Doctoral Thesis

University of Trento

School of Social Sciences

Doctoral School in Economics and Management

Subjective Probabilities in Choice Experiments' Design: Three Essays

a dissertation submitted to the doctoral school of economics and management in partial fulfillment of the requirements for the Doctoral degree (Ph.D.) in Economics and Management

Simone Cerroni

December 2012

Advisors:

Advisor:

Prof. Sandra Notaro, University of Trento.

Co-Advisor:

Prof. W. Douglass Shaw, Texas A&M University

Doctoral Committee:

Prof. Giovanni Signorello, University of Catania

Prof. Jürgen Meyerhoff, University of Berlin

ABSTRACT

This dissertation which consists of three essays investigates the influence of subjective probabilities on decision making processes under conditions of risk. In particular, it examines whether subjects adjust new risk information on their prior subjective estimates, and, to what extent this adjustment affects their choices.

In the first essay, by using an artefactual field experiment, I examine the potential correlation between incentive compatibility and validity of subjective probabilities elicited via the Exchangeability Method, an innovative elicitation mechanism which consists of several chained questions. Here, validity is investigated using de Finetti's notion of coherence under which subjective probabilities are coherent if and only if they obey all axioms and theorems of probability theory. Experimental results suggest that subjects provided with monetary incentives and randomized questions more likely express valid subjective probabilities than others because they are not aware of the chaining which undermines the incentive compatibility of the Exchangeability Method.

In the second essay, by using the same experimental data, I show that valid subjective probabilities do not significantly diverge from invalid ones, indicative of little effect of internal validity on the actual magnitude of subjective probabilities.

In the third essay, by using a field Choice Experiment, I investigate to what extent subjects adjust risk information given in the status quo alternative on their subjective probability estimates. An innovative two-stage approach that incorporates subjective probabilities into Choice Experiments' design is developed to investigate this phenomenon, known as the scenario adjustment. In the first stage, subjective probabilities that given outcomes will occur are elicited using the Exchangeability Method. In the second stage, two treatment groups are designed: in the first group, each subject is presented with a status quo alternative which incorporates her/his subjective probabilities, and, hence, no adjustment is required; in the second group, each subject

faces a status quo alternative where the presented risk is not consistent with her/his probability estimates, and, hence, a mental adjustment to the scenario might take place. By comparing willingness to pay across the treatment groups, my results suggest that, when subjects are provided with SQ alternatives in which the risk is lower than the perceived one, the mental adjustment takes place, but, when subjects are provided with SQ alternatives in which the risk is higher than their own estimates, these subjects appear to make irrational choices.

Keywords: subjective probability; discrete choice modeling; exchangeability method; choice experiment; apple; pesticide.

ACKNOWLEDGEMENTS

I would like to thank my committee co-chairs, Dr. Notaro and Dr. Shaw, and my committee members, Dr. Signorello and Dr. Meyerhoff. I would like to thank Dr. Raffaelli for her support during these years. Thanks also go to Marco Tecilla for his help in organizing sessions and running the experiment at the Computable and Experimental Economics Laboratory (CEEL) at University of Trento. I appreciate comments on my essays from Dr. Ilaria Pertot, Dr. Matteo Ploner, Dr. Mary Riddel, and Dr. Richard Woodward. This research was funded by Autonomous Province of Trento, project ENVIROCHANGE, Call for Major Project 2006.

TABLE OF CONTENTS

LIST OF TABLE 8

LIST OF FIGURES 9

LIST OF EQUATIONS 9

CHAPTER I . INTRODUCTION 10

CHAPTER II. ELICITING AND ESTIMATING VALID SUBJECTIVE PROBABILITIES: AN EXPERIMENTAL INVESTIGATION OF THE EXCHANGEABILITY METHOD 13

Introduction 13

Methods for eliciting subjective probabilities 16

Specific Objectives 21

The Experimental Design 22

The empirical application 22

The Exchangeability Method and the related game 23

Other games 25

The sample 26

Treatments 27

Hypotheses about the validity of subjective probabilities 28

Testing hypotheses 30

Results 33

Non-parametric tests 33

The validity rate 38

The econometric analysis 42

Conclusion 48

CHAPTER III. THE IMPACT OF INTERNAL VALIDTY ON SUBJECTIVE PROBABILITIES 50

Introduction 50

Subjective probabilities and pesticide residues 53

Objectives 59

The empirical application 59

The case study 59

The dataset 60

The sample 60

Methods 61

The elicitation of subjective probabilities: the Exchangeability Method *61*
The validity of subjective probabilities: the Repeated Exchangeability Game and Certainty Equivalent Game *63*
Results 64
The analysis of subjective probabilities *64*
The difference between valid and not valid subjective probabilities *68*
Factors shaping subjective probabilities *70*
Conclusion 78
CHAPTER IV. THE INCORPORATION OF SUBJECTIVE PROBABILITIES INTO CHOICE EXPERIMENTS TO TEST SCENARIO REJECTION 80
Introduction 80
Literature review 83
Realism and scenario adjustment *83*
Elicitation of Subjective probabilities *87*
Objectives and testable hypotheses 90
The empirical application 92
Methodology 93
Exchangeability Method *93*
Best-worst pivot choice experiment *95*
Experimental treatments and sampling procedure *99*
Modeling, estimation, and welfare measures 104
Selection Model *104*
Discrete Choice Modeling *106*
Estimation procedure and welfare measures *110*
Results 112
Selection Model *112*
Discrete choice models *115*
Willingness to pay *120*
Hypothesis testing *123*
Conclusions 124
CHAPTER V. CONCLUSIONS 126
REFERENCES 128
APPENDIX A 140
APPENDIX B 142

LIST OF TABLE

Table 2.1 McNemar's test of consistency 31

Table 2.2 Summary statistics of median values obtained via EG (X1/2) and REG ($X1_{/2}$') 34

Table 2.3 Results at sample level obtained via EG ($X_{1/2}$) and REG ($X_{1/2}$')................... 35

Table 2.4 Summary statistics of the Certainty Equivalents obtained via CEG 36

Table 2.5 Results at sample level obtained via the CEG.. 37

Table 2.6 Percentage of rewarded subjects based on their answers to Question 3......... 38

Table 2.7 Validity rates (*V*) for all treatments.. 39

Table 2.8 Percentage of probabilities per level of invalidity in each treatment 42

Table 2.9 Description of dependent and independent variables of Model 1 44

Table 2.10 Generalized Linear Model Estimation of Models 1and 2 47

Table 3.1 Summary statistics of percentile estimates for each variable 66

Table 3.2 Average values of variable g, a, and r considering valid and invalid observations 68

Table 3.3 Comparison of valid and invalid percentile estimates at sample level........... 69

Table 3.4 Comparison of valid and invalid percentile estimates at individual level...... 70

Table 3.5 Description of independent variables of Model 1,2,and 3 71

Table 3.6 Generalized Linear Model Estimation 77

Table 4.1 Attribute levels for the SQ.. 97

Table 4.2 Attribute level for R&D plans 98

Table 4.3 Choice Task 1 for subject *i* in the SSQ treatment 100

Table 4.4 Choice Task 1 for subject *i* in the OSQ_{LOW} treatment................................. 102

Table 4.5 Choice Task 1 for subject *i* in the OSQ_{HIGH} treatment 103

Table 4.6 Summary statistics of variables in the Subjective SQ treatment.................. 104

Table 4.7 Summary statistics of variables included in the selection model 105

Table 4.8 Logit Selection model.. 113

Table 4.9 Chi Squared Test for comparing coefficients of the OSQ_{LOW} and the OSQ_{HIGH} treatments 114

Table 4.10 Mixed Multinomial Logit estimation of discounted EUT models 119

Table 4.11 Marginal willingness to pay for risk reductions.. 122

Table 4.12 One sided t test for comparing marginal willingness to pay 124

LIST OF FIGURES

Figure 2.1 Scheme of the Exchangeability Game's bisection procedure 25

Figure 2.2 Histogram of the percentage of subjective probabilities per level of invalidity 42

Figure 3.1 The average number of days in which the infestation will occur during the blossoming period in 2030 (*a*), the average number of apples containing at least one residue in a sample of 100 apples in 2030 (*b*), and the average number of apples containing more than 1 residue in a sample of 100 apples in 2030 (*c*). 67

LIST OF EQUATIONS

Equation 2.1 43

Equation 3.1 72

Equation 4.1 107

Equation 4.2 108

Equation 4.3 108

Equation 4.4 108

Equation 4.5 111

Equation 4.6 111

Equation 4.7 112

CHAPTER I . INTRODUCTION

As many everyday choices involve future events that are surrounded by uncertainty, subjective probabilities strongly influence decision making processes (Manski, 2004). For example, households' probability estimates of future income have been shown to affect consumption and saving decisions, while students' probability estimates of returns to education to impact schooling choices (see Manski, 2004, for a review). Outside of the financial domain, subjective probabilities of future health outcomes have been demonstrated to affect the support for policies reducing mortality risk (Cameron et al., 2010), or consumption of food and bottled water (e.g., Viscusi and Evans, 1993; Williams and Hammit, 2001; Jakus et al., 2009; Shaw et al., 2012).

My dissertation which is presented in the form of three essays investigates the influence that subjective probabilities of having contaminated apples have on subjects' preferences for R&D programs that are geared to control the future spread of new apple diseases in the Province of Trento in Italy. More specifically, here, I explore strengths and limitations of a novel technique to elicit subjective probabilities, the Exchangeability Method (EM), and, more important, I develop an innovative approach which incorporates subjective probabilities into the design of choice experiments (CE).

In the first essay, by using an artefactual field experiment, I investigate subjective probabilities of having contaminated apples elicited via the EM (Baillon, 2008, Abdellaoui et al., 2011). Specifically, I explore whether incentive compatibility affects the validity of subjective probabilities elicited using this innovative technique which consists of several chained questions. As chained elicitation mechanism are not necessarily incentive compatible, four experimental designs which aim to enhance the EM's incentive compatibility are created. Afterwards, the validity of subjective probabilities elicited using each experimental design is investigated using de Finetti's notion of coherence, under which probability estimates are valid if and only if they obey all axioms of probability theory (de Finetti, 1937).

In the second essays, by drawing on the experimental investigation presented above, the potential discrepancy between valid and invalid subjective probabilities is examined to fully understand whether failure to recognize validity implies an over- or underestimation of consumers' probability estimates. Furthermore, a simple behavioral model is estimated to identify attitudinal and socio-economic factors that affect consumers' subjective probabilities that apples will contain pesticide residues.

In the third essay, by using a CE field survey, I investigate the influence that subjective probabilities of having contaminated apples have on preferences for alternative R&D programs that plan to control new apple diseases in the Province of Trento. Although subjects' are commonly pretended to fully accept the risk information provided by researchers in the status quo (SQ) alternative, here, I hypothesize that subjects adjust this information on their subjective probability estimates. This phenomenon, called scenario adjustment, might generate confounding factors that researchers cannot capture in their choice models, and, therefore, compromise the accuracy of willingness-to-pay (WTP) estimates (Cameron et al., 2010).

Here, the extent of this phenomenon and its impact on choice-behavior are investigated by using an innovative two-stage approach which compares WTP estimates elicited from subjects who might adjust the risk information given in the SQ on their subjective probabilities with WTP estimates of subjects who might not. In the first stage, subjective probabilities that given numbers of apples will contain pesticide residues are elicited by using the EM. In the second stage, WTP estimates for alternative R&D programs are elicited from subjects who belong to different treatment groups. In one treatment, each subject is presented with a SQ' risk level which differs from her/his subjective estimates. In the other treatment, each subject faces an individual-specific SQ in which the risk level is equal to her/his probability estimate. Subjective probabilities are incorporated into my CE by using a best-worst pivot experimental design. Pivot CE were developed in transport economics to generate SQ alternatives based on each subject's most recent driving experience (e.g., Hensher and Greene, 2003, Hensher et al., 2009), however, this investigation is the first attempt to use this technique for designing subject-specific SQ alternatives tailored on subjective probability estimates.

CHAPTER II. ELICITING AND ESTIMATING VALID SUBJECTIVE PROBABILITIES: AN EXPERIMENTAL INVESTIGATION OF THE EXCHANGEABILITY METHOD

Introduction

During the last two decades, many social scientists have become more interested in investigating and eliciting subjective probabilities of everyday events. The main reason to pursue this line of inquiry is because many choices in the real world involve future outcomes and take place under uncertainty. Hence, people often behave and make decisions according to their beliefs and expectations. Manski (2004) demonstrates the importance of subjective probabilities in several branches of applied economics, ranging from the influence of households' probabilistic income expectations on their consumption and saving decisions, to the impact of students' probabilistic expectations of the returns (again, in income terms) to education on schooling choices.

Expectations on risky and uncertain outcomes, which lie outside of the financial domain, are potentially complex, but also important to deal with. These have been neglected in economics until quite recently, perhaps, because they pertain to issues which are more difficult to address than financial risk and uncertainty, such as stock market activity. Early work on subjective probability pertained to another issue that is relatively simple to understand and for which outcomes are readily observable with short delays: the weather, specifically, temperature and precipitation forecasts (e.g., Brier, 1950; Baillon, 2008).

A domain where subjective probabilities have been recognized to be crucial in understanding and predict people's choice behavior is food safety, but little in this area has been done to explore subjective probability elicitation. Despite many studies have shown how consumers' probabilistic expectations of food safety might affect purchases (e.g., Buzby et al., 1998; Williams and Hammit, 2001), they often use very simple and rough methods for eliciting subjective probabilities, which often consist in directly asking subjects a guess of the probability that given outcomes will occur in the future. The key problem with issues such as food safety is that the nature of the uncertainty is less accessible to laypeople, and the primary outcome, the health effect, may be unobservable for quite some time to come[1]. However, a recent study suggests that uncertainty in food safety decisions may be quite important (Kivi and Shogren, 2010).

In this essay, I investigate and elicit consumers' perceptions of the probability that given levels of pesticide residues will be present in apples produced in the future in the Province of Trento (Italy). Pesticide residues pose health risks to people who eat apples, and, thus, people's perceptions of their presence can affect their preferences for agricultural policies that local authorities are planning to incentivize the production of healthy apples. The investigation of this topic might be very important to this region as apple production is a key sector of its economy (P.A.T., 2010). Generally, the presence of pesticides in food is quite important, as we all must eat; several studies have shown that human exposures to chemicals are associated with risks to human health, they may even produce very severe illnesses as cancer (Alavanja et al., 2004).

There are many different ways to elicit subjective probabilities and several are briefly discussed below. I use an innovative technique for eliciting probabilities, known as the Exchangeability Method (EM), recently used by Baillon (2008). He elicited subjective probabilities for future daily temperature in Paris, the euro/dollar exchange rate, and the daily variation of the French stock index CAC 40. His experimental subjects were asked to estimate these for a given day about four weeks after the experiment was conducted. The same technique was further developed by Abdellaoui et

[1] Short-term food sickness is perhaps observable after a short delay, but ethics in experiments preclude subjecting subjects to this.

al. (2011) to elicit subjective probabilities and investigate ambiguity attitudes related to similar topics[2].

The EM consists of a set of binary questions where subjects are asked to bet a certain amount of money on a given outcome rather than on an alternative outcome. In each question, the outcomes which are presented to the subject result from a bisection procedure of the whole state space of the random variable under study. When subjects become indifferent between the two outcomes, they are assumed to perceive both as equally likely and subjective probabilities can be estimated. The sequential splitting process behind the EM makes this elicitation procedure chained, in the sense that the outcomes presented in each question depends on the outcome that has been chosen in the previous one.

The incentive compatibility of the EM might be questioned because previous experimental studies have shown that chained elicitation mechanisms are not necessarily incentive compatible. In fact, the provision of monetary incentives to subjects, based on their choice behavior during the experiment, might induce them to not state their real beliefs, but, instead, to strategically behave to be better rewarded upon completion of the tasks for the experiment (e.g., Harrison, 1986).

In this essay, I investigate whether the lack of incentive compatibility of the EM due to both the presence of chained questions and no provision of real monetary incentives, affects the validity of subjective probabilities elicited by such a technique. I determine and measure the validity of subjective probabilities elicited via the EM implementing a method based on de Finetti’s notion of *coherence* (1937). By using this

[2] They elicited subjective probabilities related to the daily variation of the French stock index CAC 40, temperature in Paris and also in a randomly drawn remote country for a given day about 3 months after the experiment.

approach I essentially aim to identify the best way for eliciting subjective probabilities via the EM, in terms of validity[3].

The remainder of the essay is laid out as follows. I first highlight the main strengths and limitations of the EM by comparing it to other techniques for eliciting beliefs. Next, I describe my testable hypotheses and the methodology used to measure validity of subjective probabilities. Finally, I offer some conclusions based on the experimental results I have obtained.

Methods for eliciting subjective probabilities

The simplest way to elicit subjective probabilities consists of asking people to directly state the chance that a specific magnitude of the outcome will happen in the future (Spetzler and Stael Von Holstein, 1975). Asking simple, direct questions is common in a host of previous health-risk studies, such as those involving smoking cigarettes (e.g., Viscusi, 1990; Gerking and Khaddaria, 2011), drinking contaminated water (e.g., Jakus et al. 2009; Shaw et al., 2012), or eating unhealthy food (e.g., Buzby et al., 1998; Williams and Hammit, 2001).

However, unless subjects are asked to state a chance for each of all possible specific magnitudes of outcomes, the information gathered from such an easy question is very limited. Using a direct approach like this, I might learn about only one point, or about a very narrow range, in the individual's subjective probability distribution.

The reliability of subjective probabilities elicited via this family of techniques, called *direct methods*, have also been often questioned, particularly by psychologists, on the grounds that laypeople may be neither familiar with the notion of probability per se,

[3] Since this experiment is conducted in the lab, with a controlled environment and real monetary incentives, we only refer to the internal validity of elicited risk estimates. Hence, I cannot analyze the external validity of my results, being aware that elicited estimates in the lab might be different from those elicited in the field, where it is impossible to control for many confounding factors (for instance, background risk) (Harrison et al., 2007).

nor willing to put efforts into thinking in probabilistic terms (Manski, 2004)[4]. Some have gone as far to suggest that individuals do better in understanding risks with verbal, rather than numerical percentage or probability scales (see discussion in Weinstein and Diefenbach, 1997). Several economic studies provide supporting evidence that people have problems with open-ended questions about probability estimates (e.g., Jakus et al., 2009; Riddel and Shaw, 2006)

Other approaches, called *indirect methods*, may overcome some of the limitations that direct methods have. Here, probability measures are indirectly estimated at the points for which subjects show their indifference between choices involving lotteries or gambles. Indirect techniques have often been used for eliciting probabilities related to financial outcomes (e.g., Andersen et al., 2010; Offerman et al., 2009) because actual monetary payments for played-out bets make the elicitation mechanism incentive compatible and appear to be relatively easy for subjects to understand. Quite recently, a few scholars have used indirect methods to estimate subjective probabilities related to health and environmental outcomes (e.g., Fiore et al., 2009; Cerroni and Shaw, 2012). As noted in the introduction, the limited use of these *indirect methods*, for eliciting probabilities related to health and environmental outcomes, is due to the fact that very long term health and environmental outcomes cannot be played out at the end of experiments in the lab setting, thus again making incentive compatibility a potential issue. Fiore et al. (2009) and Cerroni and Shaw (2012) both rely on hypothetical portrayals of adverse forest impacts, and, in the former study, the authors explore the use of virtual forest fires in the experimental setting.

The most popular of the indirect methods are called "*external reference events*", in which subjects are asked to choose between a lottery characterized by an uncertain

[4] Many studies investigated different approaches for communicating probabilities to laypeople and, then, eliciting their best estimate (e.g., Gigerenzer and Hoffrage, 1995; Hammit and Graham, 1999; Corso et al., 2001).

event (*U*), whose probability needs to be estimated, and a lottery characterized by an external reference event (*K*), whose probability is known and disclosed to subjects. The probability of the known event (K) is often visually presented through probability wheels, scroll bars, or other visual aids such as risk ladders, grids, or pie charts, all of which have been tested as probability communication devices (e.g., Morgan and Henrion, 1990). Once subjects become indifferent between the two lotteries, the uncertain outcome (*U*) is assumed to have the same probability of occurrence of the familiar outcome (*K*), so that $P(U) = P(K)$ (Spetzler and Stael Von Holstein, 1975).

Although these techniques are widely used, they may produce biased probability estimates because they ask subjects to process two sources of uncertainty at the same time: the first relates to the uncertain outcome (*U*), the second relates to the external reference event (K). Previous experimental studies have shown that individual choices depend on the source of uncertainty that subjects have been asked to consider[5] (e.g., Kilka and Weber, 2001; Abdellaoui et al., 2011), and, hence, elicitation mechanisms, which combine diverse sources of uncertainty, may become too complex and generate biased subjective probabilities (Baillon, 2008).

Source dependence does not appear to be an issue within another class of indirect methods which use *internal events*. In these elicitation techniques, subjects deal with magnitudes of the outcomes, but not with their probabilities of occurrence. In fact, subjects are only asked to bet a certain amount of money on one of the several disjoint subspaces, in which the whole state space of the variable under study has been previously divided. When subjects become indifferent regarding betting on one disjoint subspace rather than on the others, subjects are assumed to perceive those subspaces as equally likely (Spetzler and Stael Von Holstein, 1975).

[5] Baillon (2008, p.77) defined a source of uncertainty as "…a set of events that are generated by a common mechanism of uncertainty".

The EM, which was first described by Raiffa (1968) and more recently implemented by Baillon (2008) and Abdellaoui et al. (2011), belongs to this latter class of probability elicitation techniques. In the specific case of the EM, each question gives subjects the chance to bet on one of two disjoint subspaces, as the whole state space of the random variable under study is sequentially divided using a bisection process. The subdividing procedure of the event space makes each binary question of the EM chained to the previous one. In fact, the sub-events that subjects face in each question depend on the sub-event that has been chosen in the precedent question.

As noted in the introduction, chained techniques for eliciting preferences or beliefs are perhaps not incentive compatible. Strategic behaviors might have strong impacts on elicited subjective probabilities (Harrison, 1986) and chained questions may propagate subjects' strategic choices made during the choice-tasks (e.g., Spetzler and Stael Von Holstein, 1975; Wakker and Deneffe, 1996). Previous investigations that rely on chained games and real monetary incentives have validated their results by using subjects' statements of unawareness about the presence of chaining in the games (Van de Kuilen et al., 2006; Abdellaoui et al., 2011).

Baillon (2008) dealt with this problem by randomizing the order of questions. The questions are not sequentially presented and, thus, the chaining is less transparent to subjects because they are no longer aware of the relationship between the disjoint subspaces they face in one question and the subspace they have chosen in the previous one. Developing this experimental design with randomized questions, one hopes that telling the truth becomes the simplest and most efficient strategy that subjects can use when they play the EM (Baillon, 2008).

The effect of real monetary incentives on the elicitation of subjective probabilities has been investigated in another recent application of the EM by Abdellaoui et al. (2011). After having tested that subjects were unaware of the chained structure of the

EM, they next compare subjective probabilities provided by two groups of subjects, one provided with monetary incentives and the other not. They did not find any substantial difference between subjective probabilities elicited from the two groups, but do not provide a logical explanation as to why subjects provided with money incentives should have greater or lower beliefs than others.[6]

In contrast, I argue that monetary incentive may affect the validity of subjective probabilities elicited via the EM depending on whether subjects are aware of the chaining or not. In particular, I believe that monetary incentives and the ordering of questions may affect the incentive compatibility of the Exchangeability Method and, therefore, the validity of subjective probabilities elicited by using this technique. Here, I don't want to confuse truth with validity, in fact, as reported below, incentive compatibility and validity are separate and distinct concepts. An elicitation mechanism is incentive compatible if subjects have an incentive to state their real beliefs (Vossler and Evans, 2009), while subjective probabilities are valid if and only if they obey all axioms and theorems of probability theory (de Finetti, 1937).

In this essay, I hypothesize that subjective probabilities elicited via incentive incompatible mechanisms, which induce subjects to not truly state their beliefs, are likely to be invalid, in the sense that they do not obey to axioms and theory of probability theory.

To test my predictions, I create a validation method based on the de Finetti's notion of coherent probability measures (1937; 1974a; 1974b) under which subjective probabilities are coherent if and only if they obey all axioms and theorems of probability theory. The choice of using the de Finetti's notion of coherence to define

[6] They found that probability distribution functions of median temperature in Paris in a given day for both groups are quite well calibrated with historical distribution of temperature in that particular day. In contrast, they found that probability distribution functions of median daily variation of the French stock index CAC 40 in a given day for both groups differ from historical distribution of CAC 40 daily variation in that particular day.

valid subjective probabilities relies on the fact that the EM is based on the assumption of exchangeability-based probabilistic sophistication (Chew and Sagi, 2006). That, in turn, is based on the idea of equal likelihoods of exchangeable events (de Finetti, 1937)

Specific Objectives

Previous applications of the Exchangeability Method have not directly investigated the effect of chaining on subject's choice-behaviors (see Baillon, 2008; Abdellaoui et al., 2011), but they have simply tried to avoid the use of the identifiable chained questions in their experimental designs. As noted before, this is due to the fact that previous experimental studies have shown how the provision of chained questions along with real monetary incentives make the elicitation mechanism incentive incompatible (Harrison, 1986).

In line with the above discussion, I hypothesize that subjective probabilities elicited via an incentive incompatible mechanism more likely turn out to be invalid. In particular, I hypothesize that subjective probabilities elicited via the EM, using sequential questions along with real monetary incentives are invalid because, when the chaining is clear to subjects, monetary incentives will encourage them to strategically behave. In contrast, when random questions are provided in the EM along with monetary incentives, subjective probabilities are valid because when the chaining is less transparent to subjects, monetary incentives induce them to state their real beliefs, or at least, to invest more cognitive effort into the elicitation process[7].

I also hypothesize that subjects provided with random questions will perform better than those provided with sequential questions in terms of validity, even in the absence of actual monetary rewards at the end of the experiment. I expect those who are not aware of the chaining will provide invalid subjective probabilities, but less so than

[7] I thank an anonymous reviewer for suggesting this possibility.

those who are aware of the chaining structure. My prediction is supported by the fact that questions related to the elicitation of the third quartile ask subjects to choose between two prospects that they have already ruled out in previous questions. The issue is evident to subjects when questions are sequentially ordered, while it is less transparent when they are randomly ordered. This may affect the validity of probabilities elicited using sequential questions as subjects may perceive questions to be meaningless and may invest less cognitive effort in playing the game.

To test my hypotheses, I first need to understand whether elicited subjective probabilities are valid or not. The empirical way I tested the validity of subjective probability elicited via the EM is described below.

The Experimental Design

The empirical application

My specific application consists of investigating uncertain outcomes related to fire blight, a bacterial disease that has threatened apple orchards in the Province of Trento, at least since 2003 (EMF, 2006). This phytopathology damages and kills apple plants resulting in substantial losses in the production of apples. The best available science predicts a future spread of the disease in apple orchards of the Province of Trento, since suitable climatic conditions for the biology of the bacterium *Erwinia amylovora* are likely to occur in the future (unpublished results by Edmund Mach Foundation).

Although Italian farmers currently control the fire blight by using pesticides, chemicals might be not efficient enough to prevent the future spread of this apple disease. Nevertheless, the future production of apples in the Province of Trento (around 420.000 tons at the present time) might not decrease if farmers start implementing new adaptation strategies against fire blight. However, the only strategy that is currently available to farmers is the introduction of new active principles, such as the antibiotic streptomycin that is currently forbidden by the Italian legislation, but that has been already used in U.S., Germany, Belgium and, Netherlands for controlling the fire blight (Németh, 2004).

In the context presented here, I focus on three diverse random variables: the percentage (or number) of days in which the infestation will occur during the blossoming period in 2030 (g)[8], the number of apples containing at least one residue in a sample of 100 apples in 2030 (a)[9], and the number of apples containing more than 1 residue in a sample of 100 apples in 2030 (r)[10]. These variables have been selected among many other possible measures of pest infestation, or apple contamination, after having interviewed approximately 20 focus group subjects.

The Exchangeability Method and the related game

Let a random variable under study in the EM game (EG) be g. The EG uses a series of binary questions to reveal an individual's underlying cumulative distribution function (CDF) over an event x that is drawn from an event space, $S_G = G_1^1$. The first step of the EG establishes the lower and upper bounds of the event space, defined as g_0 and g_1. Each subject is asked the bounds for outcomes outside of which they are essentially certain the outcome cannot happen at all — i.e., the bounds that pertain to a non-zero probability of an outcome.

