

An empirical test of some measurement-theoretic axioms for fuzzy sets

C. Desimpelaere¹, T. Marchant²

*Department of Data-Analysis, Faculty of Psychology and Educational Sciences,
Ghent University, H. Dunantlaan 1, 9000 Ghent, Belgium.*

Abstract

In the previous years some authors have been elaborating on the measurement-theoretic foundations of fuzzy set theory. A well-known problem in this approach is the difficult applicability of the deterministically formulated axioms on data obtained from an expert (or social science data in general). In this paper a statistical method is proposed which can be used for the testing of measurement axioms in a pairwise comparison design. Consequently we apply this method for the testing of two of these axioms (i.e. Transitivity and Order of Operations) used in the context of fuzzy set theory. The results clearly indicate that subjects act in agreement with Transitivity, but, for Order of Operations there is no such vivid evidence, which may have repercussions for some of the earlier theoretical work.

Key words: measurement theoretic axioms, fuzzy set theory, testing axioms

1 Introduction

Since Zadeh (1965) introduced the concept of fuzzy sets, a major concern has been the meaning and measurement of membership functions. Related to this question is the issue of which operators (e.g. max, min, bold intersection, bounded sum,...) to use to represent the union and intersection in fuzzy set theory. Nowadays a general concept for these different operators is in use (i.e. t-norms and t-conorms, concepts borrowed from the statistical metric

Email addresses: Chris.Desimpelaere@UGent.be (C. Desimpelaere),
Thierry.Marchant@UGent.be (T. Marchant).

¹ Tel.: +32-92646390; fax: +32-92646487

² Tel.: +32-92646373; fax: +32-92646487

spaces literature), though this mere abstraction does not give an answer to the question raised above: which operator should be used?

The answer to this question depends of course on what we want to represent by means of membership degrees. In some cases, for instance in fuzzy expert systems, one tries to represent the knowledge of an expert. In other cases, membership degrees are the outcome of an algorithm in artificial intelligence, e.g. a fuzzy classification algorithm, and are not meant to represent the knowledge of an expert. In this paper, we will only consider the former case.

So, if we want to represent knowledge, we need to observe it, to capture it; and if we want to represent it by means of numbers, we need to measure it. The easiest way to measure the membership is obviously to tell the expert ‘give me a number between 0 and 1 representing the membership degree $\mu_A(x)$ of object x in the set A ?’ A problem with this technique is that we do not know the properties of the numbers we obtain. For example, does $\mu_A(x) - \mu_A(y)$ represent something? Does $\mu_A(x)\mu_B(y)$ represent $\mu_{A\cap B}(x)$, i.e. the membership degree of x in the intersection of A and B ? Or is $\mu_{A\cap B}(x)$ better represented by $\min(\mu_A(x), \mu_B(y))$? Several other measurement techniques can be thought of and have been proposed in the literature.

Türksen (1991) has shown that it is possible to analyze some of these techniques within the framework of measurement theory (Krantz et al., 1971; Roberts, 1979), this way answering the kind of questions raised in the previous paragraph. In this tradition, measurement involves a mapping of an empirical structure into a numerical one, in such a way that, corresponding to each relation in the first structure, there is a relation in the second one, this mapping having the property that some objects are in relation in the empirical structure if and only if the corresponding objects are in relation in the numerical structure. In that case, one says the empirical structure is represented by the numerical one. If one wishes to represent an empirical structure by a given numerical structure, some conditions (or axioms) need to be met, depending on the desired numerical structure. One of the goals of measurement theory is to identify those conditions making measurement possible.

In the context of measurement of fuzzy sets, the empirical relational structure consists of 3 elements: (1) a set of fuzzy sets, (2) a relation indicating that the membership of an object in a fuzzy set is at least as large as the membership of the same object in another fuzzy set (the same object being used in this relation over all fuzzy sets) and (3) some operations on the fuzzy sets (such as union and intersection). This is the relational structure we will put our attention on in this paper. Another possible structure could be one consisting of different objects and a relation on these objects concerning their membership in a given fuzzy set. Other structures can be found in the literature, (e.g.

Bollman-Sdorra et al., 1993; Marchant, 2004a) but are not within the scope of this paper.

