# Imprecise Datasets as a Source for Ambiguity A Model and Experimental Evidence*

Ayala Arad† and Gabrielle Gayer‡

September 17, 2009

**Abstract**

In our model an individual forms beliefs over events based on the frequencies of occurrences of the events in past cases. However, in some cases, he might not know whether or not a specific event has occurred. Our model suggests that ambiguity may arise due to this sort of partial information and that attitude towards ambiguity can be explained by the way the individual process such imprecise cases. An individual who tends to put low weight on the possibility that an event occurred in these imprecise cases will turn out to be ambiguity averse, whereas an individual who tends to put high weight on the possibility that this event occurred will turn out to be ambiguity loving.

The model is followed by an experiment designed to test the main features of the model. It is corroborated that given precise data subjects are ambiguity neutral while given imprecise data subjects are ambiguity averse.

*We wish to thank Anat Bracha, Itzhak Gilboa, Dotan Persitz, Ariel Rubinstein, Ella Segev, Peter Wakker and participant of the PhD seminar at Tel Aviv University for many helpful comments and suggestions.

†The Eitan Berglas School of Economics, Tel Aviv University, Ramat Aviv, Tel Aviv, 69978, Israel. E-mail: aradayal@post.tau.ac.il

‡Department of Economics, University of Haifa, Haifa 31905, Israel. E-mail: gabi.gayer@gmail.com

# 1 Introduction

## 1.1 Motivation

Suppose that you are about to undergo some medical treatment. You have agreed to this treatment knowing that it was successful in 30 out of 50 past cases and it was unsuccessful in the remaining 20. Instead, assume that you were given the same information as before and were further notified that 20 additional records were lost due to a technical problem and therefore the success of the treatment in these cases is unknown. How would you feel about going through the treatment under these alternative circumstances? Is it possible that the additional information in this second scenario lowers your confidence in the treatment? In this work we explore how such vague information effects beliefs that are formed based on past data.

Many experimental studies, following Ellsberg (1961), show that people's beliefs in uncertain (ambiguous) situations cannot be quantified by a single additive probability measure. Among the most notable theories that can accommodate the observed behavior are Schmeidler's Choquet Expected Utility (CEU) model (1989), and Gilboa and Schmeidler's Maxmin Expected Utility (MMEU) model (1989). In order to account for non-neutral attitudes to ambiguity the former uses a capacity (non additive probability) and the latter uses a set of probabilities. However, none of these theories explains where beliefs come from, nor what the origin of uncertainty attitudes is. In these works as well as in others the axiomatic derivations are in the spirit of the subjective expected utility derivations of de Finetti (1937), Savage (1954), and Anscombe-Aumann (1963). "Beliefs" are derived from an "as-if" representation using only what is considered observable behavior data and above all no claim is made about the actual cognitive processes that leads to decisions.

This approach has several limitations. First, in the absence of a theory of belief formation, one cannot tell which beliefs are reasonable in a given context. This point was forcefully made by Gilboa, Postlewaite, and Schmeidler (2004), and it seems to be a major motivation for recent derivations of case-based probabilities by Billot, Gilboa, Samet, and Schmeidler (2005) and Gilboa, Lieberman, and Schmeidler (2006). We maintain that a theory of belief formation which restricts the range of reasonable beliefs is just as important when discussing non-Bayesian beliefs.

Second, the standard axiomatic approach remains silent on the causes of, say, ambiguity aversion. What are the basic mechanisms that underlie this phenomenon? Is it an almost-paranoid tendency to believe that nature is malevolent, as suggested by Maccheroni, Marinacci, and Rustichini (2006)? Is it a personality trait, in which case we may expect that entrepreneurs would be less prone to it than the population at large, or is it a feature of the environment? While there is some neurological work attempting to compare patterns of brain activation while making decisions under risk and under uncertainty (Hsu, Bhatt, Adolphs, Tranel, and Camerer 2005, and Rustichini, Dickhaut, Ghirardato, Smith and Pardo 2005), there is practically no cognitive account of uncertainty aversion in the literature.

Third, models that elicit beliefs solely from choices have been criticized for their inability to distinguish between ambiguity as an epistemic state and ambiguity aversion as a feature of preferences. A decision maker who faces a mildly uncertain situation and is very sensitive to uncertainty may behave in the same manner as a decision maker who faces an extreme uncertain situation and is almost indifferent to uncertainty. Gajdos, Hayashi, Tallon, and Vergnaud (2006) overcame this problem by introducing objective information into the objects over which preferences are defined. However, in their work the information is provided as sets of priors, whereas statistical information is rarely given in such a form. It appears that our understanding of ambiguity aversion would benefit from a cognitive model that can explicitly describe available information, as well as the process by which it is used to generate non-Bayesian beliefs.

In the present model an individual forms a belief over an event using empirical frequencies in past cases. However, some information regarding this event may be imprecise. That is, in some past cases the event is neither known to have occurred nor to have not occurred. In this model ambiguity emerges due to this feature of the data. Had data been precise no ambiguity would arise and beliefs would be determined uniquely according to the number of past cases in which this event occurred out of the total number of cases. With imprecise data the frequency of occurrences is consistent with a range of values. This brings to mind two opposite approaches individuals may follow when processing data. An individual following the first approach will tend to treat these imprecise cases as if the evaluated event had not occurred in them, resulting in the formation of "low" beliefs. An individual following the second approach will tend to treat these past cases as if the event had occurred in them, resulting in the formation of "high" beliefs. In the framework of CEU the former beliefs correspond to ambiguity aversion which traditionally is interpreted as pessimism and equivalently, the latter beliefs correspond to ambiguity loving or optimism.

The model is accompanied by an experiment which was designed to examine if actual behavior of decision makers is in accordance with that imposed by the model. In particular it tests whether imprecise data is a cause for ambiguity aversion. The results indicate that individuals who are presented with precise information form additive beliefs, while those who are presented with imprecise information form non-additive beliefs. In fact it is shown that these non-additive beliefs reflect ambiguity aversion.

## 1.2 Example

Mr. Blue is trying to predict whether the stock market will rise or fall tomorrow. When forming beliefs Mr. Blue bases his evaluation on past cases. A good starting point for his evaluation of an event is the experienced relative frequency of occurrences of this event. He can perfectly recall the exceptional days in which the stock market rose or fell dramatically, say by at least 4%. He can consequently deduce in which days the change in the stock market was less significant (we refer to this event as event $A$). His evaluation of $A$ will then equal the number of these days out of the total number of days in which he observed the stock market. Since in most days the stock market did not change radically, he may feel confident enough to assign a high probability to the event that the stock market will change by less than 4%.

In addition, on most accounts Mr. Blue cannot recall the exact change of the stock market on days in which it was not significant. Specifically he can only remember a small amount of days in which the stock market fell (rose) by less than 4% and generally he cannot be sure about what occurred in the rest. Now Mr. Blue is asked to express his belief regarding the event "the stock market will **fall** by less than 4%" (event $B$). Consider the following two extreme approaches for evaluating the likelihood of this event: 1. assuming there is no sufficient reason to believe that the stock market is more likely to rise rather than fall (and vice versa), the evaluation will be approximately half of the likelihood ascribed to event $A$. 2. Considering only days in which he is certain this event occurred (and disregarding the rest) the evaluation will be the relative frequency of such days out of the total number of days. This second approach totally neglects the possibility that the event may have occurred in some of the remaining days. Hence, Mr. Blue will most definitely be willing to say that his belief regarding event $B$ is higher than the second evaluation. Still under these circumstances, he may feel reluctant to evaluate this event at the level of the first evaluation and therefore he will lower his belief beneath this value. Consequently, his evaluation will be between the two extreme evaluations. Given that he has the same type of data with respect to the event "the stock market will **rise** by less than 4%" (event $C$), his evaluation of this event will follow a similar path. Thus, Mr. Blue's beliefs are non additive as the sum of the evaluation of $B$ and $C$ is lower than the evaluation of their union (event $A$).