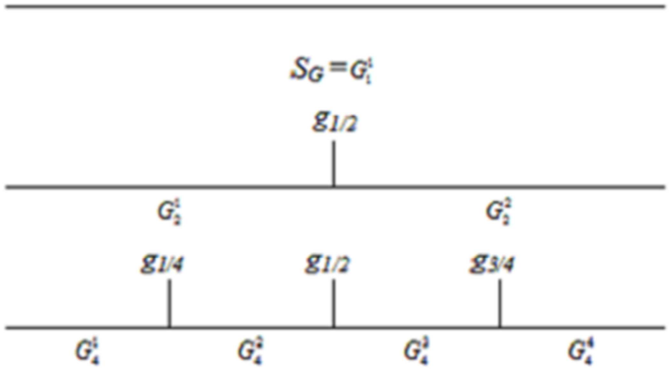
The second step of the EG involves asking a series of questions that establish the value of $g_{1/2} \in S_G$ that corresponds with the 50th percentile of the subjective CDF, in other words, the median estimate. This series of questions asks the subject to choose between binary prospects. In the first binary question, S_G is divided at a point g_a into two prospects, say $G_a=\{g_0<x<g_a\}$ and $G_a'=\{g_a \leq x<g_1\}$, where $g_a=\{g_0 + [(g_1-g_0)/2]\}$. If

[8] The blossoming period usually occurs in April in Trentino.
[9] This is the number of apples at least one residue beyond the level of 0 mg/kg.
[10] This is the number of apples containing at least two residues beyond the level of 0 mg/kg.

G_a was chosen by the individual, the implication is that the individual believes the probability of occurrence of the sub-event G_a is equal to that of the sub-event G_a', so that $P(G_a) \geq P(G_a')$ and $g_a \geq g_{1/2}$. A follow-up binary question is then asked of this same individual, using a new value g_b and two new prospects G_b and G_b'. If G_a was chosen in the first question, then $g_a > g_b$. However, if G_a' was chosen in the first question, then $g_a < g_b$. This process is repeated until the individual reaches a value g_z such that she is indifferent between G_z and G_z'. When this point is reached, it follows that $g_z = g_{1/2}$, $G_z = G_2^1$, $G_z' = G_2^2$, and $P(G_z) = P(G_z')$. This process describes the "chaining" or interdependence of these binary outcome questions.

A similar process can be followed to determine other points for the individual's subjective CDF; in theory as many as the researcher wants to identify. However, there is a limit to how many separate points can be elicited because of potential exhaustion of the subject. For example, to determine the value of $g_{1/4} \in S_G$ that corresponds with the 25th percentile, a gamble is proposed that is contingent on a value of x that is lower than $g_{1/2}$ obtained in the previous step. Once again, a sequence of values, g_a, g_b, …, g_z is used, but in this next case (the quartile) the initial upper bound is $g_{1/2}$. In the first new binary question, subjects choose between the following binary prospects, $G_a = \{g_0 < x < g_a\}$ and $G_a' = \{k_1 \leq x < g_{1/2}\}$. As above, this process is repeated until the individual is indifferent between G_z and G_z', so that $g_z = g_{1/4}$ $G_z = G_4^1$, $G_z' = G_4^2$, and $P(G_z) = P(G_z')$ (see Figure 2.1 and Appendix A). At the end of the EG, the second binary question that subjects have already answered is presented again to them in order to test the consistency of their choice behaviors.

Figure 2.1 Scheme of the Exchangeability Game's bisection procedure



Other games

The Repeated Exchangeability Game (REG) consists in eliciting a new measure of the median value of individual CDFs, say $g_{1/2}$', through a second round of Exchangeability Game. This round differs from the first one because the lower and upper bounds of the event space are now not defined by g_0 and g_1, but instead by the subjective estimates of the quartiles $g_{1/4}$ and $g_{3/4}$ elicited via the Exchangeability Game (see Example 2 in Appendix A).

The Certainty Equivalent Game (CEG) is based on the notion of certainty equivalents (CE), defined as the sure amount of money that makes subjects indifferent to gamble. For the CEG, the subjects are presented with two choice tasks, say CT1 and CT2, both containing six binary questions. In each question of the first choice task (CT1), the subject is asked to choose between a lottery (Lottery 1), in which he or she wins a monetary outcome *x* if the real outcome G_j^i will happen in the future (or a null monetary outcome otherwise), and a sure payment *z*, varying from 0 to 100€. In the same way, in the CT2, subjects are asked to choose between a lottery (Lottery 2), in which they win a monetary outcome *x* if the real outcome G_j^k will happen in the future (or a null monetary outcome otherwise), and a sure payment *z* varying from 0 to 100€. Hence, each subject is presented with two choice tasks characterized by six binary matching question where he or she has to choose between options A (bet x € on the occurrence of G_j^i in CT1 or G_j^k in CT2) and B (take the amount of money z = 0, 25, 49, 51, 75, and 100€) (see Example 3 in Appendix A). The certainty equivalent for the lottery described in option A is determined by looking at the first question of the choice task in which the subject switches from choosing option A to choose option B. Recall that G_j^i and G_j^k are the couple of sub-spaces that have been already judged to be equally likely by the subjects themselves, during the earlier Exchangeability Game. Each subject in my study was presented with this game three times for each variable of interest in the study. In the first, the two lotteries involved in the game are denoted as G_2^1 and G_2^2, in the second, they are G_4^1 and G_4^2, and in the third, they are G_4^3 and G_4^4 [11].

The sample

The sample of laboratory subjects consists of 80 individuals who were randomly recruited outside the main supermarkets of Trento and asked to come in the experimental lab of the University of Trento for a compensation of 25€ (show-up fee). Given the fact that I recruit non-students and, then, I bring them in the lab, I can define my study as an artefactual field experiment (Harrison and List, 2004). My sample consists of people between 18 and 70 years age who live in the Province of Trento and the sample is balanced regarding the gender. They are not strictly speaking, a simple random sample of the population, because they were recruited outside food markets, but as most people visit such markets to obtain food, they probably are quite representative of people living in this Province. Moreover, the random nature of the sample may be biased by subjects' motivation to participate in the experiment. For example, subjects

[11] Both games have been already used to test exchangeability in other experimental applications (e.g., Baillon, 2008; Abdellaoui et al., 2011).

may participate because they were interested in the topic or because they were in need of the show-up fee.

Treatments

Selected participants were randomly assigned to four subsamples or treatment groups, where each treatment is characterized by a different experimental design: "real incentives-random questions" (22 subjects)[12], "real incentives-sequential questions" (23 subjects), "hypothetical incentives-random questions" (19 subjects), and "hypothetical incentives-sequential questions" (16 subjects). For the "hypothetical incentives" treatments, subjects are only given a show-up fee, while in the "real incentives" treatments, subjects are told that one randomly selected individual from each group has the chance to win additional 100€ based on her/his choices during the experiment. Specifically, one subject is to be randomly selected at the end of the experiment and one of the questions she/he answers during the experiment is also randomly selected to be played out. The lucky subject is selected through the draw of a numbered chip from a bingo cage (Cage 1). The total number of chips is equal to the total number of participants in each session, so that each subject has an equal chance of being selected. The question with the potential pay-out is also selected through the draw of a numbered chip from another bingo cage (Cage 2), that contains as many numbered chips as the number of questions that the subject answered during the experiment. The drawn participant wins the additional 100€ if and only if the event she/he had chosen in the drawn question contains the value of the random variable under consideration that the best science currently predicts. This prediction is based on the research conducted by

[12] The original "real incentives-random questions" treatment had 23 subjects, however I deleted observations gathered from one particular subject who declared that she has made a mistake during the tasks. Given that subjects did not have the chance to correct their errors during the experiment and chained experimental designs propagate mistakes, my subjects were asked to declare if they unintentionally made errors answering experimental questions.

the Edmund Mach Foundation (EMF). This procedure for the determination of a “win” in the lottery situation is similar to that used by Fiore et al. (2009) in their virtual experiment on the risk of wild fires. Despite some participants already being aware of the existence of the EMF, all subjects are provided with general information about the EMF’s research that provides science-based estimate of probabilities. Note that even when all subjects receive the same information, it is a common finding that they may not form the same subjective estimates (e.g. Riddel and Shaw, 2006; Shaw et al., 2012). In all treatments subjects were provided with precise information about the values that the random variables under study had in the last ten years (from 2000 to 2010) and, then, they were asked to play the games.

In the “sequential questions” treatments subjects are asked to answer questions that allow us to elicit the percentiles of their CDFs in the following order: $g_{1/2}$, $g_{1/4}$, $g_{3/4}$, $a_{1/2}$, $a_{1/4}$, $a_{3/4}$, $r_{1/2}$, $r_{1/4}$, and $r_{3/4}$. In the “random questions” treatments this chained structure of the game is hidden through a mixed up order of questions determined once and for all. In fact, I elicit the percentiles of subjects’ CDFs in the following order: $g_{1/2}$, $a_{1/2}$, $r_{1/2}$, $g_{1/4}$, $a_{1/4}$, $r_{1/4}$, $g_{3/4}$, $a_{3/4}$, and $r_{3/4}$.

It follows that each subject, regardless of the treatment group to which she/he is randomly assigned, plays the Exchangeability and the other games three times, one for each random variable under study.

Hypotheses about the validity of subjective probabilities

To investigate the effect of sequential (or random questions) and real (or hypothetical) monetary incentives on the validity of subjective probabilities elicited via the Exchangeability Game, I first need to understand whether gathered estimates are valid or not. Given the theoretical background of the EG, I argue that subjective probabilities elicited via this technique are valid if the only if the exchangeability assumption is satisfied. Otherwise they are invalid. In fact, under the exchangeability assumption, subjective probabilities elicited via the EG satisfy all definitions, axioms and theorems of probability theory. Considering two disjoint sub-events, G_j^i and G_j^k, the exchangeability assumption is satisfied when the two sub-events are exchangeable, in the sense that the probability related to the occurrence of one must be equal to the probability of occurrence of the other (see Appendix B). When the assumption holds I fail to reject the following null hypothesis (H_0) and I consider elicited subjective probabilities valid:

H_0: $P\left(G_j^i\right) = P\left(G_j^k\right), \forall k \neq i, k \leq n$

H_1: $P\left(G_j^i\right) \neq P\left(G_j^k\right), k \neq i, k \leq n$

I test this hypothesis and, thus, the validity of subjective probabilities elicited via the EM by investigating whether subjects' choice behaviors are consistent across the Exchangeability Game, the Repeated Exchangeability Game, and the Certainty Equivalent Game. In particular, I test the following two hypotheses:

Hypothesis 1. I test whether the exchangeability assumption is satisfied or not by comparing the estimates of $g_{1/2}$ obtained from the Exchangeability Game and the estimates of $g_{1/2}$' obtained from Repeated Exchangeability Game. The exchangeability assumption is satisfied, and, thus, the subjective probability of the event $g_{1/2}$ is valid, if and only if I fail to reject the following null hypothesis (H_0):

H_0: $g_{1/2} = g_{1/2}'$

H_1: $g_{1/2} \neq g_{1/2}'$

Hypothesis 2. I test whether the exchangeability assumption is satisfied or not by comparing the certainty equivalents that subjects are willing to accept to give up the possibility to play the lotteries presented in the matched pairs of choice tasks, $\left[L\left(x:G_j^i\right)\right]$ in CT1 and $\left[L\left(x:G_j^k\right)\right]$ in CT2 (Certainty Equivalent Game). The exchangeability assumption is satisfied, and, thus, the subjective probability of the event presented in both CT1 and CT2 is valid, if and only if I fail to reject the following null hypothesis (H_0):

H_0: $CE\left[L\left(x:G_j^i\right)\right] = CE\left[L\left(x:G_j^k\right)\right]$, with $k \neq i$, $k \leq j$

H_1: $CE\left[L\left(x:G_j^i\right)\right] \neq CE\left[L\left(x:G_j^k\right)\right]$

Testing hypotheses

Before testing the hypotheses above, I first check the consistency of subjects' choice behaviors by examining their answers to the repeated binary questions presented at the end of the Exchangeability Game. In the 66.51% of cases, subjects' choices are the same in the original and repeated questions. This result is quite encouraging, given that Baillon (2008) found a consistency rate of 70.51% applying the same procedure to evaluate consistency, but investigating random variables more familiar to subjects than the ones I have examined here. Further, the McNemar test shows that subjects' choices are consistent even across treatments (Table 2.1).

Table 2.1 McNemar's test of consistency		
Treatment	**Null Hypothesis**	χ^2
Real incentives-Sequential questions	P(AB)[a] = P(BA)[b]	1.60
Real incentives-Random questions	P(AB) = P(BA)	0.31
Hypothetical incentives-Sequential questions	P(AB) = P(BA)	0.82
Hypothetical incentives-Random questions	P(AB) = P(BA)	1.32

[a] P(AB) is the probability of choosing prospect A in the original question and prospect B in the repeated question.
[b] P(BA) is the probability of choosing prospect B in the original question and prospect A in the repeated question.
*1% significance level, **5% significance level, ***10% significant level

Next, testing my hypotheses at sample level, I determine whether subjects, belonging to diverse experimental treatments, provide valid subjective probabilities or not. This allows us to test predictions presented above, in particular, the fact that subjects provided with real monetary incentives and random questions state valid subjective probabilities, while the others do not. Recall that subjects are assumed to provide valid subjective probabilities if the exchangeability assumption holds and, thus, if and only if I fail to reject the null hypotheses presented in *Hypotheses 1* and *2*.

I test *Hypotheses 1* and *2* by using non-parametric tests such as the Wilcoxon Matched-Pairs Signed-Ranks test (WMP) and the Sign Test of Matched Pairs (SMP). The SMP test is used because the assumptions behind the WMP test were not always satisfied in my sample. For example, the differences between the matched values provided by each subject were not always distributed symmetrically around the median point in my sub-samples (*symmetry assumption*).

While testing *Hypothesis 1*, I only investigate the validity of individual CDFs' medians ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$), as I rely on estimates elicited via the Exchangeability Game and Repeated Exchangeability Game, testing *Hypothesis 2*, I also examine the validity of individual CDFs' first and third quartiles ($g_{1/2}$, $a_{1/2}$, $r_{1/2}$, $g_{1/4}$, $a_{1/4}$, $r_{1/4}$, $g_{3/4}$, $a_{3/4}$,and $r_{3/4}$), as I rely on estimates elicited via the Exchangeability Game and Certainty Equivalent Game.

Further, I assess the *validity rate* (*V*) for each different experimental treatment, which is the percentage of valid subjective probabilities out of the total number of elicited estimates in each treatment. This rate allows us to quantitatively assess the validity of subjective probabilities for each treatment and test, once more, predictions presented in Paragraph 3. To compute validity rates, I first need to verify whether each observation ($g_{1/2}$, $a_{1/2}$, $r_{1/2}$, $g_{1/4}$, $a_{1/4}$, $r_{1/4}$, $g_{3/4}$, $a_{3/4}$, and $r_{3/4}$) provided by each subject (i = 1,...,80) is valid or not. For example, let's consider one specific experimental subject, who provides us with the estimate of $g_{1/2}$, I assume that this estimate is valid if and only if the certainty equivalents for Lottery 1 and 2, presented in the Certainty Equivalent Game, are equal, thus, $CE\left[L\left(x:G_2^1\right)\right]=CE\left[L\left(x:G_2^2\right)\right]$. This does not imply any statistical test, but just a simple check of the equality between $CE\left[L\left(x:G_2^1\right)\right]$ and $CE\left[L\left(x:G_2^2\right)\right]$.

In addition, by examining the dissimilarity between $CE\left[L\left(x:G_2^1\right)\right]$ and $CE\left[L\left(x:G_2^2\right)\right]$, I can also investigate how much elicited subjective probabilities are invalid. For each elicited probability, the dissimilarity is measured as the absolute value of the difference between the certainty equivalents for Lottery 1 and Lottery 2 that is given by $\Delta(CE) = \left| CE\left[L\left(x:G_2^1\right)\right] - CE\left[L\left(x:G_2^2\right)\right] \right|$.

Based on these absolute values, I create an invalidity scale consisting of five categorical level of invalidity: very low invalidity when $\Delta(CE)\in \{x < 27\}$, low invalidity when $\Delta(CE)\in \{27 \le x < 52\}$, medium invalidity when $\Delta(CE)\in \{52 \le x < 77\}$, high invalidity when $\Delta(CE)\in \{77 \le x < 101\}$, and, finally, very high invalidity when $\Delta(CE)\in \{x \ge 101\}$. These boundaries have been chosen as the absolute values, Δ(CE), are naturally grouped in five categories given the range of the sure amount of money x that subjects might accept instead of playing the lotteries presented in the Certainty Equivalent Game.

Using this classification, I calculate the percentage of invalid probability estimates that falls within each category of invalidity and, hence, I investigate how far off invalid probabilities are from being valid.

Finally, I hypothesize that, not only the features of the experimental setting may determine the validity of subjects' subjective probabilities, but also their socio-economic conditions. I econometrically test this hypothesis by estimating a model in which the discrete dependent variable captures the validity of each observation provided by each subject, while independent variables capture the characteristics of each experimental setting and other socio-economic variables which characterize subjects, allowing for some observable heterogeneity.

Results

Non-parametric tests

By testing *Hypothesis 1* for each experimental group of subjects, I identify the effect of my experimental designs on subjects' capability to provide valid estimates of the median values. In the "real incentives-sequential questions" treatment I have 24 matched pairs of observations, in the "real incentives-random questions" 40, in the "hypothetical incentives-sequential questions" 22, and in the "hypothetical incentives-random questions" 26 (Table 2.2).

Table 2.2 Summary statistics of median values obtained via EG (X1/2) and REG ($X1_{/2}$')						
Treatment	**Variable**	**Obs**	**Mean**	**St.Dev.**	**Min**	**Max**
Real incentives-Sequential questions	$X_{1/2}$	24	44.37	27.69	7	94
	$X_{1/2}$'	24	44.96	27.87	7	94
Real incentives-Random questions	$X_{1/2}$	40	44.05	26.17	2	96
	$X_{1/2}$'	40	44.17	25.98	3	96
Hypothetical incentives-Sequential questions	$X_{1/2}$	22	54.91	28.03	5	94
	$X_{1/2}$'	22	55.91	28.08	7	94
Hypothetical incentives-Random questions	$X_{1/2}$	26	40.35	28.74	3	94
	$X_{1/2}$'	26	40.65	28.27	3	96

The validity of individual CDFs' medians ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$) is determined by testing *Hypothesis 1* via both the Wilcoxon Matched-Pairs Signed-Ranks (WMP) and the Sign Test of Matched Pairs tests (SMP). Median estimates are assumed to be valid if and only if I fail to reject the null hypothesis characterizing this test. The WMP test' results suggest that "real incentives-random questions" and "hypothetical incentives-random questions" treatments provide valid estimates, while "real incentives-sequential questions" and "hypothetical incentives-sequential question" treatments do not. The SMP test almost produces the same results, except for the fact that also "hypothetical incentives-sequential question" treatment provides valid estimates (Table 2.3).

Table 2.3 Results at sample level obtained via EG ($X_{1/2}$) and REG ($X_{1/2}$')			
		Wilcoxon matched-pairs signed ranks test	**Binomial sign test**
Treatment	**Null Hypothesis**	**Z**	**P>Z**
Real incentives-Sequential questions	Median($X_{1/2}$) =Median($X_{1/2}$')	-2.234**	0.062
Real incentives-Random questions	Median($X_{1/2}$) =Median($X_{1/2}$')	-0.665	0.480
Hypothetical incentives-Sequential questions	Median($X_{1/2}$) = Median($X_{1/2}$')	-1.880***	0.125
Hypothetical incentives-Random questions	Median($X_{1/2}$) = Median($X_{1/2}$')	-1.174	0.266

*1% significance level; **5% significance level; ***10% significant level

The discrepancy between WMP and SMP's results about the "hypothetical incentives-sequential question" treatment suggests that the interpretation of these results is problematic, and thus, I conclude that only "real incentives-random questions" and "hypothetical incentives-random questions" treatments provide valid subjective estimates.

Testing *Hypothesis 2* for each experimental group of subjects allows us to investigate whether subjects, belonging to diverse experimental treatments, provide valid estimates of the median, first quartile, and third quartile values of individual CDFs or not. In the "real incentives-sequential questions" treatment I have 143 matched pairs of observations, in the "real incentives-random questions" 167, in the "hypothetical incentives-sequential questions" 136, and in the "hypothetical incentives-random questions" 115 (Table 2.4).

Table 2.4 Summary statistics of the Certainty Equivalents obtained via CEG						
Treatment	**Variable**	**Obs**	**Mean**	**St.Dev.**	**Min**	**Max**
Real incentives-Sequential questions	CE_{L1}	143	51.21	46.38	0	125
	CE_{L2}	143	76.95	44.69	0	125
Real incentives-Random questions	CE_{L1}	167	59.80	42.31	0	125
	CE_{L2}	167	68.22	41.72	0	125
Hypothetical incentives-Sequential questions	CE_{L1}	136	70.80	43.30	0	125
	CE_{L2}	136	75.86	42.14	0	125
Hypothetical incentives-Random questions	CE_{L1}	115	55.65	36.14	0	125
	CE_{L1}	115	73.17	37.11	0	125

Again, the validity of median, first quartile, and third quartile estimates of individual CDFs ($g_{1/2}$, $a_{1/2}$, $r_{1/2}$, $g_{1/4}$, $a_{1/4}$, $r_{1/4}$, $g_{3/4}$, $a_{3/4}$, and $r_{3/4}$) is determined by testing *Hypothesis 2* via both the WMP and the SMP tests. Estimates are assumed to be valid if and only if I fail to reject the null hypothesis characterizing this test. The WMP test's results show that the "real incentives-sequential questions" treatment and the "hypothetical incentives-random questions" treatments do not provide valid estimates, while the "real incentives-random questions" and the "hypothetical incentives-sequential questions" treatments do. However, the validity of WMP test's results about the "hypothetical incentives-sequential question" treatment may be compromised because all assumptions behind the test are not completely satisfied. As the SMP test's results suggest that also the "hypothetical incentives-sequential questions" treatment does not provide valid estimates, I conclude that the "real incentives-random questions" is the only treatment providing valid estimates (Table 2.5).

Table 2.5 Results at sample level obtained via the CEG			
		Wilcoxon matched-pairs signed ranks test	**Binomial sign test**
Treatment	**Null Hypothesis**	**Z**	**P>Z**
Real incentives-Sequential questions	Median(CE_{L1}) = Median(CE_{L2})	-3.713*	0.002
Real incentives-Random questions	Median(CE_{L1}) = Median(CE_{L2})	-1.513	0.304
Hypothetical incentives-Sequential questions	Median(CE_{L1}) = Median(CE_{L2})	-1.283	0.088
Hypothetical incentives-Random questions	Median(CE_{L1}) = Median(CE_{L2})	-3.005*	0.000

*1% significance level, **5% significance level, ***10% significant level

Considering the whole set of subjective estimates, and not just median estimates, my results support the hypothesis, under which only subjects provided with real monetary incentives along with random questions return valid subjective probabilities. This result demonstrates that when the chaining structure of the elicitation mechanism is not perceived by experimental subjects, monetary incentives increase the chance of eliciting valid subjective estimates. However, it does not prove that validity depends on whether subjects perceive the Exchangeability Game to be incentive compatible or not.

Above, I predicted that subjects who perceived the EM to not be incentive compatible strategically play the game and provide invalid subjective probabilities. My prediction is supported if the percentage of rewarded subjects is higher in experimental treatments where real incentives are associated with sequential questions rather than with random questions. This is due to the fact that subjects who face sequential questions, perceive the chaining and, thus, strategically play (or, at least try to) the incentive incompatible elicitation mechanism to get better rewarded at the end of the experiment. I test this hypothesis by taking into account the subjects who belong to real incentive treatments, and simulating the rewards that each subject should have gained if she/he was the randomly drawn subjects and the third question he/she answered was the randomly drawn question[13].

I found that the chance of being rewarded is 50 percent for subjects playing sequential questions and 34.78 percent for those playing random questions (Table 2.6). This finding supports the hypothesis that subjects who are aware of the chaining play the elicitation mechanism to get better rewarded, but, unfortunately, also provide invalid subjective probabilities.

Table 2.6 Percentage of rewarded subjects based on their answers to Question 3

Treatment	Number of Subjects	Number of Rewarded Subjects	Percentage of Rewarded Subjects
Real Incentives-Sequential Questions	22	11	50.00
Real Incentives-Random Questions	23	8	34.78

The validity rate

For each treatment, I calculate the validity rate (*V*) which is simply the percentage of valid estimates within each treatment. According to the previous findings, I found that “real incentives-random questions” treatment provides the highest validity rate (39.13%), then the “hypothetical incentives-random questions” (29.86%), “real incentives-sequential questions” (26.26%), and “hypothetical incentives-sequential questions” (22.22) follow (Table 2.7).

[13] I have chosen the third question because it was the randomly drawn question at the end of the “real incentives-sequential questions” session.

Table 2.7 Validity rates (*V*) for all treatments				
Treatment	**Variable**	**Number of observations**	**Number of valid observations**	***V* (%)**
Real incentives-Sequential questions	First Quartile	66	15	22.72
	Median	66	24	36.36
	Third Quartile	66	13	19.69
	Total	198	52	26.26
Real incentives-Random questions	First Quartile	69	25	36.23
	Median	69	34	49.27
	Third Quartile	69	22	31.88
	Total	207	81	39.13
Hypothetical incentives-Sequential questions	First Quartile	57	13	22.80
	Median	57	15	26.31
	Third Quartile	57	10	17.54
	Total	171	38	22.22
Hypothetical incentives-Random questions	First Quartile	48	12	25.00
	Median	48	18	37.50
	Third Quartile	48	13	27.08
	Total	144	43	29.86

Again, according to my predictions I found that subjective probabilities, elicited providing real monetary incentives and using random questions, are likely more valid than those elicited providing real monetary incentives and using sequential questions. As demonstrated above, the low validity rate I found for the “real incentives-sequential questions” treatment depends on the fact that the provision of sequential questions along with monetary incentives makes the overall incentive incompatibility of the Exchangeability Game clear.

Even when monetary incentives are not provided to subjects, I found that random questions perform better than sequential questions in terms of validity. This result may be due to the fact that, in the part of the Exchangeability Game related to the elicitation of the third quartile estimates, subjects are asked to choose between prospects that they have already ruled out in the elicitation of the first and second quartile estimates. For example, a subject who has expressed median and first quartile estimates, respectively equal to $g_{1/2}$ = 72 and $g_{1/4}$ = 68, by answering the first and second set of binary questions, is then asked to express the third quartile estimate $g_{3/4}$. She does so by answering a third set of binary questions which involve outcomes greater than 72, and, thus, in conflict with outcomes she has just chosen in previous questions.

While this is clear to subjects who belong to the “hypothetical incentives-sequential questions” treatment, this is not clear to the subjects who belong to the “hypothetical incentives-random questions” treatment. Thus, chaining may induce subjects to reduce the effort invested in the tasks, as they may believe that questions related to the elicitation of the third quartile are somewhat meaningless. My hypothesis here is supported by the fact that validity rate of third quartile estimates in the “hypothetical incentives-sequential questions” treatment (about 22 percent) is lower than that founded in the “hypothetical incentives-random questions” treatment (almost 30, see Table 2.7).

My prediction is also confirmed by the fact that while, in the “hypothetical incentives-sequential questions” treatment, the validity rate of third quartile estimates (almost 18 percent) is lower than that of first quartile (almost 23 percent), in the “hypothetical incentives-random questions” treatment, the validity rate of third quartile estimates (about 27 percent) is greater than that of first quartile (25 percent – again, see Table 2.7). The issue of meaningless sequential questions does not arise when monetary incentives are provided because subjects are assumed to put more mental effort into trying to earn as much monetary reward as they can.

Unfortunately, I found relatively low validity rates for all my treatments. However, I do not believe this is due to the elicitation mechanism per se, but rather, to a series of different issues that I discuss below. First, such low validity rates may be due to the particular uncertain outcomes I investigated. As many subjects were, at least, initially unlikely to be familiar with the pesticide risk issue addressed in the experiment, the validity of elicited probabilities may be undermined by the sense of insecurity that subjects have likely felt during the tasks (Frisch and Baron, 1988). In contrast, something simple and familiar to all, such as uncertainty about temperature, might yield higher validity.