Some authors (Bollman-Sdorra et al., 1993; Bilgiç and Türksen, 1995; Marchant, 2004a,b) have been elaborating on the conditions under which appropriate representations of membership degrees can be made, and whether this is dependent on the choice of a particular operation representing union and intersection. The focus in their papers was on the theoretical development of the conditions needed to make a particular representation.

A practical problem now arises. Thanks to measurement theory, we know the conditions guaranteeing that a given a numerical representation and the accompanying measurement technique can be used. But are these conditions met in practice? In other words, does the membership, as perceived by an expert, have properties that are compatible with the numerical representation one wishes to use? If yes, then we may safely use it. If not, then one needs to use another measurement technique.

The aim of this paper is twofold. First, we will present a new procedure for statistically testing measurement-theoretic axioms, i.e. for answering the questions of previous paragraph. Second, we will test some axioms. This will illustrate how our statistical procedure works and provide some empirical results on two key axioms for the measurement of membership.

In Section 2, we will present the axioms that we will test further in the paper. In Section 3, we will shortly review some existing statistical procedures for testing measurement-theoretic axioms and we will then present a new procedure. Section 4 will be devoted to the presentation of our experiment and its results. Section 5 will conclude.

2 Review of some measurement-theoretical work

As already mentioned we will focus in this paper on the membership of one object in several sets, in the presence of a union and intersection, a structure which can be formalised as $\langle \mathcal{F}, \succsim_x, \cup, \cap \rangle$. The primitives used are \mathcal{F} , a set whose elements can be interpreted as all fuzzy sets that are relevant in a particular context, \succsim_x , a binary relation on \mathcal{F} , \cup , a binary operation on \mathcal{F} , interpreted as the union of any two elements of \mathcal{F} , and \cap , another binary operation, interpreted as the intersection of any two elements of \mathcal{F} . The statement $A \succsim_x B$ is interpreted as “the object x belongs at least as much to A as to B ”. We will use \succ_T to denote the asymmetric part of \succsim_T (strict preference) and \sim_T to denote the symmetric part of \succsim_T (indifference).

Several papers have dealt with the measurement of $\langle \mathcal{F}, \succsim_x, \cup, \cap \rangle$. One is by Bollman-Sdorra et al. (1993), other papers are by Bilgiç, Norwich and Türksen in different combinations (Bilgiç and Türksen, 1995, 1997, 2000; Norwich and Türksen, 1982; Türksen, 1991). Bollman-Sdorra et al. (1993) present results in which \cup and \cap are represented by the max and min operations respectively. The results presented by the other authors yield representations of \cup by means of the Łukasiewicz t-conorm (or some generalizations), i.e. $S(a, b) = \min(1, a + b)$.

We present below some of the axioms that have been presented in the literature in order to characterize some numerical representations.

Completeness. For all A, B in \mathcal{F} , $A \succsim_x B$ or $B \succsim_x A$.

Transitivity For all A, B, C in \mathcal{F} , $A \succsim_x B$ and $B \succsim_x C$ implies $A \succsim_x C$.

Order of Operations. For all A, B in \mathcal{F} , $A \cup B \succsim_x A \cap B$.

Commutativity. For all A, B in \mathcal{F} ,

$$A \cup B \sim_x B \cup A \text{ and} \\ A \cap B \sim_x B \cap A.$$

Absorption. For all A, B in \mathcal{F} ,

$$A \sim_x A \cap (A \cup B) \text{ and} \\ A \sim_x A \cup (A \cap B).$$

Right Monotonicity. For all A, B, C in \mathcal{F} ,

$$A \succsim_x B \text{ implies } A \cap C \succsim_x B \cap C \text{ and} \\ A \succsim_x B \text{ implies } A \cup C \succsim_x B \cup C.$$

Etc.

It is of course impossible to test all axioms that have been discussed in the literature. We will thus test only two of them in this paper.

We do not test Completeness because of a lack of knowledge of any method to do this properly. To us it seems that, if one asks a subject to choose one of the following statements, ' $A \succ_x B$ ', ' $B \succ_x A$ ', ' $B \sim_x A$ ' or 'I cannot compare A and B ', the person will be tempted to choose the last option in situations where the first three options require some cognitive effort in order to make a decent choice. So, an experiment aimed at testing Completeness is very likely to reject Completeness.

