MANAGEMENT SCIENCE
Vol. 58, No. 1, January 2012, pp. 188–202
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

informs®
http://dx.doi.org/10.1287/mnsc.1110.1463

Imprecise Data Sets as a Source of Ambiguity: A Model and Experimental Evidence

Ayala Arad
The Eitan Berglas School of Economics, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel,
aradayal@post.tau.ac.il

Gabrielle Gayer
Department of Economics, Bar-Ilan University, Ramat Gan 52900, Israel, gabi.gayer@gmail.com

In many circumstances, evaluations are based on empirical data. However, some observations may be imprecise, meaning that it is not entirely clear what occurred in them. We address the question of how beliefs are formed in these situations. The individual in our model is essentially a "frequentist." He first makes a subjective judgment about the occurrence of the event for each imprecise observation. This may be any number between zero and one. He then evaluates the event by its "subjective" frequency of occurrence. Our model connects the method of processing imprecise observations with the individual's attitude toward ambiguity. An individual who in imprecise observations puts low (high) weight on the possibility that an event occurred is ambiguity averse (loving). An experiment supports the main assertions of the model: with precise data, subjects behave as if there were no ambiguity, whereas with imprecise data subjects turn out to be ambiguity averse.

Key words: decision analysis; theory; inference; uncertainty
History: Received July 7, 2010; accepted August 31, 2011, by Brad Barber, Teck Ho, and Terrance Odean, special issue editors.

1. Introduction

1.1. Motivation

Let us suppose you have to undergo some medical treatment. In the first scenario you have the precise information that the treatment was successful in 25 out of 50 past cases and was unsuccessful in the remaining cases. In the second scenario you have the same information as before but in addition you are notified that the outcomes of 20 other cases were lost because of some technical problem, making the overall success rate of the treatment unknown. How would you feel about undergoing the treatment in the second scenario compared to the first? Would the presence of imprecise information lower your confidence in the treatment?

In real-life situations, for example, in the fields of finance, insurance, and medicine, the available information is often given in the form of statistical data with different degrees of precision. In this work we explore how the precision of the statistical data affects beliefs. Each case in the data set is an observation of a similar instance in the past, in which an event of interest did or did not occur. However, as in the example above, some cases may be imprecise so that it is not entirely clear what occurred in them.

The individual who is forming beliefs in the present model is essentially a "frequentist." When there is precise knowledge of whether or not the event occurred in each past case, the individual evaluates the likelihood of the event by its relative frequency, that is, by the number of cases in which the event occurred out of the total number of cases. However, when faced with an imprecise case, the individual has to complete the missing information by assuming that the event did occur, did not occur, or anything in between. After making an independent judgment regarding the occurrence of the event in each imprecise case, the individual may then evaluate the event by his subjective view of its frequency of occurrence.

Two opposite methods for processing imprecise cases are studied. The first method is that of an individual who tends to treat imprecise cases as if the evaluated event did not occur. Thus, in the above example of the lost medical records, the likelihoods of both success and failure are evaluated at less than 1/2 (35/70); hence, the sum of the likelihoods of the two possibilities is less than one. We loosely use the term *low-valued beliefs* for evaluations that sum to less than one. By contrast, the second method for dealing with imprecise data is that of an individual who tends to treat imprecise cases as if the evaluated event did occur. The resulting evaluations generally sum to more than one. We refer to these as *high-valued beliefs*.

In the extreme version of the first method, the evaluator assumes that the event did not occur in all imprecise cases, leading to a belief valuation that is placed at the lower bound of the set of possible empirical frequencies. In the example above, this corresponds to evaluating both success and failure as 25/70. In the extreme version of the second method, the individual completes every imprecise case as though the event did occur. This evaluation places the belief at the upper bound of its possible empirical frequencies. In the example above, the beliefs of success and failure would then be 45/70. In our model, belief valuations may equal this lower bound or upper bound, but in general they will fall between these two extremes.

In several works that have discussed belief formation based on partial information, special attention was paid to these extreme lower and upper probabilities. In "theory of evidence," Dempster (1967, 1968) and Shafer (1976) introduced the concepts of belief and plausibility functions that, respectively, correspond to the minimum and maximum likelihoods of an event. Jaffray (1991) and Gonzales and Jaffray (1998) studied special instances of Dempster's (1967, 1968) and Shafer's (1976) theory when applied to imprecise statistical data. Walley (1991) defined lower (upper) probabilities, or, more generally, lower (upper) provisions, as the highest (lowest) price at which the decision maker is certain to buy (sell) a gamble. Finally, Mukerji (1996) described a belief formation process in the context of unawareness that again can be reduced to a belief determined by the lower bound presented above.

However, beliefs corresponding to the upper or lower probabilities are generally too restrictive to be able to explain observed behavior in experiments in which subjects are provided with information in the form of a set of probabilities (see, for example, Borghans et al. 2009, Cohen et al. 1987, Halevy 2007). The importance of allowing for a range of possible beliefs when describing actual behavior is not limited to the type of information in these experimental works; rather, it extends to our context as well in which information is given in the form of statistical data.

There is a certain theoretical symmetry between the upper and lower bounds of the empirical frequencies. Indeed, Dempster's (1967, 1968) and Shafer's (1976) belief functions have plausibility functions as their duals, and Walley's (1991) lower probabilities have the upper ones as counterparts. However, when the event in question is associated with a positive outcome, observed behavior tends to break this symmetry in the "pessimistic" direction: people tend to make decisions as if the event had a probability at the lower half of the range (as in our first method). This idea is consistent with Ellsberg (1961) and subsequent experiments[1] that showed that people are ambiguity averse, that is, they prefer betting on an outcome whose probably is known rather than on one whose probability is unknown.

In recent decades, the literature has offered several theoretical models that can explain Ellsberg-type behavior. The first such model is Schmeidler's (1989) Choquet expected utility (CEU), in which the decision maker's beliefs are nonadditive probabilities (called capacities). Roughly speaking, in the CEU model, low-valued beliefs (as in our first method) correspond to ambiguity-averse behavior. By contrast, high-valued beliefs (as in our second method) correspond to ambiguity-loving behavior. Finally, additive beliefs are identified with ambiguity neutrality.[2]

Schmeidler's (1989) seminal paper and much of the literature that followed is axiom based. These studies make assumptions about preferences that allow us to describe the decision maker as if he holds certain beliefs. This literature, like the axiomatic foundations of the classical Bayesian approach (Ramsey 1931, de Finetti 1937, Savage 1954), says little about the origin of these beliefs or about how they depend on objective data. The present work, by contrast, belongs to a tradition that attempts to model how objective data are transformed into subjective beliefs. Like Jaffray (1991), we are interested in a model where data and beliefs are both explicitly represented, leading to a better understanding of the relationship between the two. We view this approach as complementary to the axiomatic approach of Schmeidler (1989), Gilboa and Schmeidler (1989), and others. The axiomatic approach provides an interpretation of beliefs through observed behavior, whereas the data-based approach enables the prediction of when a situation is perceived as ambiguous and of which beliefs individuals are most likely to hold in those situations. In our model, imprecise information is the source of ambiguity, and reasonable beliefs fall between the extremes of the two belief formation processes described above.

Our model allows for different attitudes toward ambiguity. Which attitude is most common is an empirical question, which the present paper aims to study. We designed an experiment to examine whether imprecise data are a cause for ambiguity. The experiment involves betting about the type of ball drawn from an urn. Participants do not know the composition of the balls in the urn but are presented with information about outcomes of past draws. In Experiment 1, the participants are provided with precise data

[1] See Camerer and Weber (1992) for a review.