In the present model, an imprecise event is an event about which data is vague. That is, in some past cases it is neither known to have occurred nor to have not occurred (such as the events $B$ and $C$). A precise event is an event that is either known to have occurred or known to have not occurred in each past case (such as event $A$). In circumstances where some events are imprecise, Mr. Blue's beliefs will convey his attitude towards ambiguity. Evaluation by the second approach will exhibit extreme ambiguity aversion while evaluation by the first approach will reflect ambiguity neutrality. It is equally possible to demonstrate beliefs which exhibit an ambiguity loving attitude by modifying the belief formation process in the appropriate way.

The rest of the paper is organized as follows. In the next section two evaluation approaches which exhibit different attitudes toward ambiguity are formally introduced and studied, followed by more related literature. Proceeding, section 3 contains the experimental design and the main findings. Finally, Section 4 includes a discussion.

# 2 The Model and Results

## 2.1 Belief Formation

Let $\Omega = \{\omega_1, ..., \omega_n\}$ be a finite set of states of nature ($n \geqslant 2$) and let $\Sigma$ be an algebra of subsets of $\Omega$ called events which is given by the power set $2^\Omega$. A dataset $D$ of length $T$, is a sequence of events indexed by $i = 1, ..., T$ :

$$D = (B_1, ..., B_T)$$

where $B_i \in \Sigma \backslash \{\phi, \Omega\}$ . A dataset is also referred to as memory.

The set of all datasets or memories of length $T$ is denoted by $\mathcal{D}_T := \Sigma \backslash \{\phi, \Omega\}^T$ and the set of datasets of any length is denoted by $\mathcal{D} := \bigcup_{T \in \{1,2,...,\infty\}} \mathcal{D}_T$.

A case $i$ is the $i^{th}$ element of a dataset $D$. The event $B_i$ is interpreted as the information that the individual has regarding the occurrences in case $i$. Two types of cases are excluded, the first are cases in which $B = \phi$. Excluding these cases implies that the evaluator is aware of all the possible states of nature, thus the event that is known to have occurred cannot be empty. The second type of cases that are excluded are cases in which $B = \Omega$. These cases add no information to the evaluation of the outcomes since they do not narrow the set of states of nature that have or have not occurred. Observe that for any two cases $i$ and $j$ for which $B_i \subseteq B_j$, case $i$ is more informative than $j$, since the set of events that are known to be realized and not realized inferred by case $i$ includes the set of these events inferred by case $j$.

A capacity $v$ is a mapping from $2^{\Omega}$ into $[0,1]$ such that $v(\phi) = 0$, $v(\Omega) = 1$ and $v$ is monotone, that is $v(A) \geq v(A')$ whenever $A \supseteq A'$. Let $\mathcal{V}$ be the set of all capacities.

In this model $v$ depends on data, namely, $v : \mathcal{D} \to \mathcal{V}$. For any dataset $D$, $v_D$ is the capacity that the decision maker attaches to the dataset. Given $D = (B_1, ..., B_T)$, we define $\forall j \leq T$

$$F_j(A) = \begin{cases} 1 & \text{if } A \supseteq B_j \\ \alpha \frac{|A \cap B_j|}{|B_j|} & \text{otherwise} \end{cases} \tag{1}$$

and

$$G_j(A) = \begin{cases} 0 & \text{if } A \cap B_j = \phi \\ 1 - \alpha \frac{|A^c \cap B_j|}{|B_j|} & \text{otherwise} \end{cases} \tag{2}$$

where $A^c$ denotes $A$'s complement and $0 \leq \alpha \leq 1$. Two types of evaluations of an event are considered:

$$v_D^F(A) = \frac{\sum_{j=1}^{T} F_j(A)}{T} \tag{3}$$

and

$$v_D^G(A) = \frac{\sum_{j=1}^{T} G_j(A)}{T}. \tag{4}$$

In some cases in the dataset the individual may not be sure whether event $A$ occurred or not. Knowing that $B$ occurred in case $j$, equations (1) and (2) determine how much weight the individual puts on the possibility that $A$ occurred in this case. When $A \supseteq B$ it is obvious that $A$ occurred and the maximum weight (i.e., 1) is put on this possibility, according to both equations, while when $A \cap B = \phi$ it is obvious that $A$ had not occurred and no weight is put on this possibility. When the conditions above are not satisfied (that is, when $A \not\supseteq B$ and $A \cap B \neq \phi$) it is unclear whether $A$ occurred and the equations generally provide different answers.

When $\alpha = 1$ the equations are identical, and the weight put on the possibility that $A$ occurred equals the proportion of states in $B$ which suggest that $A$ occurred. With this parameter, the equations yield an additive probability measure and therefore this approach is referred to as neutral. For $\alpha < 1$, an individual following (1) will put a lower weight on the possibility that $A$ occurred compared to the neutral approach, whereas an individual following (2) will put a higher weight on this possibility. Roughly speaking, the former approach leads to low beliefs, while the later approach leads to high beliefs. We refer to the former as a pessimistic approach and to the later as an optimistic approach.

Beliefs increase with $\alpha$ according to Equation (3) and decrease according to Equation (4). For the extreme value of $\alpha = 0$, Equation (3) corresponds to the lowest possible beliefs given the dataset. An individual holding such beliefs presumes the event has not occurred unless he is informed otherwise. Equation (4) corresponds to the highest possible beliefs given the dataset. Holding these beliefs is in accord with assuming that the event has occurred unless informed otherwise.[1] Any belief in the range between these two extreme beliefs is considered plausible given the dataset.

Equations (3) and (4) should be treated only as a rough approximation of a belief formation process and must not be taken literally. It is their qualities we find appealing and not necessarily their exact values. According to both equations, a stronger indication that an event $A$ occurred in case $j$ increases the perception of its likelihood.[2] However, as long as this indication is vague, the belief according to (3) will tend to be

[1]The lowest value of $v_D^F(A)$ and the highest value of $v_D^G(A)$ can also be interpreted as the lowest and highest additive probabilities assigned to event $A$ in the core, respectively.

[2]This is captured by the property that both formulas increase with the expression $\frac{|A \cap B_j|}{|B_j|}$.

relatively low while the belief according to (4) will tend to be relatively high.

In the next section we study the properties of beliefs based on (3) and (4) and connect them to the literature. Among the two, we find procedure (3) much more natural and in the experimental section it is verified that participants' behavior is consistent with it.

## 2.2 Attitudes towards Ambiguity

In this section it is shown that Equation (3) represents ambiguity aversion while Equation (4) represents ambiguity loving, where the degree of ambiguity aversion or love depends on $\alpha$.

First, a few known properties of capacities that will prove useful are presented. A capacity $v$ is *convex* if for all $A$ and $A'$ $v(A) + v(A') \leq v(A \cup A') + v(A \cap A')$ and it is *concave* if the inequality is reversed.