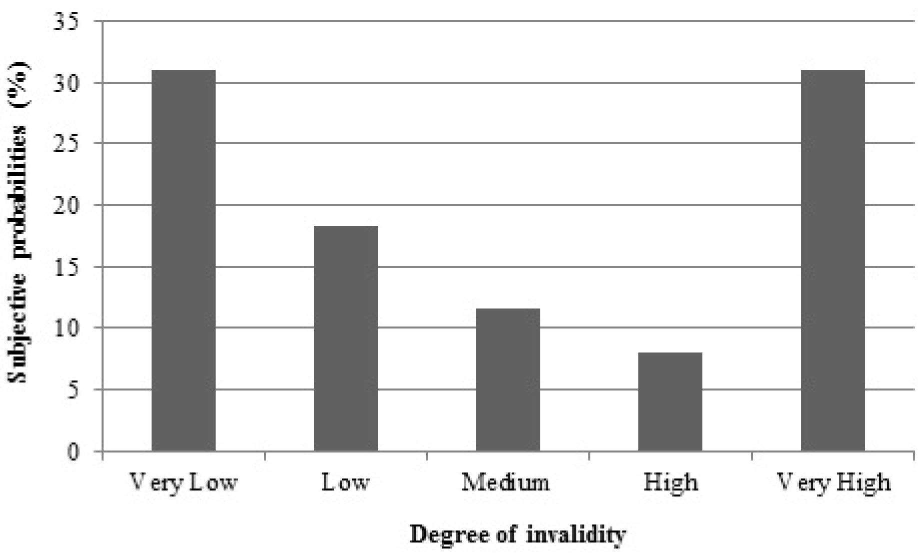
An alternative potential reason, as to why my subjects' responses have such low validity rates, involves the test I have used to investigate the validity of elicited probabilities. Recall that to calculate the validity rate, I assume that each estimate is valid if and only if the certainty equivalent for Lottery 1 was equal to that for Lottery 2. This procedure seems to be quite constraining as it does not imply any statistical test, but is just a simple check of the equality. Unfortunately, here, I cannot either measure or disentangle the effect of such influencing factors, but only speculate on them.

Given the large proportion of invalid probabilities, I also investigate their level of invalidity. Using the invalidity scale described above, I found that about 31 percent of the invalid probability measures are characterized by a very low level of invalidity, about18 percent by a low level, approximately 12 percent by a medium level, about 8 percent by a high level, and about 31 percent by a very high level (Table 2.8 and Figure 2.2). The fact that invalid observations are concentrated at the two extreme levels of invalidity emphasizes that subjects were either rather sophisticated about their probability estimates or not at all, with a smaller portion of the subjects falling in-between.

Table 2.8 Percentage of probabilities per level of invalidity in each treatment

Level of Invalidity	∆(CE)	TRC	TRU	THC	THU	Total
Very Low	< 27	23.71	31.18	30.93	40.54	31.02
Low	27-51	12.37	18.28	24.74	17.57	18.28
Medium	52-76	8.25	17.20	11.34	9.46	11.63
High	77-100	7.22	10.75	6.19	8.11	8.03
Very High	>101	48.45	22.58	26.80	24.32	31.02

Figure 2.2 Histogram of the percentage of subjective probabilities per level of invalidity



The econometric analysis

In this essay, I hypothesize that, not only experimental designs, but also socio-economics characteristics of subjects and their degree of familiarity with the problem influence individual performances in terms of validity. This hypothesis is econometrically tested by estimating a discrete model in which the dependent variable *VALID* represents the validity of each estimate provided by each subject. The dependent variable takes the value 1 if and only if the estimate is valid according to Hypothesis 2, and thus $CE\left[L\left(x:G_j^i\right)\right]=CE\left[L\left(x:G_j^k\right)\right]$, with $k \neq i$, $k \leq j$[14]. Given that each subject i provides 9 estimates ($g_{1/2}$, $a_{1/2}$, $r_{1/2}$, $g_{1/4}$, $a_{1/4}$, $r_{1/4}$, $g_{3/4}$, $a_{3/4}$, and $r_{3/4}$), I should have a panel data of 720 observations. However, I have 142 missing values for the dependent variable *VALID* because the Certainty Equivalent Game investigating the validity of each estimate was not always displayed to subjects during the experiment depending on their choice behavior.

In my model (Equation 2.1), the probability that each individual estimate is valid, depends on a set of explanatory variables available from survey-type questions given in the laboratory: the experimental treatment that subjects belong to, the socio-economics status of subjects themselves, and subjects' degree of interest in the issue of food safety (see Table 2.9 for details about the explanatory variables).

Equation 2.1

$$VALID_i = \beta_0 + \beta_1 T_i + \beta_2 RV_i + \beta_3 P_i + \beta_4 S_i + \beta_5 I_i + \beta_6 TR_i$$

I estimate this model by using the generalized linear model estimation with and without robust standard errors. Hereafter, I focus on the estimation with robust standard errors that allows for clustering effects.

[14] My dependent variable relies only on Hypothesis 2, but not on Hypothesis 1, because, while the latter only test the validity of median estimates, the former takes into account also first and third quartiles.

Table 2.9 Description of dependent and independent variables of Model 1					
Variable	**Definition**	**Mean**	**St.Dev.**	**Min**	**Max**
VALID	= 1 if valid, = 0 otherwise	.368	.482	0	1
TRS	= 1 if "Real Incentives-Sequential Questions" treatment, = 0 otherwise	.275	.446	0	1
TRR	= 1 if "Real Incentives-Random Questions" treatment, = 0 otherwise	.287	.452	0	1
THS	= 1 if "Hypo Incentives-Sequential Questions" treatment, = 0 otherwise	.237	.425	0	1
THR	= 1 if "Hypo Incentives-Random Questions" treatment, = 0 otherwise	.200	.400	0	1
G	Number of days when the infestation risk is extremely high in April	.333	.471	0	1
A	Number of apple containing at least one pesticide residue	.333	.471	0	1
R	Number of apple containing multiple pesticide residue	.333	.471	0	1
50^{th} PERCENTILE	Observations related to the median of G, A, and R	.333	.471	0	1
25^{th} PERCENTILE	Observations related to the I quartile of G, A, and R	.334	.471	0	1
75^{th} PERCENTILE	Observations related to the II quartile of G, A, and R	.333	.471	0	1
CONSUMER	= 1 if the subject eats at least 3 apples a week = 0 otherwise	.478	.500	0	1
CONS_ASS	= 1 if the subject is a member of a consumer association = 0 otherwise	.062	.242	0	1
PRODUCER	= 1 if the subject produces apples = 0 otherwise	.037	.190	0	1
TRENTINO	= 1 if the subject resides in the province of Trento = 0 otherwise	.737	.440	0	1
IPCC_TRUST	Trust in IPCC's predictions of the future temperature and precipitation[a]	2.950	.545	0	4
FEM_TRUST	Trust in FEM's predictions of fire blight's infestation risk in the future[a]	2.587	.684	0	4
SCENARIO_TRUST	Agreement with the fact that farmers will use the chemical control in the future)[b]	2.912	.778	0	4

AGE	Age in years	32.746	12.578	19	68
FEMALE	= 1 if female, = 0 otherwise	.4366	.4994	0	1
SECONDARY_SCHOOL	= 1 if the subject have this education level, = 0 otherwise	.1830	.3895	0	1
HIGH_SCHOOL	= 1 if the subject have this education level, = 0 otherwise	.5070	.5035	0	1
UNIVERSITY	= 1 if the subject have this education level, = 0 otherwise	.3098	.4657	0	1
SCIENTIFIC	= 1 if the subject have a scientific education = 0 otherwise	.487	.500	0	1

[a] From 0= very high trust to 4= very low trust
[b] From 0=strongly disagree to 4= strongly agree

My first aim is to test again whether the probability of providing valid estimates depends on the provision of monetary incentives and the ordering of questions. The set of variables *T* consists of four dummies (*TRS*, *TRR*, *THS*, and *THR*) which take the value 1 if and only if the subjects belong to the experimental treatment that the variable represents. I observe that only subjects who belong to the "real incentives-random questions" treatment (*TRR*) have a statistically significant higher probability of providing valid estimates than those who belong to the "hypothetical incentives-sequential questions" treatment (*THS*) which is used as baseline (Table 2.10). This result supports my previous findings from non-parametric testing and validity rate's analysis.

Two other sets of dummy variables have been included in my model, the first, *RV*, to capture whether the probability of providing valid estimates depends on the variable that subjects have to consider in playing the Exchangeability Game (*G, A*, or *R*), the second, *P*, to capture whether the validity of stated estimates is statistically different among median ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$), first quartile ($g_{1/4}$, $a_{1/4}$, and $r_{1/4}$), and third quartile estimates ($g_{3/4}$, $a_{3/4}$, and $r_{3/4}$). However, I found no statistical difference in terms of validity between estimates related to diverse variables and diverse percentiles (Table 2.10).

Then, I also investigate the effects of socio-economic variable *S* on the probability that subjects provide valid estimates. I take my cues from extensive psychological research on the role that several factors can play in the determination of perceived risks. The variables under study are age (*AGE*), gender (*FEMALE*), education (*SECONDARY, HIGH_SCHOOL,* and *UNIVERSITY*), and the type of education (*SCIENTIFIC*). I expected that the probability of providing valid estimates would possibly increase for high educated and younger subjects, but I found that older subjects' estimates are more likely to be valid than the others (even though at 10 percent significance level) and education does not affect the validity of individual estimates, at least in my sample (Table 2.10).

Furthermore, I consider also the interest of subjects on apples and food safety by including in the model a set of dummy variable (*I*) such as being an apple farmer (*PRODUCER*), being an apple consumer (*CONSUMER*), being a member of a consumer association (CONS_ASS), and being resident in the Province of Trento (*TRENTINO*). Although I expected to observe that subjects who reside in the Province of Trento and consume and/or produce apples perform better than the other in terms of validity, perhaps, because they are more interested than the others in the topic, my empirical results suggest no significant explanatory effects for these variables (Table 2.10).

Finally, I add in my model another set of dummy variables (*TR*) which capture whether subjects trust the predictions of IPCC about temperature and precipitation in 2030 (IPCC_TRUST), the predictions of Edmund Mach Foundation (EMF) about the fire blight's infestation risk in 2030 (EMF_TRUST), and my statement that apple farmers will continue to use the chemical control against apple disease in the future (SCENARIO_TRUST). In this case, I predict that subjects who trust the information I gave them during the experimental instructions more likely provide valid estimates than the others. This is due to the fact that the truster plays the game more carefully. Despite that my predictions are confirmed overall, I found the trust in EMF's predictions reduces the probability of providing valid estimates (Table 2.10).

Table 2.10 Generalized Linear Model Estimation of Models 1and 2

Dependent Variable: VALID

Variable	**Model 1**	**Model 2**[a]
TRS	.370**	.370
TRR	.648*	.648**
THR	.385**	.385
A	-.058	-.058
R	-.173	-.173
MEDIAN	-.077	-.077
25^{th} PERC	-.094	-.094
FEMALE	-.097	-.097
AGE	.019*	.019***
SEC_SCHOOL	-.086	-.086
HIGH_SCHOOL	-.016	-.016
SCIENTIFIC	.173	.173
PRODUCER	.584***	.584
CONSUMER	-.021***	-.021
CONS_ASS	.312	.312
TRENTINO	.067	.067
IPCC_TRUST	.359*	.359***
FEM_TRUST	-.355*	-.355**
SCEN_TRUST	.253*	.253***
CONSTANT	-2.160*	-2.160**
LOG L.HOOD	-347.702	-347.702

[a] Robust standard errors and clustering effects
*1% significance level
**5% significance level
***10% significant level

The consistency of my econometric results with those obtained from non-parametric tests and validity rate's analysis suggests that the provision of real monetary incentives along with random questions increases the validity of elicited estimates. Moreover, I found that socio-economic variables and the interest of subjects in the topic do not influence the likelihood of providing valid estimates. Only age and trust affect subjects' ability to state valid estimates.

Conclusion

This essay has considered the influence of monetary incentives and question ordering on elicitation of subjective probabilities via the Exchangeability Method. In particular, I have shown that incentive compatibility of elicitation mechanisms determines the validity of elicited beliefs, at least in my study. In fact, when subjects are provided with monetary incentive along with sequential questions, which make subjects aware of the chaining and, thus, of the incentive incompatibility of the game, they try to strategically behave in order to get better rewarded at the end of the task and provide invalid subjective probabilities. On the other hand, when subjects are provided with monetary incentive, but random questions, which make the chaining and, thus, the incentive incompatibility of the game less transparent to subjects, they state their real beliefs and return valid subjective probabilities. Non-parametric tests demonstrate that only subjects provided with real monetary incentives and random questions state valid subjective probabilities.

Although non-parametric tests have shown that subjects who are not provided with monetary incentives return invalid estimates, I demonstrated, by investigating validity rates, that subjects provided with random questions performs better than those provided with chained questions in terms of validity, given or not given monetary incentives. In fact, validity rate for "hypothetical incentives-random questions" are substantially higher than that for “hypothetical incentives-sequential questions”. This result is likely due to the fact that sequential questions generate less meaningful tasks where subjects are asked to choose between two prospects that they have just ruled out in previous questions. This in turn may affect the validity of elicited probabilities, as subjects may invest less cognitive effort in playing the game.

Those interested in using the Exchangeability Method can thus walk away with important messages here. First, incentive compatibility of elicitation mechanisms may affect the validity of elicited beliefs. Subjects are indeed more likely to provide valid estimates, over more of an entire distribution (than one measure of central tendency), if they are rewarded with real monetary incentives based on their performances and presented with experimental design where the chaining is hidden through a particular randomization of the questions. Second, and more disappointing perhaps, is that only a relatively small portion of stated estimates (almost 40%) can be considered valid under the definition I have applied here, which relates to behavioral axioms. The latter implication may be of little surprise to skeptics, but is relevant in my goal to continue to improve ways to provide reliable information about people’s subjective probabilities.

Further researches on the validity of subjective probabilities elicited via Exchangeability Method might address these issues at the individual level. Instead of investigating the validity of each single observation, one might investigate the ability of each subject in providing valid estimates. This would be possible by collecting, for each subject, a number of observations large enough to test the validity of her/his stated probabilities by using non-parametric tests.

CHAPTER III. THE IMPACT OF INTERNAL VALIDTY ON SUBJECTIVE PROBABILITIES

Introduction

Despite progress that international and national authorities have made toward ensuring food safety (e.g., food-labeling, packaging, inspections), food-related risks still get the attention of a substantial proportion of consumers. For example, approximately 30 percent of all Europeans remain concerned about health consequences of pesticide residues in food (European Commission, 2010).

As both short- and long-term health outcomes induced by food safety are often uncertain, people's own probability estimates may become crucial for understanding their choice-behavior towards food products or policies (Kivi and Shogren, 2010). In fact, probability estimates may dictate consumers' choices far more than science-based predictions would. Several empirical investigations have shown that subjective probabilities often differ from science-based ones, even when people are told what these are (e.g., Riddel and Shaw, 2006). There might be two general reasons why such a discrepancy exists. First, while science-based probabilistic estimates may be simple averages based on frequency values for homogenous populations, individual subjective probabilities may be heterogeneous, and causes for this heterogeneity may be observed or unobserved. For many individuals, their subjective probabilities might be accurate, and not truly equal to the average population probability. Second, some individuals may make mistakes in processing probability-related information, and formulate estimates that are higher or lower than the science-based predictions. Much of what economists know about subjective probabilities has been borrowed from initial work by psychologists (e.g., Slovic, 1987).

Although an extensive literature has shown that subjective probabilities related to financial outcomes affect people's choices under risk and uncertainty in several branches of applied economics (see Manski, 2004 for a review), a relatively small number of studies has investigated the influence that subjective probabilities related to health outcomes have on people's behavior related to their everyday choices. A few studies have primarily coped with probability estimates of health outcomes related to smoking behavior (e.g., Viscusi, 1990; Gerking and Khaddaria, 2011) as well as drinking contaminated water (e.g., Jakus et al., 2009; Shaw et al., 2012). Unfortunately, little has been done into investigating whether subjective probabilities of health outcomes due to food safety affect people's economic choices in their everyday life. A small number of studies have shown that consumers' subjective probabilities of health outcomes (i.e., mortality rate) due to the presence of pesticide residues in fresh fruit and vegetables drive their preferences for free-pesticide fresh fruit and vegetables in hypothetical markets. (e.g., Hammit, 1990; van Ravenswaay and Hoehn, 1991; Buzby et al., 1998).

In contrast, here, I mainly examine probability estimates of food safety outcomes themselves. In particular, I investigate consumers' subjective probabilities that given proportions of apples produced in the Province of Trento (Italy) will contain pesticide residues in 2030. Given that pesticide residues have consequences on health, consumers' expectations about the future presence of pesticide residues in apples may affect their support for agricultural policies which aim to incentivize the production of free-pesticide fruit and vegetables. This issue becomes particularly important in areas like the Italy's Province of Trento, where the saleable gross production of apple is approximately 23 percent of the entire agricultural saleable gross production in that area (P.A.T., 2010).

The bulk of the literature which has investigated subjective probabilities related to food safety has barely taken into account the fact that elicited subjective probabilities might not be valid, in the sense that subjective estimates might not obey to all axioms and theorems of Probability Theory (de Finetti, 1937; 1974a; 1974b). An exception is the experiment conducted by Cerroni et al. (2012) in which the validity of subjective probability elicited via the Exchangeability Method (EM) (Baillon, 2008; Abdellaoui et al., 2011), an innovative elicitation techniques based on the notion of exchangeable events (de Finetti, 1937), has been tested.

Investigating the validity of subjective probabilities might help to better understand people's choices under risk and uncertainty. In fact, the inclusion of invalid observations in subjective expected utility or other non-expected utility models to predict decision making, might generate biased results, especially if invalid observations systematically differ from valid ones in terms of magnitude. For example, if invalid subjective probabilities are systematically lower (or greater) then valid ones, consumers' preferences (i.e., willingness to support) for agricultural policies might be underestimated (or overestimated).

Given that, in this essay, by using Cerroni et al.'s (2012) results on the validity of subjective probabilities, I analyze the discrepancy between valid and invalid probability estimates in terms of magnitude. Furthermore, I also econometrically identify attitudinal and socio-economic factors that shape the subject's perceptions, comparing my results with previous findings.

The remainder of the essay is laid out as follows. In the next section, I review previous studies dealing with perceptions of pesticide residues and its consequences on human health. Next, I define the aims of the current study and provide detailed information about the experimental design. Finally, I offer a discussion of my results.

Subjective probabilities and pesticide residues

Many stated-preference (SP) studies have investigated the role of consumers' perceptions of health outcomes due to pesticide residues in determining food-purchasing behaviors. In general, these studies have shown a negative correlation between people's perceptions of health outcomes due to pesticide residues and willingness to purchase products which contain those chemical substances. Many food products have been considered, ranging from general unlabeled ones (e.g., Misra, et al., 1991; Eom, 1994; Rimal, et al. 2008) to specific types of fresh fruit and vegetables (e.g., Fu et al., 1999; Boccaletti and Nardella, 2000).

Most studies do not focus on subjective probabilities, but on people's concern about the severity of health consequences due to food safety[15]. For example, individuals might be asked to indicate the presence of health risks using simple descriptive labels (e.g. high, medium, or low), likert or other numerical scales.

Eom (1994) elicited subjects' concern about the presence of pesticides in general commercially grown food products by using a likert scale between 0 (no risk) and 10 (very serious risk). This study found that the average concern across consumers was quite high, around 6.6. The same approach was taken by Fu et al. (1999), but for fresh fruit and vegetables. In this case, the average level of concern was extremely high, exceeding 6, on a scale between 0 and 7**.** In their experimental auction for residue-free foods, Roosen et al. (1998) used a simple scale of concern (1 to 5) to investigate the influence of subjective perceptions on consumers' bidding behaviors. The approach recently used by Rimal et al. (2008) to elicit people's perceptions of pesticide residues in food was even simpler. Here, individuals were simply asked to state whether the

[15] In contrast, one might use observed purchases or transactions as a way of revealing individuals' sense of risk, but identification issues may easily arise in the effort to uncover the risks and sort these out from other influences on purchases, from the data.

problem of pesticides in food was serious, moderate or inexistent, and the finding was that more than half the subjects chose the serious option.

Boccaletti and Nardella (2000), improved the approach used by Misra et al. (1991), implementing a Likert Attitude Scaling Procedure, where individuals are asked several questions and, then, an individual-specific score is calculated to measure the concern about pesticide residues on fresh fruit and vegetables. The mean score across consumers was 78 on the maximum of 100, where the latter value is not a probability per se, but indicates very high concern.

While these simple efforts are appealing, they may be lacking in that they do not provide the information that would be ideal in actual modelling risky behaviours. For example, a reliable numerical estimate of probability can be directly used in either an expected utility or subjective expected utility framework, but measures of concern, or other responses, which are not probabilities cannot be used in this way (Manski, 2004).

Several scholars have questioned whether perceptions measured on some scale, as done in some of the studies above, are good indicators of probability (e.g., Viscusi and Hakes, 2003). At the very least, one would have to make strong assumptions to re-map from a 0 to 10 discrete response scale to a 0 to 1 unit interval. This could be done for example, to get a relevant probability, which is of course a continuous variable on the unit interval. Simple recoding would of course make it impossible to obtain other probability estimates than in 10% jumps (10%, 20%, 30%, etc.). Hence, many other studies have paid closer attention to the elicitation of actual numerical probability measures. In most of these studies the elicitation scheme is simple, and people are just asked to state probability estimates. The specific magnitude of the outcome that will happen is typically first presented, and individuals are then asked about the probability of this occurring to others (e.g., Viscusi 1990, asks people to guess how many smokers out of 100 will get, or die from, lung cancer), or to themselves, but many variations in presentation are possible. The techniques which directly elicit subjective probabilities are called direct methods (Spetzler and Von Holstein, 1975).

Extensive research, much of which is in the psychology literature, has shown that people do not easily understand numerical probabilities (especially small ones), and, given that, suggests different approaches (i.e., frequencies) for making people willing and able to state their best estimates (e.g., Gigerenzer and Hoffrage, 1995; Hammit and Graham, 1999; Corso et al., 2001).

Several studies suggest that mortality risks be couched as deaths per 100,000 or some other number in the population, avoiding small decimal place numbers that are confusing. Buzby et al. (1998) ask subjects their own subjective probability of dying from consuming fresh products containing pesticides in a similar manner, specifically, as the annual number of deaths per 1 million individuals. Since this probability-estimation task may be difficult for laypeople, subjects in both of these studies were provided with risk ladders showing probability of dying from more-familiar causes of death. The mean probability estimate was roughly 43 deaths per million in the population, per year.

Williams and Hammit (2001) used this same basic technique to examine the annual fatality rate per 1 million in the population of the United States for several categories of food hazards, and one of these was also the presence of pesticide residues in food. Generally, consumers perceived the probability of dying due to pesticides as being greater than either natural toxins or microbial pathogens. In particular, to conventional buyers, the annual median fatality rate because of pesticide residues on fresh products was 50 per million, while, to organic food buyers, this was 200 per million.

Although direct methods are very easy to design and implement, they have been questioned because of the quality or accuracy of the elicited subjective probabilities. In the cognitive psychology literature the ability, or more specifically, the willingness of subjects to put efforts in expressing their belief in numerical probabilities, has been extensively debated. The elicitation of numerical probabilities is of course easy and feasible, but reliable results are not guaranteed (Manski, 2004).

An alternative way of eliciting subjective probabilities consists of using subjects' choices, most often made over lotteries and gambles. In particular, probability measures are indirectly estimated by the researcher at the points for which people show their indifference between lotteries or gambles, which can be thought of as games that the subjects play. These techniques which indirectly elicit subjective probabilities are called indirect methods (Spetzler and Von Holstein, 1975). Those methods are assumed to be less demanding than direct methods from a cognitive point of view as subjects are not asked to directly express a numerical probability, but to compare risky outcomes and choose the most likely one (Spetzler and Von Holstein, 1975).

To my knowledge, the first application of an indirect technique in eliciting subjective probabilities related to the presence of pesticide residues in food is represented by the Cerroni et al. (2012)'s experimental investigation. In particular, that study has elicited numerical subjective probabilities that given proportions of apples will contain pesticide residues by using the EM, an indirect elicitation techniques in which subjects are asked to bet a given amount of money on a given outcome rather than on an alternative one. Subjective probability are indirectly inferred at the point for which subjects show their indifference for betting on one of the two outcomes. One innovative aspect of this elicitation techniques consists in asking subjects to focus on the severity of the outcome, rather than on the probability of a given outcome to occur.

This investigation into outcomes is rare, as compared to attention paid by previous studies to subjective probabilities of endpoint risks, such as human mortality or morbidity risk (Kuhn and Budescu, 1996).

The study by Cerroni et al.'s (2012) differs from the bulk of the literature which investigates subjective probabilities related to food safety in other aspects. First, while previous investigations were purely hypothetical, that study used monetary incentives to push subjects into stating their real expectations (Vossler and Evans, 2009). Given that, it must be considered the first economic experiment eliciting subjective probabilities of food safety outcomes related to pesticide residues. Second, Cerroni and colleagues (2012) have analyzed for the first time the validity of subjective probabilities related to food safety outcomes. They identify valid estimates by using a validation procedure based on the deFinetti's notion of coherent subjective probabilities (de Finetti, 1937; 1974a; 1974b).

In particular, they have tested whether validity of elicited subjective probability depends on the monetary incentives and the order of questions. To test their hypotheses, selected participants were randomly assigned to four treatment group, the real monetary incentives-sequential questions, the real monetary incentives-random questions, the hypothetical monetary incentives-sequential questions, and the hypothetical monetary incentives-random questions. In the hypothetical treatments, subjects are only given the show-up fee, while in the real monetary incentives treatments, one randomly selected subject from each treatment has the chance to win up to an additional 100€ based on her/his choices during the experimental games. The only difference between the sequential and random treatments is the order of the questions, and in fact subjects in the former treatments face sequentially ordered questions, while subjects in the latter face randomly ordered questions. Previous research within decision analysis and management science has shown that the order of questions undermines the reliability of subjective probabilities (e.g., Wakker and Deneffe, 1996).

Investigating the validity within each treatment group, Cerroni et al. (2012) found that subjects provided with real monetary incentives and random questions more likely return valid estimates. Examining the validity of each elicited subjective probability, they found that the proportion of valid estimates is 29.72 percent in the sample. In particular, they showed that the proportion of valid subjective probabilities is 39.13 percent in the real monetary incentives and random questions treatment, followed by 29.86 percent in the hypothetical monetary incentives and random questions treatment, 26.26 percent for the real monetary incentives and sequential questions treatment, and 22.22 percent for the hypothetical monetary incentives and sequential questions treatment. This suggests that in each treatment group there is a relatively small portion of valid subjective probabilities, and the real compensation with sequential responses out-performs the other treatments.

As subjective probability are often incorporated in the standard subjective expected utility or other non-standard theories of decision making under risk and uncertainty to model and predict risky behaviors, the identification of valid probability estimates becomes crucial to obtain highly predictive models, and thus, reliable findings on subjects' choice behavior. This is particularly true if valid observations systematically differ from invalid ones in terms of magnitude. In the latter case, failure to recognize valid subjective probabilities might induce us to over- or underestimate subjects' true expectations, and hence, to wrongly predicts their behavior.

Objectives

By drawing on Cerroni et al.'s (2012) investigation and using the same dataset they have used in their analysis, I first investigate subjective probability estimates that given proportions of apples will contain pesticide residues.

Second, I examine the potential discrepancy between valid and invalid subjective probabilities to fully understand whether failure to recognize validity implies an over- or underestimation of consumers' true probability estimates.

Finally, I estimate a behavioral model to identify attitudinal and socio-economic factors that affect the subject's probability estimates of pesticide residues in apples.

The empirical application

The case study

The fire blight is a bacterial disease that has damaged and killed apple plants in the Province of Trento since 2003 (EMF, 2006). The current infestation rate which is the number of days in which the infestation occurs in the blossoming period is less than 1 per cent. The infestation rate depend on climatic parameters such us temperature and precipitation. In this region of Italy, farmers currently adopt preventative measures based on pesticide usage in the form of copper compounds or Acibenzolar-S-metile to control the mild negative consequences that fire blight has on apple production. However, the future increase of the infestation rate, which is predicted to reach 17 percent in 2030, might eventually induce farmers to use new pesticides for preventative and curative control of fire blight. One candidate is the antibiotic streptomycin, currently forbidden under Italian law, but which has been already used in U.S., Germany, Belgium, and The Netherlands to control fire blight (Németh, 2004). After providing subjects with a description of the relevant scenario, as well as precise information about the values that the random variables under study had in the last ten years (from 2000 to 2009), I ask them to express their subjective probabilities of the number of apples containing pesticide residues in 2030 by playing my experimental games.

The dataset

The dataset is the same used by Cerroni et al. (2012) and consists of 1,200 probability estimates, 400 for each of the three random variables under study which are: the number of apples, *a*, containing at least one residue in a sample of 100 apples in 2030[16], the number of apples, *r*, containing at least two residues (multiple residues) in a sample of 100 apples in 2030[17], and the number of days, *g*, during which the infestation will occur during the blossoming period in 2030[18]. The latter variable *g* was added because of the potential link between the development of fire blight and the presence of pesticide residues in apples.