[2] To be precise, Schmeidler (1989) associates ambiguity aversion (loving) with convex (concave) capacities. In particular, the sum of probabilities over all the states of nature will be smaller (larger) than 1 with convex (concave) capacities. A formal definition of convex and concave capacities is presented in §2.2.

about past draws. We show that their behavior does not differ significantly from that of participants who know the proportions of the balls in their urn. These results indicate that individuals who are presented with precise information act as if there is no ambiguity. In Experiment 2, one group of participants is provided with precise data about the event in question, whereas the other is provided with imprecise data. The latter group's behavior is consistent with their forming low-valued beliefs reflecting ambiguity aversion.

1.2. Medical Example

The following example illustrates the main features of our model.

A patient suffering from a runny nose, a headache, and fatigue comes to Dr. Blue for a diagnosis. All symptoms are consistent with the following illnesses: allergy, flu, and pneumonia. For the purpose of the example it is assumed that these illnesses are both mutually exclusive and jointly exhaustive.

Dr. Blue has seen four patients in the past who suffered from these same symptoms. His experience is summarized in the following table.

	Allergy	Flu	Pneumonia
Case 1	X	X	
Case 2	X		
Case 3	X	X	
Case 4			X

Each row in this table represents a different patient (or a separate past case), and each column represents one of the possible illnesses. A cell with an X indicates that this illness may have occurred in this case, whereas an empty cell indicates that it definitely did not occur. When a single X appears in a case (such as in cases 2 and 4), the illness was perfectly identified; however, when more than one X appears in a case (such as in cases 1 and 3), the diagnosis was inconclusive (or imprecise).

Dr. Blue would like to apply the "frequentist" approach to his diagnosis. When evaluating the event "the patient has an allergy or the flu" he recalls that three out of four patients suffered from either one of these illnesses, and therefore he assigns it the probability 3/4.

When he evaluates the likelihood of the event "the patient has an allergy," he can no longer apply the frequentist approach, because for both patients 1 and 3 it is unclear whether they had an allergy or the flu. He may overcome this problem by assuming that in one case it was an allergy and in the other it was the flu, leading to an evaluation of the likelihood of an allergy as 2/4 and the likelihood of the flu as 1/4.

But Dr. Blue, who is cautious by nature, would like to lower this evaluation, especially because it relies on considerable speculation. He realizes that there is actually only one patient that suffered from an allergy with 100% certainty (case 2) and no patient who suffered from the flu to that degree of certainty. Assuming the event occurred only in cases in which there is absolute certainty reduces the probability of allergy to 1/4 and of flu to 0.

However, this approach may be too cautious because in these imprecise cases, one of the illnesses must have occurred. Thus, he will arrive at a belief about the likelihood of allergy and flu that falls between the two different evaluations; that is, the probability of allergy will be between 0.25 and 0.5, and the probability of flu will be between 0 and 0.25.

Obviously, Dr. Blue's evaluations are nonadditive. His evaluation of the event "the patient has an allergy or the flu" is 0.75. However, when it is partitioned into the events "flu" and "allergy," Dr. Blue's evaluations are relatively low, summing to less than 0.75, reflecting ambiguity aversion.[3]

Now suppose that there are two types of drugs, one cheap drug that treats only allergies and another more expensive drug the treats both allergies and the flu. With the additive probability above (of 0.5 for allergy and 0.25 for flu), Dr. Blue may prescribe the first medication for allergies only, based on the belief that the advantage of the low cost of the first drug outweighs the advantage of the high-priced drug that treats both illnesses. However, with his resulting nonadditive belief, Dr. Blue is more inclined to prescribe the second medication because the advantage of the first drug is mitigated.

In the present model, a precise event is an event that is either known to have occurred or known to not have occurred in each past case (such as "an allergy or the flu"). An imprecise event is an event about which data are vague; namely, in some past cases it is unknown whether it occurred or not (such as the events "allergy" and "flu"). Here ambiguity arises only when evaluating the likelihoods of imprecise events.

The rest of this paper is organized as follows. Section 2 introduces a formal explication of the two evaluation methods. It studies the implications of these methods and how they relate to the literature. Section 3 presents the design and the main findings of the experiment. Finally, this paper concludes with a discussion of extensions of the model.

2. The Model and Results

2.1. Belief Formation

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite set of states of nature ($n \geq 2$), and let Σ be an algebra of subsets of Ω called

[3] It is equally possible to demonstrate beliefs that exhibit ambiguity loving by modifying the belief formation process in the appropriate way.

events that is given by the power set 2^Ω. A data set D of length T is a sequence of events indexed by $i=1,\ldots,T$:

$$D=(B_1,\ldots,B_T),$$

where $B_i\in\Sigma\backslash\{\varnothing,\Omega\}$. A data set is also referred to as memory.

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

A case i is the ith element of a data set D. The event B_i is interpreted as the information available to the individual regarding the occurrences in case i. For any two cases i and j for which $B_i\subseteq B_j$, case i is more informative than j. Two types of cases are excluded, the first are cases in which $B=\varnothing$. 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 excluded cases are those in which $B=\Omega$. These cases add no information to the evaluation of the outcomes because they do not narrow the set of states of nature that have or have not occurred.

A capacity v is a mapping from 2^Ω into $[0,1]$ such that $v(\varnothing)=0$, $v(\Omega)=1$ 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, $V\colon\mathcal{D}\to\mathcal{V}$.[4] For any data set D, v_D is the capacity that the decision maker attaches to the data set. Given $D=(B_1,\ldots,B_T)$, we define $\forall\, j\leq T$

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

and

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

where $0\leq\alpha\leq1$ (the α in (1) and (2) are not necessarily the same). The event A^c denotes A's complement, and $|\ |$ denotes the cardinality of the set. 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 data set 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, whereas when $A\cap B=\varnothing$ it is obvious that A did not occur, and no weight is put on this possibility. When the conditions above are not satisfied (that is, when $A\not\supseteq B$ and $A\cap B\neq\varnothing$), it is unclear whether A occurred, and the equations generally provide different solutions. We refer to these as judgments.

When $\alpha=1$, the equations are identical, and the weight put on the possibility that A occurred equals the proportion of states in B that imply 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.

Put differently, in (1) the individual starts out with an initial judgment about the possibility that an event occurred in case j, which is an additive belief p^0.[5] However, with no firm grounds to justify this judgment he lowers it to some extent out of cautiousness (the opposite being the analog of (2)). Under this interpretation, α reflects the degree of confidence in the initial judgment.[6]

Given a data set, the belief that a particular event A will occur is formed by averaging the judgments regarding the occurrences of the event A in each of the past cases. In particular, for procedure (1) with the extreme value of $\alpha=0$, the individual presumes the event did not occur unless he is informed otherwise. Thus, Equation (3) with $\alpha=0$ corresponds to the lowest possible belief valuation given the data set. The reverse is true for procedure (2) with $\alpha=0$, and therefore Equation (4) corresponds to the highest possible belief valuation given the data set. In general, belief valuations increase with α according to Equation (3) and decrease with α according to Equation (4). Any belief valuation in the range between these two extreme valuations is considered plausible given the data set. Roughly speaking, the former approach leads

[4] This is a slight abuse of notation, because $V\colon\mathcal{D}\to\mathcal{V}$ refers to two separate mappings defined below.

[5] Here the additive probability is $p^0(A)=|A\cap B_j|/|B_j|$. In §4 we discuss how (1) and (2) can be modified to account for other additive probabilities.

[6] The way the individual in our model tries to make sense of what occurred in a single past case is closely related to the way the set of beliefs are determined in the statistical ϵ-contamination model (e.g., Huber 1973, Berger 1985, Berger and Berliner 1986, Wasserman and Kadane 1990). In the ϵ-contamination model, the lower (upper) bound of the set of beliefs P over Ω corresponds to (1) ((2)).

to low-valued beliefs, whereas the later approach leads to high-valued beliefs.