A capacity $v$ is a *belief function*[3] if it satisfies, for any collection $A_1, ..., A_n$ of subsets of $\Omega$, $v(\bigcup\limits_{i=1,...,n} A_i) \geq \sum_{\{H:H\subseteq\{1,...,n\}\}}(-1)^{|H|+1}v(\bigcap\limits_{i\in H} A_i)$. This condition is known to be equivalent to the following: for every event $A$, $\sum_{B\subseteq A}(-1)^{(|A|-|B|)}v(B) \geq 0$. Note that $n = 2$ is the convexity condition of $v$, thus every belief function is convex.

Define $\bar{v}$ by $\bar{v}(A) = 1 - v(A^c)$.[4] Then the following properties hold:

- $v$ is a capacity if and only if $\bar{v}$ is a capacity
- $v$ is concave if and only if $\bar{v}$ is convex
- if $v$ is a probability then $v = \bar{v}$.

Note that for any event $A$, $G_j(A) = 1 - F_j(A^c)$, so we get $v_D^G = \bar{v}_D^F$.

---

[3] A "belief function" is a technical term in Dempster (1967 and 1968) and Shafer's (1976) theory.

[4] In the context of cooperative game theory $\bar{v}$ is called the dual game and when $v$ is a belief function Dempster (1967 and 1968) and Shafer (1976) call $\bar{v}$ an upper probability.

The concept of ambiguity aversion is that decision makers prefer to be exposed to randomness in situations in which probabilities are known as opposed to situations in which they are unknown. In order to capture this idea Schmeidler (1989) suggested a behavioral axiom by which a decision maker who is indifferent between two alternatives will (weakly) prefer the mixture of the two. The rationale behind this notion is that when probabilities are unknown one alterative can be used as a hedge against the other, which thereby reduces the uncertainty. Similarly, the reverse preference towards mixing reflects ambiguity loving while indifference expresses ambiguity neutrality.

Schmeidler (1989) shows that in the context of CEU this notion of ambiguity aversion (loving) translates into convexity (concavity) of the capacity. A capacity which is a probability, naturally, reflects ambiguity neutrality. Not all the literature agrees that convexity (concavity) is a necessary condition for ambiguity aversion (loving), yet it generally does agree that it is a sufficient one (see, for example, Ghirardato and Marinacci, 2002 and Epstein and Zhang, 2001. See an exception to this approach in Wakker, 2008 which discusses the importance of relative convexity).

According to the above notions, using the CEU decision rule the following propositions establish the attitudes towards ambiguity of beliefs based on data in our model:

**Proposition 1** $v_D^F$ *and* $v_D^G$ *as defined in equations (3) and (4) respectively are capacities.*

**Proposition 2** $v_D^F$ *as defined in Equation (3) is a belief function.*

All proofs can be found in Appendix A. Proposition 1 establishes that $v_D^F$ and $v_D^G$ are both capacities. Proposition 2 shows that $v_D^F$ is in fact a belief function and thus convex. Moreover, since $v_D^G = \bar{v}_D^F$ the convexity of $v_D^F$ implies the concavity of $v_D^G$. Therefore, evaluating the likelihood of events according to Equation (3) leads to behavior of ambiguity aversion, while evaluating the likelihood of events according to Equation (4) leads to behavior of ambiguity loving. When $\alpha = 1$ we have neutrality towards ambiguity. Note that when there is no vagueness regarding the occurrences of events in the data (that is, data are precise) any $\alpha$ leads to an additive probability and therefore the individual's attitude towards ambiguity in these circumstances cannot be identified.

Ghirardato and Marinacci (2002) developed a notion of comparative ambiguity aversion[5] by which 1 is more ambiguity averse than 2 if for every two alternatives, one ambiguous and the other not, if 2 prefers the unambiguous alternative over the ambiguous one then so does 1. In the context of CEU the characterization of this definition is that for every event, 2's capacity of this event is larger or equal to that of 1's.[6] In the present framework it is easily seen that for a given dataset, according to procedure (3) ((4)), a smaller (larger) $\alpha$ corresponds to a more ambiguity averse individual.

## 2.3 Ambiguous Events

In recent literature there are several definitions of ambiguous events. Usually ambiguous events are derived endogenously from observed behavior. Generally speaking, an event is deemed to be ambiguous if the decision maker's preferences imply so (without going into detail about the exact definitions). The definitions of ambiguous events were translated into conditions on the capacities in the CEU model, or into conditions on the set of probabilities in the MMEU model.

Nehring (1999) showed that for MMEU preferences, unambiguous events can be identified with events on which all probabilities agree. Furthermore, since a capacity can be associated with a set of probabilities, he was able to define unambiguous events in the CEU model by the associated probability set in the same man-

[5] Following Yaari's (1969) notion of comparative risk aversion Epstein (1999) was the first to define comparative ambiguity aversion.

[6] The notion of "more ambiguity averse" also requires that the two decision makers' utilities over prizes are essentially the same. However, they later argue that even when their utilities are not identical, the same idea of capacity domination can be used to compare between attitudes towards ambiguity.

ner. Zhang (2002) and Epstein and Zhang (2001) showed that for a subclass of CEU preferences with a convex (concave) capacity the set of unambiguous events is $\{A \,| v(A) + v(A') = v(A \cup A') \; \forall A'$ such that $A' \subseteq A^c\}$, which is equivalent to $\{A \,| v(A) + v(A^c) = 1\}$. Ghirardato and Marinacci's (2002) definition of unambiguous events agrees with Epstein and Zhang's definition when expressed in terms of capacities and with Nehring's definition when expressed in terms of the set of probabilities. Klibanoff, Marinacci and Mukerji's (2005) definition is the same as Nering's when expressed in terms of the set of probabilities.

Although this study's focus is not observed behavior of a decision maker, it is still possible to ask which events are perceived in our model as unambiguous by the definitions mentioned above. The following defines imprecise information and classifies the set of imprecise events. We proceed by showing how these definitions relate to the existing definitions of unambiguous events.

Given $D = (B_1, ..., B_T)$, we define case $j$ as *precise with respect to event* $A$ if $B_j \subseteq A$ or $B_j \subseteq A^c$. Namely, in case $j$ it is known whether or not $A$ was realized. We refer to $A$ as *precise with respect to case* $j$ when the condition is satisfied, and to $A$ as *precise* when it is precise with respect to every case in the dataset. This definition extends naturally to cases and datasets. Case $j$ is regarded as *precise* when it is precise with respect to every event in $2^\Omega$, i.e., when $B_j = \{\omega_i\}$ for some $\omega_i \in \Omega$. A dataset is *precise* when all cases in it are precise and is *imprecise* otherwise.

As stated in Proposition 1 and Proposition 2, beliefs in our model are represented by convex and concave capacities. It turns out (as the next Lemma shows) that under these conditions, Nehring's definition of unambiguous events is equivalent to that of Epstein and Zhang. Therefore, we refer to an event as *unambiguous* if it satisfies any one of these two definitions.

**Lemma 1** *Let $v$ be convex (concave) and let $A$ be an event. Then $v(A) + v(A') = v(A \cup A')$ for all $A'$ disjoint from $A$ if and only if $v(A) = \bar{v}(A)$.*

**Proposition 3** *For $\alpha < 1$, event $A$ is precise given $D$ if and only if $A$ is unambiguous. Furthermore, the set of precise events forms an algebra.*

Proposition 3 states that ambiguity, in our model, is due to partial information in the data, and that should have information been precise no ambiguity would arise.

The next proposition shows that only an imprecise memory leads to a non additive probability measure.

**Proposition 4** *Let $\alpha < 1$ and let $v_D^F$ and $v_D^G$ be defined as in equations (3) and (4) respectively, then $v_D^F$ and $v_D^G$ are probabilities if and only if the dataset is precise. In this case $v_D^F = v_D^G$.*

Note that for $\alpha = 1$, $v_D^F$ and $v_D^G$ are additive probabilities with no dependence on the precision of the data.