These variables were selected after having interviewed approximately 20 focus group subjects. The year 2030 is chosen because the best available science predicts that the heavy development of new phytopathology, as the fire blight, will start approximately twenty years from now in the Province of Trento.

The sample

The pool of sample subjects consists of 80 individuals between 18 and 70 years age who live in the Province of Trento. The sample is not, strictly speaking, randomly selected because subjects were recruited outside food markets, but it is still quite generally representative of people living in this Province because most all people in the

[16] The apple containing residues are those containing at least one residue beyond the level of 0 mg/kg.
[17] The apple containing residues are those containing at least two residues beyond the level of 0 mg/kg.
[18] The blossoming period usually occurs in April in the Province of Trento.

region go shopping in those markets at some point or another. A show-up fee of 25€ was given to each participant as a compensation for agreeing to come into the experimental lab of the University of Trento to take part in the experiment.

Methods

The elicitation of subjective probabilities: the Exchangeability Method

In this section, I briefly describe the EM, the technique used by Cerroni et al. (2012) to elicit subjective probabilities. The EM consists of multiple binary questions where subjects are only asked to bet a certain amount of money on one of the two disjoint subspaces in which the whole state space of the variable under study has been previously divided based on their choices. When subjects become indifferent to bet on one disjoint subspace rather than on the other, they are assumed to perceive those subspaces as equally likely (Spetzler and Von Holstein, 1975). This method allows eliciting several point estimates of the individual cumulative distribution function (CDF) of the random variable under study for each experimental subject. Interested readers may find additional details about the EM in Abdellaoui et al. (2011), Baillon (2008), and Cerroni and Shaw (2012).

The EM is applied to elicit subjective probabilities of three random variables, a, r, and g. As the EM is formally described in Cerroni et al. (2012), for brevity's sake here I only describe my application of the EM that concerns the number of apples containing at least one residue in a sample of 100 apples in 2030 (variable a). At the beginning of the game, I ask subjects to express the lower (a_0) and upper bounds (a_1) of the event space A. In this way, I identify the individual-specific range outside of which subjects are essentially certain that the outcome cannot happen at all. Assume that subject i states that a_0 is equal to 60 apples and a_1 is equal to 76. This means that she/he believes that the probability that the portion of apples containing at least one pesticide residue in 2030 will be outside these bounds (i.e. less than 60 and greater than 76) is equal to zero.

The second step involves asking a series of questions to establish the value of $a_{1/2}$ that corresponds with the 50th percentile of the subjective CDF, the median estimate. The first binary question is generated by splitting the event space in two prospects by using the following algorithm, 60 + [(76 - 60)/2] = 68. It follows that the first binary question implies a choice between prospects $A_1=\{60<x<68\}$ and $A_2=\{68\leq x<76\}$ (see Example 1 in Appendix A). Following the first choice, the exercise is repeated using a bisection of the chosen prospect. For example, if subject i has chosen prospect $A_1=\{60<x<68\}$, the second binary question asks subjects to choose between prospects $A_3=\{60<x<64\}$ and $A_4=\{64\leq x<68\}$. The bisectioning process goes on until the subjects become indifferent between the two prospects; at this point I am able to estimate the median point $a_{1/2}$ of the subject's subjective CDF. This estimate indicates that there is a 50 per cent chance that the number of apples that will contain at least one pesticide residue in 2030 will be equal to or less than $a_{1/2}$. A similar process can be followed to determine as many other points for the individual's subjective CDF as is desired, depending on limitations of the subjects' attention spans. For this study I also elicit the 25th percentile ($a_{1/4}$) and the 75th percentile ($a_{3/4}$).

Subjective probabilities are elicited for variables a, g, and r, defined above. For each variable, I elicit 5 percentile estimates, the lower bound (g_0, a_0, and r_0), the 25th percentile ($g_{1/4}$, $a_{1/4}$, and $r_{1/4}$), the 50th percentile ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$), the 75th percentile ($g_{3/4}$, $a_{3/4}$, and $r_{3/4}$), and the upper bound (g_1, a_1, and r_1).

The validity of subjective probabilities: the Repeated Exchangeability Game and Certainty Equivalent Game

In this section, I briefly describe two additional experimental games that were implemented by Cerroni et al. (2012) to facilitate the identification of valid probability measures, the Repeated Exchangeability and Certainty Equivalent Game (REG and CEG, respectively).

The REG allows us to identify valid probability measures at the sample level by statistically comparing estimates of the 50th percentile elicited via the EM ($a_{1/2}$) with those elicited via the REG ($a_{1/2}$'). The REG differs from the standard EM as the lower and upper bounds of the event space are not defined by a_0 and a_1, but by the subjective probability estimates of $g_{1/4}$ and $g_{3/4}$ elicited via the Exchangeability Game.

The sample provides valid subjective probabilities if and only if these estimates do not significantly differ from each other. In the CEG, subjects are presented with two choice tasks, say CT1 and CT2, both containing six binary questions, each asking subjects to choose between a gamble and a certain amount of money.

Next, I provide an example of the CEG that concerns the number of apples containing at least one residue in a sample of 100 apples in 2030 (variable *a*). Assume that subject *i* provides us with an estimate of $a_{1/2}$ that is equal to 66 apples, in CT1 she/he has to choose between options A (place a bet of *x* € on the fact that *a* is lower than 66) or B (take the certain amount of money *z* = 0, 25, 49, 51, 75, and 100€). For the second choice task CT2, she/he has to choose between options A (a bet of *x* € on the fact that *a* is greater than or equal to 66) or B (take the amount of money *z* = 0, 25, 49, 51, 75, and 100€). The certainty equivalent for the lottery described in option A is determined by looking at the first question of the six in the choice task in which the subject switches from choosing option A to choose option B (the amount of money).

The CEG is played for the 25th percentile ($g_{1/4}$, $a_{1/4}$, and $r_{1/4}$), the 50th percentile ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$), and the 75th percentile ($g_{3/4}$, $a_{3/4}$, and $r_{3/4}$). The CEG allows identification of valid probability estimates at both the sample and individual level. In the former case, the sample provide valid probabilities if and only if CE estimates related to CT1 and CT2 does not statistically differ from each other. At the individual level, each specific probability observation is valid if and only if the CE estimates related to CT1 and CT2 are equal. In this case, I do not use any statistical procedure, but I only ascertain that each pair of CE measures (for CT1 and CT2) is equal.

Results

The analysis of subjective probabilities

On average, estimated bounds of variable *a* suggest that the subjects believe the number of contaminated apples out of 100 will be between 56 and 75. Using information from the estimated 25th percentile, I argue that subjects believe there is only a 25 percent chance that the number of apples containing pesticides will be lower than or equal to 66. Using average values for the 50th and the 75th percentiles it appears that the subjects attach a 50 percent chance to the fact that the number of bad apples will be lower than or equal to 69, and 75 percent chance to the fact that this number will be lower than or equal to 71 apples (Table 3.1 and Figure 3.1b). Taking into account that the number of apples with at least one pesticide residue at present (in 2009) is 63 out of 100 (Italian Ministry of Health, 2010), I conclude that subjects do not in fact perceive an increase in the number of apples containing at least one pesticide residue by the year 2030 to be particularly substantial and, very likely.

Following the same general approach, I interpret percentile estimates related of the *r* variable, which is the number of apples containing multiple residues in a sample of

100 apples in 2030. In this case, I found that the lower bound (r_0) is about 31, the 25th percentile ($r_{1/4}$) is 42, the 50th percentile ($r_{1/2}$) is 45, the 75th percentile ($r_{3/4}$) is 48, and the upper bound (r_1) is 52 (Table 3.1 and Figure 3.1c). As might be expected, the average percentile estimates of r are always smaller than those of variable a (see Figure 3.1b and 3.1c) because the number of apples with multiple residues should always be lower than the number of apples with at least one residue. However, given that 31 apples, over the 63 containing at least one residue, have multiple residues in 2009 (Italian Ministry of Health, 2010), I deduce that subjects perceive an increase in the number of apples with multiple residue to be quite significant and likely. For example, they think that there is 75 percent chance that the number of apples with multiple residues will be 48 at the worst.

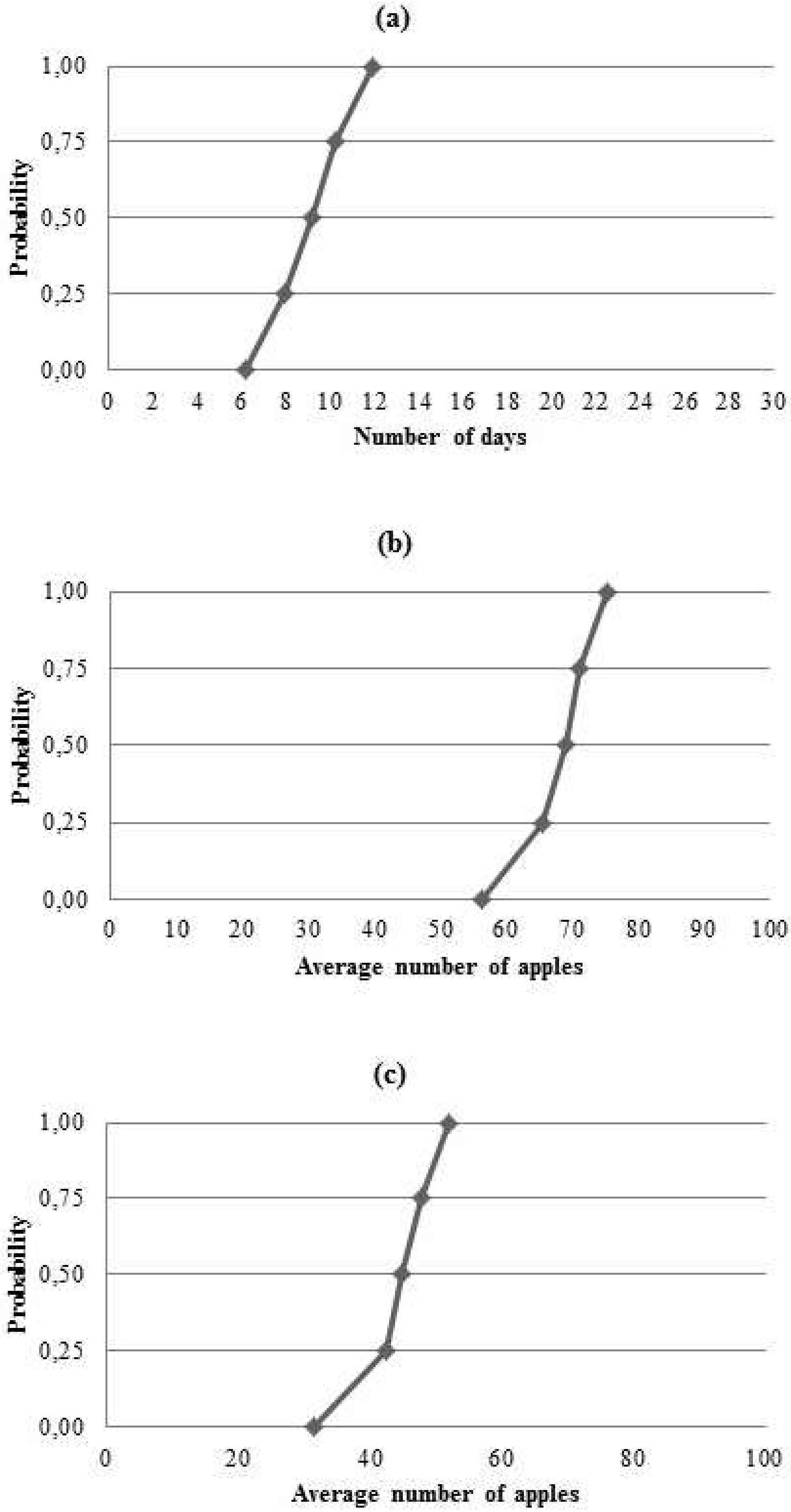
To summarize, although subjects believe that the number of apples containing one residue or more will not significantly increase by the year 2030, they predict that the number of apples containing multiple residues (more than one) will significantly increase. This means that the number of apples containing only one pesticide residue will decrease, but the number of apples with multiple residues will significantly grow by the year 2030.

Considering the infestation rate which is the number of days in which the infestation will occur during the blossoming period in 2030, I found that the lower bound (g_0) is 6, the 25th percentile ($g_{1/4}$) is 8, the 50th percentile ($g_{1/2}$) is 9, the 75th percentile ($g_{3/4}$) is 10, and the upper bound (g_1) is 12 (Table 3.1 and Figure 3.1a). Given the fact that the number of days in which the infestation actually occurred in 2000, 2005, and 2010 was very close to zero, I conclude that subjects perceive the infestation rate in 2030 as being quite high and likely.

Table 3.1 Summary statistics of percentile estimates for each variable

Variable	Obs.	Mean	Median	St.Dev.	Min	Max
g_L	80	6.176	5.000	4.677	1.000	29.000
$bleg_{1/4}$	80	7.912	6.750	5.879	0.205	29.250
$g_{1/2}$	80	9.175	7.500	6.320	0.500	29.500
$g_{3/4}$	80	10.250	9.000	6.228	0.750	29.750
g_U	80	11.925	10.500	6.072	1.000	30.000
a_L	80	56.354	60.000	20.455	4.000	90.000
$a_{1/4}$	80	65.637	68.000	21.879	5.000	96.000
$a_{1/2}$	80	69.200	72.000	21.907	6.000	98.000
$a_{3/4}$	80	71.187	74.500	21.896	8.000	99.000
a_U	80	75.450	80.000	21.706	10.000	100.000
r_L	80	31.392	32.000	16.381	4.000	82.000
$r_{1/4}$	80	42.387	38.000	19.066	5.000	90.000
$r_{1/2}$	80	44.875	41.000	18.941	6.000	92.000
$r_{3/4}$	80	47.700	43.000	19.334	8.000	93.000
r_U	80	51.825	47.000	19.241	12.000	100.000

Figure 3.1 The average number of days in which the infestation will occur during the blossoming period in 2030 (*a*), the average number of apples containing at least one residue in a sample of 100 apples in 2030 (*b*), and the average number of apples containing more than 1 residue in a sample of 100 apples in 2030 (*c*).



The difference between valid and not valid subjective probabilities

Using results on validity obtained by Cerroni et al. (2012) via the Certainty Equivalent Game, for each random variables, I compare the magnitudes of valid and invalid estimates at both the sample and individual levels[19]. At the sample level, I find here that the valid estimates are lower than invalid ones for each percentile (the 25th, the 50th, and the 75th) of each variable (a, r, and g) (Table 3.2). However, by using the Kolmogorov-Smirnov (KS) and the Mann-Whitney U (MWU) tests, I find that the discrepancy between the magnitudes of valid and invalid estimates is not statistically significant for all variables, a, g, and r (Table 3.3). Hence, even if my results suggest that failure to recognize validity may induce researchers to overestimate subjects' true probabilistic expectations, this finding is not statistically supported.

Table 3.2 Average values of variable g, a, and r considering valid and invalid observations

Variable	**Valid at sample level**		**Invalid at sample level**		**Valid at individual level**		**Invalid at individual level**	
	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs.	Mean
$g_{1/4}$	23	7.326	57	8.149	19	9.421	34	8.500
$g_{1/2}$	23	8.434	57	9.473	35	8.228	43	9.674
$g_{3/4}$	23	9.583	57	10.512	12	9.916	33	9.196
Tot.	69	-	171	-	66	-	110	-
$a_{1/4}$	23	62.691	57	66.823	21	66.476	46	68.195
$a_{1/2}$	23	67.304	57	69.964	25	63.280	55	71.890
$a_{3/4}$	23	69.652	57	71.807	19	68.157	34	68.882
Tot.	69	-	171	-	65	-	135	-
$r_{1/4}$	23	38.782	57	43.842	17	36.470	46	47.782
$r_{1/2}$	23	41.826	57	46.105	26	38.846	54	47.777
$r_{3/4}$	23	45.608	57	48.543	18	38.944	38	46.815
Tot.	69	-	171	-	61	-	138	-

[19] I used data from the Certainty Equivalent Game because it allows me to take into account first, second, and third quartile estimates, while the Repeated Exchangeability Game only deals with second quartile estimates.

Table 3.3 Comparison of valid and invalid percentile estimates at sample level

Null Hypothesis	Mann-Whitney U Test	Kolmogorov-Smirnov Test
H_0	**P-value**	**P-value**
$g_{valid} = g_{invalid}$	0.194	0.149
$g_{1/4,\ valid} = g_{1/4,\ invalid}$	0.413	0.444
$g_{1/2,\ valid} = g_{1/2,\ invalid}$	0.412	0.704
$g_{3/4,\ valid} = g_{3/4,\ invalid}$	0.466	0.444
$a_{valid} = a_{invalid}$	0.249	0.299
$a_{1/4,\ valid} = a_{1/4,\ invalid}$	0.284	0.567
$a_{1/2,\ valid} = a_{1/2,\ invalid}$	0.543	0.664
$a_{3/4,\ valid} = a_{3/4,\ invalid}$	0.733	0.534
$r_{valid} = r_{invalid}$	0.160	0.562
$r_{1/4,\ valid} = r_{1/4,\ invalid}$	0.290	0.444
$r_{1/2,\ valid} = r_{1/2,\ invalid}$	0.437	0.844
$r_{3/4,\ valid} = r_{3/4,\ invalid}$	0.503	0.923

Next, the valid and invalid estimates are compared at the individual level. For the random variables *a* and *r*, I found the same pattern as before, the 25th, the 50th, and the 75th percentiles are lower in valid estimates as compared to invalid ones (Table 3.2). Using the KS and MWU tests, I found that such a discrepancy between valid and invalid estimates is not statistically supported for variable *a*, while it is for variable *r*. In particular, valid estimates of 25th percentile ($r_{1/4}$) are statistically lower than the corresponding invalid ones (Table 3.4).

I found a different pattern for the variable *g*; valid estimates of the 25th and 75th percentiles ($g_{1/4}$ *and* $g_{3/4}$) are greater than the corresponding invalid estimates, while valid estimates of the 50th percentile ($g_{1/2}$) are lower than invalid ones (Table 3.2). However, these results are not statistically supported by the KS and the MWU tests (Table 3.4).

In general, the valid estimates are smaller than the invalid ones in variable *a* and *r*, but greater in variable *g*. However, I note that the such discrepancies are statistically supported only for variable *r*, but not for *a* and *g*. For what concern r, mistakes appear here to result in upward bias, and thus, failure to recognize validity may thus result in an overestimation of subjects' average probabilistic expectations.

Table 3.4 Comparison of valid and invalid percentile estimates at individual level

Null Hypothesis	Mann-Whitney U Test	Kolmogorov-Smirnov Test
H_0	**P-value**	**P-value**
$g_{valid} = g_{invalid}$	0.890	0.842
$g_{1/4,\ valid} = g_{1/4,\ invalid}$	0.408	0.214
$g_{1/2,\ valid} = g_{1/2,\ invalid}$	0.336	0.503
$g_{3/4,\ valid} = g_{3/4,\ invalid}$	0.463	0.250
$a_{valid} = a_{invalid}$	0.259	0.488
$a_{1/4,\ valid} = a_{1/4,\ invalid}$	0.560	0.705
$a_{1/2,\ valid} = a_{1/2,\ invalid}$	0.206	0.278
$a_{3/4,\ valid} = a_{3/4,\ invalid}$	0.933	0.928
$r_{valid} = r_{invalid}$	0.002	0.035
$r_{1/4,\ valid} = r_{1/4,\ invalid}$	0.048	0.018
$r_{1/2,\ valid} = r_{1/2,\ invalid}$	0.068	0.164
$r_{3/4,\ valid} = r_{3/4,\ invalid}$	0.199	0.676

Factors shaping subjective probabilities

To further analyze the factors that explain subjects' probabilistic expectations of both the number of apples containing pesticide residues and the fire blight's infestation rate, I estimate three empirical models. Key factors (see Table 3.5 for definitions) considered are the subjects' perceptions of climate change and farmers' usage of pesticides; their trust in science-based predictions about climate change and fire blight's infestation rate; their status of apple producers or apple consumers; their socio-economic features such as age, gender, place of residence, education, and income.

Table 3.5 Description of independent variables of Model 1,2,and 3

Variable	Definition	Mean	St.Dev.	Min	Max
VALID_S	= 1 if the observation is valid at sample level[a], = 0 otherwise	0.287	0.453	0	1
VALID_IND_G	= 1 if the observation is valid at sample level[a], = 0 otherwise	0.375	0.485	0	1
VALID_IND_A	= 1 if the observation is valid at individual level, = 0 otherwise	0.325	0.469	0	1
VALID_IND_R	= 1 if the observation is valid at individual level, = 0 otherwise	0.306	0.462	0	1
IPCC_TRUST	Trust in IPCC's predictions of temperature and precipitation (at 5 levels)[b]	2.950	.545	0	4
CC_H&N	= 1 if the subject believes that the climate change is due to both human activities and natural processes, = 0 otherwise	0.600	0.490	0	1
CC_H	= 1 if the subject believes that the climate change is mostly due to human activities, = 0 otherwise	0.337	0.473	0	1
CC_HH	= 1 if the subject believes that the climate change is only due to human activities, = 0 otherwise	0.062	0.242	0	1
PEST_NOW	Subjects' beliefs about the current usage of pesticides by farmers[c]	3.200	0.994	0	4
PEST_FUT	Subjects' beliefs about the future usage of pesticides by farmers[c]	2.912	0.779	1	4
EMF_TRUST	Trust in EMF's predictions of fire blight's infestation risk in the future[d]	2.587	0.685	1	4
CONSUMER	The number of apples consumed by the subjects in a week	3.700	5.160	0	20
CONS_ASS	= 1 if the subject belongs to a consumer association, = 0 otherwise	0.062	0.242	0	1
APP_LINK	= 1 if the subject is tied to apple production, processing and marketing. = 0 otherwise	0.212	0.409	0	1

TRENTINO	= 1 if the subject resides in the Province of Trento, = 0 otherwise	0.737	0.440	0	1
AGE	Age in years	33.625	13.213	19	68
FEMALE	= 1 if female, = 0 otherwise	0.436	0.4994	0	1
SEC_SCHOOL	= 1 if the subject have this education level,= 0 otherwise	0.183	0.3895	0	1
HIGH_SCHOOL	= 1 if the subject have this education level,= 0 otherwise	0.512	0.5035	0	1
UNIVERSITY	= 1 if the subject have this education level,= 0 otherwise	0.300	0.4657	0	1
INCOME	The yearly net income in 2010 in thousand €	18.968	19.560	0.075	0.115

[a] The observations valid at sample level belongs to the real incentive and random questions treatment
[b] I ask subjects whether IPCC's predictions will happen surely, very likely, maybe, not likely, or never.
[c] I ask people if they agree with the statement saying that farmers mostly use chemical control against apple diseases, 0=strongly disagree, 1=disagree, 2=do not know, 3=agree, 4=strongly agree.
[d] I ask subjects whether FEM's predictions about fire blight will happen surely, very likely, maybe, not likely, or never.

Given that my dependent variables are all essentially fractions, I do not estimate my models (Model 1, 2, and 3) by using a simple OLS estimator, although many apply the linear probability model to such data. Here, I use the Generalized Linear Model (GLM) along with robust standard errors (Papke and Woolridge, 1996). Observations in 80 groups are clustered because each subject provides three different percentile estimates (25^{th}, 50^{th}, and 75^{th} percentile) for each random variable under study (*g*, *a*, and *r*), and these may be correlated.

The general empirical specification common to the three models is:

Equation 3.1

$$y_{i,s} = \beta_0 + \beta_1 VALIDITY_{i,s} + \beta_2 ATTITUDE_{i,s} + \beta_3 APPLE\_LINK_{i,s} + \beta_4 SOCIOECONOMIC_{i,s}$$

In Model 1, the dependent variable (*y*) is each subject's estimates of the number of days in which the infestation will occur during the blossoming period in 2030 (*g*), in Model 2, each subject's estimates of the number of apples containing at least one residue in a sample of 100 apples in 2030 (*a*), and in Model 3, each subject's estimates of the number of apples containing multiple residue in a sample of 100 apples in 2030 (*r*).

In all models, I investigate the difference between valid and invalid estimates in terms of magnitude by creating two sets of dummy variables. The first, *VALID_S,* assumes value 1 if and only if the estimates are valid at the sample level according to the rules in Cerroni et al. (2012). The second, *VALID_IND,* equals 1 if and only if the estimates are valid at the individual level according to the rule in Cerroni et al. (2012). In Equation 3.1 these variable are grouped in the set of variables called *VALIDITY*. Considering Model 2 (*a*) and 3 (*r*), the negative signs of both coefficients are consistent with result from non-parametric testing which show that average valid estimates are lower than invalid ones (see Paragraph 5.2). However, estimated coefficients are not statistically supported in either Model 2 (*a*) or Model 3 (*r*) (Table 3.6). In contrast, in Model 1, I found that *VALID_S* and *VALID_IND* have positive signs, but their influence is not statistical significant even in this case (Table 3.6).

The composition of the indicator variable *ATTITUDE* used to explain the random variable *g* strongly differs from that used to explain the other variables, *a* and *r*. For what concerns Model 1 (*g*), *ATTITUDE* captures subjects' trust in the IPCC's predictions about climate change (*IPCC_TRUST*) and their beliefs about the human and/or natural determinants of this phenomenon (*CC_HN*, *CC_H*, and *CC_HH*). In the former case, I predict that the number of days in which the infestation will occur during the blossoming period in 2030 (*g*) increases when subjects trust the IPCC' predictions because, during the instructions, I inform subjects about the positive correlation between the fire blight's infestation rate and the increase in temperature and precipitation. The coefficient of the variable *IPCC_TRUST* has the positive and statistically significant expected sign (Table 3.6). In the latter case, my results show that subjects who believe that climate change is only due to human activities (*CC_HH*) perceive the infestation to be more likely than subjects who blame the climate change on both natural and human processes (*CC_HN*). These results are consistent with the psychology literature about perceptions of risk which has shown that technology-induced risks are strongly perceived by laypeople, more than nature-induced ones (e.g., Slovic, 1987).

For what concerns Model 2 (*a*) and Model 3 (*r*), the set of variables *ATTITUDE* captures subjects' beliefs about the current and future usage of pesticides to control apple disease (*PEST_NOW* and *PEST_FUT*) and subjects' trust in Edmund Mach Foundation's predictions about the fire blight's infestation rate (*EMF _TRUST*). As I expected, the number of apples that respondents think will contain pesticides increases with subjects' agreement on the fact that farmers mainly use pesticides (*PEST_NOW*) and they keep doing this in the future (*PEST*_FUT). However these results are statistically significant in Model 3 (*r*) (at 5% and 1% significance level), but not in Model 2 (*a*). Then, I hypothesize that the number of apples containing pesticide residues in 2030 increases when subjects trust the Edmund Mach Foundation's predictions showing that the fire blight's infestation rate will increase from the 1% of 2010 to the 17% of 2030. This hypothesis is supported by the positive and significant coefficients of the variable *EMF_TRUST* in both Model 2 and 3, at 1% and 5% significance level, respectively (Table 3.6).

The *APPLE_LINK* variable set, which consists of three diverse dummy variables, *APP_LINK*, *CONSUMER*, and *CONS_ASS*, is present in all models. In Model 1 (*g*), people who work in apple production, processing, and marketing (*APP_LINK*) provide lower estimates of the number of days in which the fire blight's infestation will occur during the blossoming period in 2030 than the others. This finding was quite predictable because farmer and, more broadly, experts have a better knowledge of the actual low infestation rate in the Province of Trento. However, this coefficient is not statistically significant. As I expected, people who are tied to the apple industry (*APP_LINK*) have generally higher estimates of pesticide residues in apples than others, and the positive coefficient is statistically significant in Model 2 and 3 at the 1% and 10% significance level, respectively (Table 3.6). In this case, farmers and people who are involved in the apple industry know much better than the others that chemicals are commonly used to control apple diseases.

While the fact that the number of apples consumed weekly (*CONSUMER*) does not affect estimates regarding the fire blight's infestation rate (*g*) is not surprising, it is striking that this variable weakly influence the consumers' perceptions of pesticide residues in apples (*a* and *r*). The variable *CONSUMER* is negative and statistically significant in Model 2 (*a*) at the 5% significance level, but it is not significant in Model 3 (*r*) (Table 3.6). In contrast, I found that members of consumer associations (*CONS_ASS*) who are assumed to be very concerned about pesticide residues have higher estimates of both *a* and *r* than the others (Table 3.6). The coefficient of this variable is positive and statistically significant at 1% and 5% level, respectively.

