginal utility functions. In Section 3, I present some previous covariance inequalities. In Section 4, I derive the main results of the chapter. In Section 5, I present the mentioned applications. I finish the chapter with concluding remarks.

2.2 Non-monotonic marginal utility functions

In this section, I give a brief introduction to different types of non-marginal utility functions. As we have seen, a utility function, u, can take on various shapes: concave, convex, S-shaped and reversed S-shaped, among others. For a further discussion of different forms of the utility function I refer to Gillen and Markowitz (2009).

Friedman and Savage (1948) are among the first to propose alternative shapes of the utility function. Instead of using a utility function that is concave in all the domain, they propose a utility function that could have convex and concave sections. This modification of the utility function would explain, among other things, why individuals buy lotteries (risk) and insurance at the same time.

Markowitz (1952) criticizes Friedman and Savage's proposal and posits an alternative model that modifies the shape of the utility function. In particular, he proposes a utility function where its domain is all the real line. It starts convex then turns concave with an inflection point at the origin turning to convex and finishing concave. The argument in the utility function is the deviation of the final wealth from the current wealth. For simplicity, many authors have used a reverse-shaped (RS) type utility function with only one inflection point at the origin, (cf., e.g., Egozcue, Fuentes García, Wong & Zitikis, 2011; Levy, 2006). Although Markowitz's proposal is appealing, there is nevertheless mixed empirical evidence with regard to this theory (e.g., Hershey and Schoemaker, 1980; Louberge and Outreville, 2001; Post and Levy, 2005; Reilly, 1986).

Egozcue *et al.* (2011) use a power function to represent a RS-shaped utility function given by

$$u(x) = \begin{cases} dx^3 & \text{when} \quad x < 0, \\ x^3 & \text{when} \quad x \geq 0, \end{cases} \tag{2.1}$$

where $d > 0$. As we see in Figures 1 and 2, the marginal utility of this RS-shaped utility function is non-monotonic.

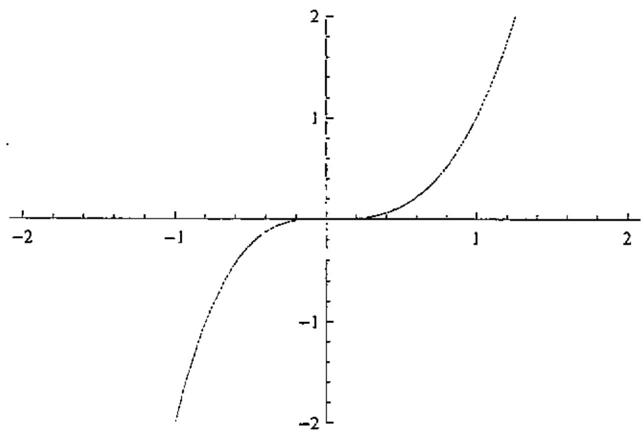

Figure 1: Utility function (2.1) for d= 0.1

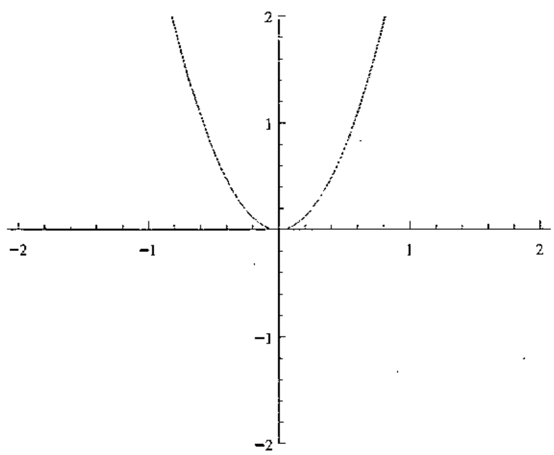

Figure 2: Corresponding marginal utility

The marginal utility is decreasing for negative values of x and increasing for positive values of x. This corresponds with the assumption that decision maker is risk averse in its negative domain and risk seeker in its positive domain.

Based on some ideas by Edwards (1954a, 1954b, 1955, 1962), prospect theory is one of the most famous alternative theories to expected utility theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). It serves to explain a wide range of phenomena that are not explained within the traditional expected utility framework. It is used in different fields such as: economics, finance, marketing and psychology (cf., e.g., Barberis, Huang & Santos, 2001; Dalal, 1983; Pennings & Smidts, 2003; Thaler, 1985, 1994, 1999; and references therein). This theory put forward arguments in favor of an S-shaped utility function that has as an argument the changes of wealth with respect to a certain reference point. Now, I pause to present a proper definition of a S-shaped utility function (Neilson, 2002).

Definition 2.2.1 *A continuous strictly non-decreasing function $u : R \to R$ is called S-shaped if there is a point x_0 such that the function is non-positive and convex to the left of x_0 and non-negative and concave to the right of x_0. The point x_0, that separates losses from gains, is frequently called the reference point, or the status quo. (Throughout the chapter, I set $x_0 = 0$).*

Specifically, Kahneman and Tversky (1979) propose the following power function

$$u(x) = \begin{cases} -\lambda\,(-x)^{\gamma_L} & \text{when} \quad x < 0, \\ x^{\gamma_G} & \text{when} \quad x \geq 0, \end{cases} \tag{2.2}$$

where $\lambda > 0$ is the degree of loss aversion, and γ_G and $\gamma_L \in (0,1)$ are degrees of diminishing sensitivity.

al-Nowaihi, Bradley and Dhami (2008) prove that (2.2) with $\gamma_G = \gamma_L$ is a proper S-shaped function that accounts for preference homogeneity and loss aversion. Nonetheless, this utility function has a mathematical tractability limitation, which is that its first derivative does not exist at $x = 0$. I shall consider this limitation in the main result.

Nevertheless, other different types of S-shaped utility functions have been proposed. For instance, De Giorgi and Hens (2006) suggest to use the following S-shaped function:

$$u(x) = \begin{cases} \lambda_L(e^{\gamma_L x} - 1) & \text{when} \quad x < 0, \\ \lambda_G(1 - e^{-\gamma_G x}) & \text{when} \quad x \geq 0, \end{cases} \tag{2.3}$$

with parameters $\gamma_L, \gamma_G \in [0,1]$ and $\lambda_L, \lambda_G \in (0,\infty)$.

Since we are interested in analyzing the marginal utility of an S-shaped utility function, then we can write them as follows:

(i) For Kahneman and Tversky (1979)[1]

$$u'(x) = \begin{cases} \lambda\gamma_L(-x)^{\gamma_L-1} & \text{when} \quad x < 0, \\ \gamma_G x^{\gamma_G-1} & \text{when} \quad x \geq 0, \end{cases} \tag{2.5}$$

and (ii) for De Giorgi and Hens (2006) as follows

$$u'(x) = \begin{cases} \lambda_L\gamma_L e^{\gamma_L x} & \text{when} \quad x < 0, \\ \lambda_G\gamma_G e^{-\gamma_G x} & \text{when} \quad x \geq 0. \end{cases} \tag{2.6}$$

Naturally, the marginal utility function u' is non-negative because it is generally assumed that the underlying utility function u is non-decreasing. Furthermore, u' in many situations is non monotonic on the entire real line. For instance, assume that $\lambda = 2$ and $\gamma_L = \gamma_G = 0.5$. As we can see in Figures 3 and 4, the marginal utility of a prospect utility function, as defined in (2.2), is non-monotonic. Indeed, it is increasing

[1]Throughout my thesis, I use u' to denote the first derivative (when it exists) of u, and the Radon-Nykodym derivative in the absolutely continuous case (when the derivative may not exist). For example, given the marginal utility $u'(x)$ as in (2.2), the utility function $u(x)$ is given by the formula

$$u(x) = \begin{cases} -\int_x^0 u'(t)dt & x < 0, \\ \int_0^x u'(t)dt & x > 0. \end{cases} \tag{2.4}$$

Notice, that $u(x)$ in (2.4) coincides with $u(x)$ in (2.2). This is in line with the frequently used in statistics notion of absolutely continuous distribution functions. For example, the uniform on $[0, 1]$ density function $f_0(x)$ is related to the uniform distribution $F_0(x)$ by the equation $F_0(x) = \int_0^x f_0(t)dt$, but $F_0(x)$ is not differentiable at the points 0 and 1.

in the loss domain and decreasing in the gains domain.

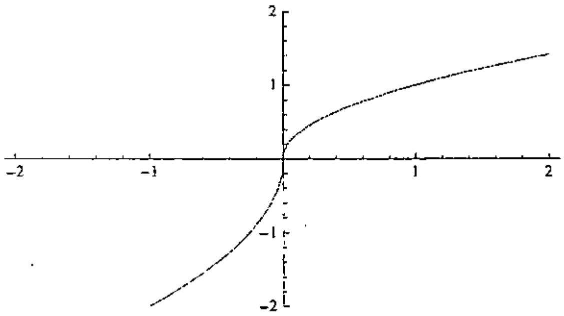

Figure 3: Kahneman and Tversky utility function

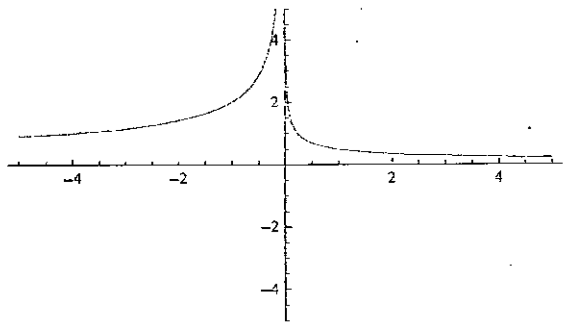

Figure 4: Corresponding marginal utility

In Figures 5 and 6, we display the graphs of the utility function and the respective marginal utility of (2.3) considering $\lambda_L = 2, \lambda_G = 1$ and $\gamma_L = \gamma_G = 0.5$.

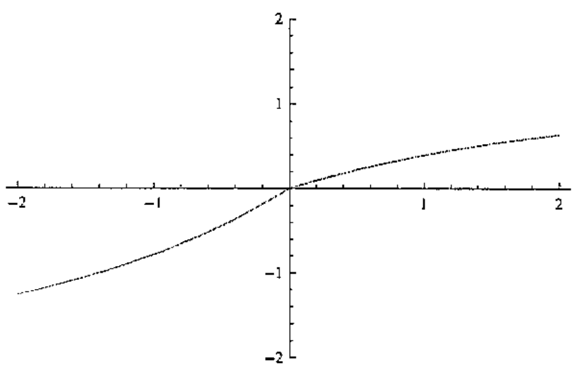

Figure 5: DiGiorgi and Hens utility function

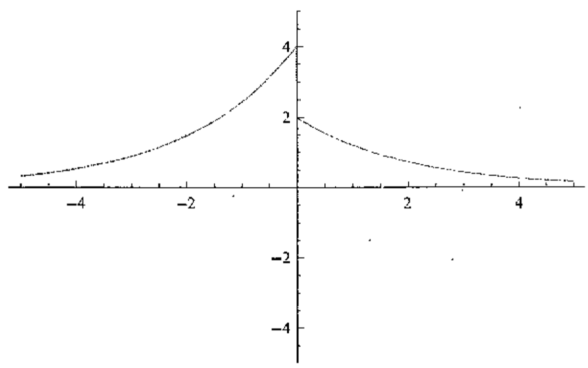

Figure 6: Corresponding marginal utility

Nevertheless, other types of S-shaped utility functions have been little explored in the literature. For instance, Berhold (1973) and LiCalzi (2000) propose the use of cumulative distribution functions to represent S-shaped utility functions. In particular, Broll, Egozcue, Wong and Zitikis (2010) and LiCalzi (2000) use S-shaped utility function of the form,

$$u(x) = F(x), \tag{2.7}$$

where $F(x)$ is the cumulative distribution function of a symmetric random variable. For instance, setting $u(x) = \Phi(x) - 1/2$, where Φ is the standard normal distribution function, it has an S-shaped form, with the reference point at the origin, $x_0 = 0$.

One of the innovative features of prospect theory is loss aversion. The basic idea is that losses loom larger than similar gains. It can be defined in different ways (cf., Neilson, 2002; Köbberling & Wakker, 2005). Therefore, I present a brief summary of the most well known definitions of loss aversion.

Definition 2.2.2 *Let u be an S-shaped utility function, with $u(0) = 0$. Then u exhibits loss aversion if it fulfills one of the following conditions:*

1. *$-u(-x) \geq u(x)$ for all $x > 0$.*
2. *$\frac{u(y)}{y} \leq \frac{u(z)}{z}$ for all $z < 0 < y$ (weak loss aversion).*
3. *$u'(y) \leq u'(z)$ for all $z < 0 < y$ (strong loss aversion).*
4. *If u is defined as in Definition (2.2.1) and $u'(0^-) > u'(0^+)$.*
5. *If u is defined as in equation (2.2) with $\gamma_G = \gamma_L$ and $\lambda > 1$.*

Some authors also define loss aversion as $u'(x) \leq u'(-x)$ for all $x \geq 0$, which is a particular case of the third condition in Definition 2.2.2. Hereafter, I shall consider utility functions that posses this last particular characterization of loss aversion, and I shall also use those utility functions with loss aversion as defined in condition 5.

Nonetheless, the evidence of the presence of loss aversion has received mixed empirical support (cf., e.g., Harinck,Van Dijk, Van Beest & Mersmann, 2007; McGraw, Larsen, Kahneman, & Schkade, 2010; Rozin & Royzman, 2001, and the references therein).

Indeed, although the idea of loss aversion is appealing, there are recent studies that have found evidence of opposite effects. For instance, Harinck et.al.(2007) and McGraw et al. (2010) find evidence that for small outcomes loss aversion is reversed, and individuals weigh more heavily gains than losses, which is referred as reverse loss aversion. In order to make the analysis as general as possible, I would also consider the case of reversed loss aversion (i.e., gains loom larger than losses). I shall call reverse loss aversion utility functions, those with the condition $u'(x) \geq u'(-x)$ for all $x \geq 0$. Note that for power utility functions as defined in (2.2), then reversed loss aversion implies that $\lambda < 1$.

2.3 Some covariance inequalities

In this section, I proceed to present a brief review of some well known covariance inequalities. I begin with the celebrated Chebyshev's integral inequality, which can be stated in its integral (original) form as follows:

Theorem 2.3.1 *Let α, $\beta : [a,b] \to R$ and $f(x) : [a,b] \to R_+$. Then we have:*

1. *If α and β are both increasing or both decreasing, then*

$$\int_a^b f(x)dx \int_a^b \alpha(x)\beta(x)f(x)\ dx \geq \int_a^b \alpha(x)f(x)\ dx \int_a^b \beta(x)f(x)\ dx\,. \tag{3.1}$$

2. *If one of the functions α and β is increasing and the other is decreasing, then the inequality is reversed.*

It is common to see this inequality in its probability form. This can be easily done supposing that $f(x)$ is a probability density function. Then Chebyshev's integral inequality in equation (3.1) can be written as follows:

$$\mathbf{Cov}\left[\alpha(X), \beta(X)\right] \geq 0\,.$$

Therefore, the above theorem, can be expressed in its most usual probabilistic form as follows.

Theorem 2.3.2 *Let X be a continuous random variable defined on $[a,b] \subset R$, with well defined expectations. Consider two real functions α and β then:*

1. *If α and β are both increasing or both decreasing, then* $\mathbf{Cov}\left[\alpha(X), \beta(X)\right] \geq 0$.

2. *If one function is non-decreasing and the other one is non-increasing, then we have* $\mathbf{Cov}\left[\alpha(X), \beta(X)\right] \leq 0$.

We note that if the random variable is non-degenerate and both functions are strictly monotonic then the inequalities in Theorem 2.3.2 are strict. The following Lemma plays an important role in proving the Chebyschev's inequality.

Lemma 2.3.1 *Let α and β be two continuous real functions and X be a continuous random variable defined on $[a, b] \subset R$. Then*

$$\mathbf{Cov}\left[\alpha(X), \beta(X)\right] = \mathbf{E}\left[\left(\alpha(X) - \alpha\left(c\right)\right)\left(\beta(X) - \beta\left(c\right)\right)\right]$$

where $c \in [a, b]$ is such that $\alpha\left(c\right) = \mathbf{E}\left[\alpha(X)\right]$.

Remark 2.3.1 *This result follows directly from applying the Second Mean Value Theorem for integrals (cf., e.g., Sahoo & Riedel, 1998), and using the definition of the covariance (see, e.g. Gurland, 1967; Schmidt, 2003). Notice that Theorem 2.3.2 can be proved invoking Lemma 2.3.1. For instance, assume both functions are increasing, then (i) If $x > c$ then as α and β are both increasing then $\left(\alpha(x) - \alpha\left(c\right)\right)\left(\beta(x) - \beta\left(c\right)\right)$ is non-negative. (ii) On the other hand, if $x < c$ then $\alpha(x) \leq \alpha\left(c\right)$ and $\beta(x) \leq \beta\left(c\right)$, which yields the same result.*

As I have noted earlier, there is an important limitation of the Chebyshev's integral inequality. It requires that both functions must be monotonic. This strong assumption might be violated on several occasions. Hence, studying the sign of (1.1) by relaxing the monotonicity assumption is not only a problem of pure mathematical interest, but it is also of interest in applied mathematics.

Steffensen (1925) proposes a non-monotonic version of Chebyshev's integral inequality. Instead of considering two monotonic functions, he relaxes the monotonicity of one of the functions, however, imposing a special condition on one of the functions. I present this result written with probabilistic notation.

Theorem 2.3.3 *Let α and $\beta : [a,b] \to R$, be differentiable real functions. Consider a random variable X with density function f and support on $[a,b]$. Assume the expectations exist. If*

$$\mathbf{E}\left[\beta(X)|X \leq x\right] \leq (\geq)\, \mathbf{E}\left[\beta(X)\right] \quad \text{for all } x \in [a,b]\,, \tag{3.2}$$

where $E[\cdot|\cdot]$ is the conditional expectation operator, then:

1. *If α is non-decreasing then $\mathbf{Cov}[\alpha(X), \beta(X)] \geq (\leq)\ 0$.*

2. *If α is non-increasing then $\mathbf{Cov}[\alpha(X), \beta(X)] \leq (\geq)\ 0$.*

Egozcue *et al.* (2009, 2011) derive some new covariance inequalities relaxing the monotonicity assumption, but they only work for symmetric random variables.

Theorem 2.3.4 *Let X be a random variable, symmetric[2] about zero with support on $[-b,b]$, and with density function f. Consider two continuous real functions α and β. Assume that β is an odd function[3] with $\beta(x) \geq (\leq) 0$ for all $x \geq 0$. We have that*

1. *if $\alpha(x)$ is increasing, then $\mathbf{Cov}\left[\alpha(X), \beta(X)\right] \geq (\leq) 0$; and*

2. *if $\alpha(x)$ is decreasing, then $\mathbf{Cov}\left[\alpha(X), \beta(X)\right] \leq (\geq) 0$.*

Note that to get consistent results, relaxing the monotonicity assumption of one functions needs the symmetry assumption of the random variable and also the odd function condition. In the next result, however, we relax the monotonicity assumption of both random variables.

Theorem 2.3.5 *Let X be a random variable symmetric about zero. Consider two real functions $\alpha(x)$ and $\beta(x)$. Let $\beta(x)$ be an odd function of bounded variation with $\beta(x) \geq (\leq) 0$ for all $x \geq 0$. We have that*

1. *if $\alpha(x) \geq \alpha(-x)$ for all $x \geq 0$, then $\mathbf{Cov}[\alpha(X), \beta(X)] \geq (\leq) 0$; and*

---

[2] *Recall that a random variable X is symmetric if $X \stackrel{d}{=} -X$, equality in distribution.*

[3] *Recall that $\beta(x)$ is an odd function if $\beta(x) = -\beta(-x)$ for all $x \geq 0$.*

2. *if* $\alpha(x) \leq \alpha(-x)$ *for all* $x \geq 0$*, then* $\mathbf{Cov}[\alpha(X), \beta(X)] \leq (\geq) 0$.

An extension of this result appears for S-shaped utility function in Broll *et al.* (2010). They show that the mean has an important role to determine the covariance sign for a particular type of S-shaped utility functions, as we can see in the following theorem.

Theorem 2.3.6 *Let* X *be symmetric around its mean* $\mu = \mathbf{E}[X]$*. If* u *is an* S*-shaped function, with* $u'(x) = u'(-x)$ *for all* $x \in R$*. Then we have the following statements:*

1. *If* $\mu \geq 0$ *then* $\mathbf{Cov}[X, u'(X)] \leq 0$.

2. *If* $\mu \leq 0$ *then* $\mathbf{Cov}[X, u'(X)] \geq 0$.

This theorem characterizes the sign of the covariance (1.2) for a non-monotonic marginal utility. However, it works only for a utility function that does not consider strict loss aversion, as it is defined in Definition (2.2.2). In the next section, I present a general result of this theorem for S-shaped with loss aversion and reversed loss aversion as well. I also extend this theorem considering reverse S-shaped utility functions.

2.4 Main results

In this section, I present the main results of this chapter. We have seen that S-shaped utility functions have non-monotonic marginal utilities. We have also seen that for some S-shaped utility functions, (e.g. (2.2)), the marginal utility u' does not exist at the reference point 0. Nevertheless, there other S-shaped utility functions with more mathematically tractable behavior such as (2.3) or (2.7). First, I shall state a general theorem where the marginal utility exists in all the real line. Second, I shall relax the assumption of existence of the marginal utility at the origin.

First, in the next result I extend Broll *et al.* (2010) findings considering general S-shaped and RS-shaped utility functions. The novelty of this result is that I shall consider loss aversion and reversed loss aversion as well.

Theorem 2.4.1 *Let X be a symmetric random variable about its mean μ. Let u be a differentiable utility function.*

1. *If u is S-shaped, then we have the following two statements:*

   (a) *If $\mu \geq 0$ and $u'(x) \leq u'(-x)$ for all $x \geq 0$, then $\mathbf{Cov}[X, u'(X)] \leq 0$.*

   (b) *If $\mu \leq 0$ and $u'(x) \geq u'(-x)$ for all $x \geq 0$, then $\mathbf{Cov}[X, u'(X)] \geq 0$.*

2. *If u is RS-shaped utility function, then we have the following two statements:*

   (a) *If $\mu \geq 0$ and $u'(x) \geq u'(-x)$ for all $x \geq 0$, then $\mathbf{Cov}[X, u'(X)] \geq 0$.*

   (b) *If $\mu \leq 0$ and $u'(x) \leq u'(-x)$ for all $x \geq 0$, then $\mathbf{Cov}[X, u'(X)] \leq 0$.*

Proof. First I prove case 1 (a). Define the random variable $Z = X - \mu$. Therefore, Z is symmetric about zero with $\mathbf{E}[Z] = 0$. Thus, we rewrite the covariance as follows

$$\begin{aligned}
\mathbf{Cov}[X, u'(X)] &= \mathbf{Cov}[Z + \mu, u'(Z + \mu)] \\
&= \mathbf{Cov}[Z, u'(Z + \mu)] \\
&= \mathbf{E}[Zu'(\mu + Z)] \\
&= \mathbf{E}[Zu'(\mu + Z) \cdot \mathbf{1}\{Z \geq 0\}] + \mathbf{E}[Zu'(\mu + Z) \cdot \mathbf{1}\{Z < 0\}] \\
&= \mathbf{E}[Z\,(u'(\mu + Z) - u'(\mu - Z)) \cdot \mathbf{1}\{Z \geq 0\}]],
\end{aligned} \tag{4.1}$$

where $\mathbf{1}\{Z \geq 0\}$ is the indicator function, which is equal to 1 whenever $Z \geq 0$ and equal to 0 otherwise.

There are two cases to consider: (i) Let $\mu - z \geq 0$. Since $z \geq 0$ implies $\mu + z \geq \mu - z$ and u' is non-increasing on $(0, \infty)$, we have that

$$u'(\mu + z) - u'(\mu - z) \leq 0. \tag{4.2}$$

(ii) Now assume that $\mu - z \leq 0$. Since $\mu \geq 0$ and $z \geq 0$, we therefore have that $\mu - z \leq 0 \leq \mu + z$. Consequently, using the assumption of $u'(x) \leq u'(-x)$ for all $x \geq 0$, we have that $u'(\mu - z) = u'(-(z-\mu)) \geq u'(z-\mu)$, and thus

$$u'(\mu + z) - u'(\mu - z) \leq u'(\mu + z) - u'(z - \mu). \tag{4.3}$$

We exploit the fact that the right-hand side of bound (4.3) is non-positive because u' is non-increasing on $(0, \infty)$ and $0 \leq z - \mu \leq z + \mu$. Consequently, $u'(z-\mu) \geq u'(z+\mu)$, and thus

$$u'(\mu + z) - u'(\mu - z) \leq u'(\mu + z) - u'(z - \mu) \leq 0. \tag{4.4}$$

Therefore, together from (4.2) and (4.4) we conclude that

$$u'(\mu + z) - u'(\mu - z) \leq 0 \text{ for all } \mu \geq 0 \text{ and } z \geq 0. \tag{4.5}$$

Multiplying in both sides of (4.5) by $\mathbf{1}\{Z \geq 0\}$, taking expectations in both sides and using equality (4.1), we obtain that $\mathbf{Cov}[X, u'(X)] \leq 0$. This finishes the proof of part 1(a).

I now prove part 2 (b). Starting with equality (4.1), we have, again, two cases:

(i) Assume $\mu + z \leq 0$, then (since $\mu \leq 0$ and $z \geq 0$) we have $\mu - z \leq \mu + z \leq 0$. And, thus, we are in the negative domain of u which as it is RS-shaped it is concave. Therefore, we conclude that

$$u'(\mu - z) \geq u'(\mu + z).$$

(ii) Now, assume that $\mu + z \geq 0$. Since we assume that $u'(x) \leq u'(-x)$ for all $x \geq 0$ then $u'(z-\mu) \leq u'(\mu - z)$. So that,

$$u'(\mu + z) - u'(z - \mu) \geq u'(\mu + z) - u'(\mu - z).$$

Notice that $z - \mu \geq \mu + z \geq 0$, thus we are in the positive domain of u, implying that it is convex, thus $u'(\mu + z) \leq u'(z - \mu)$, therefore, we conclude that

$$u'(\mu + z) - u'(\mu - z) \leq 0.$$

At the end, we conclude that for all $z \geq 0$ and $\mu \leq 0$ we have that

$$u'(\mu + z) - u'(\mu - z) \leq 0$$

Following the same steps as in the proof of part 1(a) we conclude that $\mathbf{Cov}[X, u'(X)] \leq 0$. This ends the proof of part 2(b). The other parts can be proved in the same way. ■

I shall now present a numerical illustration of Theorem 2.4.1.

Example 2.4.1 *Suppose X is continuous and uniformly distributed on $[-1, b]$ with $b > 0$. Now, consider the following S-shaped utility function*

$$u_S(x) = \begin{cases} \lambda(e^{0.5x} - 1) & \textit{when} \quad x < 0, \\ 1 - e^{-0.5x} & \textit{when} \quad x \geq 0, \end{cases}$$

and the RS-shaped utility function as defined in (2.1). Let $h(b)$ be defined as follows[4]:

$$h(b) = \mathbf{Cov}[X, u'(X)].$$

In the next figures, we display the graphs of $h(b)$ for different values of d and λ.