Transitivity is so controversial in many areas (Tversky, 1969; Fishburn, 1991; Luce, 2000, section 2.2) that we felt there was no option not testing it. Furthermore this condition is crucial for most applications of fuzzy sets. If transitivity

does not hold, then membership functions with values in totally ordered sets (like \mathbb{R}) cannot be used.

Order of Operations is another crucial condition for most applications since it is a necessary condition for all numerical representations using a t -norm and a t -conorm to represent the intersection and the union. That is why we will test it.

More research is of course needed to test other axioms and, hence, to see which numerical representations is empirically validated.

3 Statistical Method

A well-known problem in measurement theory concerns the difficult applicability of the deterministically formulated axioms on data in the social sciences which are sensitive to random error (due to a number of sources, such as inherent unreliability of human behavior, sampling error,...) and systematic error. Therefore some major effort has been spent on the development of statistical models of axiom testing. First, Iverson and Falmagne (1985) suggested a likelihood-ratio method and later Karabatsos and others (Karabatsos, 2001; Karabatsos and Sheu, 2001; Karabatsos and Ullrich, 2002; Myung et al., 2005) proposed different Bayesian statistical methods. All of these methods were based on a recasting of deterministic axioms into a probabilistic form. This approach, although extremely interesting, poses at least two problems. First, a probabilistic version of an axiom is not the same as the corresponding deterministic one: they are distinct mathematical objects. Testing the probabilistic one is therefore not equivalent to testing the deterministic one. Second, there are several ways to recast a deterministic axiom into a probabilistic form. Transitivity, for instance, has been recasted by different researchers in weak stochastic transitivity, strong stochastic transitivity, min-transitivity, transitivity w.r.t. a given t -norm, etc. Choosing a particular recasting is hence, to some extent, arbitrary.

Karabatsos (2005) also developed an approach based on an exchangeable multinomial model, in which a recasting of the deterministic axioms into a probabilistic form is no longer necessary. In this latter approach a framework is provided for testing any deterministic axiom, on discrete- or real-valued response data that can be represented as frequencies in one or more multidimensional contingency tables. More specifically, the data are modeled with a Dirichlet posterior distribution, under a multinomial sampling distribution of the data, and a Dirichlet prior distribution specified over the multinomial parameters. Consequently a Bayes factor can be directly computed (closed form solution) to provide a test for any of the deterministic axioms. Karabatsos also

explains how this model generalizes the previously mentioned approaches.

In this paper we will present a different model, developed at the same time as the method of (Karabatsos, 2005), suggesting a different data-generating process (see below), and giving us a method to analyze our data in which there is also no need to recast any deterministic axiom into a probabilistic form. A Bayes factor is computed to provide a test for an axiom.

3.1 Data

Let \mathcal{F} ($\#\mathcal{F} = n$ with $\#\mathcal{F}$ denoting the cardinality of \mathcal{F}) be a set containing fuzzy sets (as in Section 2) or, for instance, gambles if we are in decision theory, or tones, in psychophysics. The elements of \mathcal{F} will be called alternatives. Suppose that, in an experiment, we present two alternatives A and B of \mathcal{F} to a subject and we ask him to choose one of them according to some criterion. For example, we ask him to choose the fuzzy set such that the membership of some object x is maximal. Or the gamble he prefers, or the tone he finds louder. We repeat this for every pair of distinct alternatives in \mathcal{F} . As a result of such an experiment, the subject is giving us a complete and asymmetric binary relation \succ_O (O for observed) over \mathcal{F} .

Note that, by asking the subject to choose A or B , we do not let him express that he is indifferent between A and B or that he cannot compare A or B . This to avoid that the subject systematically chooses the indifference or the incomparability when the comparison is difficult.