I have the same set of socioeconomic variables in all my models. Although I found that women (*FEMALE*) have higher estimates as predicted in the literature about risk perceptions (e.g., Flynn et al., 1994; Krewski et al., 1994; Lin, 1995; Hamilton, 1985; 1995), the coefficients are not statistically significant in all of my models.

I found contrasting results for the age of subjects (*AGE*). For what concerns *g*, my results are consistent with the previous literature on health risks (e.g., Krewski et al., 1994; Williams and Hammit, 2001), as I found that elderly subjects have higher estimates of the infestation rate than the others (at the 10% significance level). In contrast, I found that the number of apples containing pesticide residue decreases with age in Model 2 (*a*) and 3 (*r*) (5% and 1% significance level, respectively) (Table 3.6). This result may be due to the fact that younger consumers are expected to be more sensitive to food-safety issues than older ones because they have all their lives in front of them.

Again, I found contrasting results about education. In Model 1 (*g*), my results support the hypothesis that more educated subjects (*UNIVERSITY*) have lower estimates of the infestation rate than the others (*SEC_SCHOOL*) as suggested by Dosman et al. (2001) and Williams and Hammit (2001). However in Model 2 (*a*) and 3 (*r*) I found that people with a master degree have higher estimates of apples containing pesticides than people with lower education levels (5% significance level) (Table 3.6). Again, this divergence may be due to the fact that highly educated subjects are expected to be more sensitive to food-safety issues.

Subjects who were born in the Province of Trento provided lower estimates of the number of apples containing at least one pesticide residue (*a*) than the others (at the 10% significance level in Model 2). I speculate that they may trust their fellow citizens or unconsciously protect their own apple products. However, the same variable was not statistically significant for what concern multiple pesticide residue in Model 3 (*r*).

Finally, I found that subjects with higher incomes have higher perceptions of the presence of pesticide than others, at least in Model 2 (*a*).

Among models which explain the perceptions of pesticides, Model 2 related to the number of apples with one or more residues (*a*) is more predictive than Model 3 related to the number of apple with multiple residues (*r*) (Table 3.6). There are various hypotheses that may explain the lower explanatory power of Model 3. First, this may be related to the discrepancy between valid and invalid probability estimates detected at individual level for variable *r*, second, boredom and fatigue may have mattered, given that half of the sample assessed the variable *r* at the end of the experiment, while in the other half the order of questions has been randomized.

Table 3.6 Generalized Linear Model Estimation

Variable	**Model 1 (*g*)**	**Model 2 (*a*)**	**Model 3 (*r*)**
VALID_S	.079	-.163	-.207
VALID_IND	.060	-.084	-.228
IPCC_TRUST	.266***	-	-
CC_H	-.196	-	-
CC_HH	.746*	-	-
EMF_TRUST	-.070	.643*	.355**
PEST_NOW	-	.082	.122**
PEST_FUT	-	-.001	.003*
APP_LINK	-.564	.653*	.400***
CONSUMER	.018	-.044**	-.005
CONS_ASS	-1.077*	1.004*	.541**
FEMALE	.018	.138	.092
AGE	.011***	-.024**	-.020*
TRENTINO	-.284	-.381***	.160
HIGH_SCHOOL	-.469	.234	.205
UNIVERSITY	-1.097*	.713**	.442**
INCOME	.001	.001**	.001
CONSTANT	-1.118*	-.578	-1.161*
LOG P.L.[§]	-73.990	-83.265	-91.752

*1% significance level, **5% significance level, ***10% significant level
[§] Log Pseudo-Likelihood

In summary, the results of my econometric analysis support many of the predictions I had about the potential factors shaping people's perceptions of the fire blight's infestation rate and the presence of pesticide residues in apples, especially those related to consumer association membership, age, and education. Moreover, by using an innovative approach, an economic experiment based on an indirect elicitation technique, I have quite consistent results with previous studies investigating the same issues with different techniques, and when I have not, I provide plausible explanations of these discrepancies.

Conclusion

Elicited subjective probabilities are important because they explain behaviors under risk and uncertainty and thus, can be used in risk-oriented behavioral models that incorporate them, such as the subjective expected utility model, or non-expected utility models. In general, empirical results in previous studies have indicated that consumers have a high level of anxiety about such contaminants in food. Using data elicited via an indirect technique applied in a laboratory experiment, I have shown that subjects are in fact not very concerned about a general increase of pesticide residues in apples at a key policy-related future date, but are more concerned about the presence of multiple residues in apples. These results have important policy implications, given the fact that consumers' subjective probabilities of pesticide residues in apples might affect their purchasing behaviors and ultimately, prices and quantities transacted in fresh fruit markets.

However, the main contribution of this essay consists of investigating the discrepancy between valid and invalid subjective probabilities. My results suggest that valid estimates are smaller than the invalid ones for what concern the number of contaminated apples (variable *a* and *r*), but greater for what concern the number of days in which the fire blight's infestation will occur in the blossoming period (*g*). However, I note that the such discrepancies are statistically supported only for variable *r* which indicates that number of apples that will contain multiple residues.

My econometric analysis has investigated factors shaping perceptions of pesticide residues in apples and it provides other useful information. For example, I found that the average consumer in my subject pool is not particularly concerned about this issue; in fact their expectations about the presence of pesticide residue do not statistically differ between apple consumers and non-consumers. In contrast, members of consumers associations are very sensitive to the problem, as they show higher probabilistic expectations. I also found that young and highly educated subjects are expected to be more sensitive to food-security issues. These results highlight, and enhance my understanding of the factors affecting perceptions and, thus, shed light on consumers' behaviors.

As a final caveat, note that my subjects were asked to answer questions about risky outcomes pertaining to a future policy period, in the year 2030. It is possible that subjects discounts the future differently than others do, which could affect each subject's probability estimates. This suggests future research to try to simultaneously estimate discount rates and subjective probabilities within the context of the EM approach that I have implemented here. To my knowledge, thus far no one has considered the elicitation of both simultaneously within the context of the EM.

CHAPTER IV. THE INCORPORATION OF SUBJECTIVE PROBABILITIES INTO CHOICE EXPERIMENTS TO TEST SCENARIO REJECTION

Introduction

In the past two decades, several stated-preference (SP) studies, mainly choice experiment (CE) applications, have shown the key influence that the specification of the status quo (SQ) alternative has on subjects' choice-behavior, and thus, on estimated welfare measures (e.g., Kontoleon and Yabe, 2003; Scarpa and Ferrini, 2005; Meyerhoff and Liebe, 2009). These studies largely focus on the scenario rejection phenomenon that occurs when subjects always reject new alternative scenarios in favor of the SQ (Cameron et al., 2010).

In essay investigates another phenomenon related to the design of the SQ alternative, a potential mental adjustment to the SQ scenario. Discrete choice' modelers have generally assumed that subjects make choices while fully accepting the attribute levels provided by the researcher in the SQ, however, recent SP studies have shown that subjects often adjust the information given in the SQ on the basis of their prior beliefs and/or expectations (e.g., Burghart et al., 2007; Cameron et al., 2010). These studies have shown that the scenario adjustment phenomenon potentially affects the reliability of SP studies because individuals may be responding to attribute levels that are not actually present in the presented scenario. Thus, if this phenomenon is not taken into account, behavioral models of decision-making might have low predictive power, and produce biased welfare estimates.

My CE application examines whether the scenario adjustment takes place when subjects are asked to make choices under risk, and, in particular, what extent subjects adjust the risk information provided in the SQ on their prior subjective risks[20]. Subjects' tendency to revise their own risk estimates once they acquire additional information has been extensively investigated in the literature within the economics of risk and uncertainty. Several studies have shown that subjects commonly update their prior subjective risks using new information, perhaps, in a Bayesian fashion (e.g., Viscusi, 1985; Viscusi, 1989). However, the weight that an individual puts on their prior versus new or experiential information is an empirical issue (Viscusi and Magat, 1992), and whether individuals are Bayesians remains controversial (Cameron, 2005b; Baker et al., 2009).

I specifically investigate to what extent a mental adjustment to the SQ scenario takes place in choices over alternative R&D programs which are geared to control the future spread of new apple diseases in the Province of Trento in Italy. As compared to the farmers' standard practice, which is to use pesticide residues, the implementation of new methods, based on natural organism and resistant varieties of apples, will reduce the number of apples containing pesticide residues by 2030. This is the year during which the spread of new diseases are predicted to occur, according to the best scientific estimates. Given the uncertainty surrounding R&D programs' outcomes, the alternatives presented in choice tasks depict the risk of having contaminated apples in 2030. Here, I refer to the "risk" of having contaminated apples as the probability that given numbers of apples will contain pesticide residues in 2030. Given this context, the scenario adjustment might easily affect subjects' choices over the alternative R&D programs. In fact subjects might either make choices by using the provided probability of having contaminated apples given by the researcher in the SQ, or, they might adjust the provided estimates based on their estimates, if the latter differs from the former. This

[20] In this Chapter, I switch to the term "risk" because SP studies commonly use this term, rather than the term "probability", for example the mortality risk literature. However, these terms are equivalent, in fact, the mortality risk is just the probability of dying.

investigation also helps to identify risk communication strategies that make people more willing to support policies that they may not initially perceive as important based on their priors.

Previous SP studies have investigated the occurrence of the scenario adjustment simulating subjects' choice-behavior and willingness-to-pay (WTP) estimates under a full acceptance of attribute levels presented in the SQ alternative (e.g., Burghart et al., 2007; Cameron et al., 2010). However, as ex-post econometric simulations may not mirror real decision making processes, and, hence, true WTP estimates (Burghart et al., 2007), in my study, I actually elicit WTP when the scenario adjustment takes place and when it does not. More specifically, I investigate the extent of this phenomenon in CEs by comparing subjects' WTP estimates when risk levels presented in the SQ alternative either coincide or do not with subjects' perceived ones. In particular, I hypothesize that when the presented risk levels are lower than expected ones, subjects positively adjust the information given in the SQ on their expectations, and, hence, they provide higher WTP estimates. In contrast, when the presented risk levels are higher than expected ones, subjects negatively adjust the information given in the SQ on their expectations, and, hence, they provide lower WTP estimates.

To investigate subjects' choice behavior when the risk level presented in the SQ coincide with subjects' perceived ones, I design subject-specific SQ alternatives based on each subject's subjective probabilities of having contaminated apples in 2030 if farmers will continue to use conventional chemical controls. This study, to my knowledge, represents the first attempt to incorporate subjective probabilities into a CE design. In order to accomplish this, I implement a best-worst Pivot CE. Pivot CE are extensively used in transport economics to generate subject-specific SQ alternatives based on the information that each subject provides about her/his most recent trip.

Afterwards, attribute levels of other alternatives are generated by pivoting them on the attribute levels of the SQ alternative by adding or subtracting given percentages or values from the baseline attribute levels (e.g., Hensher and Greene, 2003; Hensher and Rose, 2007; Hensher et al., 2009).

To generate the subject-specific SQ alternatives, subjective probabilities that given numbers of apples will contain pesticide residues in 2030 have been elicited using the Exchangeability Method, an elicitation techniques based on the de Finetti's notion of exchangeable events (de Finetti, 1937; 1974a; 1974b). This innovative method indirectly elicits subjective probabilities by asking subjects to play lotteries containing uncertain outcomes occurring in the future (Baillon, 2008; Abdellauoi et al., 2011, Cerroni and Shaw, 2012; Cerroni et al., 2012a).

In the remainder of the essay, I first review previous finding about the concept of a mental scenario adjustment. Next, I describe the CE survey, provide testable hypotheses, and present my discounted Expected Utility Theory driven models. In the final section, I offer some conclusions based on the experimental results that were obtained.

Literature review

Realism and scenario adjustment

CE studies, and more broadly SP applications, generally investigate subjects' choice-behavior in hypothetical markets that have been designed by researchers. These techniques are extremely useful to investigates the demand for goods and services which either are not yet in the market or do not have a market at all. The latter is the case of my R&D programs. However, as subjects' choices in such hypothetical markets may or may not mimic their choices in real situations, a lot of effort has been put in designing SP investigations where hypothetical markets appears as real as possible to subjects[21]. One way to create hypothetical, but realistic choice contexts, among many others[22], consists in designing a baseline scenario that subjects perceive to be real. Recently, discrete choice research, much of which is been in the field of transport economics, has advanced knowledge on how to construct realistic choice scenarios (e.g., Adamowicz et al., 1997; Hess and Rose, 2009).

A lack of realism in characterizing the SQ alternatives might undermine the credibility of the study, inducing subjects to express untruthful preferences and researchers to infer biased estimates. This has been an issue in health, transportation, and, other fields in economics, prevalently, in the context of choice under conditions of risk. For example, despite the fact that, many studies provide all subjects with the same SQ alternatives in which an average science-based estimate of mortality risk due to a given illness is presented, these risks may depend a great deal on specific ages, gender, and other factors. To overcome this issue, other SP studies have created group-specific SQ alternatives in which mortality risk estimates depend at least on age and gender (e.g., Krupnick et al., 2002; Alberini et al., 2006).

Incorporating "realism" into SP is difficult, in fact, even supposedly realistic baseline scenarios might not be credible to all subjects. An extensive research within psychology and, to some extent, in economics, has demonstrated that subjective risk perceptions widely differ from risks that subjects currently experience in their life (e.g., Slovic, 1987; Botzen et al., 2009; Jakus et al., 2009). It is easy to imagine that any discrepancy between subjective estimates and science-based ones becomes even larger when future and uncertain outcomes are taken into account (e.g., Cameron, 2005b; Cerroni et al., 2012b).

[21] This was one of the first recommendations for SP studies highlighted in the NOAA guidelines.
[22] Another way consists in making CE incentive compatible when the case study allows this.

When subjects are presented with SQ alternatives where risk estimates are not consistent with their subjective estimates, they have three main decision-making strategies that they can adopt. First, they can ignore their own beliefs by assuming that science-based estimates presented in the SQ alternative are right and credible. These subjects essentially abandon any priors they have, putting full weight on new risk information they receive. Second, they can at least partially adjust the information given in the SQ alternative on their own subjective estimates. To be clear, they mentally adjust the risk information provided in baseline scenario to better fit their risk perceptions. This phenomenon is becoming commonly known in the literature as the scenario adjustment (Bughart, 2007; Cameron et al., 2010). Third, taking this behavior to the extreme, they can completely ignore the information provided in the SQ and make choices according to their subjective estimates. These subjects essentially put zero weight on new information, clinging to their prior, which might be based on some personal knowledge or experience.

If the adjustment phenomenon occurs, the SQ used by subjects during their choices differs from the one researchers consider in the choice modeling. This might generate confounding factors that researchers are not able to capture in their models, and, therefore, compromise the accuracy of welfare estimates (Cameron et al., 2010).

Two approaches have been identified to deal with the scenario adjustment in SP studies. Both rely on the collection of additional information about subjects' beliefs or expectations of the levels in one or more key attributes that describe the SQ alternative.

The first approach investigates to what extent the scenario adjustment affects subjects' choices and, hence, their welfare estimates by using simulations. In choice models, the elicited information is interacted with utility parameters to control for the presence of the scenario adjustment. The estimated coefficients of these interaction terms indicate to what extent the adjustment takes place. This information is commonly elicited by using quite simple debriefing questions which simply ask subjects what would have been the SQ' attribute levels that they expected to face in the choice tasks. The fact that such questions are presented to subjects after they have taken their choices, might have influenced their answers. For example, compliant respondents might formulate an estimate that is more in line with the SQ's attribute than their real prior. This would reduce the effect of the scenario adjustment. Subjects' choice-behavior and WTP estimates under a full acceptance of the SQ scenario are ex-post simulated. By using this approach, some stated choice studies dealing with risk attributes have detected scenario adjustment and they have shown the substantial influence that this phenomenon has on welfare measures (Burghart et al., 2007; Cameron et al., 2010). However, as ex-post simulated choices might strongly depend on the used econometric specification, this approach might generate biased estimates of the effect that the scenario adjustment produces on subjects' behavior and WTP (Burghart et al., 2007).

In contrast, the second approach aims to avoid the scenario adjustment and all related issues, rather than investigate this phenomenon. This approach, developed in transportation studies, relies on the design of more realistic CE survey by using pivot experimental designs (e.g., Hensher and Rose, 2007; Hensher et al., 2009). As SQ alternative's attribute levels provided by researchers do not generally coincide with the trips that commuters often make to reach given destinations, in Pivot CEs, each subject, in each choice task, is presented with a specific SQ alternative where attribute levels are based on her/his most recent commuting trip. To generate such a design, attribute levels of the SQ are first elicited from subjects themselves, and, then, used to design the attribute levels of the other alternatives presented in the choice tasks. In particular, the latter attribute levels are generated by adding and/or subtracting given amounts or percentages from the corresponding attribute level in the SQ alternative. Unfortunately, Pivot CE has been shown to induce subjects to systematically prefer the realistic SQ alternative over the hypothetical generated alternatives (e.g., Hess and Rose, 2009; Rose and Hess, 2009).

To my knowledge this approach has never been implemented in other field than transportation, and, more important, has never been used to investigate risky choices. However, since a long tradition in psychology (e.g., Slovic, 1987), now spilling over into economics, has shown that subjective risks are often better predictors of choice rather than science-based estimates, Pivot CE might be used for incorporating subjective estimates into stated choice experiments to better predict choices under situations of risk. Despite the fact that economists, more than psychologists, have put a lot of effort in incorporating subjective risks into modeling behavior (see early work on smoking decisions by Viscusi (1990); more recently see Mansky, 2004; Cameron, 2005a; Riddel and Shaw, 2006; Viscusi and Zeckhauser, 2006; Botzen et al., 2009; Jakus et al., 2009; Shaw et al., 2012), to my knowledge, subjective risk estimates have never been directly incorporated in stated choice's experimental designs.

Elicitation of Subjective probabilities

As this essay investigates whether the scenario adjustment takes place in a risky choice context, the investigation of subjective probabilities becomes crucial to understand how subjects react to the risk information provided in the SQ alternative. To accomplish this, I need to elicit the subjective probability that given outcomes will occur in the future.

There is an extensive literature in decision analysis and management science, now spilling over into behavioral and experimental economics, about the elicitation of subjective probabilities related to financial outcomes. Few has been done in other fields (e.g., Viscusi, 1990; Wakker and Deneffe, 1996; Cerroni and Shaw, 2012).

In previous stated choice studies, subjective probabilities have been commonly elicited by using the so called direct techniques which consist in directly asking subjects to express the probability that given outcomes will occur in the future (e.g., Viscusi, 1990; Williams and Hammit, 2001; Riddel and Shaw, 2006). Although the latter approach is very appealing for its simplicity, it may generate biased results as subjects are often not willing and/or able to express probabilities in numerical terms (Koriat et al., 1980; Zimmer, 1983)[23].

An alternative way for eliciting subjective probabilities consists in asking subjects to play lotteries. In these techniques, called indirect, probability estimates are indirectly estimated at the point for which subjects becomes indifferent to choose playing one lottery instead of the others (Spetzler and Stael Von Holstein, 1975).There are many variations on this theme[24], but a novel approach deserves to be mentioned, the exchangeability method (EM). This elicitation technique consists of a set of binary questions in which subjects are asked to bet a certain amount of money on one of the two disjoint subspaces that come from the bisection partition of the whole state space of the variable under study. The sectioning process depends on subjects' betting-behavior, and proceeds until subjects become indifferent to bet on one disjoint subspace rather than on the other. When this point is reached, subjects are assumed to perceive those subspaces as equally likely (Spetzler and Von Holstein, 1975). This method allows eliciting several percentiles of each subject's cumulative distribution function (CDF) of

[23] One might argue that, subjective probabilities do not need to be elicited, but they can be inferred from subjects’ choices. Unfortunately, in this study, the elicitation of subjective probabilities is necessary to investigate how subjects react when provided with risk information which differs from their prior risk estimates.

[24] To keep the essay of a manageable length we refer interested readers to Cerroni et al. (2012a).

the random variable under study. This approach is particularly appealing because outcomes are not associated to probability measures, and, hence, unlike other techniques, it does not force individuals to process numerical probability estimates (e.g., Baillon, 2008; Abdellaoui et al., 2011, Cerroni and Shaw, 2012).

In Chapter 2, I have shown that the validity of percentile estimates elicited via the EM depends on the ordering of questions and the provision of monetary incentives to subjects based on their betting-behavior during the tasks[25] (Cerroni et al., 2012a). Here, validity is tested using the de Finetti's (1937) notion of coherence, under which subjective probabilities are valid if and only if they obey to Probability Theory. On the one hand, the ordering of questions affects validity only when more than one percentile estimates of the random variable under study are elicited, otherwise not[26]. This is due to the fact that the number of binary questions needed to elicit the first percentile is so small that subjects do not become aware of the chaining nature of the elicitation procedure. On the other hand, the provision of real monetary increases the validity of elicited percentile estimates, for example, when monetary incentives are hypothetical only 38 percent of the elicited first percentile estimates are valid, while, when monetary incentives are real the percentage of valid first percentile estimates is almost 50 percent (Cerroni et al., 2012a). However, in Chapter 3, I have shown that validity does not impact the magnitude of percentile estimates, and hence, valid and invalid observations do not differ from each other (Cerroni et al., 2012b).

---

[25] In experimental economics, monetary incentives are commonly provided as they are assumed to induce subjects to state their real beliefs, expectations, or preferences, at least, when incentive compatible elicitation techniques are used. Cerroni et al. (2012a) have rewarded subjects whose expectations were consistent with science-based predictions as subjective probabilities were elicited for outcomes occurring in 2030, too far in the future to wait their realization.

[26] Cerroni et al. (2012b) have elicited three percentile estimates for each of the three variables that they investigated in their experiment. Each percentile estimates of each variable was elicited by using a specific block questions. In two experimental treatments these blocks of questions were randomized to hide the chained sequence of questions.

Objectives and testable hypotheses

In this essay, I develop a novel two-stage approach to investigate the extent of the scenario adjustment by comparing marginal WTP (MWTP) estimates between subjects who might potentially adjust SQ's risk levels on their own estimates, and subjects who might not. In the first stage, subjective probabilities that given outcomes will occur are elicited, and, in the second stage, the sample is split into two treatment groups. In the Subjective SQ (SSQ) treatment, the risk presented in the SQ is consistent to each subject's probability estimate, while in the Objective SQ (OSQ) treatment, it is not. The OSQ treatment group is further split into two other sub-groups. In one, the risk depicted in the SQ is lower than each subject's estimate (OSQ_{LOW}), while, in the other, it is higher than that (OSQ_{HIGH}).

This approach implies the incorporation of subjective probabilities, elicited using the EM method, into the CE's experimental design by using the pivot approach. While previous pivot CE applications have been commonly used to generate SQ alternatives which mirror the most recent trip that subjects have experienced, my investigation represents the first attempt of using pivot CE to create SQ alternative which are consistent and coherent with subjective probabilities of future outcomes. To my knowledge, this is also the first study using subjective probabilities elicited via the exchangeability method to model subjects' choice behavior.

In this essay, the scenario adjustment is investigated by testing the following hypotheses[27]:

[27] These hypotheses allows me to test the scenario adjustment under the Expected Utility Theory framework. In other non-standard theories of decision making under risk and uncertainty, such as Cumulative Prospect Theory and Rank Dependent Utility Theory, the reference point affects subjects' choices.

Hypothesis 1.

H$_0$: $MWTP_{SSQ} \geq MWTP_{OSQ\_LOW}$

H$_1$: $MWTP_{SSQ} < MWTP_{OSQ\_LOW}$

If the null hypothesis (H_0) is rejected, subjects who belong to the OSQ_{LOW} treatment positively adjust the risk information provided in the SQ on their estimates. As they make choices having in mind SQ's attribute levels greater than those provided in the SQ, their MWTP estimates are greater than those of subjects who belong to the SSQ treatment. In contrast, if the null hypothesis (H_0) is not rejected, subjects who belong to the OSQ_{LOW} fully accept the risk information given by the researcher (i.e., no scenario adjustment), or they negatively adjust the risk information provided in the SQ because they overreact to such information.

Hypothesis 2.

H$_0$: $MWTP_{SSQ} \leq MWTP_{OSQ\_HIGH}$

H$_1$: $MWTP_{SSQ} > MWTP_{OSQ\_HIGH}$

If the null hypothesis (H_0) is rejected, subjects who belong to the OSQ_{HIGH} treatment negatively adjust the risk estimate provided in the SQ on their subjective estimates. These subjects have lower MWTP than subjects who belong to the SSQ treatment because they unwarily generate in their mind risk estimates lower than that presented in the SQ. In contrast, if the null hypothesis (H_0) is not rejected, the scenario adjustment does not occur, or it occurs in the opposite direction (i.e., positive adjustment) as they overreact to such information.

The empirical application

This essay investigates people's preferences and, therefore, implies their maximum willingness to pay (WTP) for R&D strategies proposed by the Province of Trento to control new apple diseases. These diseases will likely develop in this region by 2030, according to the best available science. An example is the fire blight, a bacterial phytopathology that has already damaged and killed some apple orchards in the Province of Trento, at least since 2003 (EMF, 2006).

Farmers might need to use other new and more effective pesticides than ones being currently used. For example, they might introduce the antibiotic streptomycin that is currently forbidden by the Italian legislation, but which is already used in U.S., Germany, Belgium and The Netherlands for controlling fire blight (Németh, 2004). The usage of the chemical control will affect the presence of pesticide residues on apples, which is already quite high today. In fact, 63 apples out of about 100 contain pesticide residues, according to scientific data (Italian Health Ministry, 2010).

Given the fact that pesticides pose health risks to people who eat apples, the Province of Trento plans to launch R&D programs which study, develop, and implement alternative methods to control the future spread of new diseases. These programs are based on both the identification of natural organisms that are antagonists of causal pathogens, and the development of resistant varieties of apples that will be unaffected by new diseases. The introduction of these new methods will have a positive effect on the number of apples containing pesticide residue in 2030. Such R&D programs are funded thanks to a specific tax that the Province of Trento will ask the population to pay an annual sum in the period between 2012 and 2030[28].

[28] Subjects might object the tax as they might believe that farmers should pay for R&D programs. However this risk is avoided as agricultural R&D programs are commonly supported by public funds in the Province of Trento.

Methodology

My survey differs from conventional SP, and, especially from previous CE, studies. After introducing the empirical scenario to subjects, another section in which subjective probabilities are elicited by using the EM is added (Baillon, 2008, Abdellaoui et al., 2011; Cerroni et al. 2012a). Afterward, subjects' preferences for alternatives R&D programs aiming to reduce the risk of having contaminated apples in 2030 by using a best-worst pivot CE are elicited. In this section, the sample is divided into treatment groups, each presented with a specific version of the CE which differs from the others in the design of the SQ alternative, and, more specifically, in the risk of having contaminated apples in 2030 presented in the SQ alternative.

Below, I describe my empirical application of the exchangeability method and the best-worst pivot CE as well as the sampling procedure and my experimental treatments.

Exchangeability Method

In my application of the EM, the random variable under study (a) is the number of apples, produced in the Province of Trento that will contain pesticide residue in 2030 if farmers will control the spread of new diseases by using pesticides. Only the 50th percentile of each subjects' CDF is elicited ($a_{1/2}$). In the first step of the EM, subjects are asked to express the lower and upper bounds of the state space of variable a (S_a), defined as a_{min} and a_{max}. These bounds contain all outcomes that have a non-zero probability to occur. For example, if subject i believes that $a_{i,min}$=70 and $a_{i,max}$=86, then, she/he implicitly assumes that only outcomes belonging to this range will occur.

In the second step of the EM, subject i is asked to answer a series of binary questions that reveal the 50th percentile of the her/his subjective CDF ($a_{i,1/2}$). In the first binary question, S_a is divided at a point a_1 into two prospects, say A_1={a_{min}<x<a_1} and A_1'={a_1≤x<a_{max}}, where a_1={a_{min}+[(a_{max}-a_{min})/2]}. To my subject i, $a_{i,1}$={70+[(86-70)/2]}=78 apples, and, thus, the first binary question asks her/him to bet on prospect A_1={70<x<78} or prospect A_1'={78≤x<86}. If prospect A_1 is chosen by the subject i, the implication is that she/he believes the probability of occurrence of the sub-event A_1 is greater than that of the sub-event A_1', so that $P(A_1)≥P(A_1')$ and $a_{i,1}≥a_{i,1/2}$, and thus, $P(70<x<78) ≥ P(78≤x<86)$ and $78≥ a_{i,1/2}$. This process is repeated until subject i reaches a value $a_{i,1+z}$ (with z=1,2,...,n) such that she/he is indifferent between A_{1+z} and A_{1+z}'. When this point is reached, it follows that $P(A_{1+z})=P(A_{1+z}')$ and $a_{i,1/2}=a_{i,1+z}$. For example, assume that subject i was indifferent between prospect A_{1+z}={70<x<74} and prospect A_{1+z}'={74≤x<76}, this implicitly means that $P(70<x<74)=P(74≤x<76)$ and $a_{i,1/2}$= 74. To conclude, my subject i believes that there is 50% chance that the number of apples containing pesticide residue will be between 70 ($a_{i,min}$) and 74 ($a_{i,1/2}$), and another 50% chance that it will be between 74 ($a_{i,1/2}$) and 86 ($a_{i,max}$). For simplicity's sake, at the end of the task, subject i is presented with a summary screen-shot in which he/she is informed that, based on her/his choice-behavior, there is 50% chance that the number of apples containing pesticide residues will be 74 ($a_{i,1/2}$), at the worst, and another 50% chance that it will be 86 ($a_{i,max}$), at the worst. As a check, each subject is asked to confirm her/his estimate[29].

