UNIVERSITY
OF TRENTO - Italy
Department of Economics

Do Monetary Incentives and Chained Questions Affect the Validity of Risk Estimates Elicited via the Exchangeability Method? An Experimental Investigation

Simone Cerroni, Sandra Notaro, W. Douglass Shaw

n. 10/2011

Do Monetary Incentives and Chained Questions Affect the Validity of Risk Estimates Elicited via the Exchangeability Method? An Experimental Investigation

Simone Cerroni[*] Sandra Notaro[†] W. Douglass Shaw[‡§]

November 15, 2011

Abstract

Using a laboratory experiment, we investigate the validity of stated risks elicited via the Exchangeability Method (EM) by defining a valuation method based on de Finetti's notion of coherence. The reliability of risk estimates elicited through the EM has been theoretically questioned because the chained structure of the game, in which each question depends on the respondent's answer to the previous one, is thought to potentially undermine the incentive compatibility of the elicitation mechanism even when real monetary incentives are provided. Our results suggest that superiority of real monetary incentives is not evident when people are presented with chained experimental design.

Keywords: lab experiment, risk elicitation, exchangeability, validity, pesticide residue

JEL classification: C44; D81; I10

[*] Department of Economics, University of Trento; Fondazione Eni Enrico Mattei. Email: simone.cerroni@untin.it. Corresponding author.

[†]Department of Economics, University of Trento. Email: sandra.notaro@unitn.it.

[‡] Department of Agricultural Economics, Texas A&M University; Hazards Reduction and Recovery Center, Texas A&M University. Email: wdshaw@tamu.edu.

[§] We thank Roberta Raffaelli and Marco Tecilla for help in desiging, organizing and running the experiment at the Computable and Experimental Economics Laboratory (CEEL) at the University of Trento. We appreciate comments on the experimental wording from Ilaria Pertot; on the approach from Matteo Ploner; and on the paper from Roberta Raffaelli, Mary Riddel, and Richard Woodward. All remaining errors and shortcomings are, of course, entirely the authors' responsibility. This research was funded by Autonomous Province of Trento, project ENVIROCHANGE, Call for Major Project 2006.

1 Introduction

During the last decade, many social scientists have become more interested in investigating perceptions of risks, and eliciting subjective estimates of probabilities. The reason is that people often behave and make decisions according to their beliefs and these do not always coincide with science-based estimates of risks[1]. Failure to recognize the existence of divergent subjective risks may create quite a puzzling interpretation of responses to the science-based risks.

There are many different ways in which to elicit subjective risks, and several are briefly discussed below. In this paper, we use an innovative risk elicitation technique known as the Exchangeability Method (EM), focusing on issues related to implementing it in a credible manner. Our application is to 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). The study is conducted using subjects in laboratory experiments. Pesticide residues pose health risks to people who eat apples, and thus people's perceptions of their presence can affect consumers' purchasing behaviors. The investigation of this issue is quite important to this region in Italy because the saleable gross production of apple production is approximately 23% of the entire agricultural saleable gross production in that Province (P.A.T., 2007).

The reliability of stated risks estimates elicited via the EM has been questioned because the chained structure of the experimental design is thought to potentially undermine the incentive compatibility of the elicitation mechanism. An elicitation mechanism is incentive compatible if subjects have an incentive to state their real preferences (Vossler and Evans, 2009). Previous studies have overcome this issue, however, by presenting people with particular experimental designs that partially hide the chained structure of the game (Baillon, 2008; Abdellaoui et al., 2011). However, to our knowledge, no study has ever tested the effect of chained questions on the validity of stated risk estimates elicited via Exchangeability Method.

[1] A very timely example of the crucial effect of perceived risks on people's behaviors is represented by the proposal of the Italian Government to postpone the referendum (June 12nd-13rd 2011) on the installation of nuclear plants in Italy. This postponement was proposed because the recent events in Fukushima (Japan) might have strongly influenced citizens' perceived risks of nuclear disasters and, thus, influenced their desire to have or not have nuclear plants in Italy.

Our laboratory experiment uses a method for determining and measuring the validity of stated risk estimates elicited via the EM. This method is based on de Finetti's notion of coherence and allows us to test the validity of stated risks at both the sample and individual levels. By using this validation method we also aim to exam the potential effect of real monetary incentives and chained questions on stated risk estimates elicited via EM. In particular, we study whether these factors affect the validity of stated risks or not[2].

The remainder of the paper is laid out as follows. In the next sections, we first highlight the main strengths and limitations of the EM by comparing it to other risk elicitation techniques. Then, we more formally define the notion of validity and describe our testable hypotheses. Finally, we offer some conclusions based on the experimental results we have obtained.

2 Chained experimental designs and monetary incentives

The simplest way to elicit risks[3] consists of asking people to directly state either the chance that a specific magnitude of the outcome will happen in the future or, the other way round, the specific magnitude of the outcome that will happen with a certain probability (Spetzler and Von Holstein 1975). Asking simple stated risk questions is common in health risk studies, such as those involving smoking cigarettes (e.g. Viscusi 1990) or drinking contaminated water (e.g. Shaw et al. forthcoming). However, the reliability of risk estimates elicited via this family of techniques, called direct methods, have been often questioned because laypeople are usually not familiar with the notion of probability (e.g., Jakus et al. 2009; Riddel and Shaw 2006 for health or mortality risks; and Baker et al. 2009 for environmental risks).

Other studies overcome the limitations of direct methods by eliciting risk measures via indirect methods, for example, from respondents' choices over lotteries and for gambles or bets. In this case, probability measures are indirectly estimated at the points for which people

[2] 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, we cannot analyze the external validity of our 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).

[3] In this paper we define risk as the probability that given outcomes occur (or that given severities of an outcome occur).

show their indifference between lotteries or gambles. These indirect techniques have been mostly used for financial risks, (e.g., Andersen et al. 2009; Offerman et al. 2009), but recently scholars have considered them in the estimation of health and environmental risks (e.g., Fiore et al. 2009; Cerroni and Shaw 2011 for environmental risks)[4].