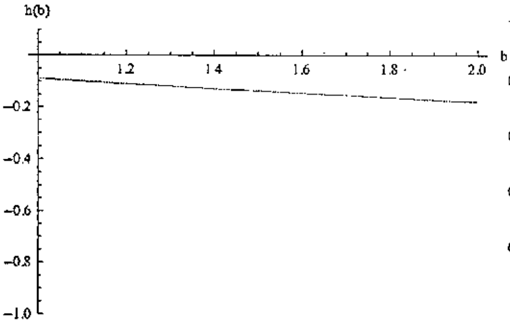

Figure 7: Case 1 (a) with $\lambda = 2$

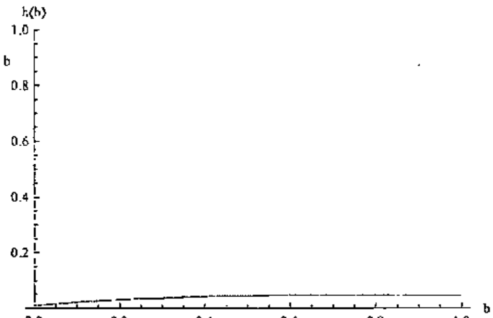

Figure 8: Case 1 (b) with $\lambda = 1/2$

[4] *For the meaning of u' and its relationship with u in the absolutely continuous case, see footnote (1).*

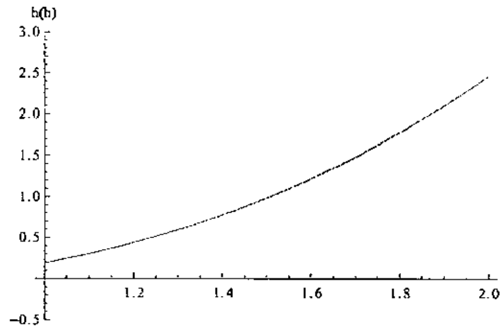

Figure 9: Case 2 (a) with $d = 1/2$

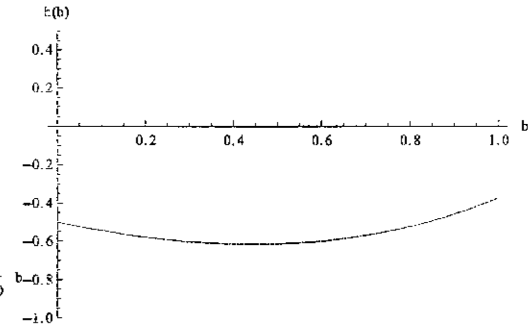

Figure 10: Case 2 (b) with $d = 2$

Figures 7 to 10 display the covariance sign for the four cases in Theorem 2.4.1.

The extension of Theorem 2.4.1 to other S-shaped such as (2.2) can be done considering certain types of symmetric random variables. For example, Theorem 2.4.1 could be adapted to those random variables, X, such as $X \neq 0$ almost surely. I present this extension in the next theorem.

Theorem 2.4.2 *Let X be a random variable symmetric about its mean μ, and such that $X \neq 0$ almost surely. Suppose u is an S-shaped utility function, as defined in (2.2), then we have the following two statements:*

1. *If $\mu \geq 0$ and $u'(x) \leq u'(-x)$ for all $x > 0$, then* $\mathbf{Cov}[X, u'(X)] \leq 0$.

2. *If $\mu \leq 0$ and $u'(x) \geq u'(-x)$ for all $x > 0$, then* $\mathbf{Cov}[X, u'(X)] \geq 0$.

Proof. I only prove the first case, the other case can be proved in the same way. The proof mimics the proof of Theorem 2.4.1. First, with the notation $Z = X - \mu$ we rewrite the covariance $\mathbf{Cov}[X, u'(X)]$ as the expectation $\mathbf{E}[Zu'(\mu+Z)]$. Since $Z \neq -\mu$ by assumption, we have also $Z \neq \mu$, by the symmetry of X. Consequently, $\mathbf{Cov}[X, u'(X)]$

is equal to $\mathbf{E}[Zu'(\mu+Z)\cdot \mathbf{1}\{Z\neq \pm\mu\}]$, and thus

$$\begin{aligned}\mathbf{Cov}[X, u'(X)] &= \mathbf{E}[Zu'(\mu+Z)\cdot \mathbf{1}\{Z\neq \pm\mu\}\cdot \mathbf{1}\{Z>0\}] \\ &\quad + \mathbf{E}[Zu'(\mu+Z)\cdot \mathbf{1}\{Z\neq \pm\mu\}\cdot \mathbf{1}\{Z\leq 0\}] \\ &= \mathbf{E}[Z(u'(\mu+Z)-u'(\mu-Z))\mathbf{1}\{Z\neq \pm\mu\}\mathbf{1}\{Z>0\}]. \qquad (4.6)\end{aligned}$$

I skip the rest of the proof, since it is similar to that of Theorem 2.4.1. ■

2.5 Applications

This section shows some applications of the main results. The range of applications is broad, but I restrict the analysis to two cases. First, I study the monotonicity condition of the mean variance indifference curve for an S-shaped utility function and RS-shaped utility function. Second, I apply the findings to the hedging policies of an enterprise that behaves according to prospect theory.

2.5.1 Mean variance indifference curves for S-shaped and RS-shaped utility functions

The expected utility approach and the mean-variance approach, which is known as (μ, σ) criterion, are in general two different approaches for decision making under uncertainty. The expected utility approach says that X is preferred to Y if and only if

$$\mathbf{E}\left[u(X)\right] > \mathbf{E}\left[u(Y)\right], \qquad (5.1)$$

where u is a concave utility function. On the other hand, the mean variance approach (sometimes also called mean variance rule) was introduced by Markowitz (1952) and states that choice X is preferred over choice Y if

$$\mu_X \geq \mu_Y \text{ and } \sigma_X \leq \sigma_Y,$$

with at least one strict inequality. Here, μ_X and μ_Y denote the mean of X and Y, and σ_X and σ_Y denote their respective standard deviations. The idea is that decision makers use only the mean and variance to make decisions. This is a common tool used by practitioners in finance (Shefrin, 2008). However, it has strong theoretical limitations. For example, does not satisfy the expected utility independence axiom (e.g., Hens and Rieger, 2010, p. 50).

Many scholars study when both approaches are equivalent. Tobin (1958) shows that the two approaches are compatible under normally distributed assets or quadratic utility functions. Moreover, under the normal distribution assumption, the mean variance rule also coincides with the expected utility approach (Hanoch and Levy, 1969). Sinn (1983) and Meyer (1987) show the equivalence of these approaches when the distributions differ only by a location and scale parameters. That is, suppose that X has a distribution that belongs to a class Ω, then $Y = \mu + \sigma X$ where, $\mu \in \mathbf{R}$ and $\sigma > 0$, also belongs to that class of distribution Ω. In other words, if the distribution of X is $F(x)$, then the distribution of Y is equal to $F(\mu + \sigma X)$. Some distributions that satisfy the location scale condition are, among others: the elliptical distributions; the normal distribution; the uniform distribution; the Cauchy distribution and the Student's t distribution.

Sinn (1983) and Meyer (1987) derive several properties of the indifference curve in a (μ, σ) space, generated by a general risk averse von Neumann-Morgenstern utility function. In particular, these studies prove that these indifference curves, represented as a function $\sigma \mapsto \mu(\sigma)$, are increasing and convex. These conditions are useful when the indifference curve is maximized over convex feasible sets. It explained, among other things, issues such as the existence of the CAPM equilibrium, as elucidated by Ormiston and Schlee (2001).

It is important to study the monotonicity of function $\mu(\sigma)$. An increasing function means that the investor is willing to take more risk in exchange of more expected return. This is a crucial assumption of portfolio theory, since larger returns associates higher risk. Therefore, as an application of the main results, I will study whether the monotonicity property still holds for S-shaped and RS-shaped utility functions.

To keep the analysis as simple as possible, I do not consider transformations of the distribution function as prospect theory suggests (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Hereafter, I assume that the random return Y belongs to the location-scale family $\{\mu + \sigma X \,:\, \mu \in \mathbf{R}, \sigma > 0\}$, where X is a random variable with mean 0 and variance 1, and whose distribution function F does not depend on μ and σ. Hence, the expected utility $\mathbf{E}[u(Y)]$ defines a two-argument function

$$V(\mu, \sigma) = \mathbf{E}[u(\mu + \sigma X)] = \int u(\mu + \sigma x) dF(x). \tag{5.2}$$

Various properties of $V(\mu, \sigma)$, its partial derivatives

$$V_\mu(\mu, \sigma) = \frac{\partial}{\partial \mu} V(\mu, \sigma),$$
$$V_\sigma(\mu, \sigma) = \frac{\partial}{\partial \sigma} V(\mu, \sigma),$$

and especially of

$$S(\mu, \sigma) = -\frac{V_\sigma(\mu, \sigma)}{V_\mu(\mu, \sigma)},$$

have been extensively investigated in the literature (see, for example, Sinn, 1983; Meyer, 1987).

The quantity $S(\mu, \sigma)$ has played a particularly prominent role. For instance, it can be viewed as the derivative with respect to the standard deviation σ of the indifference function $\sigma \mapsto \mu(\sigma)$, which, for a given constant α, can be viewed as the curve $\{(\sigma, \mu) : V(\mu, \sigma) = \alpha\}$ drawn on the (σ, μ)-plane.

Hence, if $S(\mu, \sigma)$ is positive, then the indifference function $\sigma \mapsto \mu(\sigma)$ is increasing, whereas if $S(\mu, \sigma)$ is negative, then the indifference function is decreasing. Assuming that the utility function u is differentiable and some integrability conditions are satisfied, we have the equations

$$V_\mu(\mu, \sigma) = \mathbf{E}[u'(Y)], \tag{5.3}$$

$$V_\sigma(\mu, \sigma) = \frac{1}{\sigma} \mathbf{Cov}[Y, u'(Y)], \tag{5.4}$$

where $Y = \mu + \sigma X$. Since $S(\mu, \sigma) = -V_\sigma(\mu, \sigma)/V_\mu(\mu, \sigma)$, we therefore have that

$$S(\mu, \sigma) = -\frac{1}{\sigma}\frac{\mathbf{Cov}[Y, u'(Y)]}{\mathbf{E}[u'(Y)]}. \tag{5.5}$$

We may view $V_\mu(\mu, \sigma)$ as the expected marginal utility or, in other words, the slope of the expected utility $V(\mu, \sigma)$ with respect to μ. Likewise, we may view $V_\sigma(\mu, \sigma)$ as the expected marginal utility $V(\mu, \sigma)$ with respect to σ. Finally, we may view $S(\mu, \sigma)$ as the slope of the indifference function $\sigma \mapsto \mu(\sigma)$.

This indifference curve and its various properties (e.g., monotonicity, convexity, concavity, and so forth) have received considerable attention in the literature. As we have noted above, some of the properties follow from the corresponding ones of the indifference-function $\sigma \mapsto S(\mu, \sigma)$. In particular, the following general property is well known (see, for example, Eichner, 2008; Eichner & Wagener, 2009; Meyer, 1987; and references therein).

Theorem 2.5.1 *If the distribution of Y with mean μ and variance σ^2 belongs to a location-scale family, and the twice differentiable utility function u is increasing on its domain of definition, then we have the following two statements:*

1. *If the utility function u is concave then the indifference function $\sigma \mapsto \mu(\sigma)$ is increasing and convex.*

2. *If the utility function u is convex then the indifference function $\sigma \mapsto \mu(\sigma)$ is decreasing and concave.*

It is now natural to extend formulas (5.3)–(5.5) to the case of general marginal utility functions u' and random variables Y. As before, I use the notation $\mu = \mathbf{E}[Y]$ and $\sigma^2 = \mathbf{Var}[Y]$.

Determining the sign of (5.5) is obviously equivalent related the sign of $\mathbf{Cov}[Y, u'(Y)]$. When the marginal utility is monotonic, then we know that $\mathbf{Cov}[Y, u'(Y)] \geq 0$ for every non-decreasing u' and $\mathbf{Cov}[Y, u'(Y)] \leq 0$ for every non-increasing u'. However, the marginal utility may be non-monotonic, as noted earlier. To cover such functions, I establish the following theorem that studies the monotonicity of the indifference curve generated by S-shaped utility functions.

Theorem 2.5.2 *Suppose the utility function u is S-shaped . Let* $Y = \mu+\sigma X$ *be a random variable where* X *is a symmetric random variable with zero mean and unit variance. Assume the location scale condition hold.*

1. *If* $\mu \geq 0$ *and* $u'(x) \leq u'(-x)$ *for any* $x > 0$, *then* $V_\sigma(\mu, \sigma) \leq 0$ *and thus the indifference function* $\sigma \mapsto \mu(\sigma)$ *is increasing.*

2. *If* $\mu \leq 0$ *and* $u'(x) \geq u'(-x)$ *for any* $x > 0$, *then* $V_\sigma(\mu, \sigma) \geq 0$ *and thus the indifference function* $\sigma \mapsto \mu(\sigma)$ *is decreasing.*

Proof. I only prove Part (1) of the theorem by considering the case $\mu \geq 0$. We have seen that the slope of the indifference function $\sigma \mapsto \mu(\sigma)$ is determined by the sign of $\mathbf{Cov}[Y, u'(Y)]$. Since X is symmetric about zero, then Y is also symmetric about μ. Therefore, invoking the first part of Theorem 2.4.1, we deduce that $\mathbf{Cov}[Y, u'(Y)] \leq 0$ and thus $S(\mu, \sigma) \geq 0$, which implies that the assertion in Part (1) of Theorem 2.5.2 holds. Part (2) can be proved in the same way. ■

Next, I study the monotonicity property of the indifference curve for RS-shaped utility functions.

Theorem 2.5.3 *Consider the utility function as defined in (2.1), in which case is RS-shaped. Let* $Y = \mu + \sigma X$ *be a random variable where* X *is a symmetric random variable with zero mean and unit variance.*

1. *If* $\mu \geq 0$ *and* $u'(x) \geq u'(-x)$ *for all* $x > 0$, *then* $V_\sigma(\mu, \sigma) \geq 0$ *and thus the indifference function* $\sigma \mapsto \mu(\sigma)$ *is decreasing.*

2. *If* $\mu \leq 0$ *and* $u'(x) \leq u'(-x)$ *for all* $x > 0$, *then* $V_\sigma(\mu, \sigma) \leq 0$ *and thus the indifference function* $\sigma \mapsto \mu(\sigma)$ *is increasing.*

Proof. The proof is analogous to the one in Theorem 2.5.2, but now invoking the results in the second part of Theorem 2.4.1. ■

2.5.2 Hedging policies within prospect theory

Continually changing volatilities on financial markets coupled with rises in interest rates, foreign exchange rates, and prices for goods and services have led to the development of various futures markets. These risk-oriented markets have experienced a remarkable rate of growth throughout the world and resulted in the creation of many new financial hedging instruments. These hedging instruments allow a better control of risk exposure faced by an enterprise (see, for example, Bessis, 2009; Freixas & Rochet 2008; Meyer & Robinson, 1988).

In an important contribution to the literature on futures markets and hedging, Benninga, Eldor and Zilcha (1983) address the issue of optimal hedging in the presence of unbiased futures prices. They derive conditions for the optimal hedge to be a fixed proportion of the cash position, regardless of the agent's utility function. This result is important because of the sizeable research on theoretical and empirical hedging that abstracts from the particular utility functions of risk-averse, expected utility maximizers (see, for example, Battermann *et al.* 2000; Broll & Eckwert, 2006; Dewatripont & Tirole, 1994; Freixas & Rochet, 2008 and references therein). The novelty of my application is to incorporate prospect theory into the utility function of a firm[5]. The enterprise has a prospect utility function defined over its end-of-period profit. To hedge its risk exposure, the firm trades futures contracts. I show that when the utility function is S-shaped, the main results of the previous section plays a pivotal role in determining the optimal hedging of the firm.

In this application, I follow Broll and Wahl (2006) model of a firm with one-period planning horizon.

The model set up is as follows. The enterprise that has risky assets with random return (future spot price) r. The assets are financed partially with external funds (deposits), denoted by $\mathbb{D}$, which pays a certain return (price) $r_{\mathbb{D}} > 0$. The enterprise also finances

[5] I am assuming that Fisher's separation theorem does not hold. This implies that the firm might maximize the expected utility of profits, instead of maximizing just profits. Beyond enterprises, my developed theory could be widely applicable to model and analyze decisions of individual agents, such as farmers and other individual entrepreneurs.

its assets with a fixed equity $\mathbb{K} > 0$. Therefore we can write the firm's balance sheet constraint as follows,

$$\mathbb{A} = \mathbb{D} + \mathbb{K}. \tag{5.6}$$

There are operational costs that depends on the deposits level. We represent these costs with a function $C(\mathbb{D})$, which we assume to be increasing and convex. In part of the uncertainty, the risky assets, $\mathbb{A}$, can be hedged in the forward market at a certain price $r_{\mathbb{A}}$. Let H denote the amount of the hedged assets that is determined at the beginning of the period. When H is positive means that the firm is selling assets in the future market. On the other hand, if H is negative it means that firm is purchasing assets in the future market. It is said that speculation is involved if $H \notin [0, \mathbb{A}]$; otherwise the assets are hedged without speculation. For instance, $H < 0$ it means that the firm is purchasing assets in the forward market, while $H > \mathbb{A}$ means that the firm is selling in the future market an amount greater than its current assets.

Since $\mathbb{A} = \mathbb{D} + \mathbb{K}$ is known, in this scenario, next period enterprise's profit is given by

$$\Pi(H) = r\,(\mathbb{A} - H) - r_{\mathbb{D}}\mathbb{D} - C(\mathbb{D}) + r_{\mathbb{A}}H. \tag{5.7}$$

Firm's profit is uncertain and its mean is given by

$$\mu(H) = \mathbf{E}\left[\Pi(H)\right] = \mathbf{E}\left[r\right](\mathbb{A} - H) - r_{\mathbb{D}}\mathbb{D} - C(\mathbb{D}) + r_{\mathbb{A}}H. \tag{5.8}$$

As we shall see, the value of the mean, for the reasons studied in the previous section, has an important role in determining the optimal hedging decision.

Therefore, the firm manager's problem is to find the optimal hedging that maximizes the expected utility of profits. However, instead of considering a traditional Bernoulli utility function, the firm uses an S-shaped utility function u as defined in Definition 2.2.1.

Thus, the firm wants to maximize its expected utility of profit

$$\max_{H} \mathbf{E}[u(\Pi(H))] = \mathbf{E}\left[u(r\mathbb{A} - r_{\mathbb{D}}\mathbb{D} - C(\mathbb{D}) + H(r_{\mathbb{A}} - r))\right]. \tag{5.9}$$

In other words, we want to find the H that maximizes the expected utility of profits. I denote by H^* the solution of (5.9). Since u is an S-shaped function, then there is no guarantee that $\mathbf{E}[u(\Pi(H))]$ will be concave respect to H. Therefore, I restrict the analysis to those cases where the first order condition holds and there is a global solution of (5.9).

Proposition 2.5.1 *If the first order condition of (5.9) holds then we have*

$$(r_{\mathbb{A}} - \mathbf{E}[r])(\mathbb{A} - H^*) = \frac{\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))]}{\mathbf{E}[u'(\Pi(H^*))]}. \tag{5.10}$$

Proof. Taking the first order condition of (5.9), evaluated at H^*, we have

$$\mathbf{E}[(r_{\mathbb{A}} - r)u'(\Pi(H^*))] = 0$$

The latter equation can be rewritten as follows:

$$\frac{\mathbf{E}[ru'(\Pi(H^*))]}{\mathbf{E}[u'(\Pi(H^*))]} = r_{\mathbb{A}}.$$

Now, using the covariance function we have

$$r_{\mathbb{A}} - \mathbf{E}[r] = \frac{\mathbf{Cov}[r, u'(\Pi(H^*))]}{\mathbf{E}[u'(\Pi(H^*))]}. \tag{5.11}$$

After subtracting (5.7) with (5.8) we get

$$(r - \mathbf{E}[r])(\mathbb{A} - H^*) = \Pi(H^*) - \mathbf{E}[\Pi(H^*)],$$

which implies that

$$r = \frac{\Pi(H^*) - \mathbf{E}[\Pi(H^*)]}{(\mathbb{A} - H^*)} + \mathbf{E}[r]. \tag{5.12}$$

Substituting (5.12) in the covariance term of (5.7) we have

$$\mathbf{Cov}[r, u'(\Pi(H^*))] = \mathbf{Cov}\left[\frac{\Pi(H^*) - \mathbf{E}[\Pi(H^*)]}{(\mathbb{A} - H^*)} + \mathbf{E}[r], u'(\Pi(H^*))\right]$$
$$= \frac{1}{\mathbb{A} - H^*}\mathbf{Cov}\left[\Pi(H^*), u'(\Pi(H^*))\right].$$

Therefore we have from equation (5.11) that

$$(r_{\mathbb{A}} - \mathbf{E}[r])(\mathbb{A} - H^*) = \frac{\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))]}{\mathbf{E}[u'(\Pi(H^*))]}. \tag{5.13}$$

■

Finding H^* is generally a complex task. Nevertheless, equation (5.13) has important derivations. Since $\mathbf{E}[u'(\Pi(H^*))] > 0$, the sign of the covariance $\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))]$ determines the sign of the product $(r_{\mathbb{A}} - \mathbf{E}[r])(\mathbb{A} - H^*)$. When u is more complexly shaped than being concave, then determining the sign of the covariance $\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))]$ is a challenging task.

The relation between the expected spot price and the future price will also determine the sign of $(\mathbb{A} - H^*)$. When there are more hedgers taking short positions in the future market than those that are going long then, to reach a balance, speculators must enter the market taking long positions. The speculators will do so, only if $r_{\mathbb{A}} < \mathbf{E}[r]$ (a condition named normal backwardation). Conversely, if there are more hedgers taking long positions than those that are short, speculators will enter the market if $r_{\mathbb{A}} > \mathbf{E}[r]$ (a condition named contango).

Note 2.5.1 *Notice that using the spot-futures parity relationship (which states that ratio of return on perfectly hedged stocks equals the risk-free interest rate) we can write*

$$r_{\mathbb{A}} = \mathbf{E}[r]\left(\frac{1 + r_f}{1 + k}\right)^n,$$

where r_f is the risk free interest rate, k the required rate of return and n is the number of periods (Bodie, Kane and Marcus, 1996 p. 708). Thus, $r_{\mathbb{A}}$ will be less than $\mathbf{E}[r]$ whenever $k > r_f$ (i.e., the asset has a positive beta). When the expected price equals the *forward price then the price is unbiased, this is the case when $r_f = k$. Finally, $r_{\mathbb{A}} > \mathbf{E}[r]$, whenever $k < r_f$ (i.e., the asset has a negative beta).*

Next, I present an application of Theorem 2.4.1 that studies the sign of the covariance in (5.10).

Proposition 2.5.2 *Let the distribution of r be symmetric around its mean $\mathbf{E}[r]$. Let u be an S-shaped utility function, with loss aversion defined as $u'(x) \leq u'(-x)$ for all $x > 0$. If $\mu(H) \geq 0$ then*

$$\mathbf{Cov}[\Pi(H), u'(\Pi(H))] \leq 0.$$

Proof. The proof follows directly by invoking Theorem 2.4.1. ■

We see that $\mu(H)$ plays a decisive role in determining the sign of the covariance $\mathbf{Cov}[\Pi(H), u'(\Pi(H))]$. When $H = H^*$, note that we have the following expressions for the mean:

$$\begin{aligned} \mu(H^*) &= (\mathbf{E}[r] - r_{\mathbb{A}})(\mathbb{A} - H^*) + r_{\mathbb{A}}\mathbb{D} - r_{\mathbb{D}}\mathbb{D} - C(\mathbb{D}) \\ &= -\frac{\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))]}{\mathbf{E}[u'(\Pi(H^*))]} + r_{\mathbb{A}}\mathbb{A} - r_{\mathbb{D}}\mathbb{D} - C(\mathbb{D}). \end{aligned} \tag{5.14}$$

Using Proposition 2.5.2, we obtain the following corollary, which guides the firm in deciding whether to speculate or not. More precisely, it will tell us whether H^* is smaller or greater than $\mathbb{A}$, depending on whether the expected price $E[r]$ is smaller or greater than the forward rate $r_{\mathbb{A}}$.

Corollary 2.5.1 *Let the distribution of r be symmetric around its mean $\mathbf{E}[r] > 0$. Let u be S-shaped, with loss aversion defined in (2.2.2) $u'(x) \leq u'(-x)$ for all $x > 0$. Assume that H^* is a solution of (5.13) such that $\mu(H^*) = \mathbf{E}[\Pi(H^*)] \geq 0$, then we have the following statements:*

1. *If $r_{\mathbb{A}} < \mathbf{E}[r]$, then $H^* \leq \mathbb{A}$.*

2. *If $r_{\mathbb{A}} > \mathbf{E}[r]$, then $H^* \geq \mathbb{A}$.*

Proof. Now, I prove the first part. The third part can be proved similarly. From the first order condition of (5.9) we have

$$(r_{\mathbb{A}} - \mathbf{E}[r])(\mathbb{A} - H^*) = \frac{\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))]}{\mathbf{E}[u'(\Pi(H^*))]}.$$

Since $\mu(H^*) \geq 0$ using Proposition 2.5.2 then $\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))] \leq 0$.Therefore, the sign of $(r_{\mathbb{A}} - \mathbf{E}[r])$ is the opposite to the sign of $(\mathbb{A} - H^*)$. Therefore, since

$$\frac{\mathbf{Cov}[\Pi(H^*), u'(\Pi(H^*))]}{\mathbf{E}[u'(\Pi(H^*))]} \leq 0,$$

and $r_{\mathbb{A}} < \mathbf{E}[r]$ we have $H^* \leq \mathbb{A}$. ■

This result has the following intuition.[6] In the first case, if the forward price is less than the expected spot price, then the firm will hedge an amount less than its current assets. However, if the gap between the forward price and the expected prices is large enough, then it could purchase assets in the future market. The firm will expect to sell them at a greater price in the future. In the second case, if the forward price is greater than the expected price, then the firm will speculate selling an amount greater than its assets, expecting to purchase the additional assets in the future at a lower price.

These are well known results for decision makers with strict risk aversion (Feder, Just and Schmitz, 1980; Houlthasen, 1979). At the end, under these conditions, the enterprise hedging policies with an S-shaped utility are similar as if it uses an increasing and concave utility function.

I finish this section, giving a numerical example of the enterprise hedging policies.

Example 2.5.1 *Consider* $\mathbb{A} = 10, \mathbb{D} = 1, r_{\mathbb{D}} = 0.1, C(1) = 2,$ *and* r *could be equal to* 1 *or* 0 *with equal probability. Let*

$$u(x) = \begin{cases} \sqrt{x} & \text{if } x \geq 0, \\ -2\sqrt{-x} & \text{if } x < 0. \end{cases}$$

[6] I have left the case $r_{\mathbb{A}} = \mathbf{E}[r]$ as a task for future research, because it is more involved. It requires to prove that $\mathbf{Cov}[X, u'(aX + b)] = 0$ implies $a = 0$, where a and b are real numbers.