In the next section we study the properties of belief valuations based on (3) and (4) and connect them to the literature. Of the two procedures, we find (3) much more natural. For example, if one considers the possibility that there is some underlying rule that one does not entirely understand generating the imprecision of the data, a sensible precaution would be to lower beliefs. Procedure (3) is also supported by Gilboa (1988) and Jaffray (1988), who axiomatize a non-expected utility preference over lotteries in the context of risk. Their representations assign excessive weight to the worst outcome and by adding some restrictions can be reduced to CEU with a nonadditive belief that is governed by (3). Finally, in the experimental section, it is verified that participants' behavior is consistent with (3) but not with (4).

2.2. Attitudes Toward Ambiguity

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

First, a few known properties of capacities that prove useful are as follows: 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*[7] if it satisfies the condition that, for any collection $A_1,\ldots,A_n$ of subsets of Ω, $v(\bigcup_{i=1,\ldots,n}A_i)\geq\sum_{\{H:H\subseteq\{1,\ldots,n\}\}}(-1)^{|H|+1}v(\bigcap_{i\in H}A_i)$. Note that $n=2$ is the convexity condition of v, and thus every belief function is convex.

Let $\bar{v}$ be defined by $\bar{v}(A)=1-v(A^c)$; then, the following properties hold (see Marinacci and Montrucchio 2004):

- 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 $v_D^G=\bar{v}_D^F$ if α in F is equal to that in G.[8]

The concept of ambiguity aversion is that decision makers prefer to be exposed to randomness of known probabilities as opposed to randomness of unknown probabilities. To formalize this idea Schmeidler (1989) suggested a behavioral axiom by which a decision maker who is indifferent between two uncertain alternatives will (weakly) prefer a mixture of the two. The rationale behind this axiom is that when probabilities are unknown, one alterative can be used as a hedge against the other. Their mixture thereby reduces the uncertainty. Similarly, the reverse preference toward mixing reflects ambiguity loving, whereas indifference expresses ambiguity neutrality. Often ambiguity aversion is associated with pessimism and ambiguity loving is associated with optimism.

Schmeidler (1989) shows that in the context of CEU this notion of ambiguity aversion (loving) translates into convexity (concavity) of the capacity. A capacity that is a probability, naturally, reflects ambiguity neutrality. In the literature convexity is not uniformly considered to be a necessary condition for ambiguity aversion (loving), yet it is generally accepted as a sufficient one. See, for example, Ghirardato and Marinacci (2002) and Epstein and Zhang (2001). An exception to this approach can be found in Wakker (2008), which discusses the importance of relative convexity.

In our model we use the CEU decision rule to study the implied behavior of individuals. The following proposition establishes how attitudes toward ambiguity correspond to beliefs based on our two methods:

Proposition 1. *The belief formation process v_D^F as defined in Equation* (3) *is a belief function.*

All proofs can be found the appendix. Proposition 1 establishes that v_D^F is a belief function and thus is a convex capacity. Moreover, because $v_D^G=\bar{v}_D^F$, 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 ambiguity aversion, whereas evaluating the likelihood of events according to Equation (4) leads to ambiguity loving. When $\alpha=1$ we have neutrality toward ambiguity. Note that when there is no vagueness regarding the occurrences of events in the data (that is, data are precise), any α leads to an additive probability and therefore the individual's attitude toward ambiguity in these circumstances cannot be identified.

Ghirardato and Marinacci (2002) define the notion of comparative ambiguity aversion by which Decision Maker 1 is more ambiguity averse than Decision Maker 2 as follows: if for every two alternatives, one ambiguous and the other not, Decision Maker 2 prefers the unambiguous alternative over the ambiguous one, then so does Decision Maker 1. In the context of CEU, the characterization of this definition is that for every event, Decision Maker 2's capacity of this event is larger than or equal to that of Decision Maker 1.[9] In the present framework it is easily seen that for a given data set, according to procedure (3) ((4)), a smaller (larger) α corresponds to a more ambiguity-averse individual.

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

[8] Henceforth, when we write $v_D^G=\bar{v}_D^F$ it is with the understanding the α in F equals that in G.

[9] The notion of "more ambiguity averse" also requires that the two decision makers' utilities over prizes are essentially the same. However, Ghirardato and Marinacci (2002) later argue that even when their utilities are not identical, the same idea of capacity domination can be used to compare attitudes toward ambiguity.

2.3. Ambiguous Events

It is undisputed that the presence of objective ambiguity in a problem is associated with lack of information. In our context, we identify objective ambiguity with imprecise data sets in which the missing information is in regard to outcomes of past cases. Precise and imprecise data sets and precise and imprecise events are formally defined as follows:

Given $D = (B_1, \dots, B_T)$, 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 this condition is satisfied, and we refer to A as *precise* when it is precise with respect to every case in the data set. This definition extends naturally both to single cases and to data sets. Case j is regarded as *precise* when it is precise with respect to every event in 2^{Ω}, i.e., when $B_j = \{\omega_i\}$ for some $\omega_i \in \Omega$. A data set or memory is *precise* when all cases in it are precise and is *imprecise* otherwise. We use the terms *imprecise events* and *objectively ambiguous events* interchangeably.

As it happens, there is much in common between our notion of objective ambiguity and notions of *revealed ambiguity* in the literature of decision theory. In this literature, the set of ambiguous events are usually derived endogenously from observed behavior. Generally speaking, an event is deemed to be ambiguous if the decision maker's preferences imply it (without going into detail about the exact definitions).

On its face, revealed ambiguity, which is stated in terms of preferences, is incomparable with our idea of objective ambiguity, which is defined in terms of available information. However, because revealed ambiguity can be translated into conditions over capacities, a comparison becomes possible by examining when the beliefs in our model meet these conditions. We next turn to review some of the notions of revealed ambiguity[10] in the literature and study how they relate to our notion of objective ambiguity.

Nehring (1999) showed that for maxmin expected utility preferences, unambiguous events can be identified with events on which all probabilities agree. Furthermore, because 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 manner. 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 $\{A \mid v(A) + v(A') = v(A \cup A') \forall A' \subseteq A^c\}$, which is equivalent to $\{A \mid v(A) + v(A^c) = 1\}$. Ghirardato and Marinacci's (2002) definition of unambiguous events agrees both with Epstein and Zhang's (2001) definition when expressed in terms of capacities; and with Nehring's (1999) definition when expressed in terms of the set of probabilities. Likewise, the Klibanoff et al. (2005) definition of unambiguous events is the same as Nehring's (1999) when expressed in terms of sets of probabilities.

The next lemma proves that when beliefs are represented by convex and concave capacities, Nehring's (1999) definition of unambiguous events is equivalent to that of Epstein and Zhang (2001).

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 $v(A) = \bar{v}(A)$.*

Because in the previous section it was established that in our model beliefs are either convex or concave capacities, the condition of the lemma is satisfied. This enables us to refer to an event as *revealed as unambiguous* if it satisfies either the definition of Nehring (1999) or the definition of Epstein and Zhang (2001).

In the following, the relationship between imprecise information (or objective ambiguity) and revealed ambiguity is established.

Proposition 2. *For $\alpha < 1$, event A is precise given D if and only if it is revealed as unambiguous. Furthermore, the set of precise events forms an algebra.*

Proposition 2 states that ambiguity, in our model, is due to partial information in the data, and that had information been precise no ambiguity would have arisen.