3.2 Model

In our model, we suppose the subject has a relation \succ_T (T for true) in mind, over \mathcal{F} . The relation \succ_T might be the relation \succ_x of Section 2 or, for instance, a preference relation in decision theory, or a perceived intensity relation in psychophysics. When we present two alternatives A and B to an expert and ask him to make a comparison, he chooses one of them in accordance to \succ_T , but, as humans are not always consistent in their behavior, with some ‘error’. Because of the error, \succ_T and \succ_O will probably not be identical. However we can reasonably assume, that the probability that a subject chooses A out of the set A and B , i.e. $P(A \succ_O B)$, is

- (1) $P(A \succ_O B) = p$ if $A \succ_T B$,
- (2) $P(A \succ_O B) = 1 - p$ if $B \succ_T A$
- (3) $P(A \succ_O B) = P(B \succ_O A) = 0.5$ if $A \sim_T B$,

where $p > 0.5$, otherwise our model would not make sense. We also assume p does not depend on A and B (like Harless and Camerer (1994)). In (3) we assume for a subject having an indifference in mind, an equal likelihood of choosing either A or B .

Instead of assuming that p is independent of A and B , we could use the same model with one parameter p_{AB} for each pair A and B . This would have one advantage (the model being more flexible, it would perhaps be closer to reality) but, also, one disadvantage: a considerable loss of parsimony.

If we use a model with one parameter p_{AB} for each pair A and B , it is also tempting to impose that $[A \succ_T B, B \succ_T C \text{ and } A \succ_T C]$ imply $p_{AC} \geq p_{AB}$ and $p_{AC} \geq p_{BC}$. This would reflect the intuition that errors are more frequent when the alternatives are closer to each other. In our opinion, this is not sound. The axioms we test do not imply any idea of distance. Some *systems* of axioms may eventually be equivalent to the existence of a numerical representation from which we can deduce a distance. But no one axiom, taken individually, implies a distance. Remember that we test axioms in order to know if we can use a particular numerical representation, for instance one that implies a distance. If, when testing the axioms, we use a distance, then we are using a circular reasoning. We use properties of the representation while testing axioms; but we do not know yet if this representation exists.

Besides, Carbone (1997) have found that a model where p_{AB} is a monotone function of the distance between A and B does not do a much better job than a model with p_{AB} independent of A and B although the model with p_{AB} constant is much more parsimonious. About these two models, (Luce, 2000, p.29) writes ‘[...] I know of no principled reason for choosing either.’

3.3 Odds

The set of all binary relations on \mathcal{F} satisfying some given axiom(s) is denoted by X . The hypothesis we want to test is ‘ $\succ_T \in X$ ’. The competing hypothesis is ‘ $\succ_T \notin X$ ’. We suggest to compute the odds to test the evidence of one hypothesis against the other. We begin with the data \succ_O , assumed to have arisen under one of the two hypotheses ‘ $\succ_T \in X$ ’ or ‘ $\succ_T \notin X$ ’, and this according to the respective probability distributions $P[\succ_O | \succ_T \in X]$ and $P[\succ_O | \succ_T \notin X]$. Given the prior probabilities $P[\succ_T \in X]$ and $P[\succ_T \notin X] = 1 - P[\succ_T \in X]$, reflecting our belief that the subject agrees/disagrees with some axiom(s), the data produce posterior probabilities $P[\succ_T \in X | \succ_O]$ and $P[\succ_T \notin X | \succ_O]$. From Bayes theorem, we know

$$P[\zeta_T \in X | \succ_O] = \frac{P[\succ_O | \zeta_T \in X] P[\zeta_T \in X]}{P[\succ_O | \zeta_T \in X] P[\zeta_T \in X] + P[\succ_O | \zeta_T \notin X] P[\zeta_T \notin X]}.$$

By conversion to the odds scale, we obtain the Bayes factor

$$\frac{P[\zeta_T \in X | \succ_O]}{P[\zeta_T \notin X | \succ_O]} = \frac{P[\succ_O | \zeta_T \in X]}{P[\succ_O | \zeta_T \notin X]} \frac{P[\zeta_T \in X]}{P[\zeta_T \notin X]}. \quad (1)$$