1. *Suppose that* $r_{\mathbb{A}} = 0.25 < \mathbf{E}[r] = 0.5$. *Then*

$$\begin{aligned}\mathbf{E}[u(\Pi(H))] = \frac{1}{2}\sqrt{\frac{(7.9-0.75H)\,(sgn(7.9-0.75H)+1)}{2}} \\ - \sqrt{\frac{(7.9-0.75H)\,(sgn(7.9-0.75H)-1)}{2}} \\ + \frac{1}{2}\sqrt{\frac{(0.25H-2.1)\,(sgn(0.25H-2.1)+1)}{2}} \\ - \sqrt{\frac{(0.25H-2.1)\,(sgn(0.25H-2.1)-1)}{2}},\end{aligned}$$

where $\operatorname{sgn}(x)$ *is the sign function that takes on values:* 1 *when* $x > 0$, -1 *when* $x < 0$, *and* 0 *when* $x = 0$. *Below, I display the plot of* $\mathbf{E}[u(\Pi(H))]$.

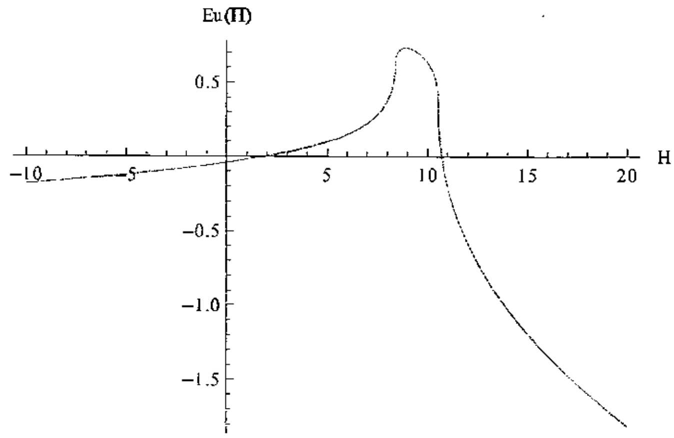

Figure 11

One can check numerically that the maximum is attained at $H^* = 8.93 < \mathbb{A} = 10$.

2. Now I assume that $r_{\mathbb{A}} = 0.55 > \mathbf{E}[r] = 0.5$. *Then*

$$\begin{aligned}\mathbf{E}[u(\Pi(H))] = &\frac{1}{2}\sqrt{\frac{(7.9-0.45H)\,(sgn(7.9-0.45H)+1)}{2}} \\ &-\sqrt{\frac{(7.9-0.45H)\,(sgn(7.9-0.45H)-1)}{2}} \\ &+\frac{1}{2}\sqrt{\frac{(0.55H-2.1)\,(sgn(0.55H-2.1)+1)}{2}} \\ &-\sqrt{\frac{(0.55H-2.1)\,(sgn(0.55H-2.1)-1)}{2}}.\end{aligned}$$

In Figure 12, I show the graph of $\mathbf{E}[u(\Pi(H))]$, *which attains a maximum value at* $H^* = 11.37 > \mathbb{A} = 10$.

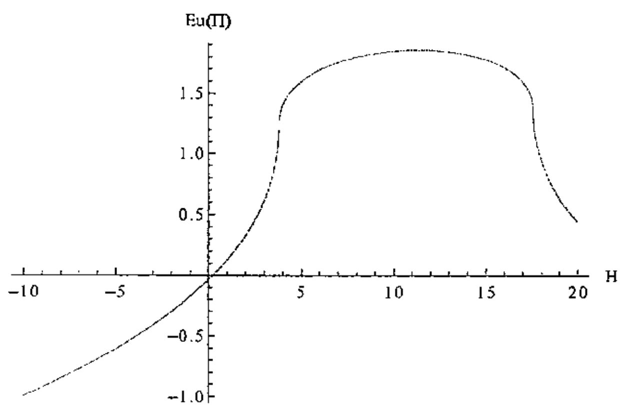

Figure 12

2.6 Concluding remarks

In this chapter, I establish new covariance inequalities that involve non-monotone functions. In particular, I derive new results to study the sign of $\mathbf{Cov}[X, u'(X)]$, when the marginal utility is non-monotonic. This is the case when the utility functions are according to Markowitz utility functions or behaves as prospect theory utility functions. I show that the sign depends on the mean of the random variable and on the degree of loss aversion.

Two applications illustrate the main results of this chapter. First, I study the monotonicity properties of the indifference curves on the (σ, μ)-plane for S-shaped and RS-shaped utility functions. My results show that the indifference curve of S-shaped utility are increasing when there is loss aversion and $\mu \geq 0$. Similar results are derived considering reverse loss avension and using RS-shaped utility functions as well. Finally, I study hedging policies of a firm that uses a utility function as postulated by prospect theory. I examine the behavior of a firm whose utility function varies with gains and losses in firm's profits. Even though, the analysis with prospect theory is more complex than assuming risk aversion, I demonstrate that similar behavior hold for symmetric random variables.

The chapter can be extended in several directions. For instance, it would be interesting to generalize Theorems (2.4.1) and (2.4.2) for skewed distributions and especially for skewed-normal distribution (Azzalini, 1985). This remains a task for future studies.

Chapter 3

Three-level recognition heuristic

3.1 Introduction

In many real life situations, individuals compare (two) objects and then choose one of them. How do individuals decide? Goldstein and Gigerenzer (1999, 2002) propose a method called the *recognition heuristic.* This theory explains some experimental results. Specifically, in one experiment American and German students were asked to rank, in pairs, German and American cities according to their population (Gigerenzer et al., 1999; Hoffrage, 1995, 2011). Surprisingly, the German students accuracy rate was higher for American cities than for German ones. This unexpected result motivated an answer to the following question: How could people have more correct answers on those issues or topics that they a priori knew less? With this in mind, Goldstein and Gigerenzer (1999, 2002) suggested the mentioned method, whose idea is based on the dictum:

> If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion. (Goldstein and Gigerenzer, 2002, p.76).

Furthermore, the recognition heuristic suggests that if neither of the two objects are recognized, then the subject must decide randomly, with equal probabilities; and if both objects are recognized then the subject should decide with the help of some additional information.

The chapter aims are fourfold. First, I extend the model by Goldstein and Gigerenzer considering three levels of recognition judgement. Goldstein and Gigerenzer's model implicitly assumes there is a direct association between higher value and recognition. Indeed, the recognition heuristic is a non-compensatory heuristic, since subjects do not use further information to decide between a recognized and an unrecognized object. In my proposal, however, the recognized objects are classified into two categories: recognizable satisfying and recognizable unsatisfying. Namely, a person might recognize an object, but the object is categorized as unsatisfying with respect to the criterion of interest and within the sample of objects. Moreover, I shall assume that unrecognized objects are preferred to recognized but unsatisfying ones. The following example shall clarify the logic behind this proposal.

> Suppose Sherlock Holmes makes pairwise comparisons of the 'nicest' person among different individuals one of whom is Professor Moriarty. I believe that Sherlock Holmes would view Professor Moriarty as a 'recognizable unsatisfying person', and thus even an unrecognizable person would be considered nicer than Professor Moriarty.

Moreover, my proposal is a generalization of the two levels recognition model. Indeed, when the person does not identify any recognizable unsatisfying objects and all the recognized objects are satisfying, we are within the original model of Goldstein and Gigerenzer (2002).

Second, I provide mathematical formulas of all the parameters involved in the model. Contrary to previous works that simply estimate the accuracy rates of the recognition heuristic (see, for instance, Goldstein & Gigerenzer, 2011, and references therein), these explicit formulas allow me to calculate exactly the accuracy rates for the two and three levels recognition heuristic. Third, I characterize the conditions under which the predictive power of the recognition heuristic is equal to choosing the objects randomly. Finally, I address the question whether less information associates with higher probability of success. In this sense, I assess whether the less is more effect (LIME) still holds within the three levels recognition heuristic.

I have organized the chapter as follows. Section 3.2, gives a brief introduction to heuristics. In Section 3.3, I put forward the three-levels model and express the probability of correct guesses. Section 3.4, gives the explicit formulas of all the parameters in the probability of correct guesses. Finally, Section 3.5, provides the calculations of the accuracy rates of both the two-levels and three-levels recognition heuristic considering a sample of three and ten objects. In addition, I study the less is more effect elucidating its implications. Concluding remarks finish the chapter.

3.2 A brief introduction to heuristics

First, we need to distinguish between various definitions of heuristic. In mathematics, Pólya (1954) uses the term heuristic as a method to solve some problems. In this science, it is also referred as a computationally fast method to get a good feasible solution to a problem (Hillier, Lieberman & Hillier, 1990). Other interpretations come from psychology and are referred as cognitive heuristics. Here, we can distinguish two different views.

On one hand, Kahneman, Slovic and Tversky (1982) define heuristics as a psychological process that might be useful, but tends to deviate from rationality, which involves what are known as biases. Thus, their use leads to significant deviations from the (mathematically) optimal (Gilboa, 2010; Gilovich &, 2002; Kahneman, Slovic & Tversky, 1982; Kahneman & Tversky, 1979; Tversky & Kahneman, 1974). Indeed, heuristics has a negative meaning, with errors in judgment and biased behavior that should be avoided (or that should be taken into account).

On the other hand, many scholars argue that individuals are not completely rational. This means that individuals do not have complete and stable preferences, and have sufficient skills that enable them to achieve the highest attainable point on their preference scale (Simon, 1955 p. 99). Moreover, rationality is limited by the information gathering process, the cognitive limitations of the mind and the available amount of time to decide, which is usually known as bounded rationality (Gigerenzer & Selten, 2001; Conlisk, 1996; Todd & Gigerenzer, 2003). For this, Gigerenzer *et.al.* (1999) define heuristics as conscious or unconscious fast and frugal strategies that search for minimal information and consist of building blocks that exploit evolved capacities and environmental structures. Furthermore, this stream of research questioned the emphasis on biases. They advocate that heuristics are faster, more frugal and more accurate methods than standard benchmark strategies (Gigerenzer & Todd, 2008). The following definition shall clarify its meaning.

Definition 3.2.1 *(Gigerenzer & Gaissmaier, 2011) Heuristic is a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods. In summary, to this view, heuristics are shortcuts that simplify the complex methods of calculating the probabilities and utilities that are required to make decisions under uncertainty, and this simple rules lead to better decisions than more complex models.*

However, as pointed out by Gilovich and Griffin (2002), the controversy between proponents and skeptics in the use of heuristics, arises as these two approaches answer different questions. For instance, Kahneman, Slovic and Tversky (1982) are interested in answering if decision makers use heuristics in their decision process. Meanwhile, Gigerenzer et al. (1999) are interested in finding if they performed better than other decision strategies.

Nevertheless, it is well-known that laypeople and practitioners often resist to use complex mathematical models such as the ones proposed by economics or finance, and instead use heuristics. Some of these heuristics appear in economic theory. For instance, Graham (1949) recommends simple investing rules to obtain abnormal returns (see also, Oppenheimer, 1984; Oppenheimer & Schlarbaum, 1981). Benartzi and Thaler (2001) show that investors do not use sophisticated models to choose their portfolio, and usually allocate their wealth with a naive strategy, which consist in investing equal shares of their wealth in each asset. Furthermore, Friedman's rule (Friedman, 1969) and Taylor's rule (Taylor, 1993) are simple interest rate strategies examples of heuristics in monetary policy.

There are many reasons why individuals use heuristics. First, decision makers may be unable to obtain all the information necessary to solve, consciously or unconsciously, a given problem. Second, even obtaining such information, they may be unaware of the optimal method to solve it. Third, often delay is not an option and decisions need to be made fast. For other reasons of the convenience of using heuristics I refer to Payne, Bettman & Johnson, 1993; Schwartz, 2010; and Thaler & Sunstein, 2008.

For a long time, it was believed that simple heuristics performed worse than more complex models. It is, however, necessary to compare whether this assumption empirically holds. Recently, these two methods have been compared in a number of problems such as forecasting the commercial success of patents (Åstebro and Elhedhli, 2006), diversifying financial portfolios (DeMiguel, Garlappi and Uppal, 2009; DeMiguel, Garlappi, Nogales, & Uppal, 2009; Huberman & Jiang, 2006; Monti, Boero, Berg, Gigerenzer, & Martignon, 2012), predicting the future purchasing behavior of past customers (Wuebben and von Wangenheim, 2008), prescribing antibiotics to children (Fischer et al., 2002), geographically profiling criminals (Bennell, Emeno, Snook, Taylor & Goodwill, 2010; Snook, Zito, Bennell & Taylor, 2005); predicting political elections (Gaissmaier & Marewski, 2011); predicting the stock and exchange market (Zaleskiewicz, 2011) and so forth.

In summary, these studies conclude that: (i) heuristics have higher predictive accuracy than optimization models when information is scarce; (ii) the opposite appears to be true when information is not scarce and (iii) each one of heuristics and more complex models can outperform the other (for a survey of these comparisons, the interested reader is referred to Katsikopolous, 2011).

As I have noted earlier, in this chapter I shall extend the two-levels recognition heuristic (Goldstein & Gigerenzer, 2002). This heuristic was proposed to explain some intriguing experiments results. Specifically, these experiments set up is as follows.

Let us posit a test in which pairs of objects are drawn randomly from the class of N objects, with n among them recognizable and $N - n$ unrecognizable by the test taker. The individual must pairwise compare and choose the object with higher value according to some criterion of interest. The objects of each pair can be: both recognized, both unrecognized, or one is recognized and the other one is not. The test score is the proportion of pairs in which the test taker has correctly identified the larger object. The recognition heuristic suggests: (i) if one object is recognized and the other is not, choose the recognized object; (ii) if neither of the two objects is recognized, then choose one of them randomly, with equal probabilities; and (iii) if both of them are recognizable, then employ a cue to decide which one to choose.

The concept of recognition is a crucial element in this heuristic, which has generated a considerable debate (e.g., Davis-Stober, Dana, & Budescu, 2010; Dougherty, Franco-Watkins, & Thomas, 2008; Marewski, Pohl, & Vitouch, 2010, 2011a, 2011b; Pohl, 2011; Tomlinson, Marewski, & Dougherty, 2011). There is, however, a certain consensus that its meaning refers to the ability of individuals to discriminate between known objects from novel ones (Pachur, Broder and Marewski, 2008). The set of objects splits in two subsets: one with recognizable objects and another one with unrecognizable objects. This framework has been criticized by Bröder & Eichler 2006; Dougherty et al, 2008; Hilbig and Pohl 2008, 2009; Newell & Fernandez, 2006; Newell & Shanks, 2004; Oppenheimer, 2003; Pachur, Broder & Marewski, 2008; Pohl, 2006; Richter & Späth, 2006; among others. Consequently, some authors have proposed distinguishing between recognizable objects (e.g., Hilbig & Pohl, 2008, 2009; Oppenheimer, 2003).

We also need to clarify the meaning of the cue. According to the recognition heuristic, a cue consists in additional information that could help the individuals to choose between recognized object. For instance, in experiments involving ranking cities according to their population, whether the the city has: an international airport, significant industries, a team in the major national soccer league, were examples of possible additional information.

The recognition heuristic has been applied for different purposes, such as comparing cities with respect to their populations (Hoffrage, 1995), choosing stocks (Andersson & Rakow, 2007; Borges, Goldstein, Ortmann & Gigerenzer, 1999; Boyd, 2001; Newell & Shanks, 2004; Ortmann, Gigerenzer, Borges & Goldstein, 2008), sports results (Andersson, Edman & Ekman., 2005; Scheibehenne & Bröder, 2007; Snook & Cullen, 2006) and choosing consumer goods (Hauser, 2011; Herzog & Hertwig, 2011; Hoffrage, 1995; Oeusoonthornwattana & Shanks, 2010; Pachur and Biele, 2007; Thoma & Williams, 2013). For instance, Borges et al. (1999) and Ortmann, et al. (2008) find evidence that constructing portfolios, in a bull market, based solely on the names of the recognized companies yields better returns than the market index. They conducted laboratory experiments where participants construct their portfolios with the most frequently recognized shares. In most of the cases, the selected portfolios outperformed the market index. These results were surprising as they are opposed to the efficient market hypothesis (Fama 1970). That is, simple investment strategies cannot consistently beat the market index. A reason of this stunning result is that recognized companies may yield higher average returns than unrecognized ones.

Boyd (2001) replicates the Borges et al. (1999) test, but, now, in a bear market, reaching different conclusions. He finds that recognition heuristic as a strategy for selecting stocks does not outperform the market as the referred work showed. A possible explanation of these opposite conclusions can be deduced from the model by Merton (1987). In this model it is assumed that investors construct their optimal portfolios only with known securities. Which implies that recognized firms will have higher demand and value. Yet, this model predicts a negative correlation between stock returns and recognition. This implies that recognized companies will yield lower returns than average, which gives a possible explanation of the results found by Boyd (2001).

Nevertheless, however, whether the recognition heuristic is a descriptive behavior in these experiments is still in debate (cf. e.g. Pachur et.al., 2008).

Finally, the recognition heuristic challenged the idea that accuracy involves effort. As experiments have shown, there are situations where a high level of accuracy is obtained with less information (recognition) (Goldstein & Gigerenzer, 1999, 2002). Indeed, more information instead of increasing the accuracy rate can decrease it. Contrary, less information might lead to higher accuracy rates.

3.3 Three-levels recognition heuristic

In this section, I put forward a three-levels recognition heuristic model. In my approach, the recognized objects are classified into two categories: recognizable satisfying and recognizable unsatisfying. My approach serves to explain some empirical evidence. For instance, in a series of experiments comparing cities populations, Oppenheimer (2003) reports that participants sometimes prefer unrecognized objects over recognized ones. Specifically, contrary to the two-level heuristic proposal, the experiments showed that participants tend to choose an unrecognized city than a recognized small one. Therefore, my proposal assumes that individuals will prefer unrecognizable objects over recognizable unsatisfying ones. I also suppose that individuals choose randomly (with equal probability) between recognizable and unsatisfying objects.

For clarity, and to keep the analysis as simple as possible, I restrict the study to the recognition heuristic with perfect memory (e.g., Smitshon, 2010) and thus, I do not consider the imperfect memory version (e.g., Katsikopoulos, 2010; Erdfelder, Küpper-Tetzel & Mattern, 2011). I also follow Smithson (2010), considering the use of a single cue with ranks and no ties.

Given two objects, we set the following recognition heuristic rules for the three-levels model:

- If one object is recognizable satisfying and the other is recognizable unsatisfying, then choose the former one.
- If both objects are recognizable satisfying, then decide according to a cue.
- If one object is recognizable satisfying and the other one is unrecognizable, then choose the recognizable one.
- If one object is recognizable unsatisfying and the other one is unrecognizable, then choose the unrecognizable one.
- If both objects are unrecognizable, then choose randomly with equal probabilities.
- If both objects are recognizable unsatisfying, then choose randomly with equal probabilities.[1]

[1] To make my theory as simple as possible, which is one of the goals of the recognition heurisitc, I would not assume the use of any cue to choose any of the two recognized unsatisfying objects.

3.3.1 Model

Now, I present the formal model of the three levels recognition heuristic model. Suppose that we are dealing with N objects, represented them as an N-dimensional vector, $\mathbf{x} = (x_1, x_2, ..., x_N)$, called the recognition vector. The position of each coordinate of $\mathbf{x}$, and thus of the underlying object to be ranked, is based on the *criterion ranking*, denoted by $\mathbf{c} = (c_1, c_2, ..., c_N) \equiv (1, 2, \ldots, N)$, which is an arrangement of the underlying objects in the decreasing order with respect to their 'size' or 'value.' For instance, according to the criterion ranking, the i^{th} object is larger in value than the j^{th} object whenever $i < j$. Each coordinate x_i of the vector can be equal to 1 if the i-th object is recognizable satisfying, 0 if the object is unrecognizable and -1 if it is recognizable unsatisfying. Hereafter, I shall use the following notation:

- N_1 is the number of recognizable satisfying objects.
- N_0 is the number of unrecognizable objects.
- N_{-1} is the number of recognizable unsatisfying objects.
- n is the number of recognizable objects, either satisfying or unsatisfying, that is, $n = N_1 + N_{-1}$.

For a recognition vector $\mathbf{x}$ we would have $\sum_{i=1}^{N} x_i^+ = N_1$, $\sum_{i=1}^{N} x_i^- = N_{-1}$ and $\sum_{i=1}^{N}(1 - |x_i|) = N_0$, where $x_i^+ = \max\{x_i, 0\}$, $x_i^- = -\min\{x_i, 0\}$ and $|x_i|$ denotes the absolute value of x_i.

The vector $\mathbf{q} = (q_1, q_2, ..., q_N)$ represents the *cue ranking* (hereafter, also the cue vector). This cue vector is used only when both objects are recognizable satisfying. Similar to Egozcue, Fuentes García, Katsikopolous and Smithson (2013) and Smithson (2010) models, the cue vector indicates the ranking of the underlying objects, which may or may not coincide with the above noted criterion ranking.

Notice that subjects must compare $\binom{N}{2}$ possible pairs combinations. That is, the comparisons are between the following pairs

$$(x_1, x_2), (x_1, x_3), ..., (x_1, x_N), ..., (x_{N-1}, x_N).$$

Correct answers are those when the subject chooses the left object of each pair. The following example will help to clarify the previous notation.

Example 3.3.1 *Consider the vector* $\mathbf{x} = (1, 0, 1, -1, 0)$*, which means that there are* $N = 5$ *objects, and the (default) criterion ranking is* $\mathbf{c} = (1, 2, 3, 4, 5)$*. Hence, objects* x_1, x_3 *and* x_4*, have been recognized, and* x_2 *and* x_5 *have not been recognized. However, object* x_4 *has an unsatisfying recognition. In addition, suppose the cue ranking is* $\mathbf{q} = (3, 1, 2, 5, 4)$*, which gives information about the ranking objects. This cue ranking is used only when the two objects have been satisfying recognized. Thus, we would only compare the first and the third element of vector* $\mathbf{q}$*. That is, when we compare objects* x_1 *and* x_3*, the individual would follow the cue ranking and "erroneously" will choose* x_3 *as the highest value of the pair, as the cue vector indicates so.*

3.3.2 The probability of success for three levels of recognition

I am interested in finding the probability of correct guesses. Let A be the event of correct guessing, and so the expected proportion of correct inferences is the probability $\mathbf{P}(A)$. In other words, $\mathbf{P}(A)$ is the proportion of correct answers in all of the pairwise comparisons. To calculate this probability, we first introduce the following mutually exclusive and exhaustive sets:

- E_{00} consists of all the pairs of different objects which are unrecognizable. The proportion of such pairs is

$$\mathbf{P}(E_{00}) = \binom{N_0}{2} \Big/ \binom{N}{2} = (N-n)\,(N-n-1)\,/N(N-1).$$

- E_{01} consists of all the pairs of different objects one of which is unrecognizable and the other one is recognizable satisfying. The proportion of such pairs is

$$\mathbf{P}(E_{01}) = \binom{N_0}{1}\binom{N_1}{1} \Big/ \binom{N}{2} = 2N_1(N-n)/N(N-1).$$

- E_{11} consists of all the pairs of different objects which are recognizable satisfying. The proportion of such pairs is

$$\mathbf{P}(E_{11}) = \binom{N_1}{2} \Big/ \binom{N}{2} = N_1 \left(N_1 - 1\right) / N(N-1).$$

- E_{-10} consists of all the pairs of different objects such as one of them is unrecognizable and the other one is recognizable unsatisfying. The proportion of such pairs is

$$\mathbf{P}(E_{-10}) = \binom{N_0}{1} \binom{N_{-1}}{1} \Big/ \binom{N}{2} = 2N_{-1}(N-n)/N(N-1).$$

- E_{-11} consists of all the pairs of different objects one of which is recognizable unsatisfying and the other one is recognizable satisfying. The proportion of such pairs is

$$\mathbf{P}(E_{-11}) = \binom{N_1}{1} \binom{N_{-1}}{1} \Big/ \binom{N}{2} = 2N_{-1}N_1/N(N-1).$$

- E_{-1-1} consists of all the pairs of different objects which are recognizable unsatisfying. The proportion of such pairs is

$$\mathbf{P}(E_{-1-1}) = \binom{N_{-1}}{2} \Big/ \binom{N}{2} = N_{-1} \left(N_{-1} - 1\right) / N(N-1).$$

Now, using the rule of total probability, we have that

$$\mathbf{P}(A) = \mathbf{P}(A \cap E_{00}) + \mathbf{P}(A \cap E_{01}) + \mathbf{P}(A \cap E_{11}) + \mathbf{P}(A \cap E_{-11}) + \mathbf{P}(A \cap E_{-10}) + \mathbf{P}(A \cap E_{-1-1}). \tag{3.1}$$

This reduces our main goal, which is calculating $\mathbf{P}(A)$, to calculating the six 'marginal' probabilities $\mathbf{P}(A \cap E_{ij})$ on the right-hand side of equation (3.1). Of course, the probability $\mathbf{P}(A \cap E_{ij})$ is equal to 0 when $E_{ij} = \emptyset$, the empty set. When, however, $E_{ij} \neq \emptyset$, then the probability can be expressed in terms of conditional probabilities by the formula

$$\mathbf{P}(A \cap E_{ij}) = \mathbf{P}(A|E_{ij})\mathbf{P}(E_{ij}). \tag{3.2}$$

Given the above formulas for the probabilities $\mathbf{P}(E_{ij})$, our task reduces to calculating the conditional probabilities $\mathbf{P}(A|E_{ij})$, which I denote them as follows:

- When $E_{00} \neq \emptyset$, then $\beta_0 := \mathbf{P}(A|E_{00})$, called the knowledge validity for unrecognizable objects. Throughout this chapter, I set $\beta_0 = 1/2$, because when facing two unrecognizable objects, we choose one of them by flipping a fair coin.

- When $E_{01} \neq \emptyset$, then $\alpha_S := \mathbf{P}(A|E_{01})$, called the recognition validity, which is the probability of scoring a correct answer when one object is satisfying recognized and the other one is not.

- When $E_{11} \neq \emptyset$, then $\beta_S := \mathbf{P}(A|E_{11})$, called the knowledge validity, which is the probability of scoring a correct answer when both objects are recognized via an additional cue (knowledge cue).

- When $E_{-11} \neq \emptyset$, then $\gamma_R := \mathbf{P}(A|E_{-11})$, called the satisfying-unsatisfying recognition validity for recognizable objects.

- When $E_{-10} \neq \emptyset$, then $\alpha_U := \mathbf{P}(A|E_{-10})$, called the unsatisfying recognition validity for recognizable unsatisfying objects.

- When $E_{-1-1} \neq \emptyset$, then $\beta_U := \mathbf{P}(A|E_{-1-1})$, called the knowledge validity for unsatisfying objects. Throughout this chapter, I set $\beta_U = 1/2$, because when facing two recognizable unsatisfying objects, we choose one of them by flipping a fair coin.

- When $E_{ij} = \emptyset$ for $i, j = -1, 0, 1$ the above parameters are undefined.

Note that under the above specified assumptions of recognition heuristics, β_S depends on $\mathbf{x}$ and $\mathbf{q}$, while α_S, α_U, and γ_R depend only on $\mathbf{x}$. To indicate these dependencies on $\mathbf{x}$ and/or $\mathbf{q}$, from now on I shall write $\alpha_S(\mathbf{x})$, $\beta_S(\mathbf{x}, \mathbf{q})$, $\alpha_U(\mathbf{x})$, and $\gamma_R(\mathbf{x})$. In view of the above, and using the notation $g(\mathbf{x}, \mathbf{q}) := \mathbf{P}(A)$ to highlight our interest in the dependence of the success probability $\mathbf{P}(A)$ on the recognition vector $\mathbf{x}$ and the cue $\mathbf{q}$, equation (3.1) reduces to

$$\begin{aligned} g(\mathbf{x},\mathbf{q}) = \beta_0 \frac{(N-n)\,(N-n-1)}{N(N-1)} + 2\alpha_S(\mathbf{x})\frac{N_1(N-n)}{N(N-1)} + \beta_S \frac{N_1\,(N_1-1)}{N(N-1)} \\ + 2\alpha_U(\mathbf{x})\frac{N_{-1}(N-n)}{N(N-1)} + 2\gamma_R(\mathbf{x})\frac{N_1 N_{-1}}{N(N-1)} + \beta_U \frac{N_{-1}\,(N_{-1}-1)}{N(N-1)}, \end{aligned} \tag{3.3}$$

with the values $\beta_0 = 1/2$ and $\beta_U = 1/2$ as noted earlier.