## 2.4 Bibliographic Note

Several other works discuss belief formation based on data. Demptser and Shafer (DS) (Dempster 1967, 1968 and Shafer 1976) presented a mathematical theory of evidence in which they modeled the connection between an individual's confidence in the belief and the evidence he possesses. They came up with a "belief function". It turns out that the evaluation according to procedure (3) satisfies their mathematical conditions.

Indeed, DS's theory and the present model share the same basic motivation. Both models attempt to capture the concept of beliefs based on judgment of evidence that need not turn out to be an additive probability measure. DS did not explain where the evidence comes from; it may be observations, rumors, or other sources of information. Nevertheless, their model mainly fits circumstances in which data contains different pieces of informations about the current situation rather than separate past cases.

In two related works, Jaffray (1991) and Gonzales and Jaffray (1998) construct a preference relation based on data which contain imprecise cases. This construction makes use of beliefs which fit our formalization with $\alpha = 0$.

Both Carnap (1952 and 1980) and Viscusi (1989) propose an updating process by which the posterior (additive) probability is a weighted average of a prior probability and the observed relative frequency, where the weight of the relative frequency depends positively on the sample size. More related to our work is Billot et al. (2005) which also presents a procedure describing how individuals form additive beliefs given available data. The main distinction between this model and ours is the type of data the individual can possess. Billot et al. considers only cases with precise information, while we allow for cases with imprecise information as well. This enables us to capture the notion of ambiguity due to imprecise information. Eichberger and Guerdjikova (2007) extends Billot et al. (2005) by considering the sample size in order to account for ambiguity.

# 3 Experimental Test of the Model

This section describes two experiments that examine whether actual behavior of decision makers is supportive of the model's main implications. The first experiment is concerned with belief formation given precise dataset. It tests whether the perception of the likelihood of an event matches the frequencies of occurrence of this event in the data. The second and the main experiment's aim is to shed light on formation of beliefs given an imprecise dataset and, in particular, to test the statement that decision makers are ambiguity averse in the presence of imprecise data.

Participants in Experiment 1 were 80 economics students in undergraduate and graduate studies at Tel-Aviv University. Participants in Experiment 2 were 292 undergraduate students both from economics at Haifa University and engineering at Ben-Gurion University.[7] The students were asked to participate in a short experiment (that lasted about 15 minutes) in the beginning of the class.

[7] Originally there were 5 more participants in the experiments whose answers were excluded, since it was clear from their forms that they misunderstood the questions.

The introductory instructions in the two experiments were the same. In the beginning of each session, before distributing the forms, the setting of the experiment was presented. The subjects were informed that some of them will be randomly chosen to participate in a lottery that will take place at the end of the experiment. This lottery, which is specified explicitly in their forms, is concerned with a drawn of a ball from an urn. Furthermore, not all of them face the same lotteries. The subjects were asked to state whether they prefer to participate in the lottery or rather to be given a certain amount of money. They were required to do so for every amount of money that appeared in their forms which included values that varied between 10 NIS to 140 NIS.

The subjects were told that those of them who would be randomly chosen at the end of the experiment (the proportion was approximately 1 subject out of 25) will get a monetary prize according to their choices. More specifically, a certain sum of money will be selected at random at the end of the experiment. If a selected student stated that he preferred this particular sum over the lottery, he would be given this sum, whereas if he stated that he preferred the lottery he would be given the amount according to its outcome (in line with the BDM procedure, Becker, DeGroot and Marschack, 1964). It was further explained that for the rest of the subjects the decisions are hypothetical, but since they all have a chance to be chosen, they all have a good reason to state their true preferences. Finally, it was emphasized that there is no correct answer and that the answer depends solely on their personal preferences. After explaining the experiment out loud the subjects were asked to read the instructions in their forms.

## 3.1 Experiment 1

In this experiment a comparison is made between the behavior of subjects who were provided with the exact proportions of different balls in an urn to that of subjects who were not informed about the proportions but were given a precise dataset of past cases (draws) in which the frequencies of occurrences of events fit the above proportions.

Two different kinds of forms were randomly distributed among all subjects within each class that participated in the experiment. The forms corresponded to two treatments: the experimental group (Treatment $a$) and the control group (Treatment $b$). The general structure of the forms in both experiments 1 and 2 can be found in Appendix B. Subjects in the experimental group were told that the urn contains a total of 90 balls of four different types with unknown proportions: yellow balls marked with $O$, white balls marked with $O$, yellow balls marked with $X$ and white balls marked with $X$. Then, they were given information concerning the outcomes in eight past draws of a ball from this urn (with replacement). This dataset appears in Table 1.[8] Subjects in the control group, who faced a different urn, were not provided with a dataset of past draws but rather were told that their urn contains exactly three yellow balls and five white balls.[9] The ratio of yellow balls in Treatment $b's$ urn was equal to proportion of observations in which a yellow ball was drawn in Treatment $a's$ dataset.

**Table 1 : Dataset of Treatment a**

| Case | **Ball Type** |
|---|---|
| 1 | Yellow with O |
| 2 | White with X |
| 3 | White with X |
| 4 | White with O |
| 5 | Yellow with X |
| 6 | White with O |
| 7 | Yellow ball |
| 8 | White with X |

Subjects in both treatments were offered to participate in the following lottery: "if at the end of the experiment, a yellow ball will be drawn from the urn, you will get

[8] All treatments in both experiments (apart from Treatment b in which there is no sample) included two versions in which the order of the cases in the dataset were shuffled.

[9] The subjects were not informed that there were two separate urns in treatments a and b.

150 NIS (around 40 USD). Otherwise, you will get nothing". Finally, subjects were asked to state whether they prefer the lottery over a sure amount of money $M$, for each $M \in \{10, 20, ..., 140\}$.

The certainty equivalent (CE) is taken to be the lowest amount $M$ that is preferred over the lottery. The average CE in treatment $i$ is denoted by $CE^i$ for $i = a, b$. Participants of both treatments were offered the same lottery, therefore a higher CE reflects a higher belief that a yellow ball will be drawn from their urn.[10] The null hypothesis of the experiment is that $CE^a = CE^b$.

**Results:**

A detailed distribution of choices in all treatments in the experiments appears in Appendix C. There were 38 subjects in Treatment $a$ and 42 subjects in Treatment $b$. The average CE of the experimental group and the control group were found to be $CE^a = 67.37$ and $CE^b = 69.52$ respectively. The results indicate that these two CEs are not significantly different ($t = -0.4$ and $p = 0.69$). Therefore, we cannot reject the null hypothesis that $CE^a = CE^b$. In addition, the samples' variances are also not significantly different ($F = 1.41$ and $p = 0.15$).

In other words, these findings support the statement that individuals form a belief over an event that matches the proportion of cases in memory in which this event had occurred.

Note that the result that there is no significant difference between the two CEs holds despite the small number of past cases in Treatment $a$'s dataset. This is in line with the evidence that people take small samples too seriously as in the phenomena of the "law of small numbers" (see Tversky and Kahneman, 1971).

[10] A discussion of this measure is postponed to Section 3.3.

### 3.2 Experiment 2

In order to demonstrate that subjects are ambiguity averse in the presence of imprecise data (that is, they behave in accordance with $F$ in Equation (3) and $\alpha < 1$),[11] it is shown that their beliefs (CEs) are lower than those induced by the neutral approach (i.e., $\alpha = 1$).