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

Proposition 3. *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 data set is precise. In this case, $v_D^F = v_D^G$.*

Notice that in both Propositions 2 and 3, α is required to be smaller than 1. When $\alpha = 1$, the proof fails because when there is no revealed ambiguity, we cannot conclude whether this results from precise data or a neutral attitude toward ambiguity on the part of the individual. It is a well-known fundamental problem in the revealed preferences approach that when the individual has a neutral attitude toward ambiguity, nothing can be inferred about the presence of ambiguity (see the discussion in Ghirardato 2004). Propositions 2 and 3 show that when revealed ambiguity is meaningful, it coincides with our notion of objective ambiguity.

[10] These definitions are very general in that they apply to many decision models, the CEU model being only one of them. Indeed, one of the objectives of Epstein and Zhang (2001) was that their definition of ambiguity would be model free.

2.4. Bibliographic Note

In two related works, Jaffray (1991) and Gonzales and Jaffray (1998), construct a representation of preferences based on beliefs formed using imprecise cases. This construction makes use of beliefs that fit our formalization with $\alpha = 0$. Jaffray and Philippe (1997) show that this representation can be reduced to that of a CEU maximizer with a capacity $\tilde{v} = \beta v_D^F + (1-\beta) v_D^G$, where v_D^F and v_D^G are as described in Equations (3) and (4), respectively, with the parameter $\alpha = 0$. Therefore, our belief formation process and that of Jaffray and Philippe (1997) intersect only when our α equals 0 and their β equals either 0 or 1. In particular, when β is strictly between 0 and 1, $\tilde{v}$ will be neither convex nor concave, which, according to our definitions in §2.2, reflects neither ambiguity aversion nor ambiguity love.

Some other works that deal with belief formation processes start out with an initial prior that is adjusted in some manner. Hogarth and Einhorn's (1990) initial prior may depend on experience or statistical data. This initial prior is modified by a process of mental simulation in which the decision maker envisions other decision weights that are higher or lower than those of the anchor. Carnap (1952, 1980) and Viscusi (1989) propose an updating process by which the posterior (additive) probabilities are a weighted average of a prior probability and observed relative frequencies, where the weight of the relative frequencies depends positively on the sample size. Billot et al. (2005) present a procedure, similar to ours, describing how individuals form additive beliefs given data on past cases. The main distinction between this model and ours is the type of data the individual can possess. Billot et al. (2005) consider only cases with precise information, whereas we allow for cases with imprecise information as well. This enables us to articulate the idea of ambiguity resulting from imprecise information. Eichberger and Guerdjikova (2010) extend the work of Billot et al. (2005) to formalize ambiguity caused by small samples.

The literature of ambiguity has also considered preferences over acts when the environment varies. In Gajdos et al. (2008), each environment is associated with a different set of probabilities, and preferences are explicitly defined over those environments. Roughly speaking, in this work an ambiguity-averse decision maker would prefer an environment that is associated with a smaller set of probabilities. This result is supported by the Ellsberg-type experiments of Becker and Brownson (1964) and Borghans et al. (2009), who find that an urn with a smaller range of the number of balls of a particular color is viewed more favorably than an urn with a larger one. Applied to our context, this idea would mean that individuals prefer a precise data set over an imprecise data set. Based on the notion of "small worlds" from Chew and Sagi (2008), Abdellaoui et al. (2011) discuss the preference of more familiar sources of ambiguity to less familiar ones. Assuming familiarity is positively correlated with precision of information again would imply the preference of a precise data set over an imprecise one.

3. Experimental Test of the Model

This section describes two experiments that were conducted to examine whether the actual behavior of decision makers is consistent with the model's main implications. The first experiment involves belief formation given a precise data set, testing whether the perception of the likelihood of an event matched the actual frequency of occurrence of this event in the data. The second and main experiment is concerned with belief formation given an imprecise data set, testing whether the subjects in the experiment are ambiguity averse in the presence of imprecise data.

Eighty economics students in undergraduate and graduate studies at Tel Aviv University participated in Experiment 1, and 292 undergraduate students from both the Economics Department at Haifa University and the Engineering Department at Ben-Gurion University participated in Experiment 2. The experiments that lasted about 15 minutes took place at the beginning of class.

The same instructions for both experiments were presented before the students received their forms. The subjects were informed that some of them (the proportion was approximately 1 subject out of 25) would be randomly chosen to participate in a lottery at the end of the experiment and that the lottery was not necessarily the same for all participants. All lotteries involved a draw of a ball from an urn, however the type of ball might differ. For example, some participants might encounter a particular bet, whereas others might encounter the opposite bet. The subjects were asked to state whether they preferred to participate in the lottery specified on their forms or be given a certain amount of money instead. They were required to state their preference for every amount of money that appeared in their forms, which varied between NIS 10 and NIS 140.

The subjects were told that those who actually participated in the lottery would be given a monetary prize according to the choices they made. More specifically, a certain randomly selected sum of money would be given to those subjects who stated that they preferred this certain sum over the lottery, whereas if a subject stated that he preferred the lottery he would be given the amount determined by its outcome (in line with the procedure introduced in; Becker et al. 1964 (BDM)). It was further explained that although the decisions for subjects not selected to participate in the lottery were hypothetical, they still had good reason to state their true preferences because all subjects had an equal chance of being chosen. Finally, it was emphasized that there was no correct answer and that their answers depended solely on their personal preferences. After receiving a verbal explanation of the experiment, the subjects were asked to carefully read the instructions on their forms.

3.1. Experiment 1

In this experiment the behavior of subjects who were provided with the exact proportions of different balls in an urn is compared with that of subjects who were not informed about the proportions, but instead were given a precise data set of past cases (draws) in which the frequency of occurrence of events fit the above proportions.

Two different kinds of forms were randomly distributed among all subjects within each class of students. The forms corresponded to two treatments: the experimental group (treatment *a*) and the control group (treatment *b*). The general structure of the forms used in both Experiments 1 and 2 can be found in Online Appendix A (online appendices are available at http://www.tau.ac.il/~aradayal/Imprecise_Appendix.pdf). Subjects in the experimental group were told that their urn contained 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 data set appears in Table 1.[11] Rather than being provided with a data set of past draws, subjects in the control group, who had a different urn, were told that their urn contained exactly three yellow balls and five white balls.[12] The ratio of yellow balls in treatment *b's* urn was equal to the proportion of observations in which a yellow ball was drawn in treatment *a's* data set.

Subjects in both treatments were offered the opportunity to participate in the following lottery: "if at the end of the experiment, a yellow ball is drawn from the urn, you will receive NIS 150 (approximately USD 40). Otherwise, you will get nothing." Finally, subjects were asked to state whether they preferred to participate in the lottery over receiving an assured amount of money M, for each $M \in \{10, 20, \ldots, 140\}$.

Table 1 **Data Set 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

The certainty equivalent (CE) is taken to be the lowest amount of cash M that is preferred over participation in the lottery. The results would not be different if the CE were defined as the highest cash amount that is not preferred over participation in the lottery (which is lower than the previous CE by NIS 10 when the participants' answers satisfy monotonicity) or the average of these two alternatives.[13] The average CE in treatment i is denoted by CE^i for $i = a, b$. In both treatments the participants were offered a chance to bet on the ball drawn from their urn being yellow. Therefore, a higher CE reflects a stronger belief that a yellow ball would be drawn from their urn.[14] To investigate the statistical differences between the CEs in treatments a and b, a Mann–Whitney U test (also known as the robust rank-order test) was performed.

3.1.1. Results. Thirty-eight subjects participated in treatment a and 42 subjects in treatment b. A detailed distribution of choices in all treatments in the experiments appears in Online Appendix B. See Figure 1 for the empirical distribution functions of the CEs in treatments a and b. The average CEs for the experimental group and the control group were $CE^a = 67.37$ and $CE^b = 69.52$, respectively. A Mann–Whitney U test indicates that the distributions of the CEs in the two treatments were not significantly different ($u = 0.8248$ and $p = 0.41$ in a two-tailed test). Therefore, the hypothesis that the CEs in treatments a and b came from the same distribution cannot be rejected.