In this application, binary questions are not randomized and monetary incentives are not provided to subjects based on their choice-behavior. While the ordering of questions does not matter in this study as only first percentile estimates are elicited, the lack of monetary incentives might have undermined the validity of elicited observations

[29] The majority of our subjects confirmed estimates inferred from their choice-behavior.

(Cerroni et al., 2012a). Nevertheless, in order to design a coherent survey, given the hypothetical nature of the choice context presented in the CE, a hypothetical version of the EM is implemented. The choice context is inevitably hypothetical because R&D programs' outcomes will be available only in 2030. However, as Cerroni et al. (2012b) have shown that validity does not affect the magnitude of percentile estimates elicited via the EM, here, I assume that, being elicited percentile estimates coherent or not with probability theory, they still mirror subjects' expectations about the number of apples that will contain pesticide residues in 2030[30].

Best-worst pivot choice experiment

After having interviewed 34 subjects during three focus-group meetings, three key attributes were selected to describe the effect of the R&D programs on the presence of pesticide residues in apples[31]. These are:

(i) the maximum number of apples containing pesticide residues in a sample of a hundred in 2030 (N),

(ii) the probability of this number N occurring (P), and

(iii) the yearly tax in euro that taxpayers of the Province of Trento must pay in the period between 2012 and 2030 if they want R&D programs to be launched in 2012 (T).

In the CE application, each subject is presented with 12 choice tasks, containing each three alternatives. Using the best-worst approach, subjects are asked to select the

[30] Although, in this study, I did not investigate the validity of elicited subjective probabilities, I am aware of the fact that subjects' choices based on valid and invalid probability measures might differ from each other. This could be another interesting topic to research on.
[31] Focus-group meetings were conducted on July 4th, 16th, 23th. 2011 at the Department of Economics, University of Trento (Italy).

most and least preferred alternatives in each choice task. In this application subject can choose to indicate first either their most preferred or least preferred alternative. If subjects indicate first their most preferred alternative, then they are asked to indicate their least preferred alternatives. In contrast, if subjects indicate first their least preferred alternative, then they are asked to indicate their most preferred alternatives. The main advantage of using best-worst CEs is in the availability of more data from each subject, enhancing the value of small samples. If subject i is presented with a choice task containing a set of alternatives J, I can assume that she/he chooses her/his most preferred alternative in each of J-1 sequential choice tasks, each containing one alternative less than the previous choice task (Scarpa et al., 2010). Another advantage of the best-worst approach, as compared to rating or simple ranking, is that subjects can more easily and consistently identify extreme options in terms of preference. In contrast, they are more cognitively demanding than standard choice experiment where subjects are asked to indicate only their preferred alternative (Marley and Louviere, 2005; Vermeulen et al., 2010).

In the SQ alternative, no R&D program is launched by the Province of Trento and, thus, farmers will control new diseases by spraying new pesticides in 2030. Given the very long time-horizon for events to evolve, the number of contaminated apples in 2030 cannot be known with certainty, thus the SQ looks like a lottery which consists of two prospects, Prospect A and B. In Prospect A, there is a given chance $P(N_{A,SQ})$ that the maximum number of contaminated apples in 2030 will be $N_{A,SQ}$; in Prospect B, there is a given chance $P(N_{B,SQ})$ = 1- $P(N_{A,SQ})$ that the maximum number of contaminated apples in 2030 will be $N_{B,SQ}$. As any R&D program is implemented, there is no tax to pay in the SQ alternative (Table 4.1).

Table 4.1 Attribute levels for the SQ		
Attribute	**Prospect A**	**Prospect B**
Maximum number of apples containing pesticide residues in 2030	$N_{A,SQ}$	$N_{B,SQ}$
Probability of occurrence	$P(N_{A,SQ})$	$1\text{-}P(N_{A,SQ})$
Yearly tax to pay in the period 2012-2030		0€

As noted above, in addition to the SQ, which allows subjects to reject the other alternatives in favor of the baseline scenario, subjects are presented with other two alternatives in every choice task. In these alternatives, the Province of Trento will launch an R&D program to develop new methods to control new disease in 2030. Such methods will reduce the number of apples containing pesticide residues in 2030, as compared to the baseline scenario depicted in the SQ alternative. In this case, not only the very long time-horizon, but also the uncertainty related to the effectiveness of R&D programs, makes an estimate of contaminated apples quite uncertain. Hence, each hypothetical alternative presented in each choice task is a lottery which consists of two prospects, Prospect A and B. In Prospect A, there is a given chance $P(N_{A,R\&D})$ that the maximum number of contaminated apples in 2030 will be $N_{A,R\&D}$; in Prospect B there is a given chance $P(N_{B,R\&D})$ =1- $P(N_{A,R\&D})$ that the maximum number of contaminated apples in 2030 will be $N_{B,R\&D}= N_{B,SQ}$. As R&D programs will reduce the presence of pesticide residues in apples, and, thus $N_{A,R\&D}<N_{A,SQ}$, I have generated three levels for $N_{A,R\&D}$ by using the pivot approach, and more specifically, the following algorithms, $N_{A,SQ}$ – 40%, $N_{A,SQ}$ – 60%, and $N_{A,SQ}$ – 80% (Table 4.2). On the other hand, as the effectiveness of R&D programs is highly uncertain, and, thus, $P(N_{A,R\&D}) \leq P(N_{A,SQ})$ and $1\text{-}P(N_{A,R\&D}) \geq 1\text{-}P(N_{A,R\&D})$, I created the pivoted four levels for $P(N_{A,R\&D})$ by using the

following algorithms, $P(N_{A,SQ})$ - 0%, $P(N_{A,SQ})$ - 50%, $P(N_{A,SQ})$ – 80%, and $P(N_{A,SQ})$ – 90% (Table 4.2).

Given that R&D programs are implemented, hypothetical alternatives implies that tax-payers of the Province of Trento must pay a specific tax to financially support such programs. The selected levels for the tax attribute (T) were the following, 15€, 30€, 50€, and 80€ (Table 4.2). These levels were determined to be appropriate based on previous related studies, as well as taking into account focus group participants' opinions and expectations about R&D programs and their costs.

Table 4.2 Attribute level for R&D plans

Attribute	**Prospect A**	**Prospect B**
Maximum number of apples containing pesticide residues in 2030	$N_{A,SQ}$ – 40% $N_{A,SQ}$ – 60% $N_{A,SQ}$ – 80%	$N_{B,SQ}$
Probability of occurrence	$P(N_{A,SQ})$ – 0% $P(N_{A,SQ})$ – 50% $P(N_{A,SQ})$ – 80% $P(N_{A,SQ})$ – 90%	1-[$P(N_{A,SQ})$ – 0%] 1-[$P(N_{A,SQ})$ – 50%] 1-[$P(N_{A,SQ})$ – 80%] 1-[$P(N_{A,SQ})$ – 90%]
Yearly tax to pay in the period 2012-2030		15€ 30€ 50€ 80€

In this study, I used a *D*-efficient homogeneous pivot design that has been generated through a two-step procedure. In the first step, by running a pre-test CE survey[32] prior coefficients of my attributes were estimated, and, then used to generate a *D*-efficient design. Given my 3×4^2 factorial design of my pre-test study, I have

[32]The pre-test CE survey was conducted in the period from November 14th and 19th, 2011. The sample consists of 80 randomly selected subjects in the Province of Trento. Subjects were interviewed by appointment in their own home.

generated a simple optimal orthogonal design with four blocks of 9 choice tasks by using Ngene 1.1.1. Reference levels and segment weights[33] of my homogeneous pivot design were obtained by examining the median percentile estimates of the number of apples containing pesticide residues in 2030 elicited via the EM by Cerroni et al. (2012a)[34]. A homogeneous pivot design was chosen, rather than a heterogeneous one, because the former allows us to generate a single design that can be used for all individual-specific SQ alternatives. As subjects face the same experimental design whatever treatments they belong to, confounding factors due to the use of different designs across treatments are avoided[35]. The final design was generated again by using Ngene 1.1.1.

Experimental treatments and sampling procedure

The final sample consists of 797 taxpayers who reside in the Province of Trento[36]. Data were collected by trained interviewers using the computer-assisted personal interviewed (CAPI) system which consists in face-to-face interviews usually conducted at respondents' home or business via a portable personal computer. Data obtained from each subject were automatically stored in a central computer. Hour, date and place for the interviews were previously arranged by phone calls during which interviewers ascertain themselves that subjects were taxpayers living in the Province of Trento.

The Subjective SQ (SSQ) treatment group consists of 487 subjects randomly selected from the full sample of 797 people, and the Objective SQ (OSQ) treatment

---

[33]Reference levels define the number of individual-specific SQ alternatives, while segment weights define the number of subjects that fall in each reference level. Weights are needed to calculate the AVC matrix of the design (ChoiceMetrics, 2011).
[34] This is a quite standard procedure, even though it is not perfect. The best would be to elicit SQ' attribute levels and generate an orthogonal design, then, based on subjects' choice in each choice task progressively generates an efficient pivot design. To my knowledge, this has never been done yet.
[35] The number of simulate respondents was 500, the number of Halton random draws was 800.
[36] The survey was conducted in the period between January 24th and March 12th, 2012.

group has 310 randomly selected subjects. To ensure comparability across treatments the same sampling procedure was used across groups, more specifically, a stratified proportional sampling for what concerns age and population of the Province of Trento's eleven valleys[37].

In the Subjective SQ, each subject *i* is presented, with a SQ alternative (No R&D Program) which specifies the risk of having contaminated apples in 2030, presuming that farmers will control new diseases using chemicals. The risk is elicited from her/him by using the exchangeability method. The Subjective SQ consists of a lottery containing two risky prospects: in Prospect A, there is a 50% chance that the maximum number of contaminated apples in 2030 will be $N_{A,SQ} = a_{i,1/2}$, the 50th percentile estimates of each subject's CDF. In prospect B, there is a 50% chance that the maximum number of contaminated apples in 2030 will be $N_{B,SQ} = a_{i,max}$, the 100th percentile estimates of each subject's CDF. Recall that, as the SQ implies that no R&D Program will be implemented, there is no tax to pay for subjects (Tables 4.3 and 4.6).

Table 4.3 Choice Task 1 for subject *i* in the SSQ treatment

	R&D Program X		**R&D Program Y**		**NO R&D Program**	
	Prospect A	**Prospect B**	**Prospect A**	**Prospect B**	**Prospect A**	**Prospect B**
Maximum number of apples containing pesticide residues in 2030	$a_{i,1/2}$–80%	$a_{i,max}$	$a_{,1/2}$-40%	$a_{i,max}$	$a_{1/2,i}$	$a_{i,max}$
Probability of occurrence	10%	90%	25%	75%	50%	50%
Yearly tax to pay in the period 2012-2030	15€		50€		0€	

[37] The Province of Trento is administratively divided in 11 valley communities.

In the Objective SQ treatment, each subject *i* is presented in the SQ alternative, with a probability of having contaminated apples in 2030 assuming that farmers will control new diseases using chemicals, and this probability differs from the one she/he expressed through the EM.

Subject *i* was assigned to one treatment subgroup, rather than to the other, based on her/his 50th percentile estimate ($a_{i,1/2}$) that has been previously elicited by using the EM. In fact, if subjects *i's* 50th percentile estimate falls between 76 and 100 apples ($76 \leq a_{i,1/2} \leq 100$), she/he belongs to the SQ_{LOW} treatment, while if it falls between 50 and 74 apples ($50 \leq a_{i,1/2} \leq 74$), she/he belongs to the SQ_{HIGH} treatment.

This “splitting” rule which aims to generate the same sample size across subgroups was defined using experimental results by Cerroni et al. (2012a) about the number of subjects who have the same 50^{th} percentile estimates of the numbers of apples containing pesticides in 2030. The reliability of this approach is supported by the fact that both treatment groups consists of 155 subjects. Unfortunately, this procedure may have affected the composition of my subsamples which, in this study, should be similar across treatment groups, as key socioeconomic variables likely affect willingness to pay for R&D programs. Fortunately, having data on these variables allows control via additional econometric modeling. To detect variables that must be included in the choice models to control their effect on WTP, a very simple logit selection model will be run as you will see below.

In the Objective SQ_{LOW} treatment, if subjects *i's* 50th percentile estimate falls between 76 and 86 apples ($76 \leq a_{i,1/2} \leq 86$), the SQ alternative’s prospect A reports that there is a 50% chance that the maximum number of contaminated apples in 2030 will be

$N_{A,SQ}$=65, which is lower than 50th percentile estimates of each subject's CDF ($a_{i,1/2}$), while prospect B informs the subject that there is a 50% chance that the maximum number of contaminated apples in 2030 will be $N_{B,SQ}$=75, which is lower than the 100th percentile estimates of each subject's CDF ($a_{\max,i}$). In contrast, if subjects *i's* 50th percentile estimate is between 88 and 100 apples ($88 \leq a_{i,1/2} \leq 100$), in Prospect A, $N_{A,SQ}$=75 apples will be contaminated, at the worst, with 50% chance, in Prospect B, $N_{B,SQ}$=85 apples will be contaminated, at the worst, with 50% chance (Table 4.4 and 4.6).

Table 4.4 Choice Task 1 for subject *i* in the OSQ_{LOW} treatment

	R&D Program X		R&D Program Y		NO R&D Program	
	Prospect A	Prospect B	Prospect A	Prospect B	Prospect A	Prospect B
Maximum number of apples containing pesticide residues in 2030	65–80% or 75–80%	75 or 85	65-40% or 75-40%	75 or 85	65 or 75	75 or 85
Probability of occurrence	10%	90%	25%	75%	50%	50%
Yearly tax to pay in the period 2012-2030	15€		50€		0€	

In the objective SQ_{HIGH} treatment, if subjects *i's* 50th percentile estimate falls between 50 and 66 apples ($50 \leq a_{i,1/2} \leq 66$), Prospect A reports that there is a 50% chance that the maximum number of contaminated apples in 2030 will be $N_{A,SQ}$ =75, which is higher than 50th percentile estimates of each subject's CDF ($a_{i,1/2}$), and Prospect B informs that there is a 50% chance that the maximum number of contaminated apples in 2030 will be $N_{B,SQ}$=85, which is higher than the 100th percentile estimates of each subject's CDF ($a_{i,max}$). On the other hand, if subjects *i's* 50th percentile estimate is between 68 and 74 apples ($68 \leq a_{i,1/2} \leq 74$), in prospect A, $N_{A,SQ}$ =90 apples will be contaminated, at the worst, with 50% chance, in prospect B, $N_{B,SQ}$ = 100 apples will be contaminated, at the worst, with 50% chance (Table 4.5 and 4.6)[38].

Table 4.5 Choice Task 1 for subject *i* in the OSQ$_{HIGH}$ treatment

	R&D Program X		**R&D Program Y**		**NO R&D Program**	
	Prospect A	**Prospect B**	**Prospect A**	**Prospect B**	**Prospect A**	**Prospect B**
Maximum number of apples containing pesticide residues in 2030	75–80% or 90–80%	85 or 100	75-40% or 90-40%	85 or 100	75 or 90	85 or 100
Probability of occurrence	10%	90%	25%	75%	50%	50%
Yearly tax to pay in the period 2012-2030	15€		50€		0€	

[38] Two diverse SQ alternatives were designed for each objective SQ treatment groups because of the deep uncertainty surrounding scientific predictions of the number of apples containing pesticides in 2030 in the Province of Trento.

Table 4.6 Summary statistics

Summary statistics of variables in the Subjective SQ treatment

Variable	N. Obs.		Mean		St. Dev		Min		Max	
	SQ	R&D	SQ	R&D	SQ	R&D	SQ	R&D	SQ	R&D
N_A	5,844	5,844	76.480	30.590	6.530	12.813	64	13	98	59
N_B	5,844	5,844	87.188	34.874	9.710	14.863	66	13	100	60
$P(N_A)$	5,844	5,844	0.5	0.225	0	0.175	0.5	0.05	0.5	0.5
$1\text{-}P(N_A)$	5,844	5,844	0.5	0.775	0	0.175	0.5	0.5	0.5	0.95
T	5,844	5,844	0	43.750	0	24.337	0	15	0	80
REDD		5,844		17,012.320		11,103.230		5,000		120,000

Summary statistics of variables in the Objective SQ_{LOW} treatment

Variable	N. Obs.		Mean		St. Dev		Min		Max	
	SQ	R&D	SQ	R&D	SQ	R&D	SQ	R&D	SQ	R&D
N_A	1,860	1,860	68.290	27.316	4.699	11.338	65	13	75	45
N_B	1,860	1,860	78.290	31.316	4.699	12.948	75	15	85	51
$P(N_A)$	1,860	1,860	0.5	0.225	0	0.175	0.5	0.05	0.5	0.5
$1\text{-}P(N_A)$	1,860	1,860	0.5	0.775	0	0.175	0.5	0.5	0.5	0.95
T	1,860	1,860	0	43.750	0	24.341	0	15	0	80
REDD		1,860		25,870.970		19,022.490		5,000		120,000

Summary statistics of variables in the Objective SQHIGH treatment

Variable	N. Obs.		Mean		St. Dev		Min		Max	
	SQ	R&D	SQ	R&D	SQ	R&D	SQ	R&D	SQ	R&D
N_A	1,860	1,860	87.967	35.187	5.134	14.539	75	15	90	54
N_B	1,860	1,860	97.967	39.187	5.134	16.155	85	17	100	60
$P(N_A)$	1,860	1,860	0.5	0.225	0	0.175	0.5	0.05	0.5	0.5
$1\text{-}P(N_A)$	1,860	1,860	0.5	0.775	0	0.175	0.5	0.5	0.5	0.95
T	1,860	1,860	0	43.750	0	24.341	0	15	0	80
REDD		1,860		25,451.610		13,931.760		5,000		120,000

Modeling, estimation, and welfare measures

Selection Model

To identify variables that affect the composition of my treatment groups, and, therefore might potentially influence WTP estimates inferred from my choice models, a simple Logit selection model is estimated where the probability of belonging to the SSQ treatment group rather than to the other treatment depends on a set of variables indicating the socioeconomic status and attitudes of subjects. This set contains variables that the literature in food choices under conditions of risk has shown to be relevant in explaining subjects' behavior such as subjects' habits that lead to consuming the good under study, consumer association membership, family size, being a parent or grandparent, and other standard socioeconomic variables such as age, gender, education, residence, income and having a life insurance.

To investigate whether these variables also affect the probability of belonging to the OSQ_{LOW} rather than to the OSQ_{HIGH} treatment group, the equality of logit coefficients estimated for both treatment groups is tested by using a Chi Squared Test. Individual data about such variables have been elicited by using debriefing questions at the end of the survey. All these explanatory variables are described into detail in Table 4.7.

Table 4.7 Summary statistics of variables included in the selection model

Variable	Description	N.Obs.	Mean	St. Dev.	Min	Max
SSQ	=1 if the subject belongs to the this treatment; 0 = otherwise	797	0.611	0.487	0	1
OSQ_{LOW}	=1 if the subject belongs to the this treatment; 0 = otherwise	797	0.194	0.396	0	1
OSQ_{HIGH}	=1 if the subject belongs to the this treatment; 0 = otherwise	797	0.194	0.396	0	1
APPLE	Number of apples eaten in a week	797	4.500	3.992	0	20
C_ASS	=1 if member of a consumer association; 0=otherwise	797	0.136	0.343	0	1
PEEL	=1 if the subject peels apples before eating them; = 0 otherwise	797	0.438	0.496	0	1
PROD	=1 if apple producer, = 0 otherwise	797	0.100	0.300	0	1
FAM	Number of family members	797	3.504	1.291	1	7
CHILD	=1 if the subject has children, = 0 otherwise	797	0.711	0.453	0	1
GCHILD	=1 if the subject has grandchildren, = 0 otherwise	797	0.249	0.433	0	1

FEM	=1 if female, = 0 otherwise	797	0.498	0.500	0	1
AGE	Age in years	797	46.436	14.878	19	75
SC_PRI	= 1 if the respondent have this education level;= 0 otherwise	797	0.326	0.177	0	1
SC_SEC	= 1 if the respondent have this education level;= 0 otherwise	797	0.084	0.277	0	1
SC_HIGH3	= 1 if the respondent have this education level;= 0 otherwise	797	0.183	0.387	0	1
SC_HIGH5	= 1 if the respondent have this education level;= 0 otherwise	797	0.503	0.500	0	1
UNI	= 1 if the respondent have this education level;= 0 otherwise	797	0.168	0.374	0	1
PHD	= 1 if the respondent have this education level or higher;= 0 otherwise	797	0.028	0.167	0	1
INC	Yearly net income	797	20,376	14,192	5,000	120,000
LIFE	= 1 if the respondent have a life insurance;= 0 otherwise	797	0.166	0.373	0	1
NON	= 1 if the respondent lives in Non Valley;= 0 otherwise	797	0.115	0.319	0	1
SOLE	= 1 if the respondent lives in Sole Valley;= 0 otherwise	797	0.031	0.174	0	1
GIUD	= 1 if the respondent lives in Giudicarie Valley;= 0 otherwise	797	0.066	0.249	0	1
ADIGE	= 1 if the respondent lives in Adige Valley;= 0 otherwise	797	0.318	0.466	0	1
GARDA	= 1 if the respondent lives in Garda-Ledro Valley;= 0 otherwise	797	0.083	0.279	0	1
GRINA	= 1 if the respondent lives in VallagrinaValley;= 0 otherwise	797	0.160	0.366	0	1
A_SUG	= 1 if the respondent lives in Alta Sugana Valley;= 0 otherwise	797	0.095	0.293	0	1
TESINO	= 1 if the respondent lives in BassaSugana Valley;= 0 otherwise	797	0.056	0.230	0	1
FASSA	= 1 if the respondent lives in Fassa Valley;= 0 otherwise	797	0.020	0.140	0	1
FIEMME	= 1 if the respondent lives in Fiemme Valley;= 0 otherwise	797	0.032	0.177	0	1
PRIM	= 1 if the respondent lives in Primiero Valley;= 0 otherwise	797	0.018	0.139	0	1

Discrete Choice Modeling

Unlike most CE applications, here, I do not use a standard Random utility model (RUMs) which assume that decision makers are certain about their choices. In contrast, as my subjects are asked to make choices over lotteries, I implement an Expected Utility Theory (EUT) driven model which assumes that subject i makes a choice over j alternatives, with $j = 1,...,J$, by using a utility maximization rule[39]. Like RUMs, my models also assume that the utility that subject i attaches to each alternative j is decomposed into two parts, $V_{i,s,j}$ that is the part of the utility observed by the researcher, and $\varepsilon_{i,s,j}$ that is the one cannot be observed by the researcher, so that, $U_{i,j} = V_{i,j} + \varepsilon_{i,j}$. While researchers can model $V_{i,j}$, they can only make assumptions of the distribution that $\varepsilon_{i,j}$ follows.

The EUT approach, following von Neuman and Morgenstern (1947), assumes that subjects have rational preferences over lotteries L implying risky outcomes x_n with $n = 1,.., N$. An outcome is risky when it occurs with a given probability, $P(x_n)<1$, such that $\sum_{1}^{N} P(x_n) = 1$. Under the EUT (in discrete form), the utility of lottery L is described as follows:

Equation 4.1

$$U(L) = \sum_{1}^{N} P(x_n) \times U(x_n)$$

As described above, in each choice task, each subject i faces three alternatives, and, in turn, each alternative j depicts a lottery involving two risky prospects. In

[39] Here, I could refer to EUT when I model choices made in the Objective SQ treatments, where lotteries described in choice task's alternatives contain probabilities given by the researchers, thus, objective probabilities. In contrast, I could refer to SEUT when I model choices made in the Subjective SQ treatments, where the lottery presented in the SQ alternative contains subjective probabilities elicited via the exchangeability methods. However, given that lotteries presented in the other alternatives contain probabilities that have been designed on the basis of such elicited probabilities, but are not purely subjective, I prefer to refer to the EUT.

Prospect A, there is a probability $P(N_{A,j})$ that the maximum number of contaminated apples in 2030 will be $N_{A,j}$, and in Prospect B the probability $P(N_{B,j})= 1-P(N_{A,j})$ that the maximum number of contaminated apples in 2030 will be N_B. In each alternative j, subject i is asked to pay an annual tax (T), i.e. a tax that is paid in each year $n=\{1,\ldots,N\}$over the period between 2012 and 2030. Each year this tax (T) is taken away from each subject's yearly income (INC_i)[40], so that, the parameter ($INC_i - T_j$) enters in the conditional indirect utility function. Given this framework, the discounted utility ($U_{i,,j}$) that subject i attaches to alternative j is the sum of the utility that she/he attaches to Prospect A ($W_{i,A,j}$) and the utility that she/he attaches to Prospect B ($W_{i,B,j}$):

Equation 4.2

$$U_{i,j} = W_{i,j,A} + W_{i,j,B} + Y_{i,j} + \varepsilon_{i,j}$$

where:

Equation 4.3

$$W_{i,j,A} = \left\{ P\left(N_{A,j}\right) \times \left[\beta_{0,j} + \beta_{N,i} \times \frac{N_{A,j}^{1-r_N}}{1-r_N} + \beta_{INC,j} \times \left(\frac{(INC_i - T_j)^{1-r_{INC}}}{1-r_{INC}} \right) \right] \right\} \times \sum_{n=1}^{N} \frac{1}{\left(1+\delta\right)^n}$$

Equation 4.4

$$W_{i,j,B} = \left\{ \left[1 - P\left(N_{A,j}\right)\right] \times \left[\beta_{0,j} + \beta_{N,i} \times \frac{N_{B,j}^{1-r_N}}{1-r_N} + \beta_{INC,j} \times \left(\frac{(INC_i - T_j)^{1-r_{INC}}}{1-r_{INC}} \right) \right] \right\} \times \sum_{n=1}^{N} \frac{1}{\left(1+\delta\right)^n}$$

[40] I assume the income to be constant over the period between 2012 and 2030.

In the model presented above, the parameter β_0 is the alternative specific constant related to each alternative j. As it is evident in Equation 4.3 and 4.4, I investigate the presence of an unobserved between-subject heterogeneity for the coefficient $\beta_{N,i}$. After having tested diverse distributional forms (normal, lognormal, SB Johnson), the triangular distribution was chosen to model this random parameter. To my knowledge, only a few CE studies have modeled a random parameter related to risky outcomes. Glenk and Colombo (2011) investigated the risk of failure of environmental policies aiming to store carbon dioxide in the soil, and, in a second study, by Hensher and Li (2012) the risk of being late for a trip is considered, and both studies allow for some heterogeneity in the model via a random parameter. Interestingly, both previous studies also used a constrained triangular distribution for their random parameter distributions[41].

The r parameters included in the modeling, r_N and r_{INC}, generally measure the utility function's curvature, and, in my EUT framework, these terms correspond to coefficients of constant relative risk aversion (CRRA). Linear-in-income specifications assume all subjects are risk neutral, which might not be desirable in modeling. Recently, risk attitudes have been empirically shown to be context-dependent (Riddel, forthcoming), hence, I estimated two different CRRA coefficient here, one for the contaminated apple outcome (r_N) and the other for the income outcome (r_{INC}). Otherwise, the usual assumption is that risk preferences are consistent across sources of risk. More specifically, the parameter r_A accounts for a subject's risk attitude with respect to the number of contaminated apples in our fruit bowls in 2030, while the parameter r_{INC} represents the subjects' risk attitudes with respect to income. The CRRA

[41] I did not need to constrain my triangular distributions as we did not have the issue of having some risk parameter positive.

coefficient's specification used in my model has been extensively implemented in economic experiments investigating risk attitudes or preferences for monetary or financial outcomes (e.g. Andersen et al., 2006; 2008). To my knowledge, only Glenk and Colombo (2011), and Hensher and Li (2012) have already incorporated the CRRA specification I used here to model utility functions in CE studies. Others take into account the utility function's curvature by incorporating exponential or log specifications of the monetary/income attribute (Cameron, 2005a; Riddel and Shaw, 2006).