As I have already noted, when the set E_{ij} is empty, then by definition, the conditional probability $\mathbf{P}(A|E_{ij})$ is undefined. Consequently, some of the parameters in equation (3.3) might be undefined. Nevertheless, the right-hand side of equation (3.3) is always well defined, because if any of the parameters are undefined, then the corresponding term in the equation vanishes. Indeed, this follows from equations (3.1) and (3.2), with the latter implying in particular that if $\mathbf{P}(A|E_{ij})$ is undefined, which implies that $\mathbf{P}(E_{ij})$ is equal zero, and thus the probability $\mathbf{P}(A \cap E_{ij})$ must be zero.

Indeed, when the there are not distinction between recognizable satisfying or unsatisfying objects, then $N_{-1} = 0$ and thus $n = N_1$, so the last three terms on the right-hand side of equation (3.3) vanish. Thus, the three level recognition heuristic collapses to the two levels recognition heuristic. Hence, we obtain the equation of Goldstein and Gigerenzer (2002, p.78) stating that the success probability, which in this two-levels case I denote by $f(\mathbf{x},\mathbf{q})$, is equal to

$$f(\mathbf{x},\mathbf{q}) = \beta_0 \frac{(N-n)(N-n-1)}{N(N-1)} + 2\alpha \frac{n(N-n)}{N(N-1)} + \beta \frac{n(n-1)}{N(N-1)}, \tag{3.4}$$

where $\beta_0 = 1/2$, $\alpha = \alpha_S(\mathbf{x})$ and $\beta = \beta_S(\mathbf{x},\mathbf{q})$.

3.4 Explicit formulas for the parameters

In this section, I extend Egozcue et al. (2013) and establish closed form solutions of the parameters in equation (3.3). These findings would allow me to calculate exactly the accuracy rate for all possible recognition and cue vectors. In addition, these calculations would be helpful in understanding a number of effects related to the recognition heuristic

(Gigerenzer & Goldstein, 2011), as well to clarify some arguments that have arisen with respect to, e.g., the observed inference accuracy and the number of recognized objects (e.g., Snook & Cullen, 2006; Pohl, 2006;, Pachur & Biele, 2007; Pleskac, 2007; Hertwig et al, 2008; Pachur et al, 2009; Katsikopoulos, 2010).

Since two of these parameters β_0 and β_U are assumed to be 1/2, we need to find the four remaining parameters, which are $\alpha_S, \beta_S, \gamma_R$ and α_U.[2] We shall see from the derived formulas that, as noted by Katsikopoulos (2010), Smithson (2010), none of these parameters remain constant when n varies, and none of them is a simple function of n.

Now, I proceed to derive the explicit formulas of the parameters in equation (3.3).

3.4.1 The recognition validity $\alpha_S(\mathbf{x})$

In the next theorem I derive the explicit formula for the recognition validity.

Theorem 3.4.1 *For N objects represented by their 'recognition vector' $\mathbf{x}$, we have that*

$$\alpha_S(\mathbf{x}) = \frac{1}{N_1 N_0} \sum_{i=1}^{N} \left(x_i^+ \sum_{j=i+1}^{N} (1 - |x_j|) \right). \tag{4.1}$$

Proof. We need to calculate the proportion of correctly guessed pairs among those with one recognized-satisfying and one unrecognized objects. Since there are N_1 recognized and satisfying objects and N_0 unrecognized objects, using the multiplication rule of counting, we obtain $N_1 N_0$ pairs with one recognized-satisfying and one unrecognized objects. This gives the denominator on the right-hand side of the first equation of (4.1). The numerator must be equal to the number of correctly guessed pairs. To confirm the assertion, we recall that we are comparing pairs where one object which is recognized and satisfying, i.e. $x_i = 1$ (I note that this element will also coincide with the same element

[2] I note the reasons behind the notations:

- S stands for "satisfying"
- R stands for "recognition"
- U stands for "unsatisfying"

of $\mathbf{x}^+$, say $x_i^+ = 1$), with another one which is not recognized, i.e., $x_j = 0$. According to the heuristic, we shall guess correctly only the pairs of the form $(1, 0)$; while the pairs of the form $(0, 1)$ will be guessed incorrectly. We start out our counting in vector $\mathbf{x}$ of the pairs $(1, 0)$ with the coordinate x_1: if it is equal to 0 or -1, we discard the case and continue discarding until we reach the first recognized object, that is, the left-most 1 among the coordinates of $\mathbf{x}$; let $x_i = x_i^+ = 1$ be this object. There are $\sum_{j=i+1}^{N}(1 - |x_j|)$ zeros (i.e., unrecognized objects) to the right of x_i. Hence, so far, we have correctly guessed $x_i^+ \sum_{j=i+1}^{N}(1 - |x_j|)$ pairs. To pick up the remaining correctly-guessed pairs, we proceed with the next 1 and count all the 0's to the right of this 1, and we proceed in the same fashion until no 1 remains. In this way, we have arrived at $\sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N}(1 - |x_j|)$ correctly guessed pairs, which are of the form $(1, 0)$. This establishes the equation

$$\alpha_S(\mathbf{x}) = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i^+(1 - |x_j|)}{N_1 N_0}. \tag{4.2}$$

■

Remark 3.4.1 *In the classical case with two levels of recognition, Goldstein and Gigerenzer (2002) define the recognition validity $\alpha = \alpha_S(x)$ by the equation*

$$\alpha_S(\mathbf{x}) = \frac{R_1}{R_1 + W_1}, \tag{4.3}$$

where R_1 is the number of correct inferences using the recognition heuristic and computed across all pairs in which one object is recognized and another one is not, and W_1 is the number of incorrect inferences under the same circumstances. I can now extend this formula to the three-outcome case. Namely, equation (4.1) implies that $R_1 = \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N}(1 - |x_j|)$ and $W_1 = \sum_{i=1}^{N} x_i^+ \sum_{j=1}^{i-1}(1 - |x_j|)$. Since $\sum_{i=1}^{N} x_i^+ = N_1$ and $\sum_{j=1}^{N}(1 - |x_j|) = N_0$, we have that $W_1 = N_1 N_0 - R_1$.

The following corollary to Theorem 3.4.1 will allow us (details in the next subsection) to easily connect formulas in the three-levels case to those already available in the literature in the two-levels case.

Corollary 3.4.1 *The recognition validity $\alpha_S(\mathbf{x})$ can be expressed by the formula*

$$\alpha_S(\mathbf{x}) = \frac{1}{2} + \frac{1}{N_1 N_0}\left[\frac{N_1(N_1+N_0+1)}{2} + \sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+ x_j^- - \sum_{i=1}^{N} i x_i\right]. \tag{4.4}$$

Proof. Starting with the earlier noted formula $R_1 = \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N}(1-|x_j|)$, we have the following equalities:

$$\begin{aligned}
R_1 &= \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N}(1-|x_j|) \\
&= \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N} 1 - \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N} |x_j| \\
&= \sum_{i=1}^{N} x_i^+ (N-i) - \sum_{i=1}^{N} x_i^+ \left(\sum_{j=i+1}^{N} x_i^+ + \sum_{j=i+1}^{N} x_i^-\right) \\
&= \sum_{i=1}^{N} x_i^+ (N-i) - \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N} x_j^+ - \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N} x_j^- \qquad (4.5)
\end{aligned}$$

Since, $\sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N} x_j^+$ is the combination of pairs with recognized-satisfying objects then is equal to $\binom{N_1}{2} = \frac{N_1(N_1-1)}{2}$, and

$$\begin{aligned}
\sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N} x_j^- &= \sum_{i=1}^{N} x_i^+ \left(N_{-1} - \sum_{j=1}^{i} x_j^-\right) \\
&= N_1 N_{-1} - \sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+ x_j^-. \qquad (4.6)
\end{aligned}$$

Hence, equation (4.5) can be rewritten as follows

$$R_1 = N\sum_{i=1}^{N} x_i^+ - \sum_{i=1}^{N} i x_i^+ - \left[\frac{N_1(N_1-1)}{2} + N_1 N_{-1} - \sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+ x_j^-\right].$$

Consequently,

$$
\begin{aligned}
R_1 &= \left[NN_1 - \frac{N_1(N_1-1)}{2} - N_1N_{-1}\right] - \sum_{i=1}^{N} ix_i^+ + \sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^- \\
&= \left[N_1(N_1+N_0+N_{-1}) - \frac{N_1(N_1-1)}{2} - N_1N_{-1}\right] - \sum_{i=1}^{N} ix_i^+ + \sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^- \\
&= N_1N_0 + \frac{N_1(N_1+1)}{2} - \sum_{i=1}^{N} ix_i^+ + \sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^-. \qquad (4.7)
\end{aligned}
$$

Substituting (4.7) into (4.3) we get

$$
\begin{aligned}
\alpha_S(\mathbf{x}) &= \frac{1}{N_1N_0}\left(N_1N_0 + \frac{N_1(N_1+1)}{2} - \sum_{i=1}^{N} ix_i^+ + \sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^-\right) \\
&= 1 + \frac{N_1(N_1+1)}{2N_1N_0} + \frac{1}{N_1N_0}\left(\sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^- - \sum_{i=1}^{N} ix\right) \\
&= \frac{1}{2} + \frac{1}{2} + \frac{N_1(N_1+1)}{2N_1N_0} + \frac{1}{N_1N_0}\left(\sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^- - \sum_{i=1}^{N} ix\right) \\
&= \frac{1}{2} + \frac{N_1N_0 + N_1(N_1+1)}{2N_1N_0} + \frac{1}{N_1N_0}\left(\sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^- - \sum_{i=1}^{N} ix\right) \\
&= \frac{1}{2} + \frac{N_1\left(N_1+N_0+1\right)}{2N_1N_0} + \frac{1}{N_1N_0}\left(\sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^- - \sum_{i=1}^{N} ix\right)
\end{aligned}
$$

■

When there are no recognizable unsatisfying objects

The following special case of Corollary 3.4.1 facilitates order relationships of the recognition validity $\alpha_S(\mathbf{x})$ for different recognition vectors, and it also connects our results with those of Pachur (2010) when there are no recognizable unsatisfying objects.

Corollary 3.4.2 *When there are no recognizable unsatisfying objects in* $\mathbf{x}$, *whose all coordinates are non-negative in this case, which implies that* $\sum_{i=1}^{N}\sum_{j=1}^{i} x_i^+x_j^- = 0$. *There-*

fore, equation (4.4) can be written as follows

$$\alpha_S(\mathbf{x}) = \frac{1}{2} + \frac{1}{N_1 N_0}\left[\frac{N_1(N_1+N_0+1)}{2} - \sum_{i=1}^{N} i x_i\right]. \tag{4.8}$$

Which is the equation that appears in Pachur (2010, p.598).

3.4.2 The knowledge validity $\beta_S(\mathbf{x}, \mathbf{q})$

As far as I know, there are no explicit formulas for the knowledge validity $\beta_S(\mathbf{x}, \mathbf{q})$ in the literature, apart from its relation with the Goodman and Kruskall measure (Goodman & Kruskal 1954; Smithson, 2010). I will give more details of this relationship in Section (3.5). As I have noted earlier, individuals decide with the help of the cue vector $\mathbf{q}$ when both objects are recognizable satisfying.

Theorem 3.4.2 *The knowledge validity $\beta_S(\mathbf{x}, \mathbf{q})$ can be expressed by the formula*

$$\beta_S(\mathbf{x}, \mathbf{q}) = \frac{1}{2} + \frac{1}{N_1(N_1-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ sgn(q_j - q_i), \tag{4.9}$$

where $\mathrm{sgn}(x)$ *is the sign function that takes on the values:* 1 *when* $x > 0$, -1 *when* $x < 0$, *and* 0 *when* $x = 0$.

Proof. Note that here we deal with the pairs of both recognized satisfying objects, that is, with pairs of the form $(1, 1)$. Hence, the knowledge validity $\beta_S(\mathbf{x}, \mathbf{q})$ can be written as the ratio

$$\beta_S(\mathbf{x}, \mathbf{q}) = \frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ \cdot \mathbf{1}\{q_i < q_j\}}{\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+}, \tag{4.10}$$

where the indicator $\mathbf{1}\{q_i < q_j\}$ is equal to 1 if the inequality $q_i < q_j$ holds, and is equal to 0 otherwise, that is, when $q_i > q_j$. (By assumption, there cannot be equality between the elements of the cue vector $\mathbf{q}$, i.e., $q_i \neq q_j$).

In the numerator we count those pairs $(1,1)$ that have been correctly recognized by the cue ranking, and this number is

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ \cdot \mathbf{1}\{q_i < q_j\} &= \sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ \cdot \mathbf{1}\{q_j - q_i > 0\} \\ &= \sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ \left(\frac{\operatorname{sgn}(q_j - q_i) + 1}{2}\right) \\ &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ \operatorname{sgn}(q_j - q_i) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+. \end{aligned} \tag{4.11}$$

Substituting the numerator on the right-hand side of equation (4.10) by the right-hand side of equation (4.11) and then, after a little simplification, we obtain the equation

$$\beta_S(\mathbf{x}, \mathbf{q}) = \frac{1}{2} + \left(\frac{1}{2}\right) \frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ \operatorname{sgn}(q_j - q_i)}{\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+}. \tag{4.12}$$

The denominator on the right-hand side of equation (4.10) is the total number of pairs $(1,1)$ that we have to deal with. We observe that

$$\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ = N_1(N_1 - 1)/2, \tag{4.13}$$

which follows from the fact that this is the number of unordered satisfactory recognized pairs, which is $\binom{N_1}{2}$. After replacing (4.13) in (4.12) we obtain

$$\beta_S(\mathbf{x}, \mathbf{q}) = \frac{1}{2} + \frac{1}{N_1(N_1 - 1)} \sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i^+ x_j^+ \operatorname{sgn}(q_j - q_i), \tag{4.14}$$

which is the desired equation. ■

3.4.3 The satisfying-unsatisfying recognition validity $\gamma_R(\mathbf{x})$

As I have noted earlier, the satisfying-unsatisfying recognition validity is the probability of correct guessing when both objects are recognized, but one of them is satisfying and the other one is unsatisfying.

Theorem 3.4.3 *The satisfying-unsatisfying recognition validity $\gamma_R(\mathbf{x})$ can be expressed by the formula*

$$\gamma_R(\mathbf{x}) = \frac{1}{N_1 N_{-1}} \sum_{i=1}^{N} \left(x_i^+ \sum_{j=i+1}^{N} x_j^- \right). \tag{4.15}$$

Proof. First we check that there are $N_1 N_{-1}$ pairs of recognized satisfying and unsatisfying objects. Here we deal with the pairs of the form $(1, -1)$ or $(-1, 1)$. Similar, to the proof in Theorem 3.4.1 there are $\sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i^+ x_j^-$ correct guesses of 1 and -1 pairs. This gives the following proportion of correct guesses

$$\gamma_R(\mathbf{x}) = \frac{1}{N_1 N_{-1}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i^+ x_j^-. \tag{4.16}$$

■

3.4.4 The unsatisfying recognition validity $\alpha_U(\mathbf{x})$

The unsatisfying recognition validity is the probability of correct guessing when one object is unrecognizable and the other is recognized and unsatisfying. The formula for this parameter is similar to the recognition validity and I derive it as follows.

Theorem 3.4.4 *The unsatisfying recognition validity $\alpha_U(\mathbf{x})$ can be expressed by the formula:*

$$\alpha_U(\mathbf{x}) = \frac{1}{N_{-1} N_0} \sum_{i=1}^{N} \left((1 - |x_i|) \sum_{j=i+1}^{N} x_j^- \right). \tag{4.17}$$

Proof. The proof resembles somewhat the proof of Theorem 3.4.1. We need to calculate the proportion of correct guesses of pairs, when one object is unrecognized and the other is recognized-unsatisfying, that is, pairs of the form $(0, -1)$ or $(-1, 0)$. The denominator is easily deduced as follows, since there are $\sum_{i=1}^{N}(1 - |x_i|) = N_0$ unrecognized objects and $\sum_{j=1}^{N} x_j^- = N_{-1}$, we obtain $N_{-1} N_0$ pairs of this type and using the multiplication rule. For the numerator, we need to count the pairs that are correctly guessed. Since I assume that the individual consider as more "valued" an unrecognized object over a recognized-unsatisfying object, then we need to count all the pairs of the form $(0, -1)$, which are the correctly guessed. The number of pairs of this form is equal to $\sum_{i=1}^{N}(1 - |x_i|) \sum_{j=i+1}^{N} x_j^-$. Therefore,

$$\alpha_U(\mathbf{x}) = \frac{\sum_{i=1}^{N}(1 - |x_i|) \sum_{j=i+1}^{N} x_j^-}{N_{-1}N_0}. \tag{4.18}$$

■

So far I have derived the explicit formulas of all the parameters in equation (3.3). This permits me to calculate (3.3) for any vector $\mathbf{x}$ and $\mathbf{q}$. Instead of running simulations that estimates the proportion of correct guesses, this formulas allows me calculate the exact accuracy rate of the recognition heuristic. These calculations and an assessment of the "less is more" effect is the objective of the next section.

3.5 Discussion

In this part of the chapter, I calculate (3.3) in different scenarios. I also compare the effectiveness of the three-levels recognition heuristic against Goldstein and Gigerenzer's proposal. First, I pause to present Goodman & Kruskal (1954) measure of association between vectors. This measure is used in determining the correlation between the cue vector and the criterion vector. Afterwards, I characterize the conditions under which the recognition heuristic expected probability of success is equal to 1/2. I also calculate the expected value $f(\mathbf{x}, \mathbf{q})$ and $g(\mathbf{x}, \mathbf{q})$ for different scenarios of the recognition and cue vector. Finally, I study the conditions under which the less is more effect can occur in the three levels recognition heuristic.

3.5.1 Goodman-Kruskal correlation measure

Goodman-Kruskal ($\mathbb{GK}$) measure of correlation between the criterion vector $\mathbf{c}$ and the cue vector $\mathbf{q}$ is commonly used in the recognition heuristic literature (see, for example, Gaissmaier & Marewski, 2011; Smithson, 2010). The values of this association measure ranges from -1 (perfect negative association) to 1 (perfect positive association). Next, I shall explain how this measure works. Suppose we have two vectors $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$, where a_i and b_j for $i, j = 1, 2, ..., n$, are positive real numbers. We compare each pair (a_i, a_j) with (b_i, b_j) for $i, j = 1, 2, ..., n$ and $i < j$. Then pairs (a_i, a_j) and (b_i, b_j) are said to be:

- Concordant if and only if $a_i < a_j$ and $b_i < b_j$ or $a_i > a_j$ and $b_i > b_j$.
- Discordant if and only if $a_i < a_j$ and $b_i > b_j$ or $a_i > a_j$ and $b_i < b_j$.

Note that the pairs with the same element's values are discarded, (i.e. those where $a_i = a_j$ or $b_i = b_j$). Finally, the Goodman and Kruskal measure, is calculated as follows

$$\mathbb{GK} = \frac{C - D}{C + D},$$

where C is the number of concordant pairs and D is the number of discordant pairs. The following example shall clarify its use.

Example 3.5.1 *Suppose we have the following vectors* $\mathbf{c} = (1, 2, 3, 4), \mathbf{x} = (1, 1, 0, 1)$ *and* $\mathbf{q} = (2, 1, 4, 3)$*, which I display in Table 1.*

Table 1

c	x	q
1	1	2
2	1	1
3	0	4
4	1	3

To calculate the $\mathbb{GK}$ *measure we conform two vectors* **a** *(which represents the criterion vector) and* **b** *(representing the cue vector) of three elements each one (after eliminating the third row that corresponds to the unrecognized object) yielding:* $\mathbf{a} = (1, 2, 4)$ *and* $b = (2, 1, 3)$. *Then, we have* $C_2^3 = 3$ *possible pairs comparisons. Table 2 shows the results* *of these comparisons*

Table 2

(a_i, a_j)	(b_i, b_j)	***Result***
$(1, 2)$	$(2, 1)$	*Discordant,*
$(1, 4)$	$(2, 3)$	*Concordant,*
$(2, 4)$	$(1, 3)$	*Concordant.*

Therefore, $C = 2, D = 1$ *and thus* $\mathbb{GK} = 1/3$.

Remark 3.5.1 *As I have noted earlier, the accuracy of the recognition heuristic is strongly linked with the correlation between the ranking vector* $\mathbf{c}$ *and the cue vector* $\mathbf{q}$. *Another reason why* $\mathbb{GK}$ *is a convenient association measure is its relation with the knowledge validity. As pointed out by Smithson (2010, p.232), the* $\mathbb{GK}$ *can be expressed as a function of* $\beta_S(\mathbf{x}, \mathbf{q})$ *as follows*

$$\mathbb{GK} = 2\beta_S(\mathbf{x}, \mathbf{q}) - 1. \tag{5.1}$$

Note that since the cue vector $\mathbf{q}$ *is used only when both objects are recognized, the Goodman and Kruskal gamma coefficient is estimated discarding the corresponding values of the unrecognized objects. For instance, if* $\mathbf{q} = \mathbf{c}^{-1} = (c_N, c_{N-1}, \ldots c_2, c_1)$, *then* $\mathbb{GK} = -1$ *and hence* $\beta_S(\mathbf{x}, \mathbf{q})$ *yields its minimum value, which is equal to* 0. *On the other hand, when* $\mathbf{c} = \mathbf{q}$ *then* $\mathbb{GK} = 1$, *and thus* $\beta_S(\mathbf{x}, \mathbf{q})$ *yields its maximum value, which is equal to* 1. *For intermediate levels of* $\mathbb{GK}$ *we would, of course, have* $0 < \beta_S(\mathbf{x}, \mathbf{q}) < 1$.

3.5.2 Expected probability of success

In this part, I discuss the expected accuracy rate of the recognition heuristic in the two versions. My aim is to find the expected probability of success of the recognition heuristic both in the two-levels and the three-levels recognition heuristic for different recognition vectors. This is done considering all the possible combinations of $\mathbf{x}$ and $\mathbf{q}$ for a fix N. As I have pointed out, equations $f(\mathbf{x}, \mathbf{q})$ and $g(\mathbf{x}, \mathbf{q})$ are both functions of vectors $\mathbf{x}$ and $\mathbf{q}$. Also, as I have noted earlier, for each $\mathbf{q}$, we would have different values of $f(\mathbf{x}, \mathbf{q})$ and

$g(\mathbf{x}, \mathbf{q})$.

Now, I pause to present some new notation. I shall denote with $\mathbf{X}$ a discrete random variable with support on all possible recognition vectors, and denote with $\mathbf{Q}$ a random variable with support on all possible cue vectors. For instance, if $N = 2$ then we have 9 possible recognition vectors, say $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_9$. Specifically, suppose $\mathbf{X}$ is uniform distributed then, obviously, $\mathbf{P}(\mathbf{X} = \mathbf{x}_1) = \mathbf{P}(\mathbf{X} = \mathbf{x}_2) = ... = \mathbf{P}(\mathbf{X} = \mathbf{x}_9) = 1/9$. In the same manner, we have 2! possible cue vectors, say $\mathbf{q}_1$ and $\mathbf{q}_2$. Thus, if $\mathbf{Q}$ is also uniform distributed then $\mathbf{P}(\mathbf{Q} = \mathbf{q}_1) = \mathbf{P}(\mathbf{Q} = \mathbf{q}_2) = 1/2$.

- As a benchmark, I shall assume the simple case when $\mathbf{X}$ and $\mathbf{Q}$ are independent random variables.

I shall denote with $\mathbf{E}\left[g(\mathbf{X}, \mathbf{Q})\right]$ the expected accuracy rate when $\mathbf{X}$ and $\mathbf{Q}$ are random. On the other hand, I denote with $\mathbf{E}\left[g(\mathbf{X}, \mathbf{Q}) | \mathbf{Q} = \mathbf{q}\right]$ the expected accuracy rate when $\mathbf{X}$ is random and $\mathbf{Q}$ is a degenerated random variable which takes value $\mathbf{q}$. I consider the following permutations of $\mathbf{x}$ and $\mathbf{q}$: $\mathbf{x}^{-1} = (x_N, x_{N-1}, ..., x_1)$ and $\mathbf{q}^{-1} = (q_N, q_{N-1}, ..., q_1)$. Likewise, $\mathbf{X}^{-1}$ and $\mathbf{Q}^{-1}$ are random variables of, respectively, all possible vectors $\mathbf{x}^{-1}$ and $\mathbf{q}^{-1}$, as defined before.

Note 3.5.1 *Notice that a double series $\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}$ can be written as the sum of the elements of a finite square matrix*

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & ... & a_{1N} \\ a_{21} & a_{22} & a_{23} & ... & a_{2N} \\ a_{31} & a_{32} & a_{33} & ... & a_{3N} \\ ... & ... & ... & ... & \\ a_{N1} & a_{N2} & a_{N3} & & a_{NN}. \end{pmatrix}$$

I shall use this fact to prove some different expressions of double series in the next propositions.

The following assertions have important implications for the expected accuracy rate of the recognition heuristic.

Proposition 3.5.1 *For any* $\mathbf{x}$ *and* $\mathbf{q}$ *the following holds:*

1. $\alpha_S(\mathbf{x}) + \alpha_S(\mathbf{x}^{-1}) = 1$

2. $\beta_S(\mathbf{x},\mathbf{q})+\beta_S(\mathbf{x}^{-1},\mathbf{q}^{-1}) =1$

3. $\alpha_U(\mathbf{x}) + \alpha_U(\mathbf{x}^{-1}) = 1$

4. $\gamma_R(\mathbf{x}) + \gamma_R(\mathbf{x}^{-1}) = 1$

Proof. I only prove the first two cases, the other statements can be proved in the same way.

Note that

$$\alpha_S(\mathbf{x}) = \frac{1}{N_1 N_0} \sum_{i=1}^{N} x_i^+ \sum_{j=i+1}^{N} (1 - |x_j|)$$

and

$$\alpha_S(\mathbf{x}^{-1}) = \frac{1}{N_1 N_0} \sum_{k=1}^{N-1} x_{N+1-k}^+ \sum_{l=k+1}^{N} (1 - |x_{N+1-l}|).$$

Now changing variables $i = N + 1 - k$ and $j = N + 1 - l$ we obtain

$$\alpha_S(\mathbf{x}^{-1}) = \frac{1}{N_1 N_0} \sum_{i=2}^{N} x_i^+ \sum_{j=1}^{i-1} (1 - |x_j|).$$

Using the notation suggested in Note 3.5.1 define $a_{ij} = x_i^+(1 - |x_j|)$, then

$$\begin{aligned}
\alpha_S(\mathbf{x}) + \alpha_S(\mathbf{x}^{-1}) &= \frac{1}{N_1 N_0} \left(\sum_{i=2}^{N} \sum_{j=1}^{i-1} a_{ij} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} a_{ij} \right) \\
&= \frac{1}{N_1 N_0} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} - \sum_{i=1}^{N} a_{ii} \right).
\end{aligned}$$

Joining the facts that $\sum_{i=1}^{N} a_{ii} = 0$ and $\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} = N_1 N_0$ then the conclusion follows.