A direct test would compare the subjects' CE to that of ambiguity neutral individuals given the same imprecise data. The problem is that the CE value of ambiguity neutral individuals is unknown. The key of the experiment is to obtain this neutral CE indirectly in a parallel treatment, which is explained in detail in the following subsection.

Eight kinds of forms, which represented eight different treatments, were randomly distributed among all subjects within each class that participated in the experiment. All the forms had the same structure of the form of the experimental group in Experiment 1, that is, all forms had a dataset of 8 cases of past draws and a proposed lottery which was defined over the type of the ball that would be drawn at the end of the experiment. The states of nature in all treatments were the different types of balls that could be drawn from the urn. That is, $\Omega = \{$white with $O$, white with $X$, yellow with $O$ and yellow with $X\}$. The treatments differed in their dataset and their proposed lottery. The dataset that appeared in the treatments was either **precise with respect to the events in question** (shown in Table 3) or **imprecise with respect to any one of these events** (shown in Table 4). All lotteries had the same structure as the one that appeared in Experiment 1: you will win 150 NIS if the type of the drawn ball is $Z$ and 0 otherwise. The lotteries differed according to the type of the ball $Z$, which could be one of the following: white ($W$), yellow ($Y$), yellow with $O$ ($YO$) or yellow with $X$ ($YX$).

[11] Both the requirement that people follow procedure $F$ and that $\alpha < 1$ are necessary for ambiguity averse behavior. Stating both conditions each time is cumbersome, therefore in the sequel the condition of following $F$ is omitted.

The treatments are denoted by $T_Z^i$, where $i \in \{P, IP\}$ and $Z \in \{W, Y, YX, YO\}$. The upper index $i$ indicates whether the dataset is precise ($P$) or imprecise ($IP$), and the lower index $Z$ indicates what is the type of ball that will grant a prize if drawn out of the urn. The average CE of each treatment is denoted by $CE_Z^i$ in the same manner. A summarized description of all 8 treatments can be found in Table 2.

**Table 2: Treatment Description**

| **Treatment No.** | **Treatment Name** | **Sample type** | **Lottery type ($Z$)** | **No. of Subjects** |
|---|---|---|---|---|
| 1 | $T_W^P$ | Precise | $W$ | 32 |
| 2 | $T_W^{IP}$ | Imprecise | $W$ | 32 |
| 3 | $T_Y^P$ | Precise | $Y$ | 46 |
| 4 | $T_Y^{IP}$ | Imprecise | $Y$ | 42 |
| 5 | $T_{YO}^P$ | Precise | $YO$ | 46 |
| 6 | $T_{YO}^{IP}$ | Imprecise | $YO$ | 38 |
| 7 | $T_{YX}^P$ | Precise | $YX$ | 33 |
| 8 | $T_{YX}^{IP}$ | Imprecise | $YX$ | 23 |

The discussion is divided into two parts. The first part, which includes treatments $1-4$ tests the main hypothesis of the experiment and the second part which includes treatments $5-8$, is a robustness test which verifies that the results hold for different events.

#### 3.2.1 Part 1

The dataset that appeared in Treatment $T_W^P$ , consisted of eight past cases which were precise with respect to the event $W$ (see Table 3). The dataset of Treatment $T_W^{IP}$ is obtained by replacing two precise cases in the dataset of Treatment $T_W^P$ - one in which $W$ had occurred (Case 1) and one in which W had not occurred (Case 8) - by two imprecise cases with respect to the event $W$. All other six past cases were practically unaffected (see the dataset of $T_W^{IP}$ in Table 4).[12] The subjects were offered the lottery "win 150 NIS if a white ball is drawn and 0 otherwise" in both treatments $T_W^P$ and $T_W^{IP}$.

**Table 3: Precise dataset**

| Case | **Ball Type** |
|---|---|
| 1 | White with O |
| 2 | Yellow with X |
| 3 | Yellow with X |
| 4 | Yellow with O |
| 5 | White with X |
| 6 | Yellow with O |
| 7 | White ball |
| 8 | Yellow with X |

**Table 4: Imprecise dataset**

| Case | **Ball Type** |
|---|---|
| 1 | A ball with O |
| 2 | Yellow with X |
| 3 | Yellow with X |
| 4 | Yellow with O |
| 5 | White with X |
| 6 | Yellow |
| 7 | White ball |
| 8 | A ball with X |

According to the model, individuals who observe the precise dataset given in Treatment $T_W^P$, regardless of their $\alpha$, hold the same belief about event $W$ . Note that $F_1(W) = 1$ and $F_8(W) = 0$ given this precise data. An individual in Treatment $T_W^{IP}$ who follows the neutral approach (i.e., $\alpha = 1$) holds the same belief about $W$ as that of individuals in Treatment $T_W^P$, since given this imprecise dataset, $F_j(W) = \alpha\frac{|W \cap B_j|}{|B_j|} = \frac{1}{2}$ for $j = 1, 8$, and $F_j$ is the same as in Treatment $T_W^P$ for any $j \neq 1, 8$. In contrast, an individual with a pessimistic approach (i.e., $\alpha < 1$) holds a lower belief regarding $W$. Therefore, the experimental findings support the hypothesis that subjects follow a pessimistic approach (or $\alpha < 1$) if subjects in Treatment $T_W^{IP}$ hold a lower belief than that of subjects in Treatment $T_W^P$.

One may be concerned that the belief of subjects in Treatment $T_W^{IP}$ regarding event

[12]Case 6 is also different in the two samples. Nevertheless, this difference should not affect the results, since this case is precise with respect to the relevant events in treatments $1-4$. The reason for this difference will become apparent in part 2 of Experiment 2.

$W$ may indeed be lower than that of subjects in Treatment $T_W^P$ but that their belief regarding the complement event $Y$ may be higher. This set of beliefs would not reflect ambiguity aversion and would be inconsistent with the model. The role of treatments $T_Y^P$ and $T_Y^{IP}$ is to verify that this does not occur.

The dataset that appeared in Treatment $T_Y^P$ was identical to that in Treatment $T_W^P$ and the dataset that appeared in Treatment $T_Y^{IP}$ was identical to that in Treatment $T_W^{IP}$. The lottery that was offered in both $T_Y^P$ and $T_Y^{IP}$ was "win 150 NIS if a yellow ball is drawn and 0 otherwise". The hypothesis of this part of Experiment 2 is that the beliefs given imprecise data are lower than those given precise data both for the event $W$ and for the complement event $Y$. Put differently, both $CE_W^{IP} < CE_W^P$ and $CE_Y^{IP} < CE_Y^P$.

**Results of Part 1:**

The average CE of Treatment $T_W^P$ is $CE_W^P = 65.3$, and of Treatment $T_W^{IP}$ is $CE_W^{IP} = 50.9$. The results indicate that $CE_W^P$ is higher than $CE_W^{IP}$, where the difference between the two is significant ($t = -2.46$ and $p = 0.008$). Likewise, the average CE of Treatment $T_Y^P$ is $CE_Y^P = 78.5$ , and of Treatment $T_Y^{IP}$ is $CE_Y^{IP} = 70$. Here again, $CE_Y^P$ is higher than $CE_Y^{IP}$, and the difference between the two is significant at the 6% level ($t = -1.59$ and $p = 0.057$).

The only difference between treatments $T_W^P$ and $T_W^{IP}$ and between treatments $T_Y^P$ and $T_Y^{IP}$ is the imprecision of the dataset. Therefore, the findings support the model's main idea that imprecise data is a source for ambiguity aversion.