As noted above, my subjects are asked to pay a yearly tax in the period between 2012 and 2030, and thus, my model incorporates a standard financial rate of discount, δ. The estimated coefficient of this parameter provides a measure of the discount rate that subjects used in their temporally dependent choices (e.g., Burghart et al., 2007).

The vector $Y_{i,j}$ consists of all socioeconomic and variables that the estimated selection model has shown to affect the composition of treatment group. They are incorporated in the model to control their potential influence on subjects' choice-behavior and, hence, on their MWTP estimates. To create differences in utilities over alternatives, each of these variables is normalized to zero when it is associated to the SQ alternative (Train, 2003). More specifically, these variables indicate subject's apple consumption habit (*APPLES*), consumer association membership (*C_ASS*), job typology (*PROD*), age (*AGE*), gender (*FEMALE*), and life insurance taker (*LIFE*).

Estimation procedure and welfare measures

As noted above, in each choice task, subjects are asked to state their best and least preferred alternatives in a set of three alternatives j, say j_1, j_2, and j_3. Such a preference elicitation procedure allows me to obtain a full ranking of the alternatives from the best preferred to the least preferred, for example, if subject *i* chooses j_1 as the best preferred and j_3 as the least preferred alternative, I might assume that, to subject *i*'s first, second, and third best are j_1, j_2, and j_3, respectively. Given that, I might assume that subject *i* made her/his chooses sequentially, in the sense that she/he first chooses j_1 in asset of three alternatives $\{j_1, j_2, j_3\}$, and then, she/he chooses the alternative j_2 in a set containing the remaining two alternatives $\{j_2, j_3\}$. Assuming this decision-making procedure, I have estimate models presented above by using a standard "exploded" MMNL, where the probability of occurrence of each ranking option is obtained as follows:

Equation 4.5

$$P\left[ranking\left(j_1, j_2, j_3\right)\right] = \frac{e^{V_{i,j_1}}}{\sum_{j=j_1,j_2,j_3} e^{v_{i,j}}} + \frac{e^{V_{i,j_2}}}{\sum_{j=j_2,j_3} e^{v_{i,j}}}$$

In my investigation, for risk reduction related to the presence of pesticide residues in apples MWTP are estimated using the following specification, which of course follows from marginal rates of substitution:

Equation 4.6

$$MWTP = \left(\frac{\partial U_{i,j}}{\partial N_{A,j}} + \frac{\partial U_{i,j}}{\partial N_{B,j}}\right) \Bigg/ \left(\frac{\partial U_{i,j}}{\partial (INC_i - T_j)}\right)$$

This specification implies that the MWTP for risk reduction is the following:

Equation 4.7

$$MWTP = \frac{\beta_{N,i}\left\{\left(P_{A,j} \times N_{A,j}^{-r_N}\right) + \left[\left(1 - P_{A,j}\right) \times N_{B,j}^{-r_N}\right]\right\}}{\beta_I \times (INC_i - T_j)^{-r_{INC}}}$$

Results

Selection Model

Results obtained from both the estimation of my logit selection model and the related Chi Squared Test suggest that the composition of the SSQ treatment group differs from that of the other treatment groups. In fact, the probability of belonging to the SSQ treatment rather than to the other treatment groups depends on several socioeconomics and attitudinal variables. In contrast, the probability of belonging to the OSQ_{LOW} rather than to the OSQ_{HIGH} treatment is affected by only a few key parameters (Table 4.8 and 4.9).

Specifically, subjects who consume many apples (*APPLE*), have large families (*FAM*), and have large incomes (*INC*) more likely belong to the OSQ_{LOW} and OSQ_{HIGH} rather than to the SSQ treatment than the others. In contrast, parents (*CHILD*) and life insurance takers (*LIFE*) more likely belong to the SSQ treatment than to the others. Note that parents (*CHILD*) have a higher probability of belonging to the OSQ_{HIGH} rather than to the OSQ_{LOW}. Older subjects (*AGE*), grandparents (GCHILD), and high educated subjects (*SC_SEC*, *SC_HIGH3*, *SC_HIGH5*, *UNI*, *PHD*) more likely belong to the OSQ_{LOW} rather than to the SSQ treatment than the others. All the coefficients on these key variables are statistically significant (Table 4.8 and 4.9). Other variables barely affect the composition of my treatment groups (Table 4.8 and 4.9).

Note that these differences would prove fatal in a conventional experiment that uses ANOVA tests of differences between randomly assigned treatment groups, but, here econometric models are used to additionally control for these influences.

Table 4.8 Logit Selection model

Baseline: SSQ	**OSQ_{LOW}**		**OSQ_{HIGH}**	
Variable	**Coefficient**	**St. error**	**Coefficient**	**St. error**
APPLE	0.210*	(0.031)	0.220*	(0.032)
C_ASS	-0.113	(0.338)	-0.270	(0.398)
PEEL	-0.331	(0.229)	-0.012	(0.226)
PROD	0.246	(0.384)	0.055	(0.453)
FAM	0.346*	(0.098)	0.474*	(0.099)
CHILD	-1.406*	(0.320)	-0.583***	(0.317)
GCHILD	1.271*	(0.346)	0.320	(0.363)
FEM	0.235	(0.227)	0.104	(0.224)
AGE	0.024**	(0.011)	0.018	(0.012)
SC_PRI	-	-	-	-
SC_SEC	3.226**	(1.503)	15.689	(1031.826)
SC_HIGH3	3.410**	(1.472)	14.973	(1031.826)
SC_HIGH5	3.987*	(1.479)	16.136	(1031.826)
UNI	4.710*	(1.506)	16.751	(1031.826)
PHD	4.152**	(1.621)	16.713	(1031.826)
INC	0.001*	(0.001)	0.001*	(0.001)
LIFE	-0.973*	(0.349)	-1.570*	(0.407)
NON	-	-	-	-
SOLE	0.581	(0.642)	0.474	(0.815)
GIUD	-2.472**	(1.080)	1.161**	(0.563)
ADIGE	0.245	(0.372)	0.451	(0.481)
GARDA	-0.103	(0.484)	0.558	(0.566)
GRINA	-0.099	(0.437)	0.706	(0.519)
A_SUG	-1.571**	(0.641)	0.785	(0.534)
TESINO	-1.354	(0.839)	1.542*	(0.588)
FASSA	0.899	(0.718)	-15.772	(1616.718)
FIEMME	-0.190	(0.649)	-0.521	(0.926)
PRIM	-16.068	(1545.315)	1.276***	(0.769)
CONS	-8.733*	(1.678)	-22.112	(1031.458)
LL				-544.123
PSEUDO R^2				0.250

*1% significance level; **5% significance level; ***10% significant level

Table 4.9 Chi Squared Test for comparing coefficients of the *OSQ_{LOW}* and the *OSQ_{HIGH}* treatments

Null Hypothesis (H_0)	χ^2
$APPLE(OSQ_{LOW})= APPLE(OSQ_{HIGH})$	0.460
$C\_ASS(OSQ_{LOW})= C\_ASS(OSQ_{HIGH})$	0.150
$PEEL(OSQ_{LOW})= PEEL(OSQ_{HIGH})$	1.440
$PROD(OSQ_{LOW})= PROD(OSQ_{HIGH})$	0.180
$FAM(OSQ_{LOW})= FAM(OSQ_{HIGH})$	1.131
$CHILD(OSQ_{LOW})= CHILD(OSQ_{HIGH})$	5.370**
$GCHILD(OSQ_{LOW})= GCHILD(OSQ_{HIGH})$	5.890**
$FEM(OSQ_{LOW})= FEM(OSQ_{HIGH})$	0.240
$AGE(OSQ_{LOW})= AGE(OSQ_{HIGH})$	0.240
$SC\_PRIM(OSQ_{LOW})= SC\_PRIM(OSQ_{HIGH})$	-
$SC\_SEC(OSQ_{LOW})= SC\_SEC(OSQ_{HIGH})$	0.000
$SC\_HIGH3(OSQ_{LOW})= SC\_HIGH3(OSQ_{HIGH})$	0.000
$SC\_HIGH5(OSQ_{LOW})= SC\_HIGH5(OSQ_{HIGH})$	0.000
$UNI(OSQ_{LOW})= UNI(OSQ_{HIGH})$	0.000
$PHD(OSQ_{LOW})= PHD(OSQ_{HIGH})$	0.000
$INC(OSQ_{LOW})= INC(OSQ_{HIGH})$	0.020
$LIFE(OSQ_{LOW})= LIFE\ (OSQ_{HIGH})$	1.790
$NON(OSQ_{LOW})= NON(OSQ_{HIGH})$	-
$SOLE(OSQ_{LOW})= SOLE(OSQ_{HIGH})$	0.020
$GIUD(OSQ_{LOW})= GIUD(OSQ_{HIGH})$	10.390***
$ADIGE(OSQ_{LOW})= ADIGE(OSQ_{HIGH})$	0.180
$GARDA(OSQ_{LOW})= GARDA(OSQ_{HIGH})$	1.230
$GRINA(OSQ_{LOW})= GRINA(OSQ_{HIGH})$	2.230
$A\_SUG(OSQ_{LOW})= A\_SUG(OSQ_{HIGH})$	10.990***
$TESINO(OSQ_{LOW})= TESINO(OSQ_{HIGH})$	10.900***
$FASSA(OSQ_{LOW})= FASSA(OSQ_{HIGH})$	0.000
$FIEMME(OSQ_{LOW})= FIEMME(OSQ_{HIGH})$	0.130
$PRIM(OSQ_{LOW})= PRIM(OSQ_{HIGH})$	0.000
$CONS(OSQ_{LOW})= CONS(OSQ_{HIGH})$	0.000

*1% significance level; **5% significance level; ***10% significant level

Discrete choice models

By using a mixed multinomial logit (MMNL) estimation procedure, three discounted Expected Utility Theory (EUT) models are estimated, one for each treatment group, the Subjective SQ (SSQ model), the Objective Low SQ (OSQ_{LOW} model), and the Objective High SQ (OSQ_{LOW} model).

As a quick preview of my estimation's results, I first note that my models perform better when applied to subjects who belong to the SSQ treatment rather than to the sample members of the OSQ ones. In fact, while the SSQ model provides coefficients which are in line with findings from previous empirical studies, the other models do not. This may suggest that subjective probabilities are better predictors of subjects' choices than probability estimates provided by researchers. This may be due to the fact that, when subjects are provided with SQ's probability measures which differ from expected ones, they adjust the former on the latter (Table 4.10).

In all the specifications, coefficients of alternative specific constants related to R&D programs *x* and *y* ($\beta_{ASC\_R\&Dx}$ and $\beta_{ASC\_R\&Dy}$) are positive and statistically significant. As R&D programs are generic and unlabelled, these coefficients do not diverge much within the same model, although $\beta_{ASC\_R\&Dx}$ is always barely greater than $\beta_{ASC\_R\&Dy}$ (Table 4.10). This result suggests that subjects consistently prefer R&D programs rather than the SQ alternative, even when they are presented with a pivot experimental design tailored on their expectations. Many transportation studies have shown that pivot CEs induce subjects to prefer the SQ rather than other alternatives (scenario rejection or status quo effect). This phenomenon is likely due to the fact that SQ is based on their experience, while other hypothetical alternatives do not (i.e., they are designed by the researcher) (Hess and Rose, 2009). Now, given that SQ alternatives in my pivot CE are not designed on real experience, but on probabilistic expectations about future outcomes, subjects likely perceive the SQ as much hypothetical as the other alternatives, and, hence, they do not have any reason to systematically prefer the SQ.

In my modeling, I investigate the presence of unobserved between-subject heterogeneity in the coefficient of the variable *N* (β_N) which indicates the number of contaminated apples in 2030. Unobserved heterogeneity is detected in all the models. The estimated mean ($\beta_{N\mu}$) of such distribution is always negative and statistically significant, while the estimated standard deviation ($\beta_{N\sigma}$) is always lower than the estimated mean, indicating that each subject's *N* parameter is negative (Table 4.10). This means that the probability of choosing one alternative rather than the others increases when the number of contaminated apples decreases. However, as the impact of the variable *N* on subjects' choices in the OSQ models is much lower than in the SSQ model, I conclude that when subjects are presented with SQ's risk levels that diverge from their expected ones, the models have a relatively low explanatory power (Table 4.10).

The coefficient of the term *INC-TAX* (β_{INC}) indicates the net yearly income: the annual income left after having paid the yearly tax in the period between 2012 and 2030 for having a R&D program. The yearly income remains intact if the SQ is chosen. Estimated coefficients are negative and statistically in all specifications, and, thus, the probability of choosing an alternative increases when the amount of money to pay decreases (Table 4.10).

In all the models, the coefficient r_N is negative and statistically significant, meaning that subjects are overall risk loving with respect to the number of contaminated apples. Specifically, subjects who belong to the SSQ are moderately risk-loving in the SSQ (r_N= -0.535), while others are extremely risk loving (r_N= -2.410 in the OSQ_{LOW} and

(r_N= -3.550 in the OSQ_{HIGH} model). The information given by the researcher in the SQ alternative appears to have a huge impact on subjects' risk preferences (Table 4.10). This result is consistent with previous empirical studies which have shown that the way of framing outcomes of a gamble strongly affects subjects' risk preferences. In particular, if outcomes are framed as a loss, subjects becomes risk seeking, while, if outcomes are framed as gains subjects are risk averse (e.g., Kahneman and Tversy, 1979; Tversky and Kahneman, 1981; Kanheman and Tversky, 1984). In this application, the fact that attribute *N* is clearly framed as a loss (number of contaminated apples out of 100 apples) likely induces subjects to "seek" the health risk: i.e., they prefer riskier gambles in this dimension[42].

The coefficient r_{INC} is positive and statistically significant, meaning that subjects are overall risk averse with respect to the income. All subjects are moderately risk averse in income, whatever group treatment they belong to (r_{INC}=0.297 in the SSQ, r_{INC}=0.054 in the OSQ_{LOW}, and r_{INC}=0.197 in the OSQ_{HIGH} model) (Table 4.10).

The coefficient δ that indicates the financial discount factor that subjects use during their choices, is statistically significant only in the SSQ model (δ =0.460). This suggests that subjects in the SSQ treatment groups have a discount factor of about 46%, while, on average, other subjects have a discount factor which is not significantly different from zero (Table 4.10)[43].

Other socioeconomic and attitudinal variables that influence subjects' choice-behavior and, hence, their MWTP estimates were incorporated in the models to control for potential differences in the subsamples. Those variables only partially affect subjects' choice behavior.

---

[42] It is likely that if the random variable under study was the number of free-pesticide apples (gain), then, subjects were risk averse.

[43] It is quite possible that for long-term decisions, the average person's discount rate is in fact close to zero, meaning that they value the future the same as the present.

More specifically, my results indicate that apple consumers and members of consumer associations more likely agree to pay for R&D programs when they are presented with their probability estimates in the SSQ (positive and statistically significant signs for *APPLES* and *C_ASS*), however, they more likely refuse R&D programs and, thus, they stuck on the SQ when they face a risk information which is not consistent with their expectations (negative and statistically significant signs for *APPLES* and *C_ASS*) (Table 4.10). Apple producers are consistently reluctant to pay for the implementation of R&D programs in all specifications (negative and statistically significant signs for *PROD*) (Table 4.10). This may be due to the fact that they perceive pesticides to be more efficient controls than microorganisms or resistant varieties in handling apples disease.

Gender affects choices in SSQ model but not in the others, in particular women are less willing to financially support R&D programs then men (negative and statistically significant sign for *FEMALE*). In contrast, elderly subjects' choices appear to be motivated by something like altruism towards future generations, in fact, they are more willing to pay for R&D programs than young subjects, at least in the OSQ models (positive and statistically significant sign for AGE) (Table 4.10). Having a life insurance policy may be an indicator of risk preferences, but barely affect subjects' choices (Table 4.10)[44]. Although all potential variables that might affect food choices, according to the related literature, are incorporated in the modeling, the discrepancy in behavior might be still affected by my failure to identify omitted variables.

[44] Variables indicating family size, children age, education and residence were introduced in the models, but they did not significantly affect choice-behavior at all.

Table 4.10 Mixed Multinomial Logit estimation of discounted EUT models

	SSQ	OSQ_{LOW}	OSQ_{HIGH}
$ASC\_R\&D_X$	0.318*	0.632**	1.54*
	(0.106)	(0.314)	(0.222)
$ASC\_R\&D_Y$	0.278*	0.546***	1.28*
	(0.0928)	(0.276)	(0.187)
N_μ^a	-0.0014*	-2.07e-6*	-8.85e-08*
	(3.22e-05)	(4.80e-08)	(2.04e-09)
N_σ^a	0.0007*	1.72e-06*	-1.81e-09*
	(0.000206)	(2.05e-07)	(1.22e-10)
r_N	-0.535*	-2.410*	-3.550*
	(0.0822)	(0.109)	(0.0139)
INC	0.257*	0.036***	0.209**
	(0.0951)	(0.0211)	(0.103)
r_{INC}	0.297*	0.054***	0.197***
	(0.0185)	(0.0351)	(0.0493)
δ	0.460**	0.919	6.91
	(0.219)	(0.793)	(1.80e+308)
$APPLE_X$	0.076*	-0.150*	-0.220*
	(0.0117)	(0.0163)	(0.0186)
$APPLE_Y$	0.080*	-0.148*	-0.194*
	(0.0115)	(0.0149)	(0.0170)
T_CON_X	0.344*	-0.228	-0.322
	(0.0874)	(0.156)	(0.206)
T_CON_Y	0.309*	-0.294***	-0.488*
	(0.0849)	(0.156)	(0.142)
$PROD_X$	-1.12e-05*	-1.76e-05***	-1.11e-5**
	(1.62e-06)	(2.88e-06)	(5.29e-06)
$PROD_Y$	-1.24e-05*	-1.66e-05*	-8.80e-6
	(1.80e-06)	(2.95e-06)	(5.63e-06)
$GENDER_X$	-0.150*	-0.074	-0.233
	(0.0558)	(0.110)	(0.142)
$GENDER_Y$	-0.250*	-0.108	-0.135***
	(0.0560)	(0.107)	(0.0727)
AGE_X	-0.001	0.008**	0.008**
	(0.000254)	(0.00396)	(0.00394)
AGE_Y	-0.001	0.010*	0.005
	(0.000317)	(0.00392)	(0.00373)
$LIFE_X$	0.335	0.383**	-1.71
	(0.0740)	(0.171)	(1.80e+308)
$LIFE_Y$	0.267	0.259	-0.488*
	(0.0743)	(0.168)	(0.142)
LL(0)	-10,471.042	-3,332.673	-3,332.673
$LL(\beta)$	-8,399.744	-2,471.542	-2,700.016
Rho	0.198	0.258	0.190
N	11,688	3,720	3,720

*1% significance level; **5% significance level; ***10% significant level

Willingness to pay

Mean yearly MWTP estimates (per taxpayer) for a marginal reduction in the risk of having contaminated apples in 2030 are estimated for each treatment group by using the formula in Equation 4.7 presented above[45]. This equation implies that MWTP estimates depend on both the number of apples containing pesticides in 2030 ($N_{A,R\&D}$ and $N_{B,R\&D}$) and the probability of this amount occurring ($P_{A,R\&D}$ and $P_{B,R\&D}$) presented in the risky prospects of R&D programs. In their turn, as a pivot experimental design is used, $N_{A,R\&D}$, $N_{B,R\&D}$, $P_{A,R\&D}$, and $P_{B,R\&D}$ depend on the SQ alternative that each subject faces, specifically, on both the number of apples containing pesticides in 2030 ($N_{A,SQ}$ and $N_{B,SQ}$) and the probability of this amount occurring ($P_{A,SQ}$ and $P_{B,SQ}$). Finally, MWTP estimates also depend on both each subject's yearly income (INC_i)and the yearly tax to pay in order to get the R&D program implemented ($T_{R\&D}$).

Given this framework, several MWTP calculations can be undertaken. However, I focus on those which allow me to investigate the scenario adjustment by testing Hypothesis 1 and 2 presented above. In particular, I only estimate MWTP that are comparable across treatment groups, more specifically, MWTP that relates to SQ alternatives which present subjects with the same risky prospects whatever treatment group they belong to. More specifically, I have estimated MWTP related to the following SQ alternatives: $SSQ_{(65\text{-}50\%,75\text{-}50\%)}$, $OSQ_{LOW(65\text{-}50\%,75\text{-}50\%)}$, $SSQ_{(75\text{-}50\%,85\text{-}50\%)}$, $OSQ_{LOW(75\text{-}50\%,85\text{-}50\%)}$, $OSQ_{HIGH(75\text{-}50\%,75\text{-}50\%)}$, $SSQ_{(90\text{-}50\%,100\text{-}50\%)}$, and, finally, $OSQ_{HIGH(90\text{-}50\%,100\text{-}50\%)}$. For example, the $SSQ_{(65\text{-}50\%,75\text{-}50\%)}$ is the estimate from subjects who belong to the SSQ treatment group and face a SQ alternative in which there is 50% chance to have 65 contaminated apples (Prospect A) and 50% chance to have 75 contaminated apples (Prospect B) (Table 4.11).

[45] MWTP is constant over time, and, hence, can be aggregated over time.

Given the fact that MWTP estimates depend also on the risk reduction that each R&D program produces, here, MWTP are inferred about 4 different risk reduction scenarios out of the 12 available for each selected SQ alternative. These reduction scenarios are the following:

i. $N_{A,SQ}$ – 40% with $P_{A,SQ}$ -90%, and $N_{B,SQ}$ with 1-($P_{A,SQ}$ -90%)
ii. $N_{A,SQ}$ – 40% with $P_{A,SQ}$ -0% chance, and $N_{B,SQ}$ with 1-($P_{A,SQ}$ -0%)
iii. $N_{A,SQ}$ – 80% with $P_{A,SQ}$ -90%, and $N_{B,SQ}$ with 1-($P_{A,SQ}$ -90%)
iv. $N_{A,SQ}$ – 80% with $P_{A,SQ}$ -0% chance, and $N_{B,SQ}$ with 1-($P_{A,SQ}$ -0%)

For example, considering the $SSQ_{(65\text{-}50\%,75\text{-}50\%)}$ these risk reductions are:

i. 39 bad apples with 5% chance, and 75 bad apples with 95% chance
ii. 39 bad apples with 50% chance, and 75 bad apples with 50% chance
iii. 13 bad apples with 5% chance, and 75 bad apples with 95 chance
iv. 13 bad apples with 50% chance, and 75 bad apples with 50% chance

Finally, because MWTP estimates involve an income effect, the estimates are assessed by assuming that the average or typical subject has a yearly income equal to €50,000 and that the R&D program yearly costs €30.

Inferred yearly MWTP estimates per taxpayer are quite reasonable. The MWTP ranges from €0.01 to €1.39 in the SSQ treatment, from €0.17 to €2.79 in the OSQ_{LOW}, and from €1.26 to €24.97 in the OSQ_{HIGH} (Table 4.11). A previous study which has investigated subjects' preferences for reducing health risks due to pesticide residues in Northern Italy, has found a MTWP per household per month of about €0.48 (lower bound €0.01 and upper bound €0.87) (Travisi and Nijkamp, 2008).

In each treatment, when the number of contaminated apples increases in the prospects of the SQ alternative, then MWTP estimate for a given risk reduction increases. For example, MWTP of a subjected presented with $SSQ_{(90,100)}$ for a risk reduction i (equal to 0.139) is greater than MWTP of a subject who faces $SSQ_{(75,85)}$ (equal to 0.126) (Table 4.11).

Second, in each treatment, when the probability of a given reduction in the number of contaminated apples increases, then MWTP increases. For example, MWTP of a subject presented with $SSQ_{(65\text{-}50\%,75\text{-}50\%)}$ for a risk reduction ii (equal to 0.179) is greater than that for a risk reduction i (equal to 0.116) (Table 4.11).

Third, in each treatment, when the reduction in the number of contaminated apples increases, being the probability of the reduction constant, then MWTP decreases. For example, MWTP of a subject who face $SSQ_{(65\text{-}50\%,75\text{-}50\%)}$ for a risk reduction i (equal to 0.179) is greater than that for a risk reduction iii (equal to 0.096) (Table 4.11). This is due to the fact the subjects are risk loving with respect to the number of contaminated apples.

Table 4.11 Marginal willingness to pay for risk reductions

SQ	$N_{A,SQ}$-40% / $N_{B,SQ}$	0.05 / 0.95	$N_{A,SQ}$-40% / $N_{B,SQ}$	0.50 / 0.50	$N_{A,SQ}$-80% / $N_{B,SQ}$	0.05 / 0.95	$N_{A,SQ}$-80% / $N_{B,SQ}$	0.50 / 0.50
$SSQ_{(65\text{-}50\%,75\text{-}50\%)}$		0.118		1.179		0.096		0.963
$OSQ_{LOW(65\text{-}50\%,75\text{-}50\%)}$		0.204		2.048		0.172		1.720
$SSQ_{(75\text{-}50\%,85\text{-}50\%)}$		0.126		1.265		0.103		1.032
$OSQ_{LOW(75\text{-}50\%,85\text{-}50\%)}$		0.278		2.787		0.232		2.237
$OSQ_{HIGH(75\text{-}50\%,85\text{-}50\%)}$		1.392		13.928		1.263		12.636
$SSQ_{(90\text{-}50\%,100\text{-}50\%)}$		0.139		1.385		0.113		1.129
$OSQ_{HIGH(90\text{-}50\%,100\text{-}50\%)}$		2.497		24.971		2.250		22.504

Hypothesis testing

Hypothesis 1 is tested by comparing MWTP inferred from subjects who face $SSQ_{(65\text{-}50\%,75\text{-}50\%)}$, $OSQ_{LOW(65\text{-}50\%,75\text{-}50\%)}$, $SSQ_{(75\text{-}50\%,85\text{-}50\%)}$, $OSQ_{LOW(75\text{-}50\%,85\text{-}50\%)}$. On the other hand, Hypothesis 2 is tested by comparing MWTP inferred from subjects presented with $SSQ_{(75\text{-}50\%,85\text{-}50\%)}$, $OSQ_{HIGH(75\text{-}50\%,75\text{-}50\%)}$, $SSQ_{(90\text{-}50\%,100\text{-}50\%)}$, and, finally, $OSQ_{HIGH(90\text{-}50\%,100\text{-}50\%)}$. My hypotheses are tested by using a simple t-test (Table 4.12).

Testing Hypothesis 1, I reject the null hypothesis that MWTPs for given risk reductions provided by subjects who expected SQ's risk levels equal to those presented in the SQ (SSQ treatment) are higher than or equal to MWTP estimates inferred from subjects who expected SQ's risk levels higher than those given in the SQ (OSQ_{LOW} treatment) (at the 1% significance level). I conclude that subjects do not fully accept the information given in the SQ, but they positively adjust this on what they expected. Subjects in the OSQ_{LOW} treatment group make choices by using a risk of having contaminated apples greater than the risk used by subjects in the SSQ treatment, and, hence, the former group has higher MWTP for risk reductions than the latter group. These results are consistent across diverse risk reductions (Table 4.12).

By testing Hypothesis 2, I fail to reject the null hypothesis that MWTP estimates for risk reductions inferred from subjects who expected SQ's risk levels equal to those given in the SQ (SSQ treatment) are lower than or equal to those obtained from subjects who expected SQ's risk levels smaller than those presented in the SQ (OSQ_{HIGH} treatment) (Table 4.12). In this case, I expected subjects who adjust the information provided in the SQ on their expectations (OSQ treatment) to have lower MWTP for risk reductions than the others. In fact, subjects should negatively adjust the information given in the SQ, on their expectations. In contrast, subjects who belong to the OSQ_{HIGH} treatment have higher MWTP than the others. Such a result is consistent across diverse risk reductions (Table 4.12). I might speculate that, when subjects find in the SQ risk of having contaminated apples substantially higher that they expected, they might feel some sort of alarm that induce them to irrationally pay more than what they would have paid if this information was not given. This would be consistent with the alarmist learning theory by Viscusi and Magat (1992).