Now the second case. Let $b_{ij} = x_i^+ x_j^+$. First, we write

$$\beta_S(\mathbf{x},\mathbf{q}) = 2\frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} b_{ij} \cdot \mathbf{1}\{q_i < q_j\}}{N_1(N_1 - 1)}.$$

Thus,

$$\beta_S(\mathbf{x}^{-1}, \mathbf{q}^{-1}) = 2\frac{\sum_{k=1}^{N-1}\sum_{l=k+1}^{N} x_{N+1-k}^{+} x_{N+1-l}^{+} \cdot \mathbf{1}\{q_{N+1-k} < q_{N+1-l}\}}{N_1(N_1-1)}$$

So that, changing variables $i = N+1-k$ and $j = N+1-l$, we obtain

$$\beta_S(\mathbf{x}^{-1}, \mathbf{q}^{-1}) = 2\frac{\sum_{i=2}^{N}\sum_{j=1}^{i-1} b_{ij} \cdot \mathbf{1}\{q_i < q_j\}}{N_1(N_1-1)}.$$

Since $\sum_{i=1}^{N}\sum_{j=i+1}^{N} b_{ij} = N_1(N_1-1)/2$, we have

$$\beta_S(\mathbf{x}, \mathbf{q}) + \beta_S(\mathbf{x}^{-1}, \mathbf{q}^{-1}) = \frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N} b_{ij} \cdot \mathbf{1}\{q_i < q_j\} + \sum_{i=2}^{N}\sum_{j=1}^{i-1} b_{ij} \cdot \mathbf{1}\{q_i < q_j\}}{\sum_{i=1}^{N}\sum_{j=i+1}^{N} b_{ij}}. \tag{5.2}$$

Notice that

$$\sum_{i=1}^{N}\sum_{j=i+1}^{N} b_{ij} = b_{12} + +b_{13} + \ldots + b_{1N} + b_{23} + \ldots + b_{N-1N}$$

and

$$\sum_{i=2}^{N}\sum_{j=1}^{i-1} b_{ij} = b_{21} + b_{31} + b_{32} + \ldots + b_{N1} + \ldots + b_{NN-1}.$$

Since each element of $\mathbf{q}$ are different (because I have assumed there are no ties in the cue vector), we have

$$\mathbf{1}\{q_i < q_j\} + \mathbf{1}\{q_i > q_j\} = 1 \text{ for } i \neq j. \tag{5.3}$$

Together with the fact that $b_{ij} = b_{ji}$ we obtain

$$b_{ij} \cdot \mathbf{1}\{q_i < q_j\} + b_{ji} \cdot \mathbf{1}\{q_j < q_i\} = b_{ij}. \tag{5.4}$$

Therefore, owing to the equalities (5.3) and (5.4), the numerator is equal to

$$\sum_{i=1}^{N}\sum_{j=i+1}^{N} b_{ij}\cdot \mathbf{1}\{q_i < q_j\} + \sum_{i=2}^{N}\sum_{j=1}^{i-1} b_{ij}\cdot \mathbf{1}\{q_i < q_j\}$$

$$= b_{12}\cdot \mathbf{1}\{q_1 < q_2\} + b_{21}\cdot \mathbf{1}\{q_2 < q_1\} + ... + b_{N-1N}\cdot \mathbf{1}\{q_{N-1} < q_N\}$$
$$+ b_{NN-1}\cdot \mathbf{1}\{q_N < q_{N-1}\}$$
$$= b_{12} + +b_{13} + ... + b_{1N} + b_{23} + ... + b_{N-1N}$$
$$= \sum_{i=1}^{N}\sum_{j=i+1}^{N} b_{ij}.$$

This shows that the numerator and the denominator in (5.2) are equal, thus the assertion follows. ■

With the aid of Proposition (3.5.1), I next establish a connection with the overall accuracy rate as it is expressed in equation (3.3).

Proposition 3.5.2 *Let* $\mathbf{x}$ *and* $\mathbf{q}$. *Then*

$$g(\mathbf{x}, \mathbf{q}) + g(\mathbf{x}^{-1}, \mathbf{q}^{-1}) = 1 \tag{5.5}$$

Proof. By Proposition 3.5.1 we have

$$\begin{aligned} g(\mathbf{x}^{-1}, \mathbf{q}^{-1}) \;=\; & (1-\beta_0)\,\frac{(N-n)\,(N-n-1)}{N(N-1)} + 2\,(1-\alpha_S(\mathbf{x}))\,\frac{N_1(N-n)}{N(N-1)} \\ & + (1-\beta_S(\mathbf{x},\mathbf{q}))\,\frac{N_1\,(N_1-1)}{N(N-1)} + 2\,(1-\alpha_U(\mathbf{x}))\,\frac{N_{-1}(N-n)}{N(N-1)} \\ & + 2\,(1-\gamma_R(\mathbf{x}))\,\frac{N_1 N_{-1}}{N(N-1)} + (1-\beta_U)\,\frac{N_{-1}\,(N_{-1}-1)}{N(N-1)}. \end{aligned}$$

Now, after some algebra, we arrive at

$$g(\mathbf{x}^{-1}, \mathbf{q}^{-1}) = 1 - g(\mathbf{x}, \mathbf{q}).$$

This concludes the proof of Proposition 3.5.2. ■

Finally, a striking implication for the expected accuracy rate of the recognition heuristic emerges from Proposition 3.5.2 which I show next.

Proposition 3.5.3 *Suppose* $\mathbf{X}$ *and* $\mathbf{Q}$ *are independent and uniformly distributed random variables. Then*

$$\mathbf{E}\left[g(\mathbf{X},\mathbf{Q})\right] = 1/2.$$

Proof. Since $\mathbf{X}$ is uniformly distributed and $\mathbf{X}$ and $\mathbf{X}^{-1}$ are bijective, it follows that $\mathbf{X}$ and $\mathbf{X}^{-1}$ are identically distributed. Likewise, we deduce that $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ are identically distributed as well. Together with the fact that $\mathbf{X}$ and $\mathbf{Q}$ are independent, we obtain that the distribution function of $(\mathbf{X},\mathbf{Q})$ is equal to that of $\left(\mathbf{X}^{-1},\mathbf{Q}^{-1}\right)$. Hence

$$\mathbf{E}\left[g(\mathbf{X},\mathbf{Q})\right] = \mathbf{E}\left[g(\mathbf{X}^{-1},\mathbf{Q}^{-1})\right]. \tag{5.6}$$

Therefore, taking expectations in both sides of (5.5) we obtain

$$\mathbf{E}\left[g(\mathbf{X},\mathbf{Q})\right] + \mathbf{E}\left[g(\mathbf{X}^{-1},\mathbf{Q}^{-1})\right] = 1.$$

By equality (5.6), the assertion follows. ■

This result has interesting implications for the expected probability of success of the recognition heuristic. It implies that, if recognition vectors are equally likely and each cue vector is available with the same probability, then the expected accuracy rate is equal to 1/2. Thus, under this condition the three levels and the two levels recognition heuristic are strategies that in, average, cannot improve the strategy of choosing pairs randomly.

However, it is a strong assumption that each recognition vector is equally likely in the class of all recognition vectors, and also each cue vector is equally likely in the class of all cue vectors. For instance, a more plausible assumption is to expect a positive correlation between $\mathbf{c}$ and $\mathbf{q}$. A reasonable assumption is to suppose that individuals would recognize the best and the worst objects in the criterion ranking. Indeed, in real life, we are usually aware of the best and the worst objects for different categories. Based on this assumption, I define the following subsets, namely $S = \{\mathbf{x} : x_1 = 1, x_N = 1\}$ and $T = \{\mathbf{x} : x_1 = 1, x_N = -1\}$. The elements of set S are those $\mathbf{x}$ such that the first and the last objects in the criterion ranking are recognized. In addition, the elements of set T consist in all recognition vectors where the first element of the criterion ranking is satisfactorily recognized and the last object of the criterion ranking is unsatisfactorily recognized.

Finally, I denote with $\mathbf{S}$ a discrete random variable with support on all possible $\mathbf{x}$ such that $\mathbf{x} \in S$ and with $\mathbf{T}$ a discrete random variable with support on all possible $\mathbf{x}$ such that $\mathbf{x} \in T$. Hereafter, I shall assume that $\mathbf{X}$, $\mathbf{S}$ and $\mathbf{T}$ are uniformly distributed and are independent with $\mathbf{Q}$.

Now, I calculate the expectation of (3.3) for vectors with three and ten objects. Notice that the number of possible combinations increases significantly with N. In the first case for $N = 3$, we would have twenty seven possible $\mathbf{x}$ vectors, and six possible $\mathbf{q}$ vectors. On the other hand, for $N = 10$ we would have $3^{10} = 59,049$ possible $\mathbf{x}$ vectors, and $10! = 3,628,800$ possible cue vectors $\mathbf{q}$.[3]

First, I start with recognition vectors with $N = 3$. As noted earlier, there are 6 possible cue vectors and 27 possible recognition vectors. As I have previously assumed, each one of the recognition vectors are chosen with the same probability. I shall also consider the cases when the recognition vectors belong to the sets S and T.

I summarize the calculations of the expected probability of success for different scenarios in Table 3.

Table 3: The expected probability of success for $N = 3$ (%)

	$\mathbf{q}$					
	$(1,2,3)$	$(1,3,2)$	$(2,1,3)$	$(2,3,1)$	$(3,1,2)$	$(3,2,1)$
$\mathbf{E}[g(\mathbf{X},\mathbf{Q})\vert\mathbf{Q}=\mathbf{q}]$	55.56	51.85	51.85	48.15	48.15	44.44
$\mathbf{E}[g(\mathbf{T},\mathbf{Q})\vert\mathbf{Q}=\mathbf{q}]$	94.44	94.44	83.33	94.44	83.33	83.33
$\mathbf{E}[f(\mathbf{X},\mathbf{Q})\vert\mathbf{Q}=\mathbf{q}]$	62.50	54.17	54.17	45.83	45.83	37.50
$\mathbf{E}[f(\mathbf{S},\mathbf{Q})\vert\mathbf{Q}=\mathbf{q}]$	83.33	66.67	66.67	33.33	33.33	16.67

[3] I have used Wolfram Mathematica software. I note that computer memory was the main limitation to run calculations with more than 10 objects.

When we consider all the possible recognition vectors, we see that $\mathbf{E}[g(\mathbf{X},\mathbf{Q})|\mathbf{Q}=\mathbf{q}]$ ranges from 55.56% when $\mathbf{q}=\mathbf{c}$ to 44.44% when $\mathbf{q}=\mathbf{c}^{-1}$. The expected accuracy rates improves significantly when we restrict our analysis to all $\mathbf{x}\in T$, yielding a minimum of $\mathbf{E}[g(\mathbf{T},\mathbf{Q})|\mathbf{Q}=\mathbf{q}]$ equal to 83.33%. For the two level recognition heuristic $\mathbf{E}[f(\mathbf{X},\mathbf{Q})|\mathbf{Q}=\mathbf{q}]$ ranges from 62.50% when $\mathbf{q}=\mathbf{c}$ to 37.50% when $\mathbf{q}=\mathbf{c}^{-1}$. However, for $\mathbf{x}\in S$ the values of $\mathbf{E}[f(\mathbf{S},\mathbf{Q})|\mathbf{Q}=\mathbf{q}]$ varies from 83.33% to 16.67%.

Now, I do the same calculations, but now considering recognition vectors with $N=10$. In this case, the number of cue vectors increases significantly. With this in mind, I have calculated the expected accuracy rate considering three different cue vectors:

1. When $\mathbf{q}=\mathbf{c}$, which implies $\mathbb{GK}=1$.
2. When $\mathbf{q}=\mathbf{c}^0=(10,9,3,4,1,2,5,7,6,8)$, which gives a correlation measure of $\mathbb{GK}=1/45$.
3. When $\mathbf{q}=\mathbf{c}^{-1}(c_N,c_{N-1},...c_2,c_1)$ which implies $\mathbb{GK}=-1$.

I present a summary of the results in (%) in Table 4.

Table 4

The expected probability of success, $N=10$

	$\mathbf{q}=\mathbf{c}$	$\mathbf{q}=\mathbf{c}^0$	$\mathbf{q}=\mathbf{c}^{-1}$
$\mathbf{E}[g(\mathbf{X},\mathbf{Q})\|\mathbf{Q}=\mathbf{q}]$	55.56	50.12	44.44
$\mathbf{E}[g(\mathbf{T},\mathbf{Q})\|\mathbf{Q}=\mathbf{q}]$	69.38	60.49	56.54
$\mathbf{E}[f(\mathbf{X},\mathbf{Q})\|\mathbf{Q}=\mathbf{q}]$	62.50	50.27	37.50
$\mathbf{E}[f(\mathbf{S},\mathbf{Q})\|\mathbf{Q}=\mathbf{q}]$	83.33	48.88	32.22

The results show that on average both the three level and the two-level recognition heuristic effectiveness is larger than 1/2 when there is a positive $\mathbb{GK}$ measure. In addition, in the three-levels recognition heuristic when the first object is recognized satisfying and the last object is recognized unsatisfying, then the accuracy rate in average is larger than choosing randomly even when the correlation measure is by perfect negatively correlated. This is not the case in the two-levels recognition heuristic, where this heuristic does better when there is a perfect positive correlation between $\mathbf{c}$ and $\mathbf{q}$, but does worse when the $\mathbf{q} = \mathbf{c}^{-1}$. When $\mathbf{q} = \mathbf{c}^0$, the average accuracy of the three levels recognition heuristic improves slightly the strategy of choosing randomly. However, the accuracy rate in the three levels recognition heuristic improves significantly if the first and the last object of the ranking are recognized. For instance, if $\mathbf{c}$ and $\mathbf{q}$ are perfect negatively correlated, which is the minimum expected probaibility of success, then $\mathbf{E}[g(\mathbf{T}, \mathbf{Q})|\mathbf{Q} = \mathbf{q}]$ equals 56.54%. Meanwhile, when $\mathbf{q} = \mathbf{c}$ the expected accuracy rate is equal to 69.38%. If we now consider all $\mathbf{x} \in S$, the two levels recognition heuristic probability of success are 37.50% and 62.50% respectively.

In summary:

- The three-level recognition heuritic outperforms the two-level recognition heuristic when the first and the last objects of $\mathbf{x}$ are recognized. For $N = 3$ the minimum expected accuracy rate for the three levels recognition heuristic is equal to 83.33%, whereas for $N = 10$ this minimum equals 56.54%.

- For $\mathbf{q} = \mathbf{c}$ and $\mathbf{q} = \mathbf{c}^0$, in average the two-level and three-level recognition heuristic outperforms the strategy of choosing randomly.

3.5.3 Less is more effect

In this subsection, I shall discuss the less is more effect. Katsikopoulos (2010) and Smithson (2010) study the LIME and characterize the conditions under which this effect could occur.[4]

This effect occurs when the recognition vector that maximizes $g(\mathbf{x}, \mathbf{q})$, which I denote with $\mathbf{x}^*$ has some unrecognized objects. That is, for this vector we will have $n < N$. Thus, subjects benefit from recognizing fewer objects. This is a controversial implication of the recognition heuristic. There is mixed empirical evidence on this effect. On the one hand, some have reported evidence of its occurrence (Goldstein and Gigerenzer, 2002;

[4]The less is more effect happens to occur also in probabilistic models that rely on the use of σ-algebras to model information (Dubra & Echenique, 2004).

Serwe and Frings, 2006; Scheibehenne and Broder, 2007). On the other hand, others have not identified its occurrence (e.g., Pachur and Biele, 2007). Nevertheless, most of these works tested the LIME in experiments and/or computer simulations. The explicit formulas of the parameters derived above allow me to find exact numerical calculations.

For this purpose, I find the recognition vectors $\mathbf{x}$ that maximizes $g(\mathbf{x}, \mathbf{q})$ for $N = 10$ considering the three $\mathbb{GK}$ that I have used before. Here I find the optimal $\mathbf{x}^*$ that solves

$$g(\mathbf{x}^*, \mathbf{q}) = \max_{\mathbf{x}} g(\mathbf{x}, \mathbf{q})$$

for three different values of $\mathbf{q}$. I summarize the results in Table 5.

Table 5: Less is more effect (LIME)

$\mathbf{q}$	$\mathbf{x}^*$	$\mathbf{q}$	$g(\mathbf{x}, \mathbf{q})$ (%)	N_1	N_0	N_{-1}	n
$\mathbf{c}$	$(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$	$\mathbf{c}$	100.00	10	0	0	**10**
$\mathbf{c}$	$(1, 1, 1, 1, 1, 1, 1, 1, 1, 0)$	$\mathbf{c}$	100.00	9	1	0	**9**
$\mathbf{c}$	$(1, 1, 1, 1, 1, 1, 1, 1, 1, -1)$	$\mathbf{c}$	100.00	9	0	1	**10**
$\mathbf{c}$	$(1, 1, 1, 1, 1, 1, 1, 1, 0, -1)$	$\mathbf{c}$	100.00	8	1	1	**9**
$\mathbf{c}^0$	$(1, 1, 0, 0, 0, 0, -1, -1, -1, -1)$	$\mathbf{c}^0$	84.44	2	4	4	**6**
$\mathbf{c}^{-1}$	$(1, 1, 0, 0, 0, 0, -1, -1, -1, -1)$	$\mathbf{c}^{-1}$	84.44	2	4	4	**6**

One can see in Table 5 that when $\mathbf{q} = \mathbf{c}$ there are four vectors (second column) that yields a 100% of accuracy rate, two of these vectors have 10 recognized objects, the other two have 9 recognized objects. Thus, we would expect that the LIME effect appears with a probability of 50%. Meanwhile, there is a unique recognition vector that maximizes $g(\mathbf{x}, \mathbf{q})$ for $\mathbf{q} = \mathbf{c}^0$ and for $\mathbf{q} = \mathbf{c}^{-1}$. This recognition vector has 6 recognized objects. Hence, in these cases the LIME appears with probability 1.

This results show that the LIME occurs (for these cue vectors) when $\mathbb{GK} < 1$. The intuition of this conclusion can be explained as follows. Since the cue vector is not perfectly correlated with the criterion vector, then the more objects we recognize the more $(1, 1)$ comparison mistakes we make. Thus, when cues are not reliable, recognizing more objects does more harm than good. In summary, the LIME appears to have greater effect when the correlation between the recognition vector and the cue vector decreases.

3.6 Concluding remarks

In this chapter, I extend Goldstein and Gigerenzer (2002) recognition heuristic model. In my proposal, instead of considering two levels of recognition, I posit three levels of recognition. To this purpose, I introduce an additional level of recognition that categorize those recognized but unsatisfying objects. This new class permit individuals to prefer an unrecognizable object over a recognizable but unsatisfying one. Also, my proposal includes as a special case the two-level recognition heuristic model.

I also derive the explicit formulas of all the parameters involved in the probability of success. These formulas allow me, instead of doing estimations of this probability, to find exact calculations of (3.3). In addition, I show that when the recognition vectors and the cue vectors are equally likely then the recognition heuristic accuracy rate is equal to 1/2.

I characterize the conditions under which the three-level recognition heuristic outperforms the two-level recognition heuristic. The three-level recognition heuristic outperforms Goldstein and Gigerenzer model when recognition includes the first and the last objects of the ranking. Finally, calculations show that less is more effect is likely to appear when the cue vector is negatively correlated with the criterion vector.

This chapter can be extended in several directions. First, one can relax the assumption of perfect memory and consider a framework of imperfect memory. Second, it could be interesting to study the heuristic expected accuracy rate for random vectors without the independence and uniform distribution assumptions. Finally, whether this heuristic is a descriptive behavior in real economic contexts remains as a task for future research.

Chapter 4

Mental accounting for multiple outcomes[1]

4.1 Introduction

Whether to keep, say, $n \in \mathbb{N}$ products segregated (e.g., unbundled) or integrate some or all of them (e.g., bundle) has been a problem of profound interest in areas such as portfolio theory in finance, risk capital allocations in insurance, and marketing of consumer products. Such decisions are inherently complex and depend on factors such as the underlying product values and consumer preferences, the latter being frequently described using value functions, also known as utility functions in economics. Quite often we want, or are required, to decide whether to combine all or only some products, objects, subjects, etc., which we call exposure units throughout the chapter – a convenient term that we borrow from the actuarial credibility theory (cf., e.g., Klugman et al., 2008).

All $n \in \mathbb{N}$ exposure units have attached to them experience values, which we simply call experiences and denote by x, y, z, x_i, and so forth. Given a value/utility function, we want to determine if all or only some exposure units should be integrated (e.g., bundled, etc.) or segregated (e.g., unbundled, etc.).

This topic is closely related to the concept of mental accounting introduced by Thaler

[1]**Jointly with Sebastien Massoni, Wing Keung Wong and Ričardas Zitikis.**

(1980, 1985). Specifically, mental accounting (Thaler, 1999) is "the set of cognitive operations used by individuals and households to organize, evaluate, and keep track of financial activities." Thaler (1980, 1985) defined a pattern of optimal behaviors depending on the type of exposure units with positive and negative experiences, concentrating on the case of two units.

The actual or perceived experiences are reflected by a value function $v : \mathbf{R} \to \mathbf{R}$, which is increasing and is also frequently assumed to be convex for non-positive experiences ($x \le 0$) and, in order to reflect the degree of risk aversion, concave for non-negative experiences ($x \ge 0$). In addition, we assume that v is continuous, which is a standard assumption. That is, unless explicitly noted otherwise, we deal with the S-shaped value function

$$v(x) = \begin{cases} v_+(x) & \text{when } \ x \ge 0, \\ -v_-(-x) & \text{when } \ x < 0, \end{cases} \tag{1.1}$$

where $v_-, v_+ : [0, \infty) \to [0, \infty)$ are continuous, increasing, and concave functions such that $v_-(0) = 0 = v_+(0)$, $v_-(x) > 0$ and $v_+(x) > 0$ for all $x > 0$. Hence, we are dealing with S-shaped functions, which are concave for gains and convex for losses. We refer to Gillen and Markowitz (2009) for a taxonomy of value/utility functions with illuminating discussions.

It has been noted (cf. al-Nowaihi *et al.*, 2008; Tversky and Kahneman, 1992) that within prospect theory, the value function v takes on the special form

$$v_{\lambda,\alpha,\beta}(x) = \begin{cases} x^\alpha & \text{when } \ x \ge 0, \\ -\lambda(-x)^\beta & \text{when } \ x < 0, \end{cases} \tag{1.2}$$

provided that the so-called preference homogeneity holds, where $\alpha, \beta \in (0, 1]$ and $\lambda > 0$ are some parameters. We refer to Wakker (2010) for a comprehensive treatment of the prospect theory.

al-Nowaihi et al. (2008) have proved that the condition of preference homogeneity is necessary and sufficient for the value function to be of the form (1.2). Furthermore, al-Nowaihi et al. (2008) have shown that under the additional and quite natural assumption of loss aversion, the parameter λ must necessarily be greater than 1, and the other two parameters $\alpha, \beta \in (0, 1]$ must be identical, i.e. $\alpha = \beta$. Thus, in this chapter we call λ the loss aversion parameter.

A natural generalization of function (1.2) under the assumption of loss aversion is therefore the following value function

$$v_\lambda(x) = \begin{cases} u(x) & \text{when} \quad x \geq 0, \\ -\lambda u(-x) & \text{when} \quad x < 0, \end{cases} \tag{1.3}$$

which features prominently in the literature (e.g., Abdellaoui et al., 2008; Broll, et al., 2010; Egozcue et al., 2011; Jarnebrant et al., 2009; Köbberling and Wakker, 2005; Wakker, 2010; and references therein), where we also find discussions concerning the loss aversion parameter λ and the base utility function $u : [0, \infty) \to [0, \infty)$. The present research also follows this line of research, and we thus mainly deal with value function (1.3). We assume that the base utility function u is continuous, increasing, concave, and such that $u(0) = 0$ and $u(x) > 0$ for all $x > 0$. The loss aversion parameter λ can be any positive real number.

Now coming back to our main discussion, we note that Thaler (1985) postulates four basic principles, known as hedonic editing hypotheses, for integration and segregation:

P1. Segregate (two) exposure units with positive experiences.

P2. Integrate (two) exposure units with negative experiences.

P3. Integrate an exposure unit carrying a smaller negative experience with that carrying a larger positive experience.

P.4 Segregate an exposure unit carrying a larger negative experience from that carrying a smaller positive experience.

Here we recall a footnote in Thaler (1985) saying that "[f]or simplicity I will deal only with two-outcome events, but the principles generalize to cases with several outcomes." When there are only two exposure units, then there can only be two possibilities: either integrate or segregate. Mathematically, if the two exposure units with experiences x and y are integrated, then their value is $v(x+y)$, but if they are kept separately (i.e., segregated), then the value is $v(x)+v(y)$. For detailed analyses of this case, we refer to Fishburn and Luce (1995), Egozcue and Wong (2010), and references therein. For example, Egozcue and Wong (2010) have found that when facing small positive experiences and large negative ones, loss averters (see, e.g., Schmidt and Zank, 2008, and references therein) sometimes prefer to segregate, sometimes to integrate, and at other times stay neutral. For a detailed analysis of the principle P4, which is known as the 'silver lining effect,' we refer to Jarnebrant et al. (2009).

In this chapter we develop decision rules for multiple products, which we generally call 'exposure units' to naturally cover manifold scenarios spanning well beyond 'products.' Our findings show, for example, that the celebrated Thaler's principles of mental accounting hold as originally postulated when the values of all exposure units are positive (i.e., all are gains) or all negative (i.e., all are losses). In the case of exposure units with mixed-sign values, decision rules are much more complex and rely on cataloging the Bell-number of cases that grow very fast depending on the number of exposure units. Consequently, in the present paper we provide detailed rules for the integration and segregation decisions in the case up to three exposure units, and partial rules for the arbitrary number of units.

We have organized the rest of the chapter as follows. In Section 4.2, we give a complete solution of the integration-segregation problem in the case of two exposure units, with experiences of any sign, whereas in Section 4.3 , we accomplish the task in the case of three exposure units. In Section 4.4, we discuss the case of the arbitrary number of exposure units by setting, naturally, more stringent assumptions than those in the previous sections. In Section 4.5 we present some applications to economics and show some illustrative examples as well. Section 4.6 finishes the chapter with the concluding remarks.

4.2 Case $n = 2$: integrate or not?

Even in the case of two exposure units (i.e., when $n = 2$), decisions whether to integrate or segregate – and there can only be these two cases – crucially depend not only on the experience values but also on the value function v. This problem has been investigated by Egozcue and Wong (2010) and Jarnebrant et al. (2009), but we shall give here a more complete picture of the matter. For illustrating examples, we refer to, e.g., Lim (2006), Gilboa (2010), and Kahneman (2011).

When we deal with only two experiences of same sign, then integration-segregation decisions are simple, as the following theorem shows. Throughout the rest of the chapter, the value maximizer means the value maximizing decision maker.