These findings support the statement that an individual forms a belief about an event that matches the proportion of cases in memory in which this

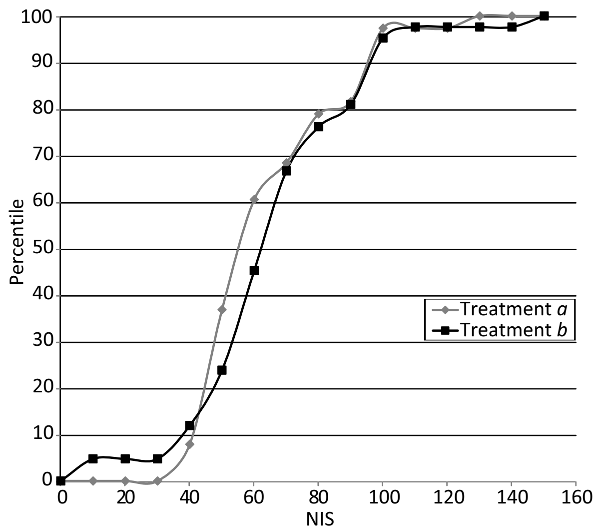
[11] Note that the balls were not drawn in front of the subjects, but rather were drawn prior to the experiment. All treatments in both experiments (apart from treatment b in which there is no sample) included two versions in which the order of the cases in the data set were shuffled.

[12] The subjects were not informed that there were two different treatments and correspondingly two separate urns.

[13] Originally there were 377 participants in the two experiments. The answers of five participants who violated monotonicity were omitted.

[14] Note that in both Experiments 1 and 2, comparisons are made only between two bets for which one stochastically dominates the other (because the only aspect that can differ between the compared treatments is the probability of the desired event). In such circumstances, criticisms of presumable preference reversals brought on by the BDM procedure (see Karni and Safra 1987, Segal 1988) do not apply. Hence, the ranking of the two compared bets is sustained through our BDM mechanism.

Figure 1 **Empirical Distribution Functions of the CEs in Treatments *a* and *b***



event occurred.[15] 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 data set. 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).

The results of Experiment 1 indicate that there is no great difference between the beliefs of drawing a yellow ball in the bets of treatments *a* and *b*, and thus a decision maker should be almost indifferent between the two of them. Nevertheless, it is perfectly possible that if instead of the task in Experiment 1 the participants were asked to choose directly between the two bets, one bet would be chosen much more frequently than the other. In other words, in this tie-breaking situation, one of the bets would prevail, despite the small differences in beliefs. We assert that in this case participants feel that the known probabilities in treatment *b* are more valid than the sample's frequencies in treatment *a*, and therefore they have more confidence in the former. Evidence of such a tendency is found in Chipman (1960) and Gigliotti and Sopher (1996), where, for example, participants preferred a bet with a known probability of 0.5 to a bet with an unknown probability tied to a sample in which the desired event occurred half the time.

3.2. Experiment 2

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 $\alpha < 1$),[16] it is shown that their beliefs (CEs) are lower than those induced by the neutral approach (i.e., $\alpha = 1$).

Table 2 **Precise Data Set**

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

A direct test would compare the subjects' CEs to those of ambiguity-neutral individuals given the same imprecise data. Such a test cannot be performed because the CE values of ambiguity-neutral individuals are unknown. Therefore, the key of the experiment is to obtain the neutral CEs indirectly by a parallel treatment. This is explained in detail in the following subsection.

Eight kinds of forms, which represent eight different treatments, were randomly distributed to all subjects within each class of students. All the forms had the same structure as that of the experimental group in Experiment 1; namely, all forms had a data set of eight cases of past draws and a proposed lottery that 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 only in their data sets and their proposed lottery. The data set that appeared in the treatments was either precise (Table 2) or imprecise with respect to the event in question (Table 3). All lotteries had the same structure as that of Experiment 1: win NIS 150 if the type of ball drawn 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). As in Experiment 1, this is a between-subjects design.[17]

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 data set is precise (P) or imprecise (IP),

[15] The experimental analysis in this between-subjects design assumes no systematic difference in risk attitudes between treatments due to the random assignment of subjects.

[16] 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, hereafter the condition $\alpha < 1$ is omitted.

[17] Ideally, one would want to compare the beliefs of the same subject given two alternative data sets. However, this raises the concern that the former data set may influence the evaluation of the event based on the later data set.

Table 3 **Imprecise Data Set**

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

and the lower index Z indicates the type of ball that yields a prize when drawn from the urn. The average CE of each treatment is denoted by CE_Z^i in the same manner.

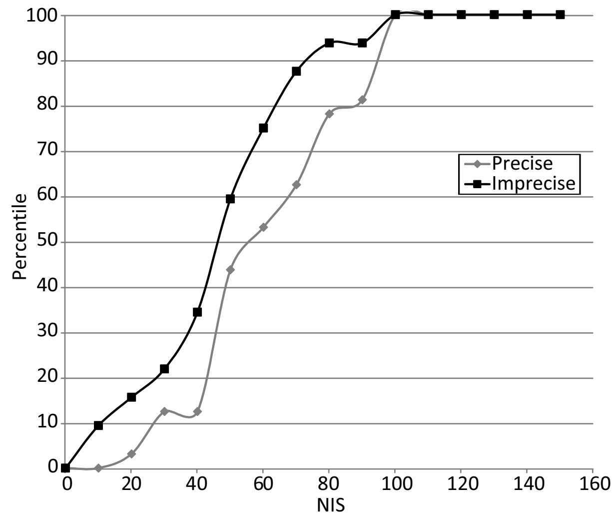
The discussion is divided into two parts. The first part (treatments T_W^P, T_W^{IP}, T_Y^P, and T_Y^{IP}) tests the main hypothesis of the experiment, and the second part (treatments T_{YO}^P, T_{YO}^{IP}, T_{YX}^P, and T_{YX}^{IP}) is a robustness test that verifies that the same results hold for other events.

3.2.1. Part 1. The data set that appeared in treatment T_W^P consisted of eight past cases, which were precise with respect to event W (see Table 2). The data set of treatment T_W^{IP} was obtained by transforming two precise cases in the data set of treatment T_W^P—one in which W did occur (case 1) and one in which W did not (case 8)—into two imprecise cases with respect to event W (by omitting the information regarding the color of the ball). The remaining six past cases were practically unaffected (see the data set of T_W^{IP} in Table 3).[18] Subjects were offered the lottery "win NIS 150 if a white ball is drawn and 0 otherwise" in both treatments T_W^P and T_W^{IP}.

According to the model, individuals who observe the precise data set given in treatment T_W^P hold the same belief about event W, regardless of their α. 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 individuals in treatment T_W^P, because given this imprecise data set, $F_j(W) = \alpha(|W \cap B_j|/|B_j|) = 1/2$ for $j = 1, 8$, and $F_j(W)$ 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-valued belief regarding W. Therefore, the experimental findings support the hypothesis of the model that subjects follow a pessimistic approach (or $\alpha < 1$) if subjects in treatment T_W^{IP} hold a lower-valued belief than subjects in treatment T_W^P.

[18] Case 6 is also different in the two samples. Nevertheless, this difference should not affect the results, because this case is precise with respect to the relevant events in treatments T_W^P, T_W^{IP}, T_Y^P, and T_Y^{IP}. The reason for this difference will become apparent in Part 2 of Experiment 2.