A simple alternative procedure of belief formation that may come to mind is that individuals form beliefs according to the frequency of occurrence of the precise cases in memory while ignoring the imprecise cases. Had this been true, $CE_W^P$ should have been higher than $CE_W^{IP}$ and $CE_Y^P$ should have been lower than $CE_Y^{IP}$and hence it is ruled out by the results.

Note that the experimental results support the model's idea that beliefs are based on the relative frequency of events in the data. To demonstrate this point a comparison between the results of Treatment $T_Y^P$ (in Experiment 2) and of Treatment $a$ in Experiment 1 is made. Recall that the only difference between the two treatments is that in the dataset of Treatment $a$, 3 out of 8 draws were yellow, while in that of Treatment $T_Y^P$ 5 out of 8 draws were yellow. In both treatments the lottery was "you will win 150 NIS if a yellow ball is drawn and 0 otherwise". It was found that $CE^a$ is significantly lower than $CE_Y^P$. This comparison indicates first, that subjects do base their beliefs on the dataset and second, that these beliefs are higher when the frequency of occurrence of an event in past cases is higher. This comparison should be treated cautiously, since the two treatments were conducted in separate classes. Nevertheless, the findings in treatments $T_W^P$ and $T_Y^P$, which were preformed in the same classes, further support this result. In Treatment $T_W^P$ the frequency of occurrence of $W$ is lower than that of $Y$ in Treatment $T_Y^P$, and at the same time $CE_W^P$ is significantly lower than $CE_Y^P$.

### 3.2.2 Part 2

Formally, in order to confirm that the subjects' beliefs exhibit ambiguity aversion in the presence of imprecise data, we need to elicit their beliefs for each event in $\Sigma$ (which amount to 14 events excluding $\phi$ and $\Omega$). Furthermore, these beliefs need to be obtained both for imprecise and precise data. This means that 28 treatments need to be preformed, which is way too large a number from any practical aspect. Therefore, in treatments $5-8$ we chose to focus on two such of these events, a draw of yellow with $X$ and a draw of yellow with $O$, both of which contain a single state of nature as opposed to the events in treatments $1-4$ which contain two. We view this as a robustness test which checks whether the previous results hold for additional events.

In this part four treatments were preformed, $T_{YO}^P$, $T_{YO}^{IP}$, $T_{YX}^P$ and $TI_{YX}^P$. The datasets in this part are the same as in the previous part. The dataset in treatments $T_{YO}^P$ and $T_{YX}^P$ is the same as in treatments $T_W^P$ and $T_Y^P$, which is precise with respect to the events $YO$ and $YX$ (see Table 3). The dataset in treatments $T_{YO}^{IP}$ and $T_{YX}^{IP}$ is the same as in treatments $T_W^{IP}$ and $T_Y^{IP}$, which is imprecise with respect to the events $YO$ and $YX$ (see Table 4). Note that the imprecise observations in Treatment $T_{YO}^{IP}$ are cases 1 and 6, while the imprecise observations in Treatment $T_{YX}^{IP}$ are cases 6 and 8. Furthermore, in the precise dataset the event $YO$ occurred in case 6 and had not in case 1, and the event $YX$ had occurred in case 8 and had not in case 6. This leads us to the same analysis and the same type of hypothesis from Part 1, that $CE_{YO}^{IP} < CE_{YO}^{P}$ and $CE_{YX}^{IP} < CE_{YX}^{P}$.

**Results of Part 2**:

The following are the main results: $CE_{YO}^{P} = 63.3$, and $CE_{YO}^{IP} = 52.4$. Evidently, $CE_{YO}^{P}$ is higher than $CE_{YO}^{IP}$ and significantly different ($t = -1.88$ and $p = 0.03$). Also $CE_{YX}^{P} = 69.1$ and $CE_{YX}^{IP} = 48.7$. Here again, $CE_{YX}^{P}$ is higher than $CE_{YX}^{IP}$, and significantly different ($t = 2.72$ and $p = 0.005$).

## 3.3 Some Remarks about the Design

Both experiments are based on a between subjects approach as opposed to a within subjects approach. Ideally, we may want to compare the belief of the same subject given alternative datasets or different information. However, one cannot hope to give the same subjects different datasets concerning the same event without worrying that the former dataset influences the evaluation of the event given the later dataset. In the between subjects approach the results indicate that on average the subjects are ambiguity averse in the presence of imprecise data. This does not rule out the possibility that some subjects are ambiguity neutral or even ambiguity loving.

In both experiments the subjects' beliefs are measured indirectly by their CEs. The reason we preferred this method over alternative techniques that elicit beliefs directly is twofold. First, the later techniques introduce a major concern that does not arise in the context of the method employed in the present work. Asking subjects about their beliefs directly may suggest to them that there exists a "correct" belief that they are expected to hold. Two such possibilities are a belief that puts equal weights on all states of nature regardless of the dataset and a belief that coincides perfectly with the dataset's frequencies. This problem is less likely to arise using our method as subjects are asked about their CE which has a much stronger association to subjective preferences. In addition, we find our experimental design much easier to convey to subjects.

Obviously, the alternative methods have the advantage that the elicited beliefs could have been compared to those predicted by the model. However, we maintain that the direction of change in beliefs given different datasets is much more important than the exact value of these beliefs. Especially, in view of the fact that we take the model only as a rough description of a belief formation process which tries to highlight one possible source of ambiguity among many. For the purpose of comparing the direction of change in beliefs both approaches are equally suitable.

# 4 Discussion

In this work we introduce a model of belief formation based on a dataset, where some observations are imprecise. The use of imprecise data leads to a belief which is non additive. Our model may be interpreted as an approximation of the actual mental process that the individual goes through while evaluating likelihood of events, which the individual may or may not be aware of. Focusing on the mental process enables to narrow down the possible beliefs that people would hold and highlight the role of imprecise information in causing ambiguity. We suggest that imprecise information is a source for non-neutral attitude towards ambiguity, in particular ambiguity aversion, and present experimental evidence which support this feature of the model.

Next, we discuss three modifications of the model, for which the main results would hold as well. Following Billot et al. (2005), we can easily enrich our model by allowing cases to vary in their characteristics.[13] Thus, an individual's evaluation of an event in a given situation will be a weighted average of the outcomes in past cases where the weights are determined by the relevance or similarity of these past cases to the current situation. When restricting cases to possess identical characteristics, the evaluation reduces to that of the basic model, i.e., to a simple average or the relative frequency of occurrences.

One of the main limitations of the present model is that the belief formation does not depend on the number of observations in the dataset. One might expect that when an individual accumulates more data about a certain situation, ambiguity will gradually disappear. Eichberger and Guerdjikova (2007) introduce ambiguity into the framework of Billot et al. (2005) by allowing the individual to hold a set of conceivable probabilities given past cases. As the dataset grows the ambiguity diminishes. It is possible to capture this element in our model by making additional assumptions about the memory accumulation process. For instance, assume memory relies on personal experience as well as other peoples' experience. Naturally, personal experience is more detailed and thus more precise than second hand experience. At an early stage of the individual's life his experience relies mostly on others' experience. With age, the proportion of personal experience of total experience grows. In this modification, a larger memory contains a higher proportion of precise cases. Therefore, beliefs based on it reflect less ambiguity.

In order to discuss another feature of our model, consider an extreme situation in which the individual has a single imprecise case in memory $B_j$. According to the model, all states in $B_j$ are perceived to be equally likely. This is in accordance with Laplace's principle of insufficient reason which is best applied to situations endowed with symmetry. This principle is not as appealing in asymmetric circumstances in which there is good reason to believe that some states are more likely than others.