Theorem 4.2.1 *The value maximizer with any value function v defined in (1.1) prefers to segregate two exposure units with positive experiences and integrate those with negative experiences.*

We skip the proof of Theorem 4.2.1 since it can be proved using majorization. Hereafter, we shall frequently use a special case of the Hardy-Littlewood-Pólya (HLP) majorization principle (e.g., Kuczma, 2009, p. 211). Namely, given two vectors (x_1, x_2) and (y_1, y_2), and also a continuous and concave function v, we have the implication:

$$\left.\begin{array}{l} x_1 \geq x_2,\ y_1 \geq y_2 \\ x_1 + x_2 = y_1 + y_2 \\ x_1 \leq y_1 \end{array}\right\} \Longrightarrow v(x_1) + v(x_2) \geq v(y_1) + v(y_2). \tag{2.1}$$

To exemplify, we may view Theorem 4.2.1 as saying that the value maximizer prefers to enjoy two positive experiences on, say, two different days, but if he faces two negative experiences and has a choice over the timing, then he prefers to get over the experiences as quickly as possible, say on the same day. Note that Theorem 4.2.1 does not impose any restriction on the value function v, except those specified in definition (1.1). Finally, we note that Theorem 4.2.1 is a special case of Theorem 4.4.1 to be established later in the chapter.

The following theorem specifies those values of the parameter λ in the value-function v_λ for which integration or segregation is preferred in the case of two exposure units having experiences of different signs.

Theorem 4.2.2 *With the value function v_λ defined in (1.3), assume that one exposure unit has a positive experience $x_+ > 0$ and another one has a negative experience $x_- < 0$. Let $\mathbf{x} = (x_+, x_-)$ and denote*

$$T(\mathbf{x}) = \frac{u(x_+) - u\big(\max\{0, x_- + x_+\}\big)}{u(-x_-) - u\big(\max\{0, -(x_- + x_+)\}\big)}. \tag{2.2}$$

Then the value maximizer prefers integrating the two experiences if and only if $T(\mathbf{x}) \leq \lambda$ and segregating them if and only if $T(\mathbf{x}) \geq \lambda$.

Theorem 4.2.2 has been established by Egozcue and Wong (2010). We shall see in Section 4.4, which deals with an arbitrary number of exposure units, that Theorem 4.2.2 is a corollary to our more general Theorem 4.4.2. Hence, we do not give a proof of Theorem 4.2.2 here.

For an illustration of Theorem 4.2.2, we suggest to think of a situation when, say, the root-canal of one of our teeth has to be done and we try to decide whether this procedure should be done on the day of an exciting concert (which would hopefully help us to forget the unpleasant experience) or on a different day (so that we would not be bothered during the concert by the earlier unpleasant experience). Personally, we find this a non trivial choice, and this is indeed reflected by the increased mathematical complexity of Theorem 4.2.2 if compared to that of Theorem 4.2.1. For more examples, one may refer to, e.g., Gilboa (2010), and Kahneman (2011).

We are now in the position to elaborate on 'our' threshold $T(\mathbf{x})$ and compare it with that used by Jarnebrant et al. (2009). In short, the two thresholds delineate two different but closely related regions: $T(\mathbf{x})$ concerns with the integration-segregation region with respect to the loss aversion parameter λ, whereas the threshold used by Jarnebrant et al. (2009) concerns with the gain region by dividing it into two parts: in one, segregation is preferred, and in the other part, integration is preferred. In more detail, Jarnebrant et al. (2009) specify conditions under which the 'silver lining effect' occurs, assuming the same value-function v_λ as in the present chapter. They show, for example, that if a gain is smaller than a certain gain-threshold, then segregation is preferred. In contrast, our λ-based threshold is related to a certain value of the loss aversion parameter λ: if it is smaller than the threshold, then segregation is preferred; otherwise integration is preferred. Note also that the threshold $T(\mathbf{x})$ has an explicit formula, whereas a formula for the threshold used by Jarnebrant et al. (2009) is more difficult to arrive at. Moreover, their definition is not yet clear for more than two exposure units, even for three units, because in this case we could have, for example, two gains and a loss and would thus be required to use a threshold-set of some kind, instead of just a threshold-parameter.

We next provide an insight into the magnitude of the threshold $T(\mathbf{x})$; namely, whether it is below or above 1. Knowing the answer is useful because if, for example, under the assumption of Theorem 4.2.2, $T(\mathbf{x}) \le 1$ and the decision maker is loss averse, that is, $\lambda \ge 1$, implies that the value maximizer prefers integration.

Theorem 4.2.3 *Assume that the conditions of Theorem 4.2.2 are satisfied, and thus among x_1 and x_2 there is one positive and one negative value. If $x_1 + x_2 \ge 0$, then $T(\mathbf{x}) \le 1$, and if $x_1 + x_2 \le 0$, then $T(\mathbf{x}) \ge 1$.*

Proof. We start with the case $x_1 + x_2 \ge 0$. Then $T(\mathbf{x}) \le 1$ is equivalent to $u(x_+) - u(x_- + x_+) \le u(-x_-)$, which using the notation $y_1 = -x_- \ge 0$ and $y_2 = x_- + x_+ \ge 0$ can be rewritten as the bound $u(y_1 + y_2) \le u(y_1) + u(y_2)$. By Theorem 4.2.1, the latter bound holds, which establishes $T(\mathbf{x}) \le 1$. When $x_1 + x_2 \le 0$, then $T(\mathbf{x}) \ge 1$ is equivalent to the bound $u(x_+) + u(-x_- - x_+) \ge u(-x_-)$. With the notation $z_1 = x_+ \ge 0$ and $z_2 = -x_- - x_+ \ge 0$, the above bound becomes $u(z_1 + z_2) \le u(z_1) + u(z_2)$. By Theorem 4.2.1, the latter bound holds, and so we have $T(\mathbf{x}) \ge 1$. This completes the proof of Theorem 4.2.3. ■

4.3 Case $n = 3$: which ones to integrate, if any?

Complete integration or complete segregation may not result in the maximal value when there are more than two exposure units, and thus a partial integration-segregation decision could be better. In this section we shall give a complete solution to this problem in the case of three exposure units (i.e., $n = 3$).

We begin with a note that the value maximizer with the value function v defined in (1.1) prefers to segregate three exposure units with positive experiences, and integrate three exposure units with negative experiences (we refer to Theorem 4.4.1 to be established later). When there are mixed experiences (i.e., at least one positive and at least one negative), then integration-segregation decisions are complex. To illustrate, we next give an example (in two parts) violating principles P3 and P4.

Example 4.3.1 *Assume the value function*

$$v_{\lambda,\gamma}(x) = \begin{cases} x^\gamma & \text{when} \quad x \geq 0, \\ -\lambda(-x)^\gamma & \text{when} \quad x < 0. \end{cases}$$

Countering P3: *Suppose that $\lambda = 1.4$ and $\gamma = 0.4$. Let $\mathbf{x} = (2, 2, -3.99)$. The sum of the experiences is $\sum x_k = 0.01$. Hence, a straightforward extension of Principle P3 with $n = 3$ would suggest integrating the three exposure units into one, but the following inequality implies the opposite: $v_{\lambda,\gamma}(\sum x_k) = 0.1584 < \sum v_{\lambda,\gamma}(x_k) = 0.2039$.*

Countering P4: *Suppose that $\lambda = 2.25$ and $\gamma = 0.88$. Let $\mathbf{x} = (0.5, -10, -20)$. The sum of the experiences is $\sum x_k = -29.5$. Hence, a straightforward extension of Principle P4 with $n = 3$ would suggest segregating the three exposure units, but the following inequality says the opposite: $v_{\lambda,\gamma}(\sum x_k) = -44.2207 > \sum v_{\lambda,\gamma}(x_k) = -47.9361$.*

Hence, we now see that neither complete segregation nor complete integration of three (or more) experiences with mixed exposures may lead to maximal values. For this reason, we next develop an exhaustive integration-segregation theory for three exposure units, which is a fairly frequent case in practice. To illustrate, the following example is borrowed from the telecommunications industry (Bell Aliant, 2012):

- TV + Internet + Home Phone: \$99.00/month (regular \$135.95)
- TV + Home Phone: \$64.95/month (regular \$98.95)
- TV + Internet: \$94.95/month (regular \$110.95)
- Internet + Home Phone: \$69.95/month (regular \$91.95)

Note from the prices that depending on factors such as the prices of individual products as well as (likely unknown but guessed) underlying value functions, there are possibilities for discounts due to bundling. Another popular example of bundling would be vacation packages (e.g., Orbitz, 2012) that usually involve flight, hotel, and car; in various combinations. Yet another popular bundle would be the office software suit, which among possibly many 'auxiliary' components, usually has the following three base components: word processor, spreadsheet, and presentation program. Note that the above examples concern with three different products, as is generally the case throughout the current chapter, but there can also be, for example, 'volume bundling' of identical products, in which case we would deal with identical $x_1, \dots, x_n$ or, specifically to this section, with identical x, y, and z, that is, $x = y = z$.

Unless explicitly noted otherwise, we shall work with the value function v_λ defined by equation (1.3). The three experiences are x, y and z, and we assume that they satisfy

$$x + y + z \geq 0. \tag{3.1}$$

Furthermore, without loss of generality, we assume that

$$x \geq y \geq z, \tag{3.2}$$

since every other case can be reduced to (3.2) by simply changing the notation. We also assume, without loss of generality, that

$$x \neq 0, \;\; y \neq 0, \;\; z \neq 0, \tag{3.3}$$

because if at least one of the three experiences is zero, then the currently investigated case $n = 3$ reduces to $n = 2$, which has been discussed earlier in this chapter and also investigated by Egozcue and Wong (2010).

Finally, we note that there are five possibilities for integration and segregation in the case of three exposure units:

A. $v_\lambda(x) + v_\lambda(y) + v_\lambda(z)$,

B. $v_\lambda(x) + v_\lambda(y + z)$,

C. $v_\lambda(y) + v_\lambda(x + z)$,

D. $v_\lambda(z) + v_\lambda(x + y)$,

E. $v_\lambda(x + y + z)$.

In summary, our goal in this section is to determine which of the above five possibilities produces the largest value (maximal). We also want to know, and Note 4.3.1 below will explain why, which of the five cases and under what conditions produces the smallest value (minimal). This is exactly what Theorems 4.3.1–4.3.5 will establish.

Note 4.3.1 *The reason for including the minimal values when only the maximal ones seem to be of interest, is due to the fact that finding the maximal ones in the case $x + y + z \leq 0$ can be reduced to finding the minimal ones under the condition $x + y + z \geq 0$. Indeed, note that $x + y + z \leq 0$ is equivalent to $x^- + y^- + z^- \geq 0$ with the notation $x^- = -x$, $y^- = -y$, and $z^- = -z$. Since $\lambda > 0$, the equation*

$$v_\lambda(x) = -\frac{1}{\lambda^*} v_{\lambda^*}(-x)$$

with $\lambda^* = 1/\lambda$ implies that finding the maximal value among (A)–(E) is equivalent to finding the minimal value among the following five cases:

$v_{1/\lambda}(x^-) + v_{1/\lambda}(y^-) + v_{1/\lambda}(z^-)$,

$v_{1/\lambda}(x^-) + v_{1/\lambda}(y^- + z^-)$,

$v_{1/\lambda}(y^-) + v_{1/\lambda}(x^- + z^-)$,

$v_{1/\lambda}(z^-) + v_{1/\lambda}(x^- + y^-)$,

$v_{1/\lambda}(x^- + y^- + z^-)$.

The minimal values among the latter five cases will be easily derived from Theorems 4.3.1–4.3.5, where we only need to replace x, y, and z by x^-, y^-, and z^-, respectively, and also replace the parameter λ by $1/\lambda$.

Since from now on we are only concerned with the case $x+y+z \geq 0$, we therefore know that at least one of the three exposure units has a non-negative experience. Furthermore, every triplet (x, y, z), falls into one of the following five cases:

$$x \geq y \geq z \geq 0 \tag{3.4}$$

$$x \geq y \geq 0 \geq z \quad \text{and} \quad y \geq -z \tag{3.5}$$

$$x \geq y \geq 0 \geq z \quad \text{and} \quad x \geq -z \geq y \tag{3.6}$$

$$x \geq y \geq 0 \geq z \quad \text{and} \quad -z \geq x \tag{3.7}$$

$$x \geq 0 \geq y \geq z \tag{3.8}$$

In the proofs of Theorems 4.3.1–4.3.5 below, we use notation such as "$\succcurlyeq$." To clarify its meaning, we take, e.g. the statement $(A) \succcurlyeq (E)$, which means that $v_\lambda(x) + v_\lambda(y) + v_\lambda(z) \geq v_\lambda(x+y+z)$. Hence, $(A) \succcurlyeq (E)$ says in a concise way that the value maximizer prefers (A) to (E). Naturally, the value minimizer – and we consider this case due to the reason given in Note 4.3.1 – prefers (E) to (A) whenever the relationship $(A) \succcurlyeq (E)$ holds.

Theorem 4.3.1 *Let the value function be v_λ, and let (3.4) hold. Then we have the following two statements:*

Max: (A) gives the maximal value among (A)–(E).

Min: (E) gives the minimal value among (A)–(E).

Proof. Since the three exposure units have non-negative experiences x, y, and z, Theorem 4.4.1 implies that complete segregation maximizes the value. Hence, (A) attains the maximal value among (A)–(E). Same theorem also implies that complete integration, which is (E), attains the minimal value. ■

The following analysis of cases (3.5)–(3.8) is much more complex. Now we are ready to formulate and prove our next theorem.

Theorem 4.3.2 *Let the value function be v_λ, and let (3.5) hold.*

Max: With the threshold $T_{AC} = T(x, z)$, the following statements specify the two possible maximal values among (A)–(E):

- *If $T_{AC} \geq \lambda$, then (A).*
- *If $T_{AC} \leq \lambda$, then (C).*

Min: With the threshold $T_{DE} = T(x + y, z)$, the following statements specify the two possible minimal values among (A)–(E):

- *If $T_{DE} \geq \lambda$, then (E).*
- *If $T_{DE} \leq \lambda$, then (D).*

Proof. Since x and y are non-negative, from Theorem 4.4.1 we have that $(A) \succcurlyeq (D)$, and since x and $y + z$ are non-negative, the same theorem implies that $(B) \succcurlyeq (E)$. The proof of $(C) \succcurlyeq (B)$ is more complex. Note that $(C) \succcurlyeq (B)$ is equivalent to

$$v_\lambda(y) + v_\lambda(x + z) \geq v_\lambda(x) + v_\lambda(y + z), \tag{3.9}$$

which we establish as follows:

- When $x + z \geq y$, then we apply the HLP principle on the vectors $(x + z, y)$ and $(x, y + z)$ and get $v_\lambda(x + z) + v_\lambda(y) \geq v_\lambda(x) + v_\lambda(y + z)$, which is (3.9).
- When $x + z \leq y$, then we apply the HLP principle on the vectors $(y, x + z)$ and $(x, y + z)$, and get $v_\lambda(y) + v_\lambda(x + z) \geq v_\lambda(x) + v_\lambda(y + z)$, which is (3.9).

This completes the proof of inequality (3.9). Hence, in order to establish the 'max' part of Theorem 4.3.2, we need to determine whether (A) or (C) is maximal, and for the 'min' part, we need to determine whether (D) or (E) is minimal.

The 'max' part.

Since $x \geq 0$ and $z \leq 0$, whether (A) or (C) is maximal is determined by the threshold T_{AC}: when $T_{AC} \leq \lambda$, then $(C) \succcurlyeq (A)$, and when $T_{AC} \geq \lambda$, then $(A) \succcurlyeq (C)$. This concludes the proof of the 'max' part.

The 'min' part.

Since $x + y \geq 0$ and $z \leq 0$, the threshold $T_{DE} = T(x + y, z)$ plays a decisive role: if $T_{DE} \leq \lambda$, then $(E) \succcurlyeq (D)$, and if $T_{DE} \geq \lambda$, then $(D) \succcurlyeq (E)$. This concludes the proof of the 'min' part and of Theorem 4.3.2 as well. ∎

Theorem 4.3.3 *Let the value function be v_λ, and let (3.6) holds.*

Max: With the thresholds $T_{AC} = T(x, z)$ and $T_{AB} = T(y, z)$, the following statements specify the two possible maximal values among (A)–(E):

- *If $T_{AC} \geq \lambda$, then (A).*
- *If $T_{AC} \leq \lambda$, then (C).*

Min: With the thresholds $T_{BE} = T(x, y + z)$, $T_{DE} = T(x + y, z)$, and

$$T_{BD} = \frac{u(x + y) - u(x)}{u(-z) - u(-y - z)},$$

the following statements specify the three possible minimal values among (A)–(E):

- *If $T_{BE} \leq \lambda$ and $T_{BD} \geq \lambda$, then (B).*
- *If $T_{DE} \leq \lambda$ and $T_{BD} \leq \lambda$, then (D).*
- *If $T_{BE} \geq \lambda$ and $T_{DE} \geq \lambda$, then (E).*

Proof. Since x and y are non-negative, we have $(A) \succcurlyeq (D)$, and since y and $x + z$ are non-negative, we have $(C) \succcurlyeq (E)$. Hence, it remains to consider only cases (A), (B), and (C) for the 'max' part of the theorem, and only (B), (D), and (E) for the 'min' part.

The 'max' part.

First we show that $T_{AC} \leq T_{AB}$. Since $x+z \geq 0$, from Theorem 4.2.3 we have $T_{AC} \leq 1$, and since $y + z \leq 0$, the same theorem implies $T_{AB} \geq 1$. Hence, $T_{AC} \leq T_{AB}$.

To establish that (A) is maximal when $T_{AC} \geq \lambda$, we check that $(A) \succcurlyeq (B)$ and $(A) \succcurlyeq (C)$. The former statement holds when $T_{AB} = T(y,z) \geq \lambda$, and the latter when $T_{AC} = T(x,z) \geq \lambda$. But we already know that $T_{AC} \leq T_{AB}$. Therefore, when $T_{AC} \geq \lambda$, then $T_{AB} \geq \lambda$. This proves that when $T_{AC} \geq \lambda$, then (A) gives the maximal value among (A), (B), (C), and thus, in turn, among all (A)–(E).

To establish that (C) is the maximal when $T_{AC} \leq \lambda$, we need to check that $(C) \succcurlyeq (A)$ and $(C) \succcurlyeq (B)$. First we note that when $T_{AC} \leq \lambda$, then $(C) \succcurlyeq (A)$. Furthermore,

$$v_\lambda(x) + v_\lambda(y+z) \leq v_\lambda(y) + v_\lambda(x+z) \iff u(x) - \lambda u(-y-z) \leq u(y) + u(x+z)$$
$$\iff T_{BC} \leq \lambda,$$

where T_{BC} is defined by the equation

$$T_{BC} = \frac{u(x) - u(x+z) - u(y)}{u(-y-z)}.$$

Hence, when $T_{BC} \leq \lambda$, then $(C) \succcurlyeq (B)$. Simple algebra shows that the bound $T_{BC} \leq T_{AB}$ is equivalent to $T_{AC} \leq T_{AB}$, and we already know that the latter holds. Hence, $T_{BC} \leq T_{AB}$ and so $T_{BC} \leq \lambda$ when $T_{AC} \leq \lambda$. In summary, when $T_{AC} \leq \lambda$, then (C) gives the maximal value among all cases (A)–(E). This concludes the proof of the 'max' part.

The 'min' part.

We first establish conditions under which (B) is minimal. We have $(E) \succcurlyeq (B)$ when $T_{BE} \leq \lambda$. To have $(D) \succcurlyeq (B)$, we need to employ the threshold T_{BD}, which is defined in the formulation of the theorem. The role of the threshold is seen from the following equivalence relations:

$$v_\lambda(x) + v_\lambda(y+z) \leq v_\lambda(z) + v_\lambda(x+y) \iff u(x) - \lambda u(-y-z) \leq -\lambda u(-z) + u(x+y)$$
$$\iff \lambda \leq T_{BD}.$$

Hence, if $T_{BD} \geq \lambda$, then $(D) \succcurlyeq (B)$. In summary, when $T_{BE} \leq \lambda$ and $T_{BD} \geq \lambda$, then (B) gives the minimal value among all (A)–(E).

We next establish conditions under which (D) is minimal. First, we have $(E) \succcurlyeq (D)$ when $T_{DE} \leq \lambda$. Next, we have $(B) \succcurlyeq (D)$ when $T_{BD} \leq \lambda$. In summary, when $T_{DE} \leq \lambda$ and $T_{BD} \leq \lambda$, then (D) gives the minimal value among all (A)–(E).

Finally, we have $(B) \succcurlyeq (E)$ when $T_{BE} \geq \lambda$, and $(D) \succcurlyeq (E)$ when $T_{DE} \geq \lambda$. Hence, when $T_{BE} \geq \lambda$ and $T_{DE} \geq \lambda$, then (E) is minimal among all (A)–(E). This finishes the proof of the 'min' part, and thus of Theorem 4.3.3 as well. ■

Theorem 4.3.4 *Let the value function be v_λ, and let (3.7) holds.*

Max: With the threshold

$$T_{AE} = \frac{u(x) + u(y) - u(x+y+z)}{u(-z)},$$

the following statements specify the two possible maximal values among (A)–(E):

- *If $T_{AE} \geq \lambda$, then (A).*
- *If $T_{AE} \leq \lambda$, then (E).*

Min: With the thresholds $T_{AC} = T(x, z)$, $T_{BE} = T(x, y+z)$, $T_{CE} = T(y, x+z)$, $T_{DE} = T(x+y, z)$, and

$$T_{BC} = \frac{u(x) - u(y)}{u(-y-z) - u(-x-z)},$$
$$T_{BD} = \frac{u(x+y) - u(x)}{u(-z) - u(-y-z)},$$
$$T_{CD} = \frac{u(x+y) - u(y)}{u(-z) - u(-x-z)},$$

the following statements specify the four possible minimal values among (A)–(E):

- *If $T_{BE} \leq \lambda$, $T_{BC} \leq \lambda$, and $T_{BD} \geq \lambda$, then (B).*
- *If $T_{CE} \leq \lambda$, $T_{BC} \geq \lambda$, and $T_{CD} \geq \lambda$, then (C).*
- *If $T_{DE} \leq \lambda$, $T_{BD} \leq \lambda$, and $T_{CD} \leq \lambda$, then (D).*

– *If* $T_{BE} \geq \lambda$, $T_{CE} \geq \lambda$, *and* $T_{DE} \geq \lambda$, *then (E).*

Proof. Since both x and y are non-negative, we have $(A) \succcurlyeq (D)$. This eliminates (D) from the 'max' part of Theorem 4.3.4 and (A) from the 'min' part.

The 'max' part.

We first eliminate (B). When $T_{BE} \leq \lambda$, then $(E) \succcurlyeq (B)$. If, however, $T_{BE} \geq \lambda$, then by Theorem 4.4.2 we have $(B) \succcurlyeq (E)$. Also when $T_{AB} \geq \lambda$, then $(A) \succcurlyeq (B)$. On the other hand, if $T_{AB} \leq \lambda$, then $(B) \succcurlyeq (A)$. We have from Theorem 4.2.3 that $T_{BE} \leq 1$. Theorem 4.2.3 also implies that $T_{AB} \geq 1$ because $x + z \leq 0$. Hence, $T_{BE} \leq T_{AB}$. Thus, we have two cases: (i) When $T_{BE} \geq \lambda$, we conclude that $T_{AB} \geq \lambda$ and thus $(A) \succcurlyeq (B)$ and (ii) when $T_{BE} \leq \lambda$ then $(E) \succcurlyeq (B)$. In either case, (B) is discarded as an optimal option. Therefore, the value maximizing decision maker will not choose (B). Analogous arguments, but with T_{CE} and T_{AC} instead of T_{BE} and T_{AB}, respectively, show that the value maximizing decision maker will not choose (C) either. Hence, in summary, we are left with only two cases: (A) and (E). Which of the two maximizes the value is determined by the equivalence relations:

$$
\begin{aligned}
v_\lambda(x) + v_\lambda(y) + v_\lambda(z) \leq v_\lambda(x+y+z) &\iff u(x) + u(y) - \lambda u(-z) \leq u(x+y+z) \\
&\iff T_{AE} \leq \lambda.
\end{aligned}
$$

This concludes the proof of the 'max' part.

The 'min' part.

To prove the 'min' part, we only need to deal with (B)–(E), because we already know that $(A) \succcurlyeq (D)$. Case (E) gives the minimal value when $T_{BE} \geq \lambda$, $T_{CE} \geq \lambda$, and $T_{DE} \geq \lambda$. If, however, there is at least one among T_{BE}, T_{CE}, and T_{DE} not exceeding λ, then the minimum is achieved by one of (B), (C), and (D). To determine which of them and when is minimal, we employ simple algebra and obtain the equivalence relationships:

$$\begin{bmatrix} (C) \succcurlyeq (B) \Longleftrightarrow T_{BC} \le \lambda \\ (D) \succcurlyeq (B) \Longleftrightarrow T_{BD} \ge \lambda \\ (E) \succcurlyeq (B) \Longleftrightarrow T_{BE} \le \lambda \end{bmatrix}, \begin{bmatrix} (B) \succcurlyeq (C) \Longleftrightarrow T_{BC} \ge \lambda \\ (D) \succcurlyeq (C) \Longleftrightarrow T_{CD} \ge \lambda \\ (E) \succcurlyeq (C) \Longleftrightarrow T_{CE} \le \lambda \end{bmatrix},$$
$$\begin{bmatrix} (B) \succcurlyeq (D) \Longleftrightarrow T_{BD} \le \lambda \\ (C) \succcurlyeq (D) \Longleftrightarrow T_{CD} \le \lambda \\ (E) \succcurlyeq (D) \Longleftrightarrow T_{DE} \le \lambda \end{bmatrix}, \begin{bmatrix} (B) \succcurlyeq (E) \Longleftrightarrow T_{BE} \ge \lambda \\ (C) \succcurlyeq (E) \Longleftrightarrow T_{CE} \ge \lambda \\ (D) \succcurlyeq (E) \Longleftrightarrow T_{DE} \ge \lambda \end{bmatrix}.$$

This finishes the proof of Theorem 4.3.4. ■

Theorem 4.3.5 *Let the value function be v_λ, and let (3.8) holds.*

Max: With the thresholds $T_{BE} = T(x, y+z)$, $T_{CE} = T(x+z, y)$, $T_{DE} = T(x+y, z)$, and

$$T_{BC} = \frac{u(x) - u(x+z)}{u(-y-z) - u(-y)},$$
$$T_{BD} = \frac{u(x) - u(x+y)}{u(-y-z) - u(-z)},$$
$$T_{CD} = \frac{u(x+y) - u(x+z)}{u(-z) - u(-y)},$$
$$T_{DE} = \frac{u(x+y) - u(x+y+z)}{u(-z)}.$$

the following statements specify the four possible maximal values among (A)–(E):

- *If $T_{BE} \ge \lambda$, $T_{BC} \ge \lambda$, and $T_{BD} \ge \lambda$, then (B).*
- *If $T_{CE} \ge \lambda$, $T_{BC} \le \lambda$, and $T_{CD} \le \lambda$, then (C).*
- *If $T_{DE} \ge \lambda$, $T_{BD} \le \lambda$, and $T_{CD} \ge \lambda$, then (D).*
- *If $T_{BE} \le \lambda$, $T_{CE} \le \lambda$, and $T_{DE} \le \lambda$, then (E).*

Min: With the thresholds $T_{AC} = T(x, z)$, $T_{AD} = T(x, y)$,

$$T_{AE} = \frac{u(x) - u(x+y+z)}{u(-y) + u(-z)},$$

and the other ones defined in the 'max' part of this theorem, the following statements specify the four possible minimal values among (A)–(E):

- *If $T_{AC} \le \lambda$, $T_{AD} \le \lambda$, and $T_{AE} \le \lambda$, then (A).*
- *If $T_{AC} \ge \lambda$, $T_{CD} \ge \lambda$, and $T_{CE} \le \lambda$, then (C).*
- *If $T_{AD} \ge \lambda$, $T_{CD} \le \lambda$, and $T_{DE} \le \lambda$, then (D).*
- *If $T_{AE} \ge \lambda$, $T_{CE} \ge \lambda$, and $T_{DE} \ge \lambda$, then (E).*

Proof. *The 'max' part.*

Since $-y \ge 0$ and $-z \ge 0$, we have from inequality (4.1) that $u(-y) + u(-z) \ge u(-(y+z))$ and thus $-\lambda u(-y) - \lambda u(-z) \le -\lambda u(-(y+z))$. The latter is equivalent to $v_\lambda(y) + v_\lambda(z) \le v_\lambda(y+z)$, which means that $(B) \succcurlyeq (A)$.