Figure 2 **Empirical Distribution Functions of the CEs in Treatments T_W^P and T_W^{IP}**



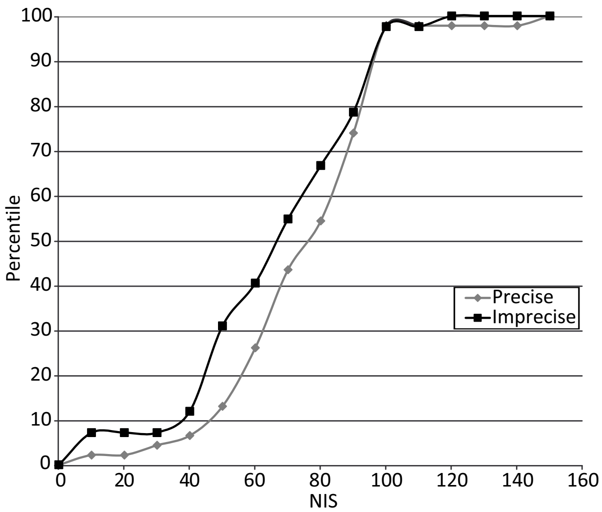
One may be concerned that although the belief valuation of subjects in treatment T_W^{IP} regarding event W may indeed be lower than that of subjects in treatment T_W^P, their belief valuation regarding the complement event Y may be higher than that of their counterparts. This set of beliefs does not reflect ambiguity aversion and is inconsistent with the model. The purpose of treatments T_Y^P and T_Y^{IP} is to test whether or not this occurs.

The data set in treatment T_Y^P is identical to that in treatment T_W^P, and the data set in treatment T_Y^{IP} is identical to that in treatment T_W^{IP}. The lottery proposed in both T_Y^P and T_Y^{IP} was "win NIS 150 if a yellow ball is drawn and 0 otherwise." The hypothesis of this part of Experiment 2 is that the belief valuation based on imprecise data is lower than that based on precise data for both the event W and for the complement event Y. In particular, both $CE_W^{IP} < CE_W^P$ and $CE_Y^{IP} < CE_Y^P$.[19]

Results of Part 1. There were 32 participants in both treatments T_W^P and T_W^{IP}, and 46 and 42 participants in treatments T_Y^P and T_Y^{IP}, respectively. The average CE of treatment T_W^P was $CE_W^P = 65.3$, and that of treatment T_W^{IP} was $CE_W^{IP} = 50.9$. The results indicate that CE_W^P is higher than CE_W^{IP}, where the difference between the two distributions of CEs is significant according to a Mann–Whitney U test ($u = -2.31$ and $p = 0.01$). Likewise, the average CE of treatment T_Y^P was $CE_Y^P = 78.5$, and that of treatment T_Y^{IP} was $CE_Y^{IP} = 70$. Here again, CE_Y^P is higher than CE_Y^{IP}, and the difference between the two distributions is significant ($u = 1.55$ and $p =$

[19] The comparisons in Experiment 2 are made only between two bets for which one stochastically dominates the other. Thus, as argued in Footnote 14, Karni and Safra's (1987) and Segal's (1988) critiques regarding the elicited CEs do not apply.

Figure 3 Empirical Distribution Functions of the CEs in Treatments T_Y^P and T_Y^{IP}



0.061). Figure 2 presents the empirical distribution functions of the CEs in treatments T_W^P and T_W^{IP}, and Figure 3 shows the empirical distribution functions of the CEs in treatments T_Y^P and T_Y^{IP}. As can be readily seen from these figures, not only are the empirical distributions different, but also the empirical distribution function that is associated with the precise data first order stochastically dominates that of the imprecise data in the corresponding treatment.

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 data set. Therefore, the findings support the premise that imprecise data are a source of ambiguity aversion.[20]

Although the model allowed for ambiguity-loving ((4) with $\alpha < 1$) and ambiguity-neutral behavior (when $\alpha = 1$ in either (3) or (4)), these experimental results provide evidence against them, because the former would entail that the CEs in treatments T_W^{IP} and T_Y^{IP} be higher than those in treatments T_W^P and T_Y^P, respectively, and the latter would imply that they should be approximately the same.

More generally, these findings undermine all additive belief formation processes (such as ignoring the sample and relying on some additive initial prior instead) because they impose that $CE_W^{IP} \leq CE_W^P$ if and only if $CE_Y^{IP} \geq CE_Y^P$.

[20] It should be noted that these comparisons do not rule out the possibility that the subjects' beliefs depend on the tasks they face. For example, they may hold different beliefs when betting on yellow as opposed to white. A test of such a conjecture should compare the belief valuation of yellow with a given data set, to that of white with the reverse data set, in which every past case in which yellow appeared is replace with white and vice versa.

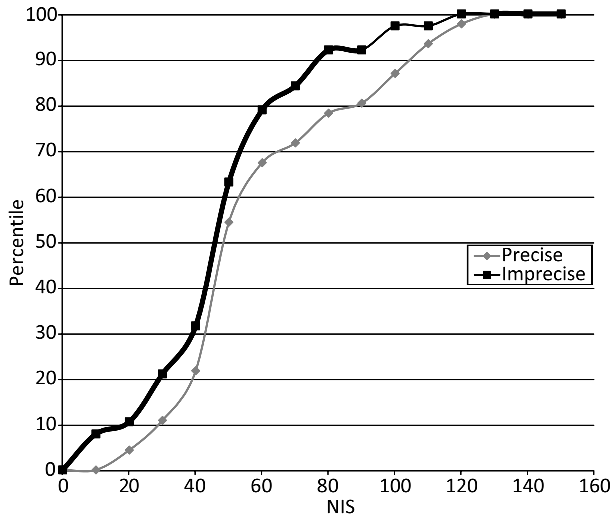
Finally, another alternative procedure of belief formation might be that individuals form beliefs according to the frequency of occurrence of precise cases in memory (with respect to the event in question), while ignoring the imprecise cases. Had this been true, CE_W^P would have been higher than CE_W^{IP}, and CE_Y^P would have been lower than CE_Y^{IP}; hence, it too is ruled out by the results of the experiment.

3.2.2. Part 2. Formally, to confirm that subjects' beliefs exhibit ambiguity aversion in the presence of imprecise data, we need to elicit their beliefs for each event in Σ (which amounts to 14 events excluding $\varnothing$ and Ω). Furthermore, these beliefs need to be obtained for both imprecise and precise data. This means that 28 treatments need to be performed, which is a number way too large from any practical perspective. Therefore, in Part 2 of this experiment, we chose to focus on two such 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 Part 1 of the experiment, which contain two states of nature. We view this as a robustness test to check whether the previous results hold for additional events.

In this part, four treatments were performed: T_{YO}^P, T_{YO}^{IP}, T_{YX}^P, and T_{YX}^{IP}. The data sets in this part are the same as in the previous part. The data set 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 2). The data set 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 3). Note that the relevant imprecise observations in treatment T_{YO}^{IP} are cases 1 and 6, whereas the relevant imprecise observations in treatment T_{YX}^{IP} are cases 6 and 8. Furthermore, in the precise data set, the event YO occurred in case 6 and did not occur in case 1, and the event YX occurred in case 8 and did not occur in case 6. This leads us to conclude that the same analysis and the same type of hypothesis as in Part 1 apply here; specifically, $CE_{YO}^{IP} < CE_{YO}^P$ and $CE_{YX}^{IP} < CE_{YX}^P$.