The most popular of the indirect methods are the so-called "external reference events" in which people are asked to choose between a lottery characterized by an uncertain event (U) whose probability needs to be estimated and a lottery characterized by an external reference event (K) whose probability is known and is disclosed to respondents. 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 risk communication devices (e.g., Morgan and Henrion 1990). Once respondents become indifferent between the two lotteries, this means that they attach to the uncertain outcome (U) the same probability with which the familiar outcome (K) will happen, so that P(U) = P(K) (Spetzler and Von Holstein 1975). Although these techniques are widely used, they have a crucial drawback, related to the notion of source dependence. Some experimental studies have recently shown that individual choices depend on the source of uncertainty that respondents have been asked to consider (Kilka and Weber 2001; Abdellaoui et al. 2011)[5]. When individuals have to process more than one source of uncertainty at the same time, the choices becomes too complex and their risk estimates might be biased. This is likely to occur in most external-reference-events approaches, i.e., those in which subjects have to deal with uncertainties related to both outcomes and probabilities represented through external devices.

The source dependence problem does not appear in another class of indirect methods, which use internal events. In these, subjects deal with magnitudes of the outcomes, but not with probabilities of the outcomes. In fact, subjects or survey respondents 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 respondents become indifferent to bet on one disjoint subspace rather than on the others, they are assumed to perceive those subspaces as equally likely (Spetzler and Von Holstein 1975). The EM that

[4] The limited use of these indirect methods for eliciting health and environmental risks is due to the fact that health outcomes and very long term environmental outcomes cannot be played out at the end of experiments in the lab setting, thus making incentive compatibility again an issue.

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

was formally described by Raiffa (1968) and more recently implemented by Baillon (2008) and Abdellaoui et al. (2011) belongs to this class of risk elicitation techniques.

As noted above, the EM unfortunately is criticized for potential failure to be incentive compatible, even when chained question structures are used with real monetary incentives. Questions are defined as chained when one question is constructed depending on respondents' answers to the previous one. For instance, because of sub-dividing event spaces, the two sub-events that respondents face in one question of an EM task depend on respondents' choices during the previous question. In their empirical application of EM, Abdellaoui et al. (2011) pointed out that:

> "…one may be concerned about it being advantageous for subjects not to answer according to their true preferences in a question but instead to seek to improve the stimuli that will occur in future questions" (p. 44).

Previous investigations that develop games with chained structures and real monetary incentives, have taken this issue very seriously. For instance, some of them have validated their results by using respondents' statements of unawareness about the presence of chained questions in the game (Van de Kuilen et al. 1981; Abdellaoui et al. 2011). In his own recent application of the Exchangeability Method, Baillon (2008) dealt with this problem by randomizing or resorting the order of questions and making the chaining unclear to respondents, such that they are no longer aware of the relationship between the disjoint subspaces they face in one question with those of the previous question.

It is clear however, that in all previous studies that utilize chained games in presence of real monetary incentives, the authors have tried to avoid the use of identifiable chained questions in their experimental designs, but have not investigated their presumed potential negative effect on subject's choice-behaviours. Hence, our study also empirically tests the presence of a potential "chaining effect" by comparing the validity of stated risk estimates elicited via EM with and without chained questions.

Baillon (2008), states that telling the truth is the simplest and most efficient strategy respondents can use when they play the games that constitute the EM tasks. This means that

---

mechanism of uncertainty".

subjects would not respond differently to tasks whether real monetary incentives were provided or not, because they are already consistent with incentive compatibility. In fact, in their recent application of Exchangeability Game, Abdellaoui et al. (2011) have tested the effect of real monetary incentives on people's choice-behaviours by comparing stated risk estimates given by two groups of respondents, one provided with monetary incentives and the other not. They concluded that the former group provides less noisy risk estimates than the latter group. However, given that their analysis uses a between–subjects investigation, the slight difference or discrepancy in their results may be due to different compositions of samples.

Below, we describe a method to definitively test possible superiority of stated risk estimates elicited via EM when people are rewarded with unchained questions. As we explain below, our validation method is based on the de Finetti's notions of exchangeability and coherence.

3 The notions of valid risk estimates and validity rate

Taking inspiration from the de Finetti's notion of coherent probability measures (de Finetti 1937; 1974a; 1974b)[6], we consider risk measures elicited via EM as valid if and only if they obey to all axioms and theorems of probability theory[7]. As noted above, we also construct a sample validity rate which is the percentage of respondents providing valid risk estimates in the sample.

The choice of using the de Finetti's notion of coherence to define valid risk measures 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)[8].

---

[6] de Finetti (1937) stated that *"...a complete class of incompatible events E_1, E_2,..., E_n being given, all the assignments of probability that attribute to p_1, p_2,..., p_n any values whatever, which are non-negative and have a sum equal to unity, are admissible assignment: each of these evaluations corresponds to a coherent opinion, (...), and every individual is free to adopt that one of these opinions (...) which he feels*".

[7] de Finetti's (1937, 1974) definition of "coherence" is related to the notion of probability. We extend his definition to the notion of risk because we define risk as the probability that a given event occurs.

[8] Exchangeability implies that the probability that each event belonging to the set occurs is the same, without depending on the order of the events, but only on the number n of events. Hence, even the joint probability of all events belonging to a set of n events is always the same and does not depend on the order of the events (de Finetti 1937).

Chew and Sagi defined two events as comparable, under a probabilistic point of view, only when a sub-event of one is exchangeable with the other event. This way of comparison is intuitively straightforward considering that a sub-event is logically less likely than the event in which it is contained. In other words, for probabilistically sophisticated subjects playing exchangeability games, two disjoint sub-events are exchangeable, and thus they have the same probability of occurrence, when they are indifferent to betting on one sub-event rather than on the other one.

4 Predictions

We first hypothesize that the provision of real monetary incentives to respondents do not have additional beneficial effects on the validity of stated risk estimates because telling the truth is the simplest and most efficient strategy respondents can use when they play the Exchangeability Game (EG) (Baillon 2008).