Table 4.12 One sided t test for comparing marginal willingness to pay

H_0	$N_{A,SQ}$-40% 0.05 $N_{B,SQ}$ 0.95	$N_{A,SQ}$-40% 0.50 $N_{B,SQ}$ 0.50	$N_{A,SQ}$-80% 0.05 $N_{B,SQ}$ 0.95	$N_{A,SQ}$-80% 0.50 $N_{B,SQ}$ 0.50
$MWTP\_SSQ_{(65\text{-}50\%,75\text{-}50\%)} \geq MWTP\_OSQ_{LOW(65\text{-}50\%,75\text{-}50\%)}$	-368.392***	-369.561***	-385.725***	-383.320***
$MWTP\_SSQ_{(75\text{-}50\%,85\text{-}50\%)} \geq MWTP\_OSQ_{LOW(75\text{-}50\%,85\text{-}50\%)}$	-487.493***	-487.225***	-497.112***	-499.776***
$MWTP\_SSQ_{(75\text{-}50\%,85\text{-}50\%)} \leq MWTP\_OSQ_{HIGH(75\text{-}50\%,85\text{-}50\%)}$	-12,999.920	-12,991.150	-14,327.540	-14,324.360
$MWTP\_SSQ_{(90\text{-}50\%,100\text{-}50\%)} \leq MWTP\_OSQ_{HIGH(90\text{-}50\%,100\text{-}50\%)}$	-19,730.590	-19,736.340	-21,247.120	-21,183.070

*1% significance level; **5% significance level; ***10% significant level

Conclusions

In this essay, I investigated to what extent the scenario adjustment occurs in choice experiments by using an innovative two-stage approach that relies on the comparison of willingness-to-pay estimates obtained in different treatment groups. In the first, subjects are presented with a status quo alternative where the risk of having contaminated apples in 2030 is consistent with their subjective estimates, while, in the second, they are presented with a status quo alternative where the risk of having contaminated apples in 2030 is not consistent with their probability estimates.

To implement this approach, I incorporated subjective probabilities, elicited via a novel approach such as the exchangeability method, into my choice experiment’s design by using a pivot experimental design. As previous stated-preference investigations have only introduced subjective probability estimates in econometric modeling, but never into their choice context designs, this investigation introduces a new way to investigate the role of subjective probabilities on choice behavior.

My discounted Expected Utility Theory model predicts choice behavior of subjects who belong to the first group quite well, while it poorly explains choices of subjects who belong to the second groups. This highlights that subjective probabilities strongly affect the decisions under conditions of risk, and that risk information has a strong impact on choice behavior.

I found that subjects when provided with risk that are lower than perceived ones, adjust attribute levels on their expectations, and express marginal willingness to pay for risk reduction higher than those that they would have provided taking choices by using status quo's risk information. In contrast, subjects who face a risk of having contaminated apples higher than the expected one, do not negatively adjust attribute levels on their expectations, but, they, driven by some sort of panic, overreact to this information and irrationally pay more than what they would have paid if they fully accepted the SQ's information.

My investigation has shown that information provided by researchers in the status quo alternative substantially affects subjects' choices. This might have very crucial policy implication, in the sense that financial support for public policies might be driven by the strategy used to communicate new information, in this case risk information. This is not necessarily a bad thing, for example, stated-preference studies might become very helpful in identifying the most effective way to communicate risk information that makes people willing to support policies that are not perceived to be important yet.

CHAPTER V. CONCLUSIONS

My dissertation has considered the influence of subjective probabilities on choice-behavior, in particular, the influence that subjective probabilities of having contaminated apples have on preferences for R&D programs that will improve apple safety in the future.

This work has contributed to the literature in food choices in two ways. First, an innovative indirect elicitation technique such as the Exchangeability Method (EM) has been used to elicit probabilities related to food outcomes. Unlike direct elicitation techniques, commonly used to investigate subjective probabilities in stated-preference studies, the EM does not ask subjects to express the probability that given outcomes will occur in the future, but infers probability estimates at the point for which subjects become indifferent to bet a certain amount of money on a given outcome rather than on an alternative one. In particular, the main EM's advantage is that it does not force individuals to process numerical probability estimates. The presumed superiority of this elicitation technique has been explored by testing the internal validity of subjective probabilities elicited via the EM in an artefactual field experiment.

Second, subjective probabilities have been incorporated in the experimental design of a stated-choice technique to test to what extent subject adjust the risk information provided by researchers in the status quo alternative on their probabilistic expectations. This dissertation has represented the first attempt to incorporate subjective probabilities into the design of stated-choice experiments to investigate subjects' behavior under conditions of risk.

In the first essay, experimental results suggest that the EM is not necessarily incentive compatible because chained questions might induce subjects to strategically behave when associated with monetary incentives. In addition, results show that incentive compatibility determines the validity of subjective probabilities elicited via the EM. In fact, subjects are more likely to provide valid subjective probabilities when they are rewarded with real monetary incentives and presented with experimental designs where the chaining is hidden through a particular randomization of the questions.

In the second essay, my experimental results show that valid subjective probabilities elicited via the EM do not significantly diverge from invalid ones, and, hence, internal validity does not affect the actual magnitude of subjective probabilities. This suggests that failure to recognize validity does not imply an over- or underestimation of consumers' probability estimates.

In the third essay, the occurrence of the scenario adjustment was detected by using an innovative approach which implies the incorporation of subjective probabilities into the design of my CE. Results suggest that when subjects are provided with risk levels that are lower than their subjective estimates, they adjust the former on the latter, and provide marginal willingness to pay estimates that are higher than those that they would have provided under fully acceptance of the status quo alternative. In contrast, when subjects are presented with status quo alternative in which risk levels are higher than their subjective estimates, they do not negatively adjust the former on the latter, but, motivated by a sense of alarm, they overreact to such an information and provide marginal willingness to pay estimates that are greater than the ones they would have expressed if they fully accepted the status quo scenario. The latter result is in line with the alarmist learning theory developed by Viscusi and Magat (1992). These results might have very interesting policy implications as they suggest that stated preference investigations might be used to identify risk communication strategies that efficiently inform subjects about the true magnitude of risks.

REFERENCES

Abdellaoui, M., Baillon, A., Placedo, L., Wakker, P.P., 2011. The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation. *American Economic Review* 101, 695-723.

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., Williams, M., 1997. Perceptions Versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation. *Journal of Environmental Economics and Management* 32 (1), 65–84.

Alberini, A., Hunt, A., Markandya, A., 2006. Willingness to pay to reduce mortality risks: Evidence from a three-country contingent valuation study. *Environmental & Resource Economics* 33, 251-264.

Anderson, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2006. Elicitation using multiple price list formats. *Experimental Economics* 9, 383-405.

Anderson, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2008. Eliciting risk and time preferences. *Econometrica* 76(3), 583-618.

Andersen, S., Fountain, J., Harrison, G.W., Rutström, E.E., 2010. Estimating Subjective Probabilities. Working Paper 09-01, Department of Economics, College of Business Administration, University of Central Florida.

Alavanja, M.C.R., Hoppin, J.A., Kamel, F., 2004. Health Effects of Chronic Pesticide Exposure: Cancer and Neurotoxicity. *Annual Review of Public Health* 25, 155-197.

Baillon, A., 2008. Eliciting Subjective Probabilities Through Exchangeable Events: an Advantage and a Limitation. *Decision Analysis* 5(2), 76-87.

Baker, J., Shaw, W.D., Bell, D., Brody, S., Riddel, M., Woodward, R.T., Neilson, W., 2009. Explaining Subjective Risks of Hurricanes and the Role of Risks in

Intended Moving and Location Choice Models. *Natural Hazards Review* 10 (3), 102-112.

Bierlaire, M., 2003. BIOGEME: A free package for the estimation of discrete choice models , Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Brier, G.W., 1950. Verification of Weather Forecasts In Terms of Probability. *Monthly Weather Review* 78(1), 1-3.

Boccaletti, S., Nardella, M., 2000. Consumer willingness to pay for pesticide-free fresh fruit and vegetables in Italy. *The International Food and Agribusiness Management Review* 3, 297–310.

Botzen, W.J.W., Aerts, J.C.J.H., van den Bergh, J.C.J.M., 2009. Dependence of flood risk perceptions on socioeconomic and objective risk factors. Water Resource Research, 45, W10440, doi:10.1029/2009WR007743.

Burghart, D.R., Cameron T.A., Gerdes G.R., 2007. Valuing publicly sponsored research projects: Risks, scenario adjustments and inattention. *The Journal of Risk and Uncertainty* 35, 77-105.

Buzby, J.C., Fox, J.A., Ready, R.C., Crutchfield, S.R., 1998. Measuring Consumer Benefits of Food Safety Risk Reductions. *Journal of Agricultural and Applied Economics* 30(1), 69–82.

Cameron, Trudy Ann, 2005a. Individual Option Prices for Climate Change Mitigation. *Journal of Public Economics* 89, 283-301.

Cameron, Trudy Ann, 2005b. Updating subjective risks in the preference of conflicting information: an application to climate change. *Journal of Risk and Uncertainty* 30(1), 63-97.

Cameron, T.A., DeShazo, J.R., Johnson, E.H., 2010. Scenario adjustment in stated preference research. *Journal of Choice Modelling* 4(1), 9-43

Cerroni, S., Notaro, S., Shaw, W.D., 2012a. Eliciting and estimating valid subjective probabilities: An experimental investigation of the exchangeability method. *Journal of Economic Behavior & Organization* 84(1), 201-215.

Cerroni, S., Notaro, S., Shaw, W.D., 2012b. How many bad apples are in a bunch? An experimental investigation of subjective probabilities. Unpublished manuscript.

Cerroni, S., Shaw, W.D., 2012. Does Climate Change Information Affect Stated Risks of Pine Beetle Impacts on Forests? An Application of the Exchangeability Method. *Forest Policy and Economics* 22, 72-84.

Chew S.H., Sagi, J., 2006. Event exchangeability: Probabilistic sophistication without continuity or monotonicity. *Econometrica* 74, 771–786

ChoiceMetrics, 2011. Ngene 1.1 User manual and reference guide. http://www.choice-metrics.com/download.html [accessed 11.23.2012]

Corso P.S., Hammit J.K., Graham J.D., 2001. Valuing mortality-risk reduction: Using visual aids to improve the validity of contingent valuation. *Journal of Risk and Uncertainty* 23, 165-184.

de Finetti, B, 1937. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'IHP* 7, 1–68

de Finetti, B., 1974a. The Value of Studying Subjective Evaluations of Probability. In: Stael Von Holstein C.A.S. (Eds.). The Concept of Probability in Psychological Experiments. Dordrecht-Holland/Boston-U.S.A: Reidel Publishing Company, 1-14.

de Finetti, B., 1974b. The True Subjective Probability Problem. In: Stael Von Holstein C.A.S. (Eds.). The Concept of Probability in Psychological Experiments. Dordrecht-Holland/Boston-U.S.A: Reidel Publishing Company, 15-23.

de Palma, A., Ben_Akiva, M., Brownstone, D., Holt, C., Magnac, T., McFadden, D., Mofat, P., Picard, N., Train, K., Wakker, P., Walker, J., 2008. Risk, uncertainty and discrete choice models. *Marketing Letters* 19(3-4), 269-285.

DeShazo, J. R., Cameron, T.A., 2005. The Effect of Health Status on Willingness to Pay for Morbidity and Mortality Risk Reductions. Department of Economics Working Paper, University of Oregon. Eugene, OR.

Dosman, D.M., Adamowicz, W.L., Hrudey, S.E., 2001. Socioeconomic Determinants of Health and Food Safety Related Risk Perceptions. *Risk Analysis* 21(2), 307-317.

Edmund Mach Foundation, 2006. Colpo di Fuoco. IASMA Notizie 5, 1-4. [accessed 10.31.2011].

Eom, Y.S., 1994. Pesticide Residue Risk and Food Safety Valuation: A Random Utility Approach. *American Journal of Agricultural Economics* 76(4), 760-771.

European Commission, 2010. Food-related risks. Special Eurobarometer 354/Wave73.5 –TNS Opinion&Social. http://www.efsa.europa.eu/en/factsheet/docs/sreporten.pdf [accessed 10.28.2011]

Fiore, S.M., Harrison, G.W., Hughes, C.E., Ruström, E.E., 2009. Virtual experiments and environmental policy. *Journal of Environmental Economics and Management* 57(1), 65-86.

Flynn, J., Slovic, P., Mertz, C. K., 1994. Gender, race, and perception of environmental health risks. *Risk Analysis* 14, 1101–1108.

Flynn, T. N., Louviere, J. J., Peters T. J., Coast, J.,2007. Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics* 26, 171–189.

Frisch, D., Baron, J., 1988. Ambiguity and rationality. *Journal of Behavioral Decision making* 1, 149-157.

Fu, T.-T., Liu, J.T., Hammitt, J.K., 1999. Consumer Willingness to Pay for Low-Pesticide Fresh Produce in Taiwan. *Journal of Agricultural Economics* 50(2), 220-233.

Gerking, S., Khaddaria, R., 2011. Perceptions of Health Risk and Smoking Decisions of Young People. *Health Economics* 21 (7), 865-877.

Gigerenzer, G., Hoffrage, U., 1995. How to improve Bayesian reasoning without instruction - frequency formats. *Psychological Review* 102, 684-704.

Glenk, K., Colombo, S., 2010. How sure can you be? A framework for considering delivery uncertainty in benefit assessment based on stated preference methods. *Journal of Agricultural Economics* 62(1), 25-46.

Hamilton, L. C., 1985. Concern about toxic wastes: Three demographic predictors. *Sociological Perspectives* 28, 463–486.

Hamilton, L. C., 1995. Who cares about water pollution? Opinions in a small-town crisis. *Sociological Inquiry* 55, 170–181.

Hammitt, J. K., 1990. Risk perceptions and food choice: An exploratory analysis of organic- versus conventional-produce buyers. *Risk Analysis* 10, 367–374.

Hammit, J.K., Graham J.D., 1999. Willingness to Pay for Health Protections: Inadequate Sensitivity to Probability? *Journal of Risk and Uncertainty* 8, 33-62.

Harrison, G.W., 1986. An experimental test for risk aversion. *Economic Letters* 21, 7-11.

Harrison, G.W., List, J.A., 2004. Field Experiment. *Journal of Economic Literature* 42(4), 1009-1055.

Harrison, G.W., List, J.A., Towe, C., 2007. Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study for Risk Aversion. *Econometrica* 75(2), 433-458.

Harrison, G.W., Rutström, E.E., 2008. Risk aversion in the laboratory. *Research in Experimental Economics* 12, 141-196.

Hensher, D.A., 2001. Measurement of the valuation of travel time savings. *Journal of Transport Economics and Policy* 35(1), 71-98.

Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. *Transportation* 30 (2),133–176.

Hensher, D.A., Li, Z., 2012. Valuing travel time variability within a rank-dependent utility framework and an investigation of unobserved taste heterogeneity. *Journal of Transport Economics and Policy* 46(2), 293-312.

Hensher, D.A., Rose J.M., 2007. Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study. *Transportation Research Part A* 41, 428-443.

Hensher, D.A., Rose J.M., Ortúzar J.d.D., Rizzi L.I., 2009. Estimating the willingness to pay and value of risk reduction for car occupants in the road environment. *Transportation Research Part A* 43, 692-707.

Hess, S., Rose, J.M., 2009. Should Reference Alternatives in Pivot Design SC Surveys be Treated Differently? *Journal of Environmental Resource Economics* 42, 297-317.

Ida, T., Goito, R., 2009. Simultaneous measurement of time and risk preferences: Stated preference discrete choice modeling analysis depending on smoking behavior. *International Economic Review* 50(4), 1169-1182.

Italian Ministry of Health, 2010. Controllo ufficiale sui residui di prodotti fitosanitari negli alimenti di origine vegetale. http://www.salute.gov.it/imgs/C_17 pubblicazioni_1305_allegato.pdf [accessed 04.26.2012]

Jakus, P.M., Shaw, W.D., Nguyen, T.N., Walker, M., 2009. Risk Perceptions of Arsenic in Tap Water and Bottled Water Consumption. *Water Resource Research* 45, W05405, doi:10.1029/2008WR007427.

Kahneman, D., Tversky, A., 1979. Prospect Theory: an analysis of decision under risk. *Econometrica* 47 (2), 263–291.

Kahneman, D., Tversky, A., 1984. Choices, values, and frames. *American Psychologist* 39 (4), 341–350.

Kilka, M., Weber, M., 2001. What Determines the Shape of the Probability weighting function under uncertainty. *Management Science* 47, 1712–1726.

Kivi, P.A., Shogren, J.F., 2010. Second-order Ambiguity in Very Low Probability Risks: Food Safety Valuation. *Journal of Agricultural and Resource Economics* 35(3), 443-56.

Kontoleon, A., Yabe, M., 2003. Assessing the Impacts of Alternative ‘Opt-out’ Formats in Choice Experiment Studies: Consumer Preferences for Genetically Modified Content and Production Information in Food. *Journal of Agricultural Policy and Resources* 5, 1-43.

Koriat, A., Lichtenstein, S., Fischhoff, B., 1980. Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory* 6, 107-118.

Krewski, D., Slovic, P., Bartlett, S., Flynn, J., Mertz, C., 1994. Health risk perceptions in Canada (ERC 94-3, Environmental Risk Management Working Paper). Edmonton, Alberta, Canada: University of Alberta.

Krupnick, A., Alberini, A., Cropper, M., Simon, N., O’Brien, B., Goeree, R., Heintzelman, M., 2002. Age, health and willingness to pay for mortality risk reductions: A contingent valuation survey of Ontario residents. *Journal of Risk and Uncertainty* 24(2), 161-186.

Kuhn, Kristine M., Budescu, David V. 1996. The relative importance of probabilities, outcomes, and vagueness in hazard risk decisions. *Organizational Behavior and Human Decision Processes* 68 (3), 301-17.

Lin, C.-T. J., 1995. Demographic and socioeconomic influences on the importance of food safety in food shopping. *Agricultural and Resource Economics Review* 24(2), 190–198.

Manski, C., 2004. Measuring Expectations. *Econometrica* 72 (5), 1329-1376.

Marley, A.A.J., Louviere, J. J., 2005. Some Probabilistic Models of Best, Worst, and Best-Worst Choices. *Journal of Mathematical Psychology* 49, 464–480.

Meyerhoff, J., Liebe, U., 2009. Status Quo Effect in Choice Experiments: Empirical Evidence on Attitudes and Choice Task Complexity. *Land Economic* 85(8), 515-528.

Misra, S.K., Huang, C.L., Ott, S.L., 1991. Consumer Willingness to Pay for Pesticide-Free Fresh Produce. *Western Journal of Agricultural Economics* 16(2), 218-227.

Morgan, M.G., Henrion, M., 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. New York: Cambridge University Press.

Németh, J., 2004. Practice of Applying Streptomycin to Control Fireblight in Hungary. Bull. OEEP/EPPO 34, 381-382.

Offerman, T., Sonnemans, J., Van de Kuilen, G., Wakker, P.P., 2009. A Truth Serum for NON-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. *Review of Economic Studies* 76, 1461-1489.

Papke, L.E., Woolridge, J.M., 1996. Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics* 11(6), 619-632.

Provincia Autonoma di Trento, 2010. La Produzione Lorda Vendibile dell'Agricoltura e della Selvicoltura nella Provincia di Trento nel 2006 e nel 2007. http://www.statistica.provincia.tn.it/binary/pat_statistica/novita/Pat_Agricoltura_bassa.1282903703.pdf [accessed 26.04.2012]

Raiffa, H., 1968. Decision Analysis. London: Addison-Wesley.

Riddel, M., fortcoming. Comparing risk preferences over finacial and environmental lotteries. *Journal of Risk and Uncertainty*.

Riddel, M., Shaw, W.D., 2006. A theoretically-consistent empirical model of non-expected utility: An application to nuclear-waste transport. *Journal of Risk and Uncertainty* 32, 131-150.

Rimal, A., Fletcher, S.R., McWatters, K.H., Misra, S.K., Deodhar, S., 2008. Perceptions of food safety and changes in food consumption habits: a consumer analysis. *International Journal of Consumer Studies* 25(1), 43-52.

Roberts, D.C., Boyer, T.A. and Lusk, J.L., 2008. Preferences for environmental quality under uncertainty, *Ecological Economics*, 66, 584-93.

Roosen, J., Fox, J.A., Hennessy, D.A., Schreiber, A., 1998. Consumers' Valuation of Insecticide Use Restrictions: An Application to Apples. *Journal of Agricultural and Resource Economics* 23(2), 367-384.

Rose, J.M., Hess, S., 2009. Dual-Response Choices in Pivoted Stated Choice Experiments. *Transportation Research Record: Journal of the Transportation Research Board* 2135, 25-33.

Rose, J.M., Masiero, L., 2010. A comparison of the impacts of aspects of prospect theory on WTP/WTA estimated in preference and WTP/WTA space. *European Journal of Transport and Infrastructure* 10(4), 330-346.

Scarpa, R., Ferrini, S. and Willis, K., 2005. Performance of Error Component Models for Status-Quo Effects in Choice Experiments. In: Applications of Simulation Methods in Environmental and Resource Economics. The Economic of Non-Market Goods and Resource, Volume 6. Springer Netherlands, 247-273.

Scarpa, R., Notaro, S., Louviere, J., Raffaelli, R., 2010. Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons. *American Journal of Agricultural Economics* 93(3), 813-828.

Scarpa, R., Marley, A.A.J., Notaro, S., Raffaelli, R., 2012. An empirical test of several best and/or worst ranking procedures. Unpublished manuscript.

Shaw, W.D., Jakus, P.M., Riddel, M., 2012. Perceived Arsenic-Related Mortality Risks for Smokers and Non-smokers. *Contemporary Economic Policy* 30(3), 417-429.

Slovic, P., 1987. Perceptions of risk. *Science* 236, 280–285.

Spetzler, C.S., Stael Von Holstein, C.-A.S., 1975. Probability encoding in decision analysis. *Management Science* 22, 340–358.

Train, K., 2003. *Discrete Choice Methods with Simulation.* Cambridge, UK: University Press.

Travisi, C.M., Nijkamp, P., 2008. Valuing environmental and health risk in agriculture: A choice experiment approach to pesticide in Italy. *Ecological Economics* 67, 598-607.

Tsuge, T., Kishimoto, A., Takeuchi, K., 2005. A Choice Experiment Approach to the Valuation of mortality. *Journal of Risk and Uncertainty* 31(1), 73-95.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Scence* 211, 453-458.

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

Van de Kuilen, G., Wakker, P.P., Zou, L., 2006. A Midpoint Technique for Easily Measuring Prospect Theory's Probability Weighting. CREED, University of Amsterdam, The Netherlands.

van Ravenswaay, E., Hoehn, J. P., 1991. Consumer willingness to pay for reducing pesticide residues in food: Results of a nationwide survey. Unpublished manuscript.

Vermeulen, B., Goos, P., Vanderbroek, M., 2010. Obtaining more information from conjoint experiments by best-worst choices. *Computational Statistics and Data Analysis* 54, 1426-1433.

Viscusi, W. K., 1985. Are individuals Bayesian decision makers? *The American Economic Review* 75(2), 381-385.

Viscusi, W. K., 1989. Prospective reference theory: toward an explanation of the paradoxes. *Journal of Risk and Uncertainty* 2, 235-264.

Viscusi, V.K., 1990. Do Smokers Underestimate Risks? *Journal of Political Economy* 98 (6), 1253-68.

Viscusi, V.K., Magat, W.A., 1992. Bayesian decisions with ambiguous belief aversion *Journal of Risk and Uncertainty* 5(4), 371-387.

Viscusi, W.K,. Hakes, J. K., 2003. Risk ratings that do not measure probabilities. *Journal of Risk Research* 6 (1), 23-43.

Viscusi, W. K., Zeckhauser R.J., 2006. The Perception and Valuation of the Risks of Climate Change: A Rational and Behavioural Blend Climate Change. *Climatic Change* 77, 151-177.

von Neumann, John, and Morgenstern Oskar , 1944. Theory of games and economic behavior. Princeton, U.S.: Princeton University Press.

Vossler, C.A., Evans, M.F., 2009. Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management* 58(3), 338-345.

Wakker, P., 2008. Explaining the characteristics of the power (CRRA) utility family. *Health Economics* 17, 1329-1344.

Wakker, P., 2010. Prospect theory for risk and ambiguity. Cambridge, UK: University Press.

Wakker P., Deneffe D., 1996. Eliciting von Neumann-Morgenstern Utilities When Probabilities are Distorted or Unknown. *Management Science* 42(8), 1131-1150.

Weinstein, N.D., Diefenbach, M.A., 1997. Percentage and Verbal Category Measures of Risk Likelihood. *Health Education Research* 12 (1), 139-41.

Whittington, D., Adamowicz, W., 2011. The Use of Hypothetical Baselines in Stated Preference Surveys. Environment for Development. Discussion Paper Series.

Williams, P.R.D., Hammitt, J.K., 2001. Perceived Risks of Conventional and Organic Produce: Pesticides, Pathogens, and Natural Toxins. *Risk Anal*ysis 21(12), 319-330.

Zimmer, A., 1983. Verbal vs. numerical processing of subjective probabilities. *Advances in Psychology*16, 159-182.

APPENDIX A

Example 1. First question of the Exchangeability Game for the variable g

I prefer to bet 100€ on the fact that the number of days of April in which the *fire blight* infestation will occur with certainty in 2030 is:

□	□
smaller than g_a[a]	greater than or equal to g_a[a]

[a] $g_a=\{g_0 + [(g_1-g_0)/2]\}$

Example 2. First question of the Repeated Exchangeability Game Test for the variable $g_{1/2}$'

I prefer to bet 100€ on the fact that the number of days of April in which the *fire blight* infestation will occur with certainty in 2030 is:

□	□
greater than $g_{1/4}$ and smaller than $g_{1/2}$	greater than or equal to $g_{1/2}$ and smaller than $g_{3/4}$

Example 3. A question of the Certainty Equivalent Game for $g_{1/2}$

In each of the following question, do you prefer to play the lottery presented in Option A or do you prefer to take the amount of money presented in Option?

<table>
<tr><th colspan="2">Option A</th><th colspan="2">Option B</th></tr>
<tr><td rowspan="6">You win 100€ if the number of days of April in which the fire blight infestation will occur with certainty in 2030 is SMALLER THAN $g_{1/2}$

0€, otherwise</td><td>□</td><td>□</td><td>0€</td></tr>
<tr><td>□</td><td>□</td><td>25€</td></tr>
<tr><td>□</td><td>□</td><td>49€</td></tr>
<tr><td>□</td><td>□</td><td>51€</td></tr>
<tr><td>□</td><td>□</td><td>75€</td></tr>
<tr><td>□</td><td>□</td><td>100€</td></tr>
</table>

In each of the following question, do you prefer to play the lottery presented in Option A or do you prefer to take the amount of money presented in Option?

<table>
<tr><th colspan="2">Option A</th><th colspan="2">Option B</th></tr>
<tr><td rowspan="6">You win 100€ if the number of days of April in which the fire blight infestation will occur with certainty in 2030 is GREATER THAN OR EQUAL TO $g_{1/2}$

0€, otherwise</td><td>□</td><td>□</td><td>0€</td></tr>
<tr><td>□</td><td>□</td><td>25€</td></tr>
<tr><td>□</td><td>□</td><td>49€</td></tr>
<tr><td>□</td><td>□</td><td>51€</td></tr>
<tr><td>□</td><td>□</td><td>75€</td></tr>
<tr><td>□</td><td>□</td><td>100€</td></tr>
</table>

APPENDIX B

Let G_j^i be disjoint events with i = {1,.., n} and j = n and S_G be a sample space, then:

<u>*Statement 1.*</u> $P(S_G)=1$

Consider the sample space S_G, I impose that $S_G = G_1^1$ = 1 by telling respondents that the probability associated to the entire sample space is equal to 1, say $S_G = G_1^1$ = 1.

<u>*Statement 2.*</u> $P(G_j^i)\geq 0$

Consider $P(G_2^1)$ and $P(G_2^2)$, I impose that $P(G_2^1)\geq 0$ and $P(G_2^2)\geq 0$ by asking respondents to the lower (g_0) and upper (g_1) bounds of the event space outside of which they are essentially certain the outcome cannot happen at all. This is basically the first question of Exchangeability Game.

<u>*Statement 3.*</u> If $\{G_j^i\}$ is a sequence of disjoint sets in S_G, then

$$P\left(\bigcup_{i=1}^{n} G_j^i\right) = \sum_{i=1}^{n} P(G_j^i)$$

Consider $P(G_2^1)$ and $P(G_2^2)$, "*exchangeability*" assumption imposes that

$$P\left(G_2^1 \bigcup G_2^2\right) = P(G_2^1) + P(G_2^2) = 0.5$$

<u>*Statement 4.*</u> $P(G_j^i)=1-P({G_j^i}^C)$

Consider $P(G_2^1)$ and $P(G_2^2)$, "*exchangeability*" assumption imposes that

$$P(G_2^1) = 1 - P(G_2^2) = 0.5 = 1 - 0.5$$

Statement 5. $P(\phi)=0$

See Statement 2.

Statement 6. For each $G_j^i \in S_G$, then $0 \le P(G_j^i) \le 1$

See Statements 1 and 2.

Statement 7. If $G_j^i \subset G_n^i$ with $n = jk, k \in N, k \neq 0$, then $P(G_j^i) \ge P(G_n^i)$

Consider G_4^1 and G_2^1, "*exchangeability*" assumption imposes that $P(G_2^1) = 0.5 \ge P(G_4^1) = 0.25$