Results of Part 2. The numbers of participants in treatments T_{YO}^P and T_{YO}^{IP} were 46 and 38, respectively, and in treatments T_{YX}^P and T_{YX}^{IP}, 33 and 23, respectively. The main results are $CE_{YO}^P = 63.3$ and $CE_{YO}^{IP} = 52.4$. Evidently, CE_{YO}^P is higher than CE_{YO}^{IP}, and the two distributions of CEs are significantly different according to a Mann–Whitney U test ($u = -1.44$ and $p = 0.076$). Also $CE_{YX}^P = 69.1$ and $CE_{YX}^{IP} = 48.7$. Here again, CE_{YX}^P is higher than CE_{YX}^{IP}, and the distributions are significantly different ($u = -2.73$ and $p = 0.003$). Thus, subjects hold lower-valued beliefs than those induced by the neutral approach within the event Y, both for the event YX and for the event $Y \setminus YX$ (namely, YO). The

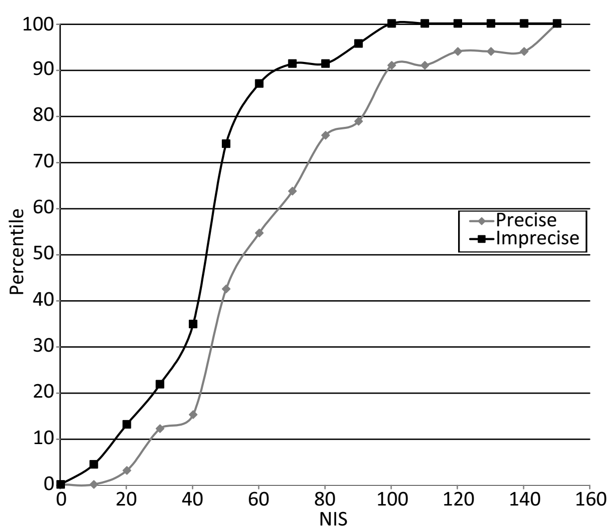
Figure 4 **Empirical Distribution Functions of the CEs in Treatments T_{YO}^{P} and T_{YO}^{IP}**



empirical distribution functions of the CEs in treatments T_{YO}^{P} and T_{YO}^{IP} can be found in Figure 4, and the empirical distribution functions of the CEs in treatments T_{YX}^{P} and T_{YX}^{IP} can be found in Figure 5. As can be seen from these figures, the empirical distribution function that is associated with the precise data first order stochastically dominates that of the imprecise data in the corresponding treatment.

We should recall that the experiments are based on a between-subjects approach. Therefore, the results indicate that subjects are ambiguity averse in the presence of imprecise data only on an aggregate level. This does not rule out the possibility that some subjects are ambiguity neutral or even ambiguity loving. In a within-subjects design it is possible to identify an individual's exact attitude toward ambiguity by eliciting his beliefs over events given imprecise data sets and then estimating his $v_D^F(A)$ or $v_D^G(A)$ value.

Figure 5 **Empirical Distribution Functions of the CEs in Treatments T_{YX}^{P} and T_{YX}^{IP}**



4. Discussion

In this work we introduce a model of belief formation based on a data set where some observations are imprecise. The use of imprecise data leads to a belief that is nonadditive. Our model may be interpreted as an approximation of the actual mental process undergone by an individual to evaluate the likelihood of events, which the individual may or may not be aware of. Focusing on the mental process enables us to narrow the possible beliefs that people might hold and to highlight the role of imprecise information in causing ambiguity. We suggest that imprecise information is a source for nonneutral attitudes toward ambiguity, ambiguity aversion in particular, and present experimental evidence to support this feature of the model.

We now 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.[21] 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 those possessing identical characteristics, the evaluation reduces to that of the basic model, i.e., to the relative frequency of occurrence.

One of the main limitations of the present model is that the belief formation studied does not depend on the number of observations in the data set. One might expect that as an individual accumulates more data about a certain situation, his ambiguity would gradually disappear. Eichberger and Guerdjikova (2010) introduced ambiguity into the framework of Billot et al. (2005) by allowing the individual to hold a set of conceivable probabilities given past cases. They showed that as the data set grows, the ambiguity diminishes. It is possible to encapsulate this element in our model by making additional assumptions about the memory accumulation process, for instance, by assuming memory relies both on first- and second-hand experience. Naturally, first-hand experience is more detailed and thus more precise than second-hand experience. At an early stage of an individual's life, his experience may rely more heavily on other peoples' experience; however, with age, the proportion of personal experience grows. Thus, in this modification, a larger memory contains a higher proportion

[21] In an earlier version of this paper, we did include this feature. That version included only extreme attitudes toward ambiguity (i.e., $\alpha = 0$), and it also had an axiomatization for those cases.

of precise cases. Therefore, beliefs based on this larger data set reflect less ambiguity.

To discuss another feature of our model, let us 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 a priori more likely to occur than others. 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 ω_i is perceived to be more likely than ω_j, it is most reasonable to set $L(\omega_i) > L(\omega_j)$. It then follows that given the same information in memory regarding ω_i and ω_j the evaluation of the former will be weakly higher than that of the latter. In symmetric situations, $L(\omega_i)$ will equal $L(\omega_j)$ for every i and j, and hence Equation (5) will be reduced to Equation (1). In this modified version of the model based on Equation (5), this paper's main theoretical results continue to obtain.

The settings in the urn experiments reported in this paper have the underlying symmetry property to which Equation (1) applies. The following is a short description of an experiment that we conducted taht better fits the circumstances of the modified model. The experiment concerned 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 common than others, such as blue compared to purple. Some 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, whereas others were provided with imprecise data on the subject. The findings of the experiment indicate that imprecise data are a source of ambiguity aversion also in such asymmetric circumstances.

Acknowledgments
The authors thank Itzhak Gilboa and Ariel Rubinstein for many helpful comments and valuable suggestions. They also thank Anat Bracha, Offer Lieberman, Dotan Persitz, Ella Segev, Peter Wakker, and Tali Ziv, as well as the participants of the Risk, Uncertainty, and Decision 2010 Conference in Paris and the participants of the Ph.D. seminar at Tel Aviv University. Finally, they thank the anonymous referees and the associate editor of this journal.

Appendix

Proof of Proposition 1. First it is established that v_D^F is a capacity. Take any data set $D = (B_1, \ldots, B_T)$. $v_D^F(\varnothing) = 0$ because for any j, $\varnothing \cap B_j = \varnothing$, and thus $F_j(\varnothing) = 0$. 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, $\{\omega \mid \omega \in A \cap B_j\} \subseteq \{\omega \mid \omega \in A' \cap B_j\}$; thus, $F_j(A) \le F_j(A')$ and $v_D^F(A) \le v_D^F(A')$.

To show that v_D^F as defined in Equation (3) is a belief function, we start by proving that the F_js for $j = 1, \ldots, T$ are belief functions; that is, for any j and any collection $A_1, \ldots, A_n$ of subsets of Ω,

$$F_j\left(\bigcup_{i=1,\ldots,n} A_i\right) \ge \sum_{\{H: H \subseteq \{1,\ldots,n\}\}} (-1)^{|H|+1} F_j\left(\bigcap_{k \in H} A_k\right). \tag{6}$$

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

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

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

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

hence (6) holds. If, on the other hand, $\bigcup_{i=1,\ldots,n} A_i \supseteq B_j$, then

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

hence (6) still holds.