Providing real monetary incentives and in contrast, not providing such real incentives, we want to test whether the use of chained questions per se, affects the validity of risk estimates. We hypothesize that chained experimental designs have negative effects on the validity of stated risk estimates because they not only undermine the incentive compatibility of the game (Baillon 2008), but also generate meaningless questions where subjects are asked to choose between two prospects that they have already ruled out in previous questions. This may happen when subjects play the part of the game related to the elicitation of the second quartile.

5 The Experimental Design

5.1 The empirical application

Our specific application consists of investigating stated risks related to fire blight, a bacterial disease that has threatened apple orchards in the Province of Trento, at least since 2003 (IASMA 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 (Edmund Mach Foundation).

Italian farmers currently control the fire blight and the negative consequences that this has on apple production by using some preventative measures which consist in spraying pesticides based on copper compounds or Acibenzolar-S-metile on orchards. Unfortunately, these measures might be not efficient enough to prevent the future spread of fire blight and consequent reductions in the production of apples. 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. The only strategy that is easily available from a technical point of view to farmers is the introduction of new active principles for preventative and curative control of fire blight 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, we 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)[9], the number of apples containing at least one residue in a sample of 100 apples in 2030 (a)[10], and the number of apples containing more than one residue in a sample of 100 apples in 2030 (r)[11]. These variables have been selected among many other possible measures of pest infestation, or apple contamination, after having interviewed approximately 20 focus group subjects.

5.2 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 we recruit non-students and, then, we bring them in the lab, we can define our study as an artefactual field experiment (Harrison and List, 2004). Our sample consists of people between

[9] The blossoming period usually occurs in April in Trentino.
[10] The apple containing residues are those containing at least one residue beyond the level of 0 mg/kg.
[11] The apple containing residues are those containing at least two residues beyond the level of 0 mg/kg.

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 leaving in this Province. Moreover, the random nature of the sample may be biased by subjects' motivation to participate in the experiment. For example, subjects may participate because they were interested in the topic or because they were in need of the show-up fee. Selected participants were randomly assigned to four subsamples or treatment groups, where each treatment is characterized by a different experimental design: "real incentives-unchained questions" (22 subjects), "real incentives-chained questions" (23 subjects), "hypothetical incentives-unchained questions" (19 subjects), and "hypothetical incentives-chained questions" (16 subjects). Next, the specific EM games or tasks are described.

5.3 The Exchangeability Method and the related game

Let a random variable under study in the EM game be g. The EM game 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 EM 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. These might be individual-specific, reflecting heterogeneity that allows formation of a set of possibilities a subject believes are feasible.

The second step of the EM 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 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/he 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 1 and Appendix A). At the end of the Exchangeability Game, the second binary question that respondents have already answered is presented again to them in order to test the consistency of their choice behaviours.

5.4 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 EG (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 people 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, 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, the respondent is asked to choose between a lottery, in which she/he wins 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 questions 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 EM game. Each subject in our 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[12].

5.5 Treatments

Recall from above that the validity of risk estimates are investigated by implementing four experimental treatments: the real monetary incentives-chained questions (TRC), the real monetary incentives-unchained questions (TRU), the hypothetical monetary incentives-chained questions (THC), and the hypothetical monetary incentives-unchained questions (THU).

Hereafter R refers to real monetary, H to hypothetical, C to chained, and U to unchained. For the H treatments, subjects are only given a show-up fee, while in the R 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

[12] Both games have been already used to test exchangeability in other experimental applications (e.g., Baillon, 2008; Abdellaoui et al., 2010).

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 respondent 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 the Edmund Mach Foundation (EMF). This procedure for 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 research that EMF has done that provides that science-based estimate of probabilities. Note that even when all subjects receive the same risk information, it is a common finding that they may not form the same subjective estimates (e.g. Riddel and Shaw 2006; Shaw et al. forthcoming). 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 C 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 U treatments, this chained structure of the game is hidden through a mixed up order of questions determined once and for all. In fact, we elicit the percentiles of respondents' 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 respondent, regardless of the treatment group to which she/he is randomly assigned, plays exchangeability games and lotteries three times, one for each random variables under study.

6 Hypotheses

Given the theoretical background of the EM, all definitions, axioms and theorems of probability theory are satisfied under the exchangeability assumption. Considering two disjoint sub-events, G_j^i and G_j^k, this 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 we fail to reject the following null hypothesis (H_0):

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

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

We test this first assumption, and thus the validity of stated risk estimates elicited via the EM by investigating whether respondents' choice behaviours are consistent across the EG, the REG, and CEG. In particular, we test two hypotheses:

Hypothesis 1. We test whether the exchangeability assumption is satisfied or not by comparing the estimates of $g_{1/2}$ obtained from the EG and the estimates of $g_{1/2}$' obtained from repeated version of the game (REG). The exchangeability assumption is satisfied if and only if we 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. We test whether the exchangeability assumption is satisfied or not by comparing the certainty equivalents that respondents are willing to accept to give up the possibility to play the lotteries presented in the matched pairs of choice tasks, $[L(x : G_j^i)]$ in CT1 and $[L(x : G_j^k)]$ in CT2. The exchangeability assumption is satisfied if and only if we fail to reject the following null hypotheses (H_0):

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

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

7 Testing hypotheses

Before testing these hypotheses, we check the consistency of subjects' choice behaviours by examining their answers to the repeated binary questions presented at the end of the Exchangeability Game. The McNemar test shows that subjects' choices are stable across treatments[13].

Now, testing our hypotheses at sample level, we examine the role of monetary incentives and chained questions in affecting the validity of stated risk estimates and we identify the experimental design providing the highest percentage of valid risk measures. We determine whether respondents belonging to diverse experimental treatments provide valid risk estimates or not. Recall that respondents provide valid stated risk estimates if and only if we fail to reject the null hypotheses presented in *Hypotheses 1* and *2*.

We test *Hypotheses 1* and 2 by using nonparametric 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 of the possibility that the assumptions behind the WMP test are not always satisfied in our sample. For example, the differences between the matched values provided by each subject are not always distributed symmetrically around the median point in our sub-samples (this is the symmetry assumption).