[13] In an earlier version of this paper we do include this feature. This version includes only extreme attitudes toward ambiguity, i.e., $\alpha = 0$, and it also has an axiomatization for this case.

Our model could be easily modified to incorporate such an element by replacing $F_j(A)$ (Equation (1)) with:

$$\tilde{F}_j(A) = \alpha \frac{\sum_{\omega \in A \cap B_j} L(\omega)}{\sum_{\omega \in B_j} L(\omega)} \tag{5}$$

where $L(\omega)$ are all positive weights. If initially $\omega_i$ is perceived to be more likely than $\omega_j$, it is most reasonable to set $L(\omega_i) > L(\omega_j)$. It follows that given the same information in memory regarding $\omega_i$ and $\omega_j$ the evaluation of the former will be weakly higher than the later. In symmetric situations $L(\omega_i)$ equals $L(\omega_j)$ for every $i$ and $j$, and hence Equation (5) is reduced to Equation (1). In this modified version of the model based on Equation (5), the paper's main theoretical results continue to follow through.

The settings in the urn experiments reported in this paper have the underlying symmetry property in which Equation (1) applies. The following is a short description of an experiment that we have conducted, which fits the circumstances of the modified model better. The experiment was concerned with subjects' beliefs about the color of a car at a randomly chosen spot in a parking lot nearby. It is clear that some car colors are perceived to be more ordinary than others, such as blue compared to purple. Part of the subjects in the experiment were provided with the actual frequencies of the various colors of cars that were parked in the lot three days earlier, while others were provided with coarser data which were imprecise in the model's terms. The findings of the experiment indicate that imprecise data is a source for ambiguity aversion also in such asymmetric circumstances.

**Apendix A**

**Proof of Proposition 1**

Let $D = (B_1, ..., B_T)$. $v_D^F(\phi) = 0$ since for any $j$, $\phi \cap B_j = \phi$ and thus $F_j(\phi) = 0$. For any $j$, $\Omega \supseteq B_j$, thus $F_j(\Omega) = 1$ and $v_D^F(\Omega) = 1$. Finally, take any $A \subseteq A'$, for all $B_j$, $\{\omega | \omega \in A \cap B_j\} \subseteq \{\omega | \omega \in A' \cap B_j\}$ thus $F_j(A) \leq F_j(A')$ and $v_D^F(A) \leq v_D^F(A')$. Therefore $v_D^F$ is a capacity. Since $v_D^G(A) = \bar{v}_D^F(A)$, $v_D^G$ is a capacity as well. ■

**Proof of Proposition 2**

Take any dataset $D = (B_1, ..., B_T)$.First we will prove for any $j$, and any collection $A_1, ..., A_n$ of subsets of $\Omega$

$$F_j(\bigcup_{i=1,...,n} A_i) \geq \sum_{\{H:H \subseteq \{1,...,n\}\}} (-1)^{|H|+1} F_j(\bigcap_{k \in H} A_k) \quad (*)$$

holds i.e. $F_j$ is a belief function.

Note that for any probability measure $p$ the following rule holds:

$$p(\bigcup_{i=1,...,n} A_i) = \sum_{\{H:H \subseteq \{1,...,n\}\}} (-1)^{|H|+1} p(\bigcap_{k \in H} A_k).$$

Take any $j$ and any collection $A_1, ..., A_n$ of subsets of $\Omega$. Let $p$ be the probability measure defined by $p(A) = \frac{|A \cap B_j|}{|B_j|}$ for all $A \in \Omega$. Then if $B_j \subseteq A$, $F_j(A) = p(A)$ and otherwise $F_j(A) = \alpha p(A) \leq p(A)$ (for $\alpha \leq 1$). If $\bigcup_{k=1,...,n} A_k \not\supseteq B_j$, it follows that $\bigcap_{i \in H} A_i \not\supseteq B_j$ for all $H$, and therefore

$$\begin{aligned} F_j(\bigcup_{i=1,...,n} A_i) &= \alpha p(\bigcup_{i=1,...,n} A_i) = \sum_{\{H:H \subseteq \{1,...,n\}\}} (-1)^{|H|+1} \alpha p(\bigcap_{k \in H} A_k) = \\ &\sum_{\{H:H \subseteq \{1,...,n\}\}} (-1)^{|H|+1} F_j(\bigcap_{k \in H} A_k) \end{aligned}$$

hence $(*)$ holds. If, on the other hand, $\bigcup_{i=1,...,n} A_i \supseteq B_j$, then

.

$$\begin{aligned} F_j(\bigcup_{i=1,...,n} A_i) &= p(\bigcup_{i=1,...,n} A_i) = \sum_{\{H:H \subseteq \{1,...,n\}\}} (-1)^{|H|+1} p(\bigcap_{k \in H} A_k) \geq \\ &\sum_{\{H:H \subseteq \{1,...,n\}\}} (-1)^{|H|+1} F_j(\bigcap_{k \in H} A_k) \end{aligned}$$

hence $(*)$ holds as well.

Since $(*)$ holds for all cases it also holds for the average, that is,

$$v_D^F(\bigcup_{i=1,\dots,n} A_i) = \frac{\sum_{j=1}^T F_j(\bigcup_{i=1,\dots,n} A_i)}{T} \geq \frac{\sum_{j=1}^T \sum_{\{H:H\subseteq\{1,\dots,n\}\}}(-1)^{|H|+1}F_j(\bigcap_{k\in H} A_k)}{T}$$

and hence

$$v_D^F(\bigcup_{i=1,\dots,n} A_i) \geq \sum_{\{H:H\subseteq\{1,\dots,n\}\}} (-1)^{|H|+1} v_D^F(\bigcap_{k\in H} A_k)$$

Furthermore, $v_D^F(\Omega) = 1$ and $v_D^F(\phi) = 0$ therefore $v_D^F$ is a belief function which concludes the proof ■.

**Proof of Lemma**

Assume $v(A)+v(A') = v(A\cup A')$ for all $A'$ such that $A\cap A' = \phi$, then in particular, $v(A) + v(A^c) = v(\Omega)$. Therefore $v(A) = \bar{v}(A)$.

Let $v$ be convex and assume $v(A) = \bar{v}(A)$. Take $A'$ such that $A \cap A' = \phi$, then $v(A) + v(A') \leq v(A \cup A')$. Assume by negation that $v(A) + v(A') < v(A \cup A')$. Then by the definition of $\bar{v}$ we have $\bar{v}(A) + (v(\Omega) - \bar{v}(A'^c)) < v(\Omega) - \bar{v}((A \cup A')^c)$. Therefore, $\bar{v}(A) + \bar{v}((A \cup A')^c) < \bar{v}(A'^c)$ which is a contradiction since $\bar{v}$ must be concave. A similar proof could be applied for a concave capacity. ■

**Proof of Proposition 3**

Let memory be precise with respect to $A$. Then for every $j$ either $B_j \subseteq A$ or $B_j \subseteq A^c$. If $B_j \subseteq A$ (then $B_j \cap A \neq \phi$) then both $F_j(A) = 1$ and $G_j(A) = 1$. If $B_j \subseteq A^c$ (then $B_j \cap A = \phi$) both $F_j(A) = 0$ and $G_j(A) = 0$. It follows that $v_D^F(A) = v_D^G(A)$. Thus, $v_D^F(A) = \overline{v}_D^F(A)$ and $v_D^G(A) = \overline{v}_D^G(A)$.