Because (6) holds for all cases, it also holds for the average, that is,

$$\begin{aligned} v_D^F\left(\bigcup_{i=1,\ldots,n} A_i\right) &= \frac{\sum_{j=1}^T F_j(\bigcup_{i=1,\ldots,n} A_i)}{T} \\ &\ge \frac{\sum_{j=1}^T \sum_{\{H: H \subseteq \{1,\ldots,n\}\}} (-1)^{|H|+1} F_j(\bigcap_{k \in H} A_k)}{T}, \end{aligned}$$

and hence

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

which concludes the proof. □

Proof of Lemma 1. Assume $v(A) + v(A') = v(A \cup A')$ for all A' such that $A \cap A' = \varnothing$; 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 $A \cap A' = \varnothing$; then $v(A) + v(A') \le v(A \cup A')$. 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 because $\bar{v}$ must be concave. A similar proof could be applied for a concave capacity. □

Proof of Proposition 2. 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 $B_j \cap A \neq \varnothing$), then both $F_j(A) = 1$ and $G_j(A) = 1$. If $B_j \subseteq A^c$ (then $B_j \cap A = \varnothing$), 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 $\varnothing$ and Ω 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: $A \cap A'$, $(A \setminus A')$, $(A' \setminus A)$, or $(A \cup A')^c$. 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 3. Take data set $D = (B_1, \ldots, B_T)$. 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 $v_D^F = v_D^G$. 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 $A \cup A' \nsupseteq B_j$, then both $A \nsupseteq B_j$ and $A' \nsupseteq B_j$. 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 that 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

Abdellaoui, M., A. Baillon, L. Placido, P. Wakker. 2011. The rich domain of uncertainty: Source functions and their experimental implementation. *Amer. Econom. Rev.* **101**(2) 695–723.

Becker, G. M., M. H. DeGroot, J. Marashak. 1964. Measuring utility by a single-response sequential method. *Behav. Sci.* **9**(3) 226–232.

Becker, S. W., F. O. Brownson. 1964. What price ambiguity? Or the role of ambiguity in decision-making. *J. Political Econom.* **72**(1) 62–73.

Berger, J. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

Berger, J., M. Berliner. 1986. Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14**(2) 461–486.

Billot, A., I. Gilboa, D. Samet, D. Schmeidler. 2005. Probabilities as similarity-weighted frequencies. *Econometrica* **73**(4) 1125–1136.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, H. Meijers. 2009. Gender differences in risk aversion and ambiguity aversion. *J. Eur. Econom. Assoc.* **7**(2–3) 649–658.

Camerer, C., M. Weber. 1992. Recent developments in modeling preferences: Uncertainty and ambiguity, *J. Risk Uncertainty* **5**(4) 325–370.

Carnap, R. 1952. *The Continuum of Inductive Methods*. University Press, Chicago.

Carnap, R. 1980. A basic system of inductive logic, Part II. C. J. Richard, ed. *Studies in Inductive Logic and Probability*, Vol. II. University of California Press, Berkeley, 7–155.

Chew, S. H., J. S. Sagi. 2008. Small worlds: Modeling attitudes toward sources of uncertainty. *J. Econom. Theory* **139**(1) 1–24.

Chipman, J. S. 1960. Stochastic choice and subjective probability. D. Willner, ed. *Decisions, Values and Groups*, Vol. 1. Symposium Publications Division, Pergamon Press, New York, 70–95.

Cohen, M., J. Y. Jaffray, T. Said 1987. Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organ. Behav. Human Decision Processes* **39**(1) 1–22.

de Finetti, B. 1937. La pr´evision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincar´e* **7** 1–68.

Dempster, A. P. 1967. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38**(2) 325–339.

Dempster, A. P. 1968. A generalization of Bayesian inference. *J. Royal Statist. Soc. B* **30** 205–247.

Eichberger, J., A. Guerdjikova. 2010. Case-based belief formation under ambiguity. *Math. Soc. Sci.* **60**(3) 161–177.

Ellsberg D. 1961. Risk, ambiguity, savage axioms. *Quart. J. Econom.* **75**(4) 643–669.

Epstein, L. G., J. Zhang. 2001. Subjective Probabilities on subjective unambiguous events. *Econometrica* **69**(2) 265–306.

Gajdos, T., T. Hayashi, J. M. Tallon, J. C. Vergnaud. 2008. Attitude toward imprecise information. *J. Econom. Theory* **140**(1) 27–65.

Ghirardato, P. 2004. Defining ambiguity and ambiguity attitude. I. Gilboa, ed. *Uncertainty in Economic Theory: A Collection of Essays in Honor of David Schmeidler's 65th Birthday*. Routledge, New York, 36–45.

Ghirardato, P., M. Marinacci. 2002. Ambiguity made precise: A comparative foundation. *J. Econom. Theory* **102**(2) 251–289.

Gigliotti, G., B. Sopher. 1996. The testing principle: Inductive reasoning and the Ellsberg paradox. *Thinking Reasoning* **2**(1) 33–49.

Gilboa, I. 1988. A combination of expected utility and maxmin decision criteria. *J. Math. Psych.* **32**(4) 405–420.

Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *J. Math. Econom.* **18**(2) 141–153.

Gonzales, C., J. Y. Jaffray. 1998. Imprecise sampling and direct decision making. *Ann. Oper. Res.* **80** 207–235.

Halevy, Y. 2007. Ellsberg revisited: An experimental study. *Econometrica* **75**(2) 503–536.

Hogarth, R. M., H. J. Einhorn. 1990. Venture theory: A model of decision weights. *Management Sci.* **36**(7) 780–803.

Huber, P. J. 1973. The use of Choquet capacities in statistics. *Bull. Internat. Statist. Inst.* **45** 181–191.

Jaffray, J. Y. 1988. Choice under risk and the security factor: An axiomatic model. *Theory Decision* **24**(2) 169–200.

Jaffray, J. Y. 1991. Belief functions, convex capacities and decision making. J.-P. Doignon, J.-C. Falmagne, eds. *Mathematical Psychology: Current Developments*. Springer Verlag, New York, 127–134.

Jaffray, J. Y., F. Philippe. 1997. On the existence of subjective upper and lower probabilities. *Math. Oper. Res.* **22**(1) 165–185.

Karni, E., Z. Safra. 1987. Preference reversal' and the observability of preferences, by experimental methods. *Econometrica* **55**(3) 675–685.

Klibanoff, P., M. Marinacci, S. Mukerji. 2005. A smooth model of decision making under ambiguity. *Econometrica* **73**(6) 1849–1892.

Marinacci, M., L. Montrucchio. 2004. Introduction to the mathematics of ambiguity. I. Gilboa, ed. *Uncertainty in Economic Theory: A Collection of Essays in Honor of David Schmeidler's 65th Birthday*. Routledge, New York, 46–107.

Mukerji, S. 1996. Understanding the nonadditive probability decision model. *Econom. Theory* **9**(1) 23–46.

Nehring, K. 1999. Capacities and probabilistic beliefs: A precarious coexistence. *Math. Soc. Sci.* **38**(2) 197–213.

Ramsey, F. P. 1931. Truth a probability. *The Foundations of Mathematics and Other Logical Essays*. Harcourt, Brace, New York.

Savage, L. J. 1954. *The Foundations of Statistics*. John Wiley and Sons, New York.

Schmeidler, D. 1989. Subjective probability without additivity. *Econometrica* **57**(3) 571–587.

Segal, U. 1988. Does the preference reversal phenomenon necessarily contradict the independence axiom? *Amer. Econom. Rev.* **78**(1) 233–236.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.

Tversky, A., D. Kahneman. 1971. The belief in the law of small numbers. *Psych. Bull.* **76**(2) 105–110.

Viscusi, W. K 1989. Prospective reference theory: Toward an explanation of the paradoxes. *J. Risk Uncertainty* **2**(3) 235–264.

Wakker, P. 2008. Uncertainty. L. Blume, S. N. Durlauf, eds. *The New Palgrave: A Dictionary of Economics.* MacMillan Press, London, 6780–6791.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Wasserman, L. A., J. Kadane. 1990. "Bayes" theorem for Choquet capacities. *Ann. Statist.* **18**(3) 1328–1339.

Zhang, J. 2002. Subjective ambiguity, probability and capacity. *Econometric Theory* **20**(1) 159–181.