Testing *Hypothesis 1*, we only investigates the validity of median risk measures ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$) elicited via the EG and REG. Testing *Hypothesis 2*, we also examine the validity of quartile risk estimates since this hypothesis relates to observations of median and quartile values 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 $r_{3/4}$) elicited via the EG and CEG.

Further, we assess the *validity rate* (V) for each different experimental treatment, where V is the percentage of respondents in each group providing valid risk estimates. In this case, we 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 respondent (i = 1, 80) is valid or not. For example, let's consider one specific experimental subject who provides us with the estimate of $g_{1/2}$. We assume that this risk estimates is valid if and only if $CE\left[L\left(x:G_2^1\right)\right]=CE\left[L\left(x:G_2^2\right)\right]$. 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 $CE\left[L\left(x:G_2^2\right)\right]$.

Finally, since we hypothesize that not only the features of the experimental setting may determine the validity of respondents' stated risk estimates, but also their socio-economic

[13] Results are available from the corrisponding author under request.

conditions, we econometrically test this hypothesis by estimating a model in which the discrete dependent variable captures the validity of each observation provided by each respondent. A set of explanatory dummy variables captures the characteristics of each experimental setting in terms of chaining and monetary incentives, and other socio-economic variables characterizing the subjects, allowing for some observable heterogeneity.

8 Results

8.1 Non-parametric tests

By testing *Hypothesis 1* for each experimental group of respondents, we identify the effect of our experimental designs on respondents' capability to provide valid estimates of the median values. In the TRC we have 24 matched pairs of observations; in the TRU 40; in the THC 22; and in the THU 26 (Table 1).

The validity of median estimates of individual CDFs ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$) is determined by testing *Hypothesis 1* via both the WMP and the SMP tests. Median estimates are assumed to be valid if and only if we fail to reject the null hypothesis characterizing this test. Based on the WMP test, TRU and THU groups provide valid stated risk estimates, while TRC and THC do not, however the validity of WMP test about the THC group may be compromised because all assumptions behind the test are not completely satisfied. The SMP test almost produces the same results except for the fact that also THC group provides valid estimates (Table 2). The discrepancy between WMP and SMP results about the THC group suggest that the interpretation of these results is problematic, and thus, we conclude that only TRU and THU groups provide valid risk estimates.

Testing *Hypothesis 2* for each experimental group of respondents allows us to investigate whether respondents belonging to diverse experimental treatments provide valid risk estimates of the median and quartile values of individual CDFs or not. In the TRC we have 143 matched pairs of observations; in the TRU 167; in the THC 136; and in the THU 115 (Table 3). Again, the validity of median, first quartile, and second 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 we fail to reject the null hypothesis characterizing this test. The WMP test shows that the

TRC and the THU groups do not provide valid risk estimates, while the TRU and the THC do. However, the SMP test suggests that also the THC do not provide valid risk estimates, and thus the TRU is the only group providing valid risk measures (Table 4). Again, dissimilar results obtained by the WMP and SMP tests do not allow us to express reliable findings about the validity of risk estimates obtained from the THC group. Hence, we conclude that the only group providing valid estimates is the TRU.

8.2 The validity rate and the econometric analysis

For each treatment, we calculate the validity rate (*V*) which is simply the percentage of valid risk estimates within each treatment group. According to the previous findings, we found that TRU provides the highest validity rate (39.13%), then the THU (29.86%), TRC (26.26%), and THC (21.64) follow. Comparing the validity rates of THU (29.86%) and TRC (26.26%), we conclude that the usage of chained experimental design totally undoes the beneficial effect of using real monetary incentives (Table 5).

Further, we hypothesize that not only experimental designs, but also socio-economics characteristics of respondents and their degree of familiarity with the problem influence individual performances in terms of validity. This hypothesis is econometrically tested by estimating diverse discrete models in which the dependent variable *VALID* represents the validity of each risk estimates provided by each respondents. The dependent variable takes the value 1 if and only if the stated risk estimate is valid, and thus $CE[L(x : G_j^i)] = CE[L(x : G_j^k)]$, with $k \neq i$, $k \leq j$.

The probability that the risk estimate is valid depends on a set of explanatory variables available from survey-type questions given in the laboratory, the experimental treatment that respondents belong to, the socio-economics status of respondents themselves, and respondents' degree of interest in the issue of food security (see Table 6 for details about the explanatory variables).

Given that each respondent *i* provides 9 risk 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 $r_{3/4}$), we should have a panel data of 720 observations. However, we have 142 missing values for the dependent variable *VALID* because the game investigating the validity of risk estimates was not always displayed to respondents depending on the choices they made during the experiment.

We estimate five diverse models (Model 1, 2, 3, 4, and 5) by using the generalized linear model estimation with and without robust standard errors. Hereafter, we focus on the estimation with robust standard errors that allows for clustering effects.

In Model 1, the probability of providing valid risk estimates only depends on the features of the experimental treatment respondents belong to. The influence of real monetary incentives and chained questions is captured by the following set of dummy variables (*T*), *TRC*, *TRU*, *THC*, and *THU*. Each dummy takes the value 1 if and only if the respondents belong to the treatment that the variable represents. The THC is used as baseline.

Model 1 $VALID_i = \beta_0 + \beta_1 T_i$

We observe that respondents who belong to the TRC, TRU, and THU have higher probability of providing valid risk estimates than respondents who belong to the THC (Table 8). However, this result is statistically significant only for the TRU dummy variable.

In Model 2, we also include two sets of dummy variables. The first (*RV*) captures whether the probability of providing valid risk estimates depends on the variable that respondents have to consider in playing the EG, the percentage (or number) of days in which the infestation will occur during the blossoming period in 2030 (*G*), the number of apples containing at least one residue in a sample of 100 apples in 2030 (*A*), and the number of apples containing more than one residue in a sample of 100 apples in 2030 (*R*). G is used as baseline. The second (*P*) aims to capture whether the validity of stated risk estimates is statistically different among median estimates ($g_{1/2}$, $a_{1/2}$, and $r_{1/2}$), first-quartile estimates ($g_{1/4}$, $a_{1/4}$, and $r_{1/4}$), and second-quartile estimates ($g_{3/4}$, $a_{3/4}$, and $r_{3/4}$), where the latter is the baseline (Table 8).