Assume memory is imprecise with respect to event $A$. Then there exists a case $j$ such that $B_j \nsubseteq A$ and $B_j \nsubseteq A^c$. Therefore, $F_j(A) < G_j(A)$ (for $\alpha < 1$). By the definitions of $F$ and $G$, $F_j(A) \leq G_j(A)$ for every $j$ and thus, $v_D^F(A) \neq v_D^G(A)$.

We turn to prove that precise events form an algebra. Observe that $\phi$ and $\Omega$ are precise events. Furthermore, by the definition of a precise event, if $A$ is precise then so is $A^c$. Therefore all we have to show is that if $A$ and $A'$ are precise events then so is

$A \cup A'$. Assume $A$ and $A'$ are precise, then for all $j$, $B_j$ is contained in one the following events: $A \cap A'$, $(A \setminus A')$, $(A' \setminus A)$, $(A \cup A')^c$. But this means that $B_j$ is contained in $(A \cup A')$ or $(A \cup A')^c$. Therefore $A \cup A'$ is a precise event. ■

**Proof of Proposition 4**

Take dataset $D = (B_1, ..., B_T)$. Assume it is precise (i.e. $B_j = w_i$ for some $i$). In this case $F_j(A) = G_j(A)$ for every $A$ and $j$. Thus $v_D^F = v_D^G$. In order to show that $v_D$ is a probability it is sufficient to show that it is additive. Take any two disjoint events $A$ and $A'$. For every $j$, if $A \cup A' \supseteq B_j$ then either $A \supseteq B_j$ or $A' \supseteq B_j$ and not both. Furthermore if $A \cup A' \not\supseteq B_j$ then both $A \not\supseteq B_j$ and $A' \not\supseteq B_j$. Thus $F_j(A \cup A') = F_j(A) + F_j(A')$ for every $j$. Therefore $v_D(A \cup A') = v_D(A) + v_D(A')$.

Let $v_D$ be additive and assume by negation that there exists a case $j$ in memory which is imprecise. Take $A$ and $A'$ disjoint such that $A \cup A' = B_j$. Then $\alpha = F_j(A) + F_j(A') < F_j(A \cup A') = 1$ (similarly, $2 - \alpha = G_j(A) + G_j(A') > G_j(A \cup A') = 1$). For every $i$, $F_i(A) + F_i(A') \leq F_i(A \cup A')$ (similarly, $G_i(A) + G_i(A') \geq G_i(A \cup A')$) thus, $v_D(A \cup A') \neq v_D(A) + v_D(A')$ which is a contradiction. ■

# References


[1] Anscombe F. J. and R.J. Aumann (1963), "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199-205.

[2] Becker, G.M., Degroth, M.H., Marashak, J (1964), "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9, 226–232.

[3] Billot, A., I. Gilboa, D. Samet, and D. Schmeidler (2005), "Probabilities as Similarity-Weighted Frequencies," *Econometrica*, 73, 1125 - 1136.

[4] Carnap, Rudolf (1952), "The Continuum of Inductive Methods," University Press, Chicago.

[5] Carnap, Rudolf (1980), "A Basic System of Inductive Logic, Part II." In Richard C. Jeffrey (Ed.), Studies in Inductive Logic and Probability, Vol. II, 7-155.

[6] de Finetti, B. (1937), "La pr´evision: Ses Lois Logiques, ses Sources Subjectives," *Annales de l'Institut Henri Poincar´e*, 7:1–68.

[7] Demster A. P. (1967), "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Mathematical Statistics*, 38, 325-339.

[8] Dempster A. P. (1968), " A Generalization of Bayesian Inference," *Journal of the Royal Statistical Society* (B), 30, 205-247.

[9] Eichberger J. and A. Guerdjikova (2007) "Multiple Priors as Similarity-Weighted Frequencies" (2007)

[10] Ellsberg D. (1961), "Risk, Ambiguity, and Savage Axioms," *Quarterly Journal of Economics*, 75, 643-669.

[11] Epstein L. G (1999), "A Definition of Uncertainty Aversion," Review of Economic Studies, 66(3), 579-608.

[12] Epstein, L. G. and J. Zhang (2001), "Subjective Probabilities on Subjective Unambiguous Events," *Econometrica*, 69, 265-306.

[13] Gajdos T., T. Hayashi, J. M. Tallon, and J. C. Vergnaud (2006), "Attitude Toward Imprecise Information," mimeo.

[14] Ghirardato, P., and M. Marinacci (2002), "Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory*, 102, 251-289.

[15] Gilboa, I., O. Lieberman, and D. Schmeidler (2006), "Empirical Similarity," *Review of Economics and Statistics*, 88 (2006), 433-444.

[16] Gilboa, I., A. Postlewaite, and D. Schmeidler (2004), "Rationality of Belief," mimeo.

[17] Gilboa, I., and D. Schmeidler (2001), *A Theory of Case-Based Decisions*. Cambridge University Press.

[18] Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141-153.

[19] Gonzales, C and J.Y. Jaffray (1998), "Imprecise Sampling and Direct decision makeing," *Annals of Operations Rasearch*, 80, 207-235.

[20] Hsu M.,M. Bhatt ,R. Adolphs,D. Tranel, and C. F. Camerer (2005), "Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making," *Science*, 310, 1680-1683.

[21] Jaffray, J.Y. (1991), "Belief Functions, Convex Capacities and Decision Making," dans Doignon, Falmagne (eds), *Mathematical Psychology: Current Developments*, New York, Springer Verlag, 127-134.

[22] Klibanoff, P., M. Marinacci, and S. Mukerji (2005), "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73(6), 1849-1892.

[23] Maccheroni, F., M. Marinacci, and A. Rustichini (2006), "Ambiguity aversion, robustness, and the variational representation of preferences", *Econometrica,* 74, 1447-1498.

[24] Nau, R.F. (2006), "Uncertainty Aversion with Second-Order Utilities and Probabilities," *Management Science*, 52:1, 136-145.

[25] Nehring, K. (1999), "Capacities and Probabilistic Beliefs: A Precarious Coexistence," *Mathematical Social Sciences*, 38, 197-213.

[26] Rustichini A., J. Dickhaut, P. Ghirardato,K. Smith and J. V. Pardo (2005), "A brain imaging study of the choice procedure," *Games and Economic Behavior*, 52, 257-282.

[27] Savage, L. J. (1954): The Foundations of Statistics. New York: John Wilely and Sons.

[28] Seo K., (2006), "Ambiguity and Second-Order Belief," mimeo.

[29] Shafer, G. (1976), A Mathematical Theory of Evidence, Princeton University Press, Princeton.

[30] Schmeidler, D. (1989), "Subjective Probability without Additivity," *Econometrica* 57, 571-587.

[31] Tversky, A. and Kahneman, D. (1971), "The belief in the law of small numbers," *Psychological Bulletin* 76, 105-110.

[32] Viscusi, W. K (1989), "Prospective Reference Theory: Toward an Explanantion of the Paradoxes," *Journal of Risk and Uncertainty* 2, 235-264

[33] Wakker, P. (2008), "Uncertainty," In Lawrence Blume & Steven N. Durlauf (Eds.), The New Palgrave: A Dictionary of Economics, 6780-6791. The MacMillan Press, London.

[34] Yaari M. E. (1969), "Some remarks on measures of risk aversion and on their uses," *Journal of Economic Theory* 1, 315–329.

[35] Zhang, J. (2002), "Subjective Ambiguity, Probability and Capacity," *Econometric Theory*, 20, 159-181.