We have four cases (B)–(E) to deal with. To determine which of them and when is maximal among (B)–(E), we employ simple algebra and obtain the equivalence relationships:

$$\begin{bmatrix} (B) \succcurlyeq (C) \Longleftrightarrow T_{BC} \ge \lambda \\ (B) \succcurlyeq (D) \Longleftrightarrow T_{BD} \ge \lambda \\ (B) \succcurlyeq (E) \Longleftrightarrow T_{BE} \ge \lambda \end{bmatrix}; \begin{bmatrix} (C) \succcurlyeq (B) \Longleftrightarrow T_{BC} \le \lambda \\ (C) \succcurlyeq (D) \Longleftrightarrow T_{CD} \le \lambda \\ (C) \succcurlyeq (E) \Longleftrightarrow T_{CE} \ge \lambda \end{bmatrix};$$

$$\begin{bmatrix} (D) \succcurlyeq (B) \Longleftrightarrow T_{BD} \le \lambda \\ (D) \succcurlyeq (C) \Longleftrightarrow T_{CD} \ge \lambda \\ (D) \succcurlyeq (E) \Longleftrightarrow T_{DE} \ge \lambda \end{bmatrix}.$$

This finishes the proof of the 'max' part.

The 'min' part.

To prove the 'min' part of the theorem, we verify the following four sets of orderings:

$$\begin{bmatrix} (C) \succcurlyeq (A) \Longleftrightarrow T_{AC} \le \lambda \\ (D) \succcurlyeq (A) \Longleftrightarrow T_{AD} \le \lambda \\ (E) \succcurlyeq (A) \Longleftrightarrow T_{AE} \le \lambda \end{bmatrix}, \quad \begin{bmatrix} (A) \succcurlyeq (C) \Longleftrightarrow T_{AC} \ge \lambda \\ (D) \succcurlyeq (C) \Longleftrightarrow T_{CD} \ge \lambda \\ (E) \succcurlyeq (C) \Longleftrightarrow T_{CE} \le \lambda \end{bmatrix},$$

$$\begin{bmatrix} (A) \succcurlyeq (D) \Longleftrightarrow T_{AD} \geq \lambda \\ (C) \succcurlyeq (D) \Longleftrightarrow T_{CD} \leq \lambda \\ (E) \succcurlyeq (D) \Longleftrightarrow T_{DE} \leq \lambda \end{bmatrix}, \quad \begin{bmatrix} (A) \succcurlyeq (E) \Longleftrightarrow T_{AE} \geq \lambda \\ (C) \succcurlyeq (E) \Longleftrightarrow T_{CE} \geq \lambda \\ (D) \succcurlyeq (E) \Longleftrightarrow T_{DE} \geq \lambda \end{bmatrix}.$$

This concludes the proof of the 'min' part and of Theorem 4.3.5 as well. ■

4.4 Arbitrary number of exposure units

We already know that when $n = 3$, then we have five cases to analyze. This number 5 – in the context of the present chapter – turns out to be the fourth member of the Bell sequence. Indeed, the number of possible cases to integrate or segregate n outcomes is related to the Bell number (Bell, 1934). This number is denoted by B_n, and is defined as the number of partitions of a set with n members. It satisfies the following recursion formula,

$$B_{n+1} = \sum_{k=0}^{n} C_k^n B_k.$$

The sequence of the first Bell numbers is $B_0 = B_1 = 1, B_2 = 2, B_3 = 5, B_4 = 15,\ B_5 = 52,\ B_6 = 203, B_7 = 877,\ B_8 = 4.140,\ \ldots$

Various partial scenarios, however, are quite reasonable to look at even for general n, and we shall next discuss some of them. For this we first observe that from the mathematical point of view, the integration-segregation rules are about the super- and sub-additivity of value functions. Decision makers, however, tend to 'visualize' the functions in terms of their shapes, such as concavity or convexity. A link between the additivity and concavity notions is accomplished by functional inequalities, such as Petrović's inequality (see, e.g., Kuczma, 2009), which says that for every $n \geq 2$ and for every continuous and concave function $v : [0, \infty) \to \mathbf{R}$ such that $v(0) = 0$, the inequality

$$v\left(\sum_{k=1}^{n} x_k\right) \leq \sum_{k=1}^{n} v(x_k) \tag{4.1}$$

holds for all $x_1, \ldots, x_n \in [0, \infty)$. In other words, inequality (4.1) says that the value function v is subadditive on $[0, \infty)$. This implies that the value maximizer prefers to segregate positive experiences. In the domain $(-\infty, 0]$ of losses, the roles of integration and segregation are reversed. Collecting the above observations, we have the following general theorem.

Theorem 4.4.1 *The value maximizer with any value function v defined in (1.1) prefers to segregate any number of exposure units with positive experiences, and integrate any number of exposure units with negative experiences.*

Theorem 4.4.1 rules out mixed experiences. We shall next relax this assumption, but at the expense of generality. First, we restrict ourselves to the value function v_λ. Second, we restrict our attention to learning if it is preferable to integrate all exposure units or to keep them all segregated, and no other option is available, or of interest, to us. The number of exposure units $n \geq 2$ remains arbitrary.

Theorem 4.4.2 *With $\mathbf{x} = (x_1, \ldots, x_n)$, we define the threshold $T(\mathbf{x})$ by*

$$T(\mathbf{x}) = \frac{\sum_{k \in \mathcal{K}_+} u(x_k) - u\Big(\max\Big\{0, \sum_{k=1}^{n} x_k\Big\}\Big)}{\sum_{k \in \mathcal{K}_-} u(-x_k) - u\Big(\max\Big\{0, -\sum_{k=1}^{n} x_k\Big\}\Big)} \, ;$$

which is always non-negative, where $\mathcal{K}_+ = \{k : x_k > 0\}$ and $\mathcal{K}_- = \{k : x_k < 0\}$ are two subsets of $\{1, \ldots, n\}$. The threshold $T(\mathbf{x})$ splits the values of the loss aversion parameter λ into two regions – integration and segregation – as follows: assuming that there is at least one exposure unit with a positive experience and at least one with a negative experience, and given that either complete integration or complete segregation of all exposure units is possible, then the value maximizer prefers

- *integrating the exposure units if and only if $T(\mathbf{x}) \leq \lambda$, and*
- *segregating the exposure units if and only if $T(\mathbf{x}) \geq \lambda$.*

Proof. We start with the case $\sum_{k=1}^{n} x_k \geq 0$. The inequality $v_\lambda(\sum_{k=1}^{n} x_k) \leq \sum_{k=1}^{n} v_\lambda(x_k)$ is equivalent to

$$u\left(\sum_{k=1}^{n} x_k\right) \leq -\lambda \sum_{k\in\mathcal{K}_-} u(-x_k) + \sum_{k\in\mathcal{K}_+} u(x_k),$$

which, in turn, is equivalent to

$$\lambda \leq T_+(\mathbf{x}) := \frac{\displaystyle\sum_{k\in\mathcal{K}_+} u(x_k) - u\left(\sum_{k=1}^{n} x_k\right)}{\displaystyle\sum_{k\in\mathcal{K}_-} u(-x_k)}. \tag{4.2}$$

Since $\sum_{k=1}^{n} x_k \geq 0$, we have $T_+(\mathbf{x}) = T(\mathbf{x})$. To show that $T(\mathbf{x})$ is non-negative, we first note that since the function u is non-decreasing and $\sum_{k\in\mathcal{K}_-} x_k \leq 0$, we have

$$\begin{aligned}\sum_{k\in\mathcal{K}_+} u(x_k) - u\left(\sum_{k=1}^{n} x_k\right) &= \sum_{k\in\mathcal{K}_+} u(x_k) - u\left(\sum_{k\in\mathcal{K}_+} x_k + \sum_{k\in\mathcal{K}_-} x_k\right) \\ &\geq \sum_{k\in\mathcal{K}_+} u(x_k) - u\left(\sum_{k\in\mathcal{K}_+} x_k\right). \end{aligned} \tag{4.3}$$

In addition, since the function $u : [0,\infty) \to \mathbf{R}$ is continuous, concave, and $u(0) = 0$, the right-hand side of bound (4.3) is non-negative. Hence, $T_+(\mathbf{x}) \geq 0$.

Considering now the case $\sum_{k=1}^{n} x_k \leq 0$, we find that $v_\lambda(\sum_{k=1}^{n} x_k) \leq \sum_{k=1}^{n} v_\lambda(x_k)$ is equivalent to

$$\lambda \sum_{k\in\mathcal{K}_-} u(-x_k) - \lambda u\left(-\sum_{k=1}^{n} x_k\right) \leq \sum_{k\in\mathcal{K}_+} u(x_k). \tag{4.4}$$

Since the function u is non-decreasing and $\sum_{k\in\mathcal{K}_+} x_k \geq 0$, we have that

$$\begin{aligned}\sum_{k\in\mathcal{K}_-} u(-x_k) - u\left(-\sum_{k=1}^{n} x_k\right) &= \sum_{k\in\mathcal{K}_-} u(-x_k) - u\left(-\sum_{k\in\mathcal{K}_-} x_k - \sum_{k\in\mathcal{K}_+} x_k\right) \\ &\geq \sum_{k\in\mathcal{K}_-} u(-x_k) - u\left(-\sum_{k\in\mathcal{K}_-} x_k\right). \end{aligned} \tag{4.5}$$

Since the function $u : [0,\infty) \to \mathbf{R}$ is continuous, concave, and $u(0) = 0$, the right-hand side of bound (4.5) is non-negative. Hence, inequality (4.4) is equivalent to

$$\lambda \leq T_{-}(\mathbf{x}) := \frac{\displaystyle\sum_{k\in\mathcal{K}_+} u(x_k)}{\displaystyle\sum_{k\in\mathcal{K}_-} u(-x_k) - u\Big(-\sum_{k=1}^{n} x_k\Big)}. \tag{4.6}$$

Given the above, we have $T_{-}(\mathbf{x}) \geq 0$. Furthermore, since $\sum_{k=1}^{n} x_k \leq 0$, we have $T_{-}(\mathbf{x}) = \check{T}(\mathbf{x})$. This completes the proof of Theorem 4.4.2. ∎

4.5 Applications and numerical illustrations

In this section, we give a brief account of possible practical applications of our results to a variety of disciplines such as economics, finance, marketing, political science and taxation. Numerical examples that we shall present in the second subsection below are designed to illustrate our earlier theoretical results, especially their optimality.

Shefrin and Statman (1984) address the question why firms pay dividends. Since dividends have been taxed at a higher rate than capital gains, an investor would prefer that the firm repurchase shares instead of paying dividends. They propose a mental accounting explanation of this behavior. The rationale for dividends is that this will make easier for investors to segregate gains from losses, hence increasing their value function. For example, suppose a stock has increased in value by \$10. If there are no dividends the investor will code this gain as $v(10)$. Alternatively, suppose the firm pays a dividend of \$2, with a capital gain of \$8, this will be segregated as $v(2)+v(8)$, which will be result in a higher value, because of the concavity of the value function in the domain of gains. Similarly, consider the case of a stock that has lost \$10 in its value, against a loss of \$12 and a dividend of \$2, the investors will mentally compare $v(-10)$ against $v(-12)+v(2)$, which by Thaler's fourth principle, will enable investors to show the silver lining, would be preferred to the no dividend loss option.

Linville and Fischer (1991) extend mental integration and segregation and examine the hedonic editing rules not only for outcomes in the financial domain, but also outcomes in non-financial domain and across different domains. For example, they consider events such as: having a nice dinner with a friend, losing a $10 bill or receiving a traffic ticket. They find evidence for mental accounting to occur within and across different domains.

Hirst, Joyce and Schadewald (1994) show how segregating and integrating the utility of purchased goods and the disutility of how this purchased are financed affect consumer borrowing decision. Indeed, financing a good creates a stream of benefits and losses. Integrating or segregating these streams of gains and losses affects the utility of credit purchases. They find that consumers would prefer to associate loans with long-lived assets. For example, most subjects prefer to finance a furniture purchase (long-lived asset) over a two week vacation (short lived asset). They find evidence that supports that loan repayment is associated with the existence of future asset benefits which could be integrated with the loan payments. This means that a consumer would be willing to prepay a loan that relates to an expired asset than a loan that relates to an unexpired asset. Finally, they find evidence that individuals would be willing to incur in additional costs to enhance the likelihood that loan costs and benefits would co-occur in time.

Another application relates investors' selling and purchasing stocks strategy. There is a mixed evidence whether investors realize gains and losses jointly or separately (Lim 2006; Lehenkari 2009). Lim (2006) documents that investors prefer to bundle sales of stocks that are trading below their purchase price (losers) on the same day than sales of stocks above their purchase price (winners). The reason is that selling losers on the same day allow investors to integrate their losses. On the other hand, selling winners on different days makes easier to segregate gains. Contrary to Lim's findings, Lehenkari (2009) finds that investors in the Finnish stock market do not integrate losses and segregate gain as Thaler's principles predict. The bias toward concentration might be a possible explanation of this inconclusive evidence (Koszegi and Szeidl, 2013).

Sul, Kim and Choi (2013) investigate the relationship between subjective well-being and hedonic editing for mixed events. They find that happy individuals displayed a stronger preference for integrating a positive social event against a loss. That is, social gains are used as a cross domain buffers, where happy individuals displayed stronger preferences for social events.

Milkman *et* al. (2012) propose a policy bundling technique in which related bills involving both losses and gains are combined to offset separate bills' cost while preserving their net benefits. For instance, Stiglitz (1998) has pointed out the failure of passing legislation with high net positive expected value. Thus, this method can transform unpopular individual pieces of legislation into more popular choices. For example, a legislation can be seen as bundling spending cuts (gains) with spending increases (losses) with a net spending cut. The next example, taken from Milkman et al. (2012), might clarify this point.

Example 4.5.1 *Suppose a legislative faces two unpopular pieces of legislation during an economic downturn period:*

A. A bill that increases government spending by 10 million dollars at a time when the deficit is soaring but would create 100 new permanent jobs.

B. A bill eliminating 90 government jobs that would reduce the deficit by 12 million dollars.

Now suppose that both bills are combined into one single bill:

C. A bill reducing the deficit in 2 million dollars and an increase of 10 new permanent jobs.

Milkman et al. (2012) have found evidence that supports the combined bill (option C) better than either of its component bills (option A or B). Hence, bundling together two unpopular bills could indeed become popular.

Mental accounting has also received great attention in Marketing science (Drumwright, 1992; Heath, Chatterjee and France, 1995). Bundling an attractive product with a less attractive product, is a direct application of Thaler's third principle. The reason is that the seller bundles these products so that the consumer surplus of the attractive product will compensate the consumer deficit with the less attractive product.

Finally, our results can also be applied to taxation. For instance, a typical framework of mental accounting appears when there is a tax refund at the end of a fiscal year. In this case, mental accounting is a useful theory to analyze whether taxpayers shall prefer having a tax refund at the end of the fiscal year, but making, in advance, large monthly tax payments against making smaller monthly tax payments, but with no tax refund.

We shall next give some numerical illustrations of our earlier developed theory. In the following examples we use the S-shaped value function (al-Nowaihi et al., 2008)

$$v_{\lambda,\gamma}(x) = \begin{cases} x^{\gamma} & \text{when} \quad x \geq 0, \\ -\lambda(-x)^{\gamma} & \text{when} \quad x < 0. \end{cases} \tag{5.1}$$

Obviously, $v_{\lambda,\gamma} = v_{\lambda}$ with $u(x) = x^{\gamma}$.

The following two numerical examples illustrate the validity of principles P1 and P2.

Example 4.5.2 (illustrating principle P1) *Let the value function be $v_{\lambda,\gamma}$ with the parameters $\lambda = 2.25$ and $\gamma = 0.88$. Suppose that we have three exposure units with positive experiences 5, 10, and 20. Principle P1 suggests segregating them, and this is mathematically confirmed by the inequality: $v_{\lambda,\gamma}(\sum x_k) = 22.8444 < \sum v_{\lambda,\gamma}(x_k) = 25.6683$. (We use $\sum$ instead of $\sum_{k=1}^{3}$ to simplify notation.) Our general results say that the value maximizing decision maker prefers segregating any number of positive exposures.* $\square$

Example 4.5.3 (illustrating principle P2) *Let the value function be $v_{\lambda,\gamma}$ with the parameters $\lambda = 2.25$ and $\gamma = 0.88$. Suppose that we have three exposure units with negative experiences -5, -10, and -20. Principle P2 suggests integrating them, and this is confirmed by the inequality: $v_{\lambda,\gamma}(\sum x_k) = -51.3999 > \sum v_{\lambda,\gamma}(x_k) = -57.7537$. Our general results say that the value maximizing decision maker prefers integrating any number of negative exposures.* $\square$

The following two examples show that principles P3 and P4 can be violated.

Example 4.5.4 (illustrating principle P3) *Let the value function be $v_{\lambda,\gamma}$ with the parameters $\lambda = 2.25$ and $\gamma = 0.88$. Suppose that we have three exposure units with mixed* *experiences* -0.5, 10, *and* 20, *whose total (positive) experience is* $\sum x_k = 29.5$. *Principle P3 would suggest integrating the exposure units into one, but the following inequality implies the opposite:* $v_{\lambda,\gamma}(\sum x_k) = 19.6537 < \sum v_{\lambda,\gamma}(x_k) = 20.3239$. *In fact, we see from our theoretical analysis of the case* $n = 3$ *that neither complete segregation nor complete integration of three (or more) experiences with mixed exposures may lead to a maximal value, which may be achieved only by a partial integration and segregation.* $\square$

Example 4.5.5 (illustrating principle P4) *Let the value function be* $v_{\lambda,\gamma}$ *with the parameters* $\lambda = 2.25$ *and* $\gamma = 0.88$. *Suppose that we have three exposure units with mixed experiences* 0.5, -10, *and* -20, *whose total (negative) experience is* $\sum x_k = -29.5$. *Principle P4 suggests segregating the exposure units, but the following inequality says the opposite:* $v_{\lambda,\gamma}(\sum x_k) = -44.2207 > \sum v_{\lambda,\gamma}(x_k) = -47.9361$. *Our theory developed above says that neither complete segregation nor complete integration may lead to a maximal value when* $n \geq 3$. $\square$

Now we shall provide some numerical illustrations of our main theorems.

Example 4.5.6 (illustrating Theorem 4.3.2) *Assume that the value function is* $v_{\lambda,\gamma}$ *with* $\gamma = 0.88$. *With the experiences* $x = 5$, $y = 3$, *and* $z = -2$, *we have* $T_{AC} = 0.8109$ *and* $T_{ED} = 0.7575$. *Hence, the following statements hold:*

Max: When $\lambda \leq 0.8109$, *then (A) is maximal, and when* $0.8109 \leq \lambda$, *then (C) is maximal.*

Min: When $\lambda \leq 0.7575$, *then (E) is minimal, and when* $0.7575 \leq \lambda$, *then (D) is minimal.*

Example 4.5.7 (illustrating Theorem 4.3.3) *Assume that the value function is* $v_{\lambda,\gamma}$ *with* $\gamma = 0.88$. *With the experiences* $x = 10$, $y = 1$, *and* $z = -2$, *we have* $T_{AC} = 0.7349$, $T_{BD} = 0.7897$, *and* $T_{BE} = 0.6717$. *Hence, the following statements hold:*

Max: When $\lambda \leq 0.7349$, *then (A) is maximal, and when* $0.7349 \leq \lambda$, *then (C) is maximal.*

Min: When $\lambda \leq 0.6717$, *then (E) is minimal, when* $0.6717 \leq \lambda \leq 0.7897$, *then (B) is minimal, and when* $0.7897 \leq \lambda$, *then (D) is minimal.*

Example 4.5.8 (illustrating Theorem 4.3.4) *Assume that the value function is* $v_{\lambda,\gamma}$ *with* $\gamma = 0.88$. *With the experiences* $x = 4$, $y = 3$, *and* $z = -5$, *we have* $T_{AE} = 0.8404$, $T_{BD} = 0.9446$, $T_{CE} = 0.7890$, *and* $T_{CB} = 0.9014$. *Hence, the following statements hold:*

Max: When $\lambda \leq 0.8404$, *then (A) is maximal, and when* $0.8404 \leq \lambda$, *then (E) is maximal.*

Min: When $\lambda \leq 0.7890$, *then (E) is minimal, when* $0.7890 \leq \lambda \leq 0.9014$, *then (C) is minimal, when* $0.9014 \leq \lambda \leq 0.9446$, *then (B) is minimal, and finally when* $0.9446 \leq \lambda$, *then (D) is minimal.*

Example 4.5.9 (illustrating Theorem 4.3.5) *Assume that the value function is* $v_{\lambda,\gamma}$ *with* $\gamma = 0.88$. *With the experiences* $x = 36$, $y = -2$, *and* $z = -14$, *we have* $T_{BE} = 0.8243$, $T_{AC} = 0.8074$, *and* $T_{CE} = 0.6636$. *Hence, the following statements hold:*

Max: When $\lambda \leq 0.8243$, *then* B *is maximal, and when* $0.8243 \leq \lambda$, *then* E *is maximal.*

Min: When $\lambda \leq 0.6636$, *then* E *is minimal, when* $0.6636 \leq \lambda \leq 0.8074$, *then* C *is minimal, and when* $0.8074 \leq \lambda$, *then* A *is minimal.*

The following two examples illustrate Theorem 4.4.2 in the case of three exposure units and assuming that the decision maker is given only two options: either integrate all exposure units or keep them segregated.

Example 4.5.10 (illustrating Theorem 4.4.2) *Let* $x_1 = 25$, $x_2 = 10$, *and* $x_3 = -0.5$, *with the positive total sum* $x_1 + x_2 + x_3 = 34.5$. *Let the value function be* $v_{\lambda,\gamma}$ *with* $\gamma = 0.88$. *The threshold* $T(\mathbf{x}) = 3.7149$. *Thus, facing the dilemma of integrating or segregating all exposure units (we are not dealing with any partial integration or partial segregation in this example), the decision maker prefers segregating them when* $\lambda \leq 3.7149$ *and integrating them when* $\lambda \geq 3.7149$.

4.6 Concluding remarks

Our theoretical study has shown that within the class of value functions specified by prospect theory, the validity of Thaler's principles can be established rigorously in the case of only non-negative experiences, or only non-positive experiences, and irrespectively of the number of exposure units. When exposure units carry both negative and positive experiences, then the principles may break down. Our theory provides a complete solution to the integration/segregation problem in the case of three exposure units and demonstrates in particular that the transition from two to three, or more, exposure units increases the complexity of decisions enormously, thus showing the challenges that the decision maker encounters when dealing with multiple exposure units.

As far as we know, there has not been a detailed *theoretical* analysis of decision maker's behavior in the case of multiple exposure units. In this chapter, we have provided such an analysis, concentrating on two and three exposure units, and we have also noted possible results in the case of an arbitrary number of exposure units. Our theoretical analysis has shown that the number of integration-segregation options for more than three exposure units is so large that, generally, a well-informed integration-segregation decision becomes quite an unwieldy task.

Naturally, under such circumstances, we may think of employing computer-based search algorithms, but this computational approach would require us to specify the underlying value function, which is usually unknown in practice, except that it belongs to a certain class of functions depending on the problem. Hence, in this chapter we have aimed at deriving integration and segregation decisions that are qualitative in nature and applicable to classes of value functions pertaining to fairly general groups of individuals.

The present work can be extended in several directions. First, we have considered only a special case of S-shaped utility functions, which is conveniently linear in λ. Whether segregating or integrating multiple outcomes, but considering utility functions that are not linear on this parameter remains as an interesting line of future research. Second, we study a special case for any arbitrary number of exposure units. A more general analysis with multiple outcomes remains as a task for future studies.

References

Abdellaoui, M., Bleichrodt, H. & L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty, 36, 245–266.*

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica: Journal of the Econometric Society, 21, 503-546.*

Allais, M., & Hagen, G. M. (Eds.). (1979). Expected Utility Hypotheses and the Allais Paradox: Contemporary Discussions of the Decisions Under Uncertainty with Allais' Rejoinder (Vol. 21). Springer.

al-Nowaihi, A., Bradley, I., & Dhami, S. (2008). A note on the utility function under prospect theory. *Economics Letters, 99, 337–339.*

Andersson, P., Edman, J. & Ekman, M. (2005). Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting, 21, 565-576.*

Andersson, P., & Rakow, T. (2007). Now you see it now you don't: The effectiveness of the recognition heuristic for selecting stocks. *Judgment and Decision Making, 2(1), 29-39.*

Arrow, K. J. (1971). *Essays in the theory of risk-bearing (Vol. 1).* Chicago: Markham Publishing Company.

Åstebro, T., & Elhedhli, S. (2006). The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science, 52(3), 395-409.*

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12, 171-178.*

Balakrishnan, N., & Lai, C. D. (2009). *Continuous bivariate distributions: Theory and Methods*, Springer-Verlag, New York.

Barberis, N., Huang, M., & Santos, T. (2001). Prospect theory and asset prices. *Quarterly Journal of Economics, 116(1), 1-53.*

Baron, J. (2000). *Thinking and deciding.* Cambridge University Press.

Battermann, H. L., Braulke, M., Broll, U., & Schimmelpfennig, J. (2000). The preferred hedge instrument. *Economics Letters, 66(1), 85-91.*

Basu, A. K., Lal, R., Srinivasan, V., & Staelin, R. (1985). Salesforce compensation plans: An agency theoretic perspective. *Marketing Science, 4(4), 267-291.*

Batra, R. N., & Russell, W. R. (1974). Gains from trade under uncertainty. *American Economic Review, 64(6), 1040-1048.*

Beattie, J., Baron, J., Hershey, J. C., & Spranca, M. D. (1994). Psychological determinants of decision attitude. *Journal of Behavioral Decision Making, 7(2), 129-144.*

Bell Aliant (2012). *Bundle options: choose the bundle that meets your needs.* Viewed on 25 June 2012 at the web address
`http://productsandservice.bellaliant.net/PS/en/shop/bundles/bundles.jsp?prov=ns`

Bell, D.E.,(1982). Regret in decision making under uncertainty. *Operations Research 30(5), 961–981.*

Bell, E. T. (1934). Exponential numbers. *American Mathematical Monthly, 41(7), 411-419.*

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review, 79-98.*

Bennell, C., Emeno, K., Snook, B., Taylor, P., & Goodwill, A. (2009). The precision, accuracy and efficiency of geographic profiling predictions: a simple heuristic versus mathematical algorithms. *Crime Mapping: A Journal of Research and Practice, 1(2), 65-84.*

Benninga, S., Eldor, R., & Zilcha, I. (1983). Optimal hedging in the futures market under price uncertainty. *Economics Letters, 13(2), 141-145.*

Berhold, M. H. (1973). The use of distribution functions to represent utility functions. *Management Science, 19(7), 825-829.*

Bessis J (2009). *Risk management in banking.* Chichester, England, 3/e.