Model 2 $VALID_i = \beta_0 + \beta_1 T_i + \beta_2 RV_i + \beta_3 P_i$

We found no statistical difference in terms of validity between risk estimates related to diverse variables and diverse percentiles.

In Model 3, we also investigate the effects of socio-economic variable (*S*) on the probability that respondents provide valid risk estimates. We take our cues from extensive psychological research on the role that several factors can play in the determination of perceived risk. The variables under study are age (*AGE*), gender (*FEMALE*), education (*SECONDARY, HIGH_SCHOOL,* and *UNIVERSITY*), and the type of education (*SCIENTIFIC*) (Table 8).

Model 3 $VALID_i = \beta_0 + \beta_1 T_i + \beta_2 RV_i + \beta_3 P_i + \beta_4 S_i$

We expected that the probability of providing valid risk estimates would possibly increase for high educated and younger respondents, but we found that older respondents' estimates are more likely to be valid than the others and that education does not affect the validity of individual risk estimates (Table 8).

In Model 4, we consider also the interest of respondents on apples and food security 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*).

Model 4 $VALID_i = \beta_0 + \beta_1 T_i + \beta_2 RV_i + \beta_3 P_i + \beta_4 S_i + \beta_5 I_i$

Although we expected to observe that people who reside in the Province of Trento and consume and/or produce apples perform better than the others in terms of valid risk estimates, perhaps because they are more interested in the topic, our empirical results suggest no significant explanatory effects for these variables (Table 8).

In Model 5, we add another set of dummy variables (*T*) which capture whether subjects trust the predictions of IPCC about temperature and precipitation in 2030 (IPCC_TRUST), the predictions of EMF about the fire blight's infestation risk in 2030 (EMF_TRUST), and our statement that apple farmers will continue to use the chemical control against apple disease in the future (SCENARIO_TRUST).

Model 5 $VALID_i = \beta_0 + \beta_1 T_i + \beta_2 RV_i + \beta_3 P_i + \beta_4 S_i + \beta_5 I_i + \beta_5 T_i$

In this case, we predict that subjects who trust the information we gave them during the experimental instructions more likely provide valid risk estimates, as we suppose that the truster plays the game more carefully. Despite our predictions are confirmed overall, we found the trust in EMF's predictions reduces the probability of providing valid risk estimates. However, the estimation of Model 2, 3, 4, and 5 provides us with results about the effect of our experimental designs that are very similar to those obtained by the estimation of Model 1.

The consistency of our econometric results with those obtained from non-parametric tests suggests that real monetary incentives increase the validity of stated risk estimates and that chained questions have the opposite effect. Moreover, we found that socio-economic variables and the interest of respondents in the topic do not influence the likelihood of providing valid risk estimates. Only the age of respondents affects their ability to state valid estimates.

9 Conclusions

The paper has considered the influence of real monetary incentives and chained ordering of questions on risk elicitation. Based on median risk estimates, our statistical analysis shows that unchained treatments provide valid risk estimates, while chained do not. This finding suggests that the chained questions undermine the incentive compatibility of the game even when respondents are provided with real monetary incentives (Baillon, 2008; Abdellaoui et al., 2011).

Furthermore, when a treatment group is presented with a design with sorted questions, so that the chained structure is hidden, these subjects provide valid risk estimated even when they are not paid based on their performances. This supports Baillon's (2008) contention that regardless of being given actual monetary incentives or not, respondents play the games by just telling the truth about their beliefs. A caveat is that this result only takes subjects' median risk estimates in account, without considering observations related to the first and second quartiles.

Considering the whole set of stated risk estimates and not just the median estimates, we found that the only treatment group providing valid estimates received real money payments and unchained questions. When more of the distribution is being considered, real monetary incentives strongly affect respondents' performances in terms of validity. However, the beneficial effect of real monetary incentives on the validity of stated risk estimates is negated when subjects are presented with the experimental design of the game clearly chained. This finding is confirmed by our measures of the validity rate (V). The percentage of valid risk estimates is almost 40% when subjects are presented with real monetary incentives and the experimental design where the chaining is hidden. The validity rate falls to 26% with hypothetical monetary incentives and the experimental design where the chaining is clear and to 29% with real monetary incentives and the experimental design where the chaining is hidden.

Those interested in using this risk elicitation methodology can thus walk away with two important messages here. First, subjects are indeed more likely to provide valid risk 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 if they are 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 risk estimates (40%) can be considered valid under the definition we have applied here, which relates to behavioral axioms. The latter implication may be of little surprise to skeptics, but is relevant in our goal to continue to improve ways to provide reliable information about people's risk perceptions and subjective probabilities.

Further researches on the validity of stated risk estimates elicited via the 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 risk estimates. This would be possible by collecting, for each subject, a number of observations large enough to test the validity of her/his stated risks by using non-parametric tests.

Tables and Figures

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

Table 2. 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-Chained questions	Median($X_{1/2}$) =Median($X_{1/2}$')	-2.234**	0.0625
Real incentives-Unchained questions	Median($X_{1/2}$) =Median($X_{1/2}$')	-0.665	0.4807
Hypothetical incentives-Chained questions	Median($X_{1/2}$) = Median($X_{1/2}$')	-1.880***	0.1250
Hypothetical incentives-Unchained questions	Median($X_{1/2}$) = Median($X_{1/2}$')	-1.174	0.2668

*1% significance level
**5% significance level
***10% significance level

Table 3. Summary statistics of the Certainty Equivalents obtained via CEG

Treatment	Variable	Obs	Mean	St.Dev.	Min	Max
Real incentives-Chained questions	CE_{L1}	143	51.21	46.38	0	125
	CE_{L2}	143	76.95	44.69	0	125
Real incentives-Unchained questions	CE_{L1}	167	59.80	42.31	0	125
	CE_{L2}	167	68.22	41.72	0	125
Hypothetical incentives-Chained questions	CE_{L1}	136	70.80	43.30	0	125
	CE_{L2}	136	75.86	42.14	0	125
Hypothetical incentives-Unchained questions	CE_{L1}	115	55.65	36.14	0	125
	CE_{L1}	115	73.17	37.11	0	125

Table 4. 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-Chained questions	Median(CE_{L1}) = Median(CE_{L2})	-3.713*	0.0027
Real incentives-Unchained questions	Median(CE_{L1}) = Median(CE_{L2})	-1.513	0.3049
Hypothetical incentives-Chained questions	Median(CE_{L1}) = Median(CE_{L2})	-1.283	0.0886
Hypothetical incentives-Unchained questions	Median(CE_{L1}) = Median(CE_{L2})	-3.005*	0.0000

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

Table 5. Validity rates (*V*) for all treatments			
Treatment	Number of observations	Number of valid observations	*V* (%)
Real-Chained	192	52	26.26
Real-Unchained	207	81	39.13
Hypothetical-Chained	171	37	21.64
Hypothetical-Unchained	144	43	29.86

Table 6a. Description of dependent and independent variables of Model 1, 2, 3 and 4

Variable	**Definition**	**Mean**	**St.Dev.**	**Min**	**Max**
VALID	= 1 if valid, = 0 otherwise	.368	.482	0	1
TRC	= 1 if "Real Incentives-Chained Questions" treatment, = 0 otherwise	.275	.446	0	1
TRU	= 1 if "Real Incentives-Unchained Questions" treatment, = 0 otherwise	.287	.452	0	1
THC	= 1 if "Hypo Incentives-Chained Questions" treatment, = 0 otherwise	.237	.425	0	1
THU	= 1 if "Hypo Incentives-Unchained 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 apples containing at least one pesticide residue	.333	.471	0	1
R	Number of apples containing multiple pesticide residue	.333	.471	0	1
50^{th} PERC.	Observations related to the median of G, A, and R	.333	.471	0	1
25^{th} PERC.	Observations related to the I quartile of G, A, and R	.334	.471	0	1
75^{th} PERC.	Observations related to the II quartile of G, A, and R	.333	.471	0	1

Table 6b. Description of dependent and independent variables of Model 1,2, 3 and 4					
Variable	**Definition**	**Mean**	**St.Dev.**	**Min**	**Max**
CONSUMER	= 1 if the respondent eats at least 3 apples a week = 0 otherwise	.478	.500	0	1
CONS_ASS	= 1 if the respondent is a member of a consumer association = 0 otherwise	.062	.242	0	1
PRODUCER	= 1 if the respondent produces apples = 0 otherwise	.037	.190	0	1
TRENTINO	= 1 if the respondent 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 (at 5 levels)[a]	2.950	.545	0	4
EMF_TRUST	Trust in EMF's predictions of fire blight's infestation risk in the future (at 5 levels)[a]	2.587	.684	0	4
SCENARIO_ TRUST	Agreement with the fact that farmers will use the chemical control in the future (at 5 levels)[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 respondent have this education level, = 0 otherwise	.1830	.3895	0	1
HIGH_SCHOOL	= 1 if the respondent have this education level, = 0 otherwise	.5070	.5035	0	1
UNIVERSITY	= 1 if the respondent have this education level, = 0 otherwise	.3098	.4657	0	1
SCIENTIFIC	= 1 if the respondent 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

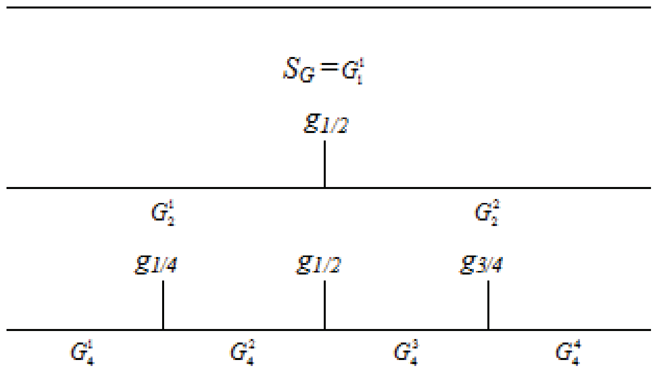
Table 7. Generalized Linear Model Estimation of Models 1,2, 3, and 4					
Dependent Variable: VALID					
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
TRC	.218	.220	.251	.382**	.370**
TRU	.520*	.505*	.545*	.575*	.648*
THU	.251	.238	.278	.319***	.385**
A	-	-.052	-.036	-.045	-.058
R	-	-.131	-.130	-.143	-.173
MEDIAN	-	-.085	-.088	-.090	-.077
25th PERC	-	-.124	-.109	-.110	-.094
FEMALE	-	-	-.131	-.126	-.097
AGE	-	-	.015*	.013**	.019*
SEC_SCHOOL	-	-	-.185	-.179	-.086
HIGH_SCHOOL	-	-	-.055	-.037	-.016
SCIENTIFIC	-	-	-.009	.084	.173
PRODUCER	-	-	-	.595***	.584***
CONSUMER	-	-	-	-.025**	-.021***
CONS_ASS	-	-	-	.184	.312
TRENTINO	-	-	-	.273***	.067
IPCC_TRUST	-	-	-	-	.359*
EMF_TRUST	-	-	-	-	-.355*
SCEN_TRUST	-	-	-	-	.253*
CONSTANT	-.606*	-.468*	-.901*	-1.086*	-2.160*
LOG L.HOOD	-374.018	-373.110	-365.030	-359.370	-347.702