Bleichrodt, H., Cillo, A., & Diecidue, E. (2010). A quantitative measurement of regret theory. *Management Science, 56(1), 161-175.*

Bodie, Z., A. Kane, & A. J. Marcus, (1996). *Investments, 3rd ed.* New York, Boston, London.

Borges, B., Goldstein, D. G., Ortmann, A., & Gigerenzer, G.(1999). Can ignorance beat the stock market? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuris-*

tics that make us smart (pp. 59–72). New York: Oxford University Press.

Boyd, M. (2001). On ignorance, intuition, and investing: A bear market test of the recognition heuristic. *Journal of Psychology and Financial Markets, 2(3), 150-156.*

Braun, M., & Muermann, A. (2004). The impact of regret on the demand for insurance. *Journal of Risk and Insurance, 71(4), 737-767.*

Bröder, A., & Eichler, A. (2006). The use of recognition information and additional cues in inferences from memory. *Acta Psychologica, 121, 275–284.*

Broll U, & Eckwert B (2006). Transparency in the interbank market and the volume of bank intermediated loans. *International Journal of Economic Theory 2, 123–133*

Broll, U., Egozcue, M., Wong, W. K., & Zitikis, R. (2010). Prospect theory, indifference curves, and hedging risks. *Applied Mathematics Research Express, 2010(2), 142-153.*

Broll U & Wahl, J.E. (2006). Banking and the advantage of hedging. *Annals of Financial Economics 1, 24-41*

Brumelle, S., (1974). When does diversification between two investments pay? *Journal of Financial and Quantitative Analysis , 9: 473-483.*

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature, 34, 669–700.*

Dalal, J.A.(1983) On the use of a covariance function in a portfolio model. *Journal of Financial and Quantitative Analysis, 18, 2, 223-227.*

Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). Why recognition is rational: Optimality results on single-variable decision rules. *Judgment and Decision Making, 5, 216–229.*

De Giorgi, E., & Hens, T. (2006). Making prospect theory fit for finance. *Financial Markets and Portfolio Management, 20, 339–360.*

Dekel, E. (1989). Asset demand without the independence axiom. *Econometrica: Journal of the Econometric Society, 57, 163-169.*

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies, 22(5), 1915-1953.*

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, 55, *Management Science, 798-812.*

Dewatripont M, & Tirole J (1994). *The prudential regulation of banks.* Cambridge, Mass., London, England.

Drumwright, M.E. (1992). A demonstration of anomalies in evaluations of bundling. *Marketing Letters, 3, 311–321.*

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review, 115(1), 199-213.*

Dubra, J., & Echenique, F. (2004). Information is not about measurability. *Mathematical Social Sciences, 47(2), 177-185.*

Edwards, W. (1954a). The theory of decision making. *Psychological Bulletin, 51, 380-417.*

Edwards, W. (1954b). Probability-preferences among bets with differing expected values. *American Journal of Psychology, 67(1), 56-67.*

Edwards, W. (1955). The prediction of decisions among bets. *Journal of Experimental*

Psychology, 50(3), 201.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological review,69(2),109.*

Egozcue, M. (2012). Gains from diversification: a regret theory approach. *Economics Bulletin, 32 (1), 204-219.*

Egozcue, M., Fuentes Garcıa, L. & Wong, W. K. (2009). On some covariance inequalities for monotonic and non-monotonic functions. *Journal of Inequalities in Pure and Applied Mathematics,10, 1–7.*

Egozcue, M., Fuentes García, L., Wong, W. K., & Zitikis, R. (2010). Gruss-Type Bounds for the Covariance of Transformed Random Variables. *Journal of Inequalities and Applications, 2010 (1).*

Egozcue, M., Fuentes García, L., Wong, W. K., & Zitikis, R. (2011). Do investors like to diversify? A study of Markowitz preferences. *European Journal of Operational Research, 215(1), 188-193.*

Egozcue, M, Fuentes García, L., Katsikoupolos, K. & Smithson, M. (2013) A mathematical analysis of the two levels recognition heuristic (mimeo).

Egozcue, M. & Wong, W.K. (2010). Segregation and integration: a study of the behaviors of investors with extended value functions. *Advances in Decision Sciences, 2010, Article ID 302895, 8 pages.*

Eichner, T. (2008). Mean variance vulnerability. *Management Science, 54, 586–593.*

Eichner, T., & Wagener, A. (2009). Multiple risks and mean-variance preferences. *Operations Research, 57, 1142–1154.*

Erdfelder, E., Küpper-Tetzel, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making, 6, 7–22.*

Esary, J. D., Proschan, F., & Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics, 38(5), 1466-1474.*

Fama, E., 1970. Effcient capital markets: a review of theory and empirical work. *Journal of Finance 25, 383–417.*

Farley, J. U. (1964). An optimal plan for salesmen's compensation. *Journal of Mar-*

keting Research, 1 (2), 39-43.

Feder, G., Just, R. E., & Schmitz, A. (1980). Futures markets and the theory of the firm under price uncertainty. *Quarterly Journal of Economics, 94(2), 317-328.*

Fischer, J. E., Steiner, F., Zucol, F., Berger, C., Martignon, L., Bossart, W., Altwegg, M., & Nadal, D. (2002). Use of simple heuristics to target macrolide prescription in children with community-acquired pneumonia. *Archive of Pediatrics and Adolescent Medicine, 156, 1005-1008.*

Fishburn, P. C. (1982). Nontransitive measurable utility. *Journal of Mathematical Psychology, 26(1), 31-67.*

Fishburn P.C. & Luce, R.D. (1995). Joint receipt and Thaler's hedonic editing rule. *Mathematical Social Science, 29, 33–76.*

Freixas X., & Rochet J.C. (2008). *Microeconomics of banking.* Cambridge, Mass., London, England, 2/e.

Friedman, Milton (1969). The Optimum Quantity of Money. In *The Optimum Quantity of Money and Other Essays. Chicago: Aldine, 1969,* pp. *1-50*

Friedman, M., & Savage, J.L. (1948). The utility analysis of choices involving risk. *Journal of Political Economy, 56, 279–304.*

Gaissmaier, W., & Marewski, J. N. (2011). Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. *Judgment and Decision Making, 6(1), 73-88.*

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62, 451-482.*

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making, 6, 100–121.*

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: the adaptive toolbox.* London: MIT Press

Gigerenzer, G., Todd, P. M., & The ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gigerenzer, G., & Todd, P. M. (2008). Rationality the fast and frugal way: introduction. *Handbook of Experimental Economics Results, 1, 976-986.*

Gilboa, I. (2010). *Making better decisions: decision theory in practice.* Wiley-Blackwell, Chichester.

Gillen, B. & Markowitz, H.M. (2009). A taxonomy of utility functions. In: *Variations in economic analysis: Essays in honor of Eli Schwartz* (Eds.: J.R. Aronson, H.L. Parmet, and R.J. Thornton). Springer, New York.

Gilovich, T., & Griffin, D. (2002). Introduction-heuristics and biases: Then and now, 1-18. in Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: the psychology of intuitive judgment.* Cambridge University Press.

Goldstein, D. & Gigerenzer, G. (1999) The recognition heuristic: How ignorance makes us smart. in Gigerenzer, Gerd Todd, Peter M. , (1999). Simple heuristics that make us smart. Evolution and cognition., (pp. 37-58). New York, NY, US: Oxford University Press, The ABC Research Groupxv.

Goldstein, D. G. & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109, 75–90.*

Goldstein, D. G. & Gigerenzer, G. (2011). The beauty of simple models: Themes in recognition heuristic research. *Judgment and Decision Making, 6, 5, 392–395.*

Gollier, C. (2004). *The economics of risk and time.* The MIT Press.

Goodman,L.A. & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association,49, 732–764.*

Graham, B. (1949). *The intelligent investor: The definitive book on value investing.* HarperBusiness.

Gurland, J. (1967). The Teacher's Corner: An Inequality Satisfied by the Expectation of the Reciprocal of a Random Variable. *American Statistician, 21(2), 24-25.*

Hadar, J., & Russell, W. R. (1971). Stochastic dominance and diversification. *Journal of Economic Theory, 3(3), 288-305.*

Hadar, J., & Russell, W. R. (1974). Diversification of interdependent prospects. *Journal of Economic Theory, 7(3), 231-240.*

Hadar, J., Russell, W.R., & Seo, T.K., (1977). Gains from diversification. *Review of*

Economic Studies, 44 (2), 363-368.

Hanoch, G., & Levy, H. (1969). The efficiency analysis of choices involving risk. *The Review of Economic Studies, 36(3), 335-346.*

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica: Journal of the Econometric Society, 62, 6, 1251-1289.*

Harinck, F., Van Dijk, E., Van Beest, I., & Mersmann, P. (2007). When gains loom larger than losses reversed loss aversion for small amounts of money. *Psychological Science, 18(12), 1099-1105.*

Hauser, J. (2011). A marketing science perspective on recognition-based heuristics (and the fast-and-frugal paradigm). *Judgment and Decision Making, 6, 5, 396–408.*

Heath, T.B., Chatterjee, S. & France, K.R. (1995). Mental accounting and changes in price: the frame dependence of reference dependence. *Journal of Consumer Research, 22, 90–97.*

Hens, T., & Rieger, M. O. (2010). *Financial economics: A concise introduction to classical and behavioral finance.* Springer.

Hershey, J. C., & Schoemaker, P. J. (1980). Risk taking and problem context in the domain of losses: An expected utility analysis. *Journal of Risk and Insurance, 46, 111-132.*

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34, 1191-1206*

Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision Making, 6, 58–72.*

Hilbig, B. E. (2010). Precise models deserve precise measures: A methodological dissection. *Judgment and Decision Making, 5, 300–309.*

Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010). One-reason decision-making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 123–134.*

Hilbig, B. E., & Pohl, R. F. (2008). Recognizing users of the recognition heuristic. *Experimental Psychology, 55, 394–401.*

Hilbig, B. E., & Pohl, R. F. (2009). Ignorance- vs. evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 1296–1305.*

Hilbig, B. E., Pohl, R. F., & Bröder, A. (2009). Criterion knowledge: A moderator of using the recognition heuristic? *Journal of Behavioral Decision Making, 22, 510–523.*

Hillier, F. S., Lieberman, G. J., & Hillier, M. (1990). *Introduction to operations research (Vol. 6).* New York: McGraw-Hill.

Hildreth, C. (1974). Expected utility of uncertain ventures. *Journal of the American Statistical Association, 69(345), 9-17.*

Hildreth, C., & Tesfatsion, L. (1977). A note on dependence between a venture and a current prospect. *Journal of Economic Theory, 15(2), 381-391.*

Hirst, D. E., Joyce, E. J., & Schadewald, M. S. (1994). Mental accounting and outcome contiguity in consumer-borrowing decisions. *Organizational Behavior and Human Decision Processes, 58(1), 136-152.*

Hoffrage, U. (1995). Zur Angemessenheit subjektiver Sicherheits-Urteile. Eine Exploration der Theorie der probabilistischen mentalen Modelle. (The adequacy of subjective confidence judgments: Studies concerning the theory of probabilistic mental models). Doctoral dissertation, University of Salzburg, Austria.

Hoffrage, U. (2011). Recognition judgments and the performance of the recognition heuristic depend on the size of the reference class. *Judgment and Decision Making, 6, 43–57.*

Holthausen, D.M. (1979). Hedging and the competitive firm under price uncertainty hedging and the competitivefirm under price uncertainty. *American Economic Review, 69, 989–995.*

Huberman, G., & Jiang, W. (2006). Offering versus choice in 401 (k) plans: Equity exposure and number of funds. *Journal of Finance, 61(2), 763-801.*

Jarnebrant, P., Toubia, O. & Johnson, E. (2009). The silver lining effect: formal analysis and experiments. *Management Science, 55, 1832–1841.*

Kahneman, D. (2011) *Thinking, Fast and Slow.* New York: Farrar, Straus & Giroux.

Kahneman, D., Tversky, A. (1979). Prospect theory of decisions under risk. *Econometrica, 47, 263–291.*

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

Katsikopoulos, K. V. (2010) The less-is-more effect: Predictions and tests. *Judgment and Decision Making, 5, 244–257.*

Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice, *Decision Analysis, 8, 10-29.*

Klugman, S.A., Panjer, H.H. & Willmot, G.E. (2008). *Loss models: from data to decisions.* (Third edition.) Wiley, New Jersey.

Kira, D., & Ziemba, W. T. (1980). The demand for a risky asset. *Management Science, 26(11), 1158-1165.*

Köbberling, V. & Wakker, P.P. (2005). An index of loss aversion. *Journal of Economic Theory, 122, 119–131.*

Kőszegi, B., & Szeidl, A. (2013). A model of focusing in economic choice. *Quarterly Journal of Economics, 128(1), 53-104.*

Kuczma, M. (2009). *An introduction to the theory of functional equations and inequalities. Cauchy's equation and Jensen's inequality.* Second edition. Birkhäuser, Basel.

Laciana, C. E., & Weber, E. U. (2008). Correcting expected utility for comparisons between alternative outcomes: A unified parameterization of regret and disappointment. *Journal of Risk and Uncertainty, 36(1), 1-17.*

Landsberger, M., & Meilijson, I. (1990). Demand for risky financial assets: A portfolio analysis. *Journal of Economic Theory, 50(1), 204-213.*

Lehenkari, M. (2009). The hedonic editing hypothesis: Evidence from the Finnish stock market. *Journal of Behavioral Finance, 10(1), 9-18.*

Lehmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics, 37, 1137–1153.*

Levy, H. (Ed.). (2006). *Stochastic dominance: investment decision making under uncertainty (Vol. 12)*. Springer.

LiCalzi, M. (2000). Upper and lower bounds for expected utility. *Economic Theory, 16(2), 489-502.*

Lim, S.S. (2006). Do investors integrate losses and segregate gains? Mental accounting and investor trading decisions. *Journal of Business, 79, 2539–2573.*

Linville, P. W., & Fischer, G. W. (1991). Preferences for separating or combining events. *Journal of Personality and Social Psychology, 60(1), 5.*

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92(368), 805-824.*

Loomes, G., & Sugden, R. (1987). Some implications of a more general form of regret theory. *Journal of Economic Theory, 41(2), 270-287.*

Loomes, G., Starmer, C., & Sugden, R. (1992). Are preferences monotonic? Testing some predictions of regret theory. *Economica, 59, 17-33.*

Loubergé, H., & Outreville, J. F. (2001). Risk taking in the domain of losses: Experiments in several countries. *Journal of Risk Research, 4(3), 227-236.*

Ma, C. (2010). *Advanced Asset Pricing Theory (Vol. 2)*. World Scientific.

Marewski, J. N., Pohl, R. F., & Vitouch, O. (2010). Recognition-based judgments and decisions: Introduction to the special issue (I). *Judgment and Decision Making, 5, 4, 207-215*

Marewski, J. N., Pohl, R. F., & Vitouch, O. (2011a). Recognition-based judgments and decisions: Introduction to the special issue (II). *Judgment and Decision Making, 6, 1–6.*

Marewski, J. N., Pohl, R. F., & Vitouch, O. (2011b). Recognition-based judgments and decisions: What we have learned (so far). *Judgment and Decision Making, 6, 359–380.*

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy, 60, 151–156.*

McEntire, P. L. (1984). Portfolio theory for independent assets. *Management Science,* *30(8), 952-963.*

McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological science, 21(10), 1438-1445.*

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance, 42(3), 483-510.*

Meyer, J. (1987). Two-moment decision models and expected utility maximization. *American Economic Review, 77, 421–430.*

Meyer, J., & Robison, L. J. (1988). Hedging under output price randomness. *American Journal of Agricultural Economics, 70(2), 268-272.*

Milkman, K. L., Mazza, M. C., Shu, L. L., Tsay, C. J., & Bazerman, M. H. (2012). Policy bundling to overcome loss aversion: A method for improving legislative outcomes. *Organizational Behavior and Human Decision Processes, 117(1), 158-167.*

Mitrinovic, D. S., & Vasic, P.(1970). *Analytic Inequalities.* Springer-Verlag Berlin, Heidelberg. New York.

Monti, M., Boero, R., Berg, N., Gigerenzer, G., & Martignon, L. (2012). How do common investors behave? Information search and portfolio choice among bank customers and university students. *Mind & Society, 11(2), 203-233.*

Mossin, J. (1968). Aspects of rational insurance purchasing. *Journal of Political Economy, 553-568.*

Muermann, A., Mitchell, O. S., & Volkman, J. M. (2006). Regret, portfolio choice, and guarantees in defined contribution schemes. *Insurance: Mathematics and Economics, 39(2), 219-229.*

Mulaudzi, M. P., Petersen, M. A., & Schoeman, I. M. (2008). Optimal allocation between bank loans and treasuries with regret. *Optimization Letters, 2(4), 555-566.*

Neilson, W. S. (2002). Comparative risk sensitivity with reference-dependent preferences. *Journal of Risk and Uncertainty, 24(2), 131-142.*

Newell, B. R. & D. Fernandez (2006). On the binary quality of recognition and the inconsequentiality of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19, 333–346.*

Newell, B. R. & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 923–935.*

Oeusoonthornwattana, O., & Shanks, D. R. (2010). I like what I know: Is recognition a noncompensatory determiner of consumer choice? *Judgment and Decision Making, 5, 310–325.*

Orbitz (2012). *Vacation Packages: Packages Save You Moolah.* Viewed on 25 June 2012 at the web address `http://www.orbitz.com/shop/packageSearch?type=ahc`

Ormiston, M. B., & Schlee, E. E. (2001). Mean-Variance Preferences and Investor Behaviour. *Economic Journal, 111(474), 849-861.*

Ortmann, A., Gigerenzer, G., Borges, B., & Goldstein, D. G. (2008). The recognition heuristic: A fast and frugal way to investment choice? In C. R. Plott & V. L. Smith (Eds.), *Handbook of experimental economics results: Vol. 1 (Handbooks in Economics No. 28) (pp. 993–1003).* Amsterdam: North-Holland.

Oppenheimer, H. R. (1984). A test of Ben Graham's stock selection criteria. *Financial Analysts Journal, 40, 5, 68-74.*

Oppenheimer, D. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition, 90, B1–B9.*

Oppenheimer, H. R., & Schlarbaum, G. G. (1981). Investing with Ben Graham: an ex ante test of the efficient markets hypothesis. *Journal of Financial and Quantitative Analysis, 16(03), 341-360.*

Pachur, T. (2010). Recognition-based inference: When is less more in the real world? *Psychonomic Bulletin and Review, 17, 589-598.*

Pachur, T. (2011). The limited value of precise tests of the recognition heuristic. *Judgment and Decision Making, 6, 413–422.*

Pachur, T. & Biele, G. (2007). Forecasting from ignorance: the use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica, 125, 99- 116.*

Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: is recognition a non-compensatory cue?. *Journal of Behavioral Decision Making, 21(2), 183-210.*

Pachur, T., Mata, R., & Schooler, L. J. (2009). Cognitive aging and the adaptive use of recognition in decision making. *Psychology and Aging, 24(4), 901.*

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J. & Goldstein, D. G. (2012). When is the recognition heuristic an adaptive tool? In: *P. Todd, G.Gigerenzer, & the ABC Research Group, Ecological rationality: Intelligence in the world. (pp. 113-143)* New York: Oxford University Press.

Paroush, J., & Venezia, I. (1979). On the theory of the competitive firm with a utility function defined on profits and regret. *European Economic Review, 12(3), 193-202.*

Payne J. W., Bettman, & Johnson, E. J. (1993). *The adaptive decision maker*, Cambridge: Cambridge University Press.

Pellerey, F., & Semeraro, P. (2005). A note on the portfolio selection problem. *Theory and Decision, 59(4), 295-306.*

Pennings, J. M., & Smidts, A. (2003). The shape of utility functions and organizational behavior. *Management Science, 49(9), 1251-1263.*

Pleskac, T. J. (2007). A signal detection analysis of the recognition heuristic. *Psychonomic Bulletin and Review, 14, 379–391*

Pohl, R. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19, 251- 271.*

Pohl, R. (2011). On the use of recognition in inferential decision making: An overview of the debate. *Judgment and Decision Making, 6, 5, 423–438.*

Pólya, G. (1954). *Mathematics and plausible reasoning: vol 1: Induction and analogy in mathematics.* Oxford University Press.

Post, T., & Levy, H. (2005). Does risk seeking drive stock prices? A stochastic dominance analysis of aggregate investor preferences and beliefs. *Review of Financial Studies, 18(3), 925-953.*

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica: Journal of the Econometric Society, 122-136.*

Quiggin, J. (1994). Regret theory with general choice sets. *Journal of Risk and Uncertainty 8(2),153–165.*

Reilly, R. J. (1986). The Markowitz Utility Function and Some Experimental Evidence for Small Speculative Risks. *Journal of Risk and Insurance, 53(4), 724-733.*

Richter, T., & Späth, P. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 150–162.*

Ross, S.A. (1981). Some stronger measures of risk aversion in the small and the large with applications. *Econometrica 49,621-638.*

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review, 5(4), 296-320.*

Sahoo, P., & Riedel, T. (1998). *Mean value theorems and functional equations.* World Scientific Publishing Company.

Samuelson, P.(1967). General Proof that Diversification Pays. *Journal of Financial and Quantitative Analysis 2, 1-13*

Sandmo, A. (1971). On the theory of the competitive firm under price uncertainty. *American Economic Review, 65-73.*

Sarver, T. (2008). Anticipating regret: why fewer options may be better. *Econometrica, 76, 2, 263–305.*

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical association, 46(253), 55-67.*

Scheibehenne, B. & A. Bröder (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting 23, 415–426*

Schmidt, K. D. (2003). On the covariance of monotone functions of a random variable. Professoren des Inst. für Math. Stochastik.

Schmidt, U. & Zank, H. (2008). Risk aversion in cumulative prospect theory. *Management Science, 54, 208–216.*

Schwartz, H. (2010) Heuristics or rules of thumb, in *Behavioral finance: investors, corporations, and markets (eds H. K. Baker and J. R. Nofsinger)*, John Wiley & Sons, Inc., Hoboken, NJ, USA.

Serwe, S., & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making, 19(4), 321-332.*

Shefrin, H. (2008). *A behavioral approach to asset pricing.* Academic Press.

Shefrin, H. M., & Statman, M. (1984). Explaining investor preference for cash dividends. *Journal of Financial Economics, 13(2), 253-282.*

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69, 99–118.*

Simon, C. P., & Blume, L., (1994). *Mathematics for economists.* Vol. 7. New York: Norton.

Simonovits, A. (1995). Three economic applications of Chebyshev's Algebraic Inequality. *Mathematical Social Sciences 30, 207–220.*

Sinn, H.-W. (1983). *Economic decisions under uncertainty.* North-Holland, Amsterdam.

Smithson, M. (2010). When less is more in the recognition heuristic. *Judgment and Decision Making, 5, 230–244.*

Snook, B., & Cullen, R. M. (2006). Recognizing National Hockey League greatness with an ignorance-based heuristic. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 60(1), 33.*

Snook, B., Zito, M., Bennell, C., and Taylor, P. J. (2005). On the complexity and accuracy of geographic profiling strategies, *Journal of Quantitative Criminology, 21, 1-16.*

Solnik, B. (2008). Equity home bias and regret: an international equilibrium model. Working paper HEC, Paris.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38(2), 332-382.*

Steffensen, J.F.,(1925). En Ulighed mellem Middelve edier. *Matematisk Tidsskrift 49–53.*

Stiglitz, J. (1998). The private uses of public interests: Incentives and institutions. *Journal of Economic Perspectives, 12, 3-22.*

Sugden, R. (1993). An axiomatic foundation for regret theory. *Journal of Economic*

Theory, 60(1), 159-180.

Sugden, R. (2004). Alternatives to Expected Utility: Foundations, *in Handbook of Utility Theory: Volume 2: Extensions, eds Salvador Barbera, Peter Hammond and Christian Seidl, Kluwer, Dordrecht, pp. 685-755.*

Sul, S., Kim, J., & Choi, I. (2013). Subjective Well-Being and Hedonic Editing: How Happy People Maximize Joint Outcomes of Loss and Gain. *Journal of Happiness Studies, 14, 1409-1430.*

Taylor, J. B. (1993). Discretion versus policy rules in practice. In Carnegie-Rochester conference series on public policy (Vol. 39, pp. 195-214). North-Holland.

Tesfatsion, L. (1976). Stochastic dominance and the maximization of expected utility. *Review of Economic Studies, 43(2), 301-315.*

Thaler, R. (1980). Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization 1, 39–60.*

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science, 4, 199–214.*

Thaler, R. H. (1994). *Quasi rational economics.* Russell Sage Foundation.

Thaler, R.H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12, 183–206.*

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Thoma, V., & Williams, A. (2013). The devil you know: The effect of brand recognition and product ratings on consumer choice. *Judgment and Decision Making, 8(1), 34-44.*

Tobin, J. (1958). Liquidity preference and behavior towards risk. *Review of Economic Studies, 25, 65–86.*

Todd, P.M.& G. Gigerenzer (2003) Bounding rationality to the world. *Journal of Economic Psychology 24, 143-165.*

Tomlinson, T., Marewski, J. N., & Dougherty, M. (2011). Four challenges for cognitive research on the recognition heuristic and a call for a research strategy shift. *Judgment* *and Decision Making, 6, 89–99.*

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5, 297–323.*

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185, 1124 –1130.*

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review, 204-217.*

Tversky, A., & Thaler, R. H. (1990). Anomalies: preference reversals. *Journal of Economic Perspectives, 4(2), 201-211.*

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton University Press..

Wagener, A. (2006). Chebyshev's algebraic inequality and comparative statics under uncertainty. *Mathematical Social Sciences, 52, 217–221.*

Wakker, P.P. (2010). *Prospect theory: for risk and ambiguity.* Cambridge University Press, Cambridge.

Wright, R. (1987). Expectation dependence of random variables, with an application in portfolio theory. *Theory and Decision 22, 111-124.*

Wong, K.P., (2011). Regret theory and the banking firm: The optimal bank interest margin. *Economic Modelling 28, 2483–2487.*

Wübben, M., & Wangenheim, F. V. (2008). Instant customer base analysis: Managerial heuristics often "get it right". *Journal of Marketing, 72(3), 82-93.*

Zaleskiewicz, T. (2011). Financial forecasts during the crisis: Were experts more accurate than laypeople? *Journal of Economic Psychology 32, 384–390.*

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making, 12, 93–106.*

Zeelenberg, M., & Beattie, J. (1997). Consequences of regret aversion 2: Additional evidence for effects of feedback on decision making. *Organizational Behavior and Human Decision Processes, 72(1), 63-78.*

Zeelenberg, M., Beattie, J., van der Pligt, J., & de Vries, N. K. (1996). Consequences of regret aversion: effects of expected feedback on risky decision making. *Organizational Behavior and Human Decision Processes, 65,* 148–158.