*1% significance level
**5% significance level
***10% significance level

Table 8. Generalized Linear Model Estimation of Models 1,2, 3, and 4 with robust standard errors and clustering effects					
Dependent Variable: VALID					
Variable	**Model 1 (Robust)**	**Model 2 (Robust)**	**Model 3 (Robust)**	**Model 4 (Robust)**	**Model 5 (Robust)**
TRC	.218	.220	.251	.382	.370
TRU	.520**	.505***	.545***	.575**	.648**
THU	.251	.238	.278	.319	.385
A	-	-.052	-.036	-.045	-.058
R	-	-.131	-.130	-.143	-.173
MEDIAN	-	-.085	-.088	-.090	-.077
25^{th} PERC	-	-.124	-.109	-.110	-.094
FEMALE	-	-	-.131	-.126	-.097
AGE	-	-	.015**	.013	.019***
SEC_SCHOOL	-	-	-.185	-.179	-.086
HIGH_SCHOOL	-	-	-.055	-.037	-.016
SCIENTIFIC	-	-	-.009	.084	.173
PRODUCER	-	-	-	.595	.584
CONSUMER	-	-	-	-.025	-.021
CONS_ASS	-	-	-	.184	.312
TRENTINO	-	-	-	.273	.067
IPCC_TRUST	-	-	-	-	.359***
EMF_TRUST	-	-	-	-	-.355**
SCEN_TRUST	-	-	-	-	.253***
CONSTANT	-.606*	-.468*	-.901**	-1.086**	-2.160**
LOG L.HOOD	-374.018	-373.110	-365.030	-359.370	-347.702

*1% significance level
**5% significance level
***10% significance level

Figure 1: Structure of the experimental design



Appendix A: Games' Examples

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:

Example 2. First question of the Repeated

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

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

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. One question from 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 B?

Option A		**Option B**	
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	□	□	0€
	□	□	25€
	□	□	49€
	□	□	51€
	□	□	75€
	□	□	100€

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 B?

Option A		**Option B**	
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	□	□	0€
	□	□	25€
	□	□	49€
	□	□	51€
	□	□	75€
	□	□	100€

Appendix B

Definition, axioms and theorems of the probability theory

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

Statement 1 $P(S_G) = 1$

Consider the sample space S_G, we 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 $S_G = G_1^1 = 1$.

Statement 2 $P(G_j^i) \geq 0$

Consider $P(G_2^1)$ and $P(G_2^2)$, we 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.

Statement 3 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

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

Statement 4. $P(G_j^i) = 1 - P\left({G_j^i}^C\right)$

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 $0 \le P\left(G_j^i\right) \le 1$

See Statements 1 and 2.

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

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

References

ABDELLAOUI, M., BAILLON, A., PLACEDO, L. and WAKKER, P.P. (2011). The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation. *American Economic Review*, **101**, 695-723.

ANDERSEN, S., FOUNTAIN, J., HARRISON, G.W. and RUTSTRÖM, E.E. (2010). Estimating Subjective Probabilities. Working Paper 09-01, Department of Economics, College of Business Administration, University of Central Florida.

ANDERSEN, S., FOUNTAIN, J., HARRISON, G.W. and RUTSTRÖM, E.E. (2009). Estimating Aversion to Uncertainty. Working Paper 07-09, Department of Economics, Copenhagen Business School, University of Copenhagen.

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., RIDDEL, M., and Woodward, R.T. (2009). Explaining Changes in Subjective Hurricane Risks as Time Passes: An Analysis of a Sample of Katrina Evacuees. *Journal of Risk Research*, **12** (1), 59-74.

CERRONI, S. and SHAW, W.D. (2011). Does Climate Change Information Affect Elicited Risks Related to Pine Beetle Impacts on Forests? An Application of the Exchangeability Method. III Workshop in Valuation Methods in Agro-food and Environmental Economics: Decision and Choices under Uncertainty in Agro-food and Natural Resource Economics. July 1st-2sd, 2010, Castelldefels (Spain).

CHEW, S.H. and SAGI, J. (2006). Event exchangeability: Probabilistic sophistication without continuity or monotonicity. *Econometrica* **74**, 771–786.

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: *The Concept of Probability in Psychological Experiments*, Stael Von Holstein C.-A.S, D. Reidel Publishing Company, Dordrecht-Holland/Boston-U.S.A. pp. 1-14.

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

FIORE, S.M., HARRISON, G.W., HUGHES, C.E. and RUSTRÖM, E.E. (2009). Virtual experiments and environmental policy. *Journal of Environmental Economics and Management* **57** (1), 65-86.

HARRISON, G.W. (1986). An experimental test for risk aversion. *Economic Letters* **21**, 7–11.

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

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

JAKUS, P.M., SHAW, W.D., NGUYEN, T.N and WALKER, M. (2009). Risk Perceptions of Arsenic in Tap Water and Bottled Water Consumption. *Water Resource Research,* Doi: 10.1029/2008WR007427

KILKA, M. and WEBER, M. (2001). What Determines the Shape of the Probability weighting function under uncertainty? *Management Science,* **47**, 1712–1726.

MACHINA, M.J. and SCHMEIDLER, D. (1992). A More Robust Definition of Subjective probability. *Econometrica,* **60**, 745-780.

MORGAN, M.G. and 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. *Bulletin OEEP/EPP,* **34**, 381-382.

OFFERMAN, T., SONNEMANS, J., VAN DE KUILEN, G. and WAKKER, P.P. (2009). A Truth Serum for NON-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. *Review of Economic Studies,* **76**, 1461-1489.

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/produzione_lorda_vendibile/Pat_Agricoltura_bassa.1282904118.pdf [accessed 02.10.2011]

RAIFFA, H. (1968). *Decision Analysis*. Addison-Wesley, London.

RIDDEL, M. and SHAW, W.D. (2006). A Theoretically-Consistent Empirical Non-Expected Utility Model of Ambiguity: Nuclear Waste Mortality Risk and Yucca Mountain. *Journal of Risk and Uncertainty,* **32** (2), 131-150.

SHAW, W.D., JAKUS, P.M. and RIDDEL, M. (2011). Perceived Arsenic-Related Mortality Risks for Smokers and Non-smokers. *Contemporary Economic Policy*, forthcoming, available at SSRN: http://ssrn.com/abstract=1804937.

SPETZLER, C.S. and STAEL VON HOLSTEIN, C.-A.S. (1975). Probability encoding in decision analysis. *Management Science,* **22**, 340–358.

TVERSKY, A. and KOEHLER, D.J. (1994). Support Theory: A Nonextensional Representation of Subjective Probability. *Psychological review,* **4**, 547-567.

VAN DE KUILEN, G., WAKKER, P.P. and ZOU, L. (2006). A Midpoint Technique for Easily Measuring Prospect Theory's Probability Weighting. CREED, University of Amsterdam, The Netherlands.

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

Elenco dei papers del Dipartimento di Economia
pubblicati negli ultimi due anni

2009.1 *Income Shocks, Coping Strategies, and Consumption Smoothing. An Application to Indonesian Data*, Gabriella Berloffa e Francesca Modena

2009.2 *Clusters of firms in space and time*, Giuseppe Arbia, Giuseppe Espa, Diego Giuliani e Andrea Mazzitelli

2009.3 *A note on maximum likelihood estimation of a Pareto mixture*, Marco Bee, Roberto Benedetti e Giuseppe Espa

2009.4 *Job performance and job satisfaction: an integrated survey*, Maurizio Pugno e Sara Depedri

2009.5 *The evolution of the Sino-American co-dependency: modeling a regime switch in a growth setting*, Luigi Bonatti e Andrea Fracasso

2009.6 *The Two Triangles: What did Wicksell and Keynes know about macroeconomics that modern economists do not (consider)*? Ronny Mazzocchi, Roberto Tamborini e Hans-Michael Trautwein

2009.7 *Mobility Systems and Economic Growth: a Theoretical Analysis of the Long-Term Effects of Alternative Transportation Policies*, Luigi Bonatti e Emanuele Campiglio

2009.8 *Money and finance: The heterodox views of R. Clower, A. Leijonhufvud and H. Minsky*, Elisabetta de Antoni

2009.9 *Development and economic growth: the effectiveness of traditional policies*, Gabriella Berloffa, Giuseppe Folloni e Ilaria Schnyder

2009.10 *Management of hail risk: insurance or anti-hail nets?*, Luciano Pilati, Vasco Boatto

2010.1 *Monetary policy through the "credit-cost channel". Italy and Germany pre- and post-EMU*, Giuliana Passamani, Roberto Tamborini

2010.2 *Has food price volatility risen*? Chrisopher L. Gilbert e C. Wyn Morgan

2010.3 *Simulating copula-based distributions and estimating tail probabilities by means of Adaptive Importance Sampling*, Marco Bee

2010.4 *The China-US co-dependency and the elusive costs of growth rebalancing*, Luigi Bonatti e Andrea Fracasso

2010.5 *The Structure and Growth of International Trade*, Massimo Riccaboni e Stefano Schiavo

2010.6 *The Future of the Sino-American co-Dependency*, Luigi Bonatti e Andrea Fracasso

2010.7 *Anomalies in Economics and Finance*, Christopher L. Gilbert

2010.8 *Trust is bound to emerge (In the repeated Trust Game)*, Luciano Andreozzi

2010.9 *Dynamic VaR models and the Peaks over Threshold method for market risk measurement: an empirical investigation during a financial crisis*, Marco Bee e Fabrizio Miorelli

2010.10 *Creativity in work settings and inclusive governance: Survey based findings from Italy*, Silvia Sacchetti, Roger Sugden e Ermanno Tortia

2010.11 *Food Prices in Six Developing Countries and the Grains Price Spike*, Christopher L. Gilbert

2010.12 *Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration*, Giuseppe Espa, Diego Giuliani, Giuseppe Arbia

2010.13 *I Would if I Could: Precarious Employment and Childbearing Intentions in Italy*, Francesca Modena e Fabio Sabatini

2010.14 *Measuring industrial agglomeration with inhomogeneous K-function: the case of ICT firms in Milan (Italy)*, Giuseppe Espa, Giuseppe Arbia e Diego Giuliani

2010.15 *Dualism and growth in transition economies: a two-sector model with efficient and subsidized enterprises*, Luigi Bonatti e Kiryl Haiduk

2010.16 *Intermediaries in International Trade: Direct Versus Indirect Modes of Export,* Andrew B. Bernard, Marco Grazzi e Chiara Tomasi

2010.17 *Production and Financial Linkages in inter-firm Networks: Structural Variety, Risk-sharing and Resilience,* Giulio Cainelli, Sandro Montresor e Giuseppe Vittucci Marzetti

2010.18 *Organizational Capital and Firm Performance. Empirical Evidence for European Firms,* Claudia Tronconi e Giuseppe Vittucci Marzetti

2010.19 *The Fund-flow Approach. A Critical Survey,* Giuseppe Vittucci Marzetti

2010.20 *Is Open Source about Innovation? How Interaction with the Open Source Community Impact on the Innovative Performances of Entrepreneurial Ventures,* Evila Piva, Francesco Rentocchini e Cristina Rossi Lamastra.

2011.1 *Leaving home and housing prices. The Experience of Italian youth emancipation,* Francesca Modena e Concetta Rondinelli.

2011.2 *Pareto versus lognormal: a maximum entropy test,* Marco Bee, Massimo Riccaboni e Stefano Schiavo.

2011.3 *Does a virtuous circle between social capital and CSR exist? A "network of games" model and some empirical evidence,* Giacomo Degli Antoni e Lorenzo Sacconi

2011.4 *The new rules of the Stability and Growth Pact. Threats from heterogeneity and interdependence,* Roberto Tamborini

2011.5 *Chinese reserves accumulation and US monetary policy: Will China go on buying US financial assets?* Luigi Bonatti e Andrea Fracasso

2011.6 *Taking Keller seriously: trade and distance in international R&D spillovers*, Andrea Fracasso e Giuseppe Vittucci

2011.7 *Exchange Rate Exposure under Liquidity Constraints,* Sarah Guillou e Stefano Schiavo

2011.8 *Global Networks of Trade and Bits,* Massimo Riccaboni, Alessandro Rossi e Stefano Schiavo

2011.9 *Satisfaction with Creativity: A Study of Organizational Characteristics and Individual Motivations*, Silvia Sacchetti e Ermanno Tortia

2011.10 *Do Monetary Incentives and Chained Questions Affect the Validity of Risk Estimates Elicited via the Exchangeability Method? An Experimental Investigation*, Simone Cerroni, Sandra Notaro e W. Douglass Shaw

PUBBLICAZIONE REGISTRATA PRESSO IL TRIBUNALE DI TRENTO