Let us compute the numerator of the right-hand side of (1). We have

$$\begin{aligned} P[\succ_O | \zeta_T \in X] &= \frac{P[\succ_O \cap (\zeta_T \in X)]}{P[\zeta_T \in X]} \\ &= \frac{P[\bigcup_{R \in X} (\succ_O \cap (\zeta_T = R))]}{P[\zeta_T \in X]} \\ &= \frac{\sum_{R \in X} P[\succ_O | \zeta_T = R] P[\zeta_T = R]}{P[\zeta_T \in X]}. \end{aligned}$$

So,

$$P[\zeta_T \in X] P[\succ_O | \zeta_T \in X] = \sum_{R \in X} (P[\succ_O | \zeta_T = R] P[\zeta_T = R]).$$

Using a similar reasoning for the denominator, expression (1) can therefore be rewritten as

$$\frac{P[\zeta_T \in X | \succ_O]}{P[\zeta_T \notin X | \succ_O]} = \frac{\sum_{R \in X} (P[\succ_O | \zeta_T = R] P[\zeta_T = R])}{\sum_{R \notin X} (P[\succ_O | \zeta_T = R] P[\zeta_T = R])}. \quad (2)$$

If we further assume that all relations in X have the same probability (i.e. $P[\zeta_T = R] = P[\zeta_T \in X] / \#X$, for $R \in X$) and that all relations in \bar{X} (meaning the complement of X) have the same probability (i.e. $P[\zeta_T = R] = P[\zeta_T \in \bar{X}] / \#\bar{X}$, for $R \notin X$), then we obtain

$$\frac{P[\zeta_T \in X | \succ_O]}{P[\zeta_T \notin X | \succ_O]} = \frac{P[\zeta_T \in X]}{P[\zeta_T \notin X]} \frac{\#\bar{X}}{\#X} \frac{\sum_{R \in X} P[\succ_O | \zeta_T = R]}{\sum_{R \notin X} P[\succ_O | \zeta_T = R]}. \quad (3)$$

In order to compute this expression, we have to compute $P[\succ_O | \zeta_T = R]$. Let $\alpha(\succ_O, R)$ be the number of pairs such that \succ_O and R agree, $\beta(\succ_O, R)$ the number of pairs with an inversion of the preference and $\gamma(R)$ ³ the number of

³ Note that γ only depends on R and not on \succ_O , since the latter is asymmetric.

pairs where an indifference has been broken. Of course,

$$\alpha(\succ_O, R) + \beta(\succ_O, R) + \gamma(R) = n(n - 1)/2.$$

We then have

$$P[\succ_O \mid \succ_{\mathcal{F}} = R] = \hat{p}(\succ_O, R)^{\alpha(\succ_O, R)} (1 - \hat{p}(\succ_O, R))^{\beta(\succ_O, R)} 0.5^{\gamma(R)}, \quad (4)$$

where $\hat{p}(\succ_O, R)$ is an estimation of p and is given by

$$\hat{p}(\succ_O, R) = \max\left(\frac{1}{2}, \frac{\alpha(\succ_O, R)}{\alpha(\succ_O, R) + \beta(\succ_O, R)}\right).$$

The $\max(1/2, \cdot)$ operator is used here because $p < .5$ does not make sense in our model (see section 3.2).

3.4 Practical considerations

When \mathcal{F} is small, it is possible to compute the sums in (3) in an exact way. But when \mathcal{F} is large (6 is already large, since with 6 elements in \mathcal{F} , more than 1.4348×10^7 different relations are possible), we must resort to a Monte Carlo procedure. We randomly generate k relations belonging to X (and l relations belonging to its complement for the denominator), denoted by R_1, R_2, \dots, R_k (or S_1, S_2, \dots, S_l). We then approximate the odds ratio by

$$\frac{P[\succ_{\mathcal{F}} \in X \mid \succ_O]}{P[\succ_{\mathcal{F}} \notin X \mid \succ_O]} \approx \frac{P[\succ_{\mathcal{F}} \in X]}{P[\succ_{\mathcal{F}} \notin X]} \frac{\#\bar{X}}{\#X} \frac{\#X}{\#\bar{X}} \frac{\sum_{i=1}^k P[\succ_O \mid \succ_{\mathcal{F}} = R_i]}{\sum_{i=1}^l P[\succ_O \mid \succ_{\mathcal{F}} = S_i]}. \quad (5)$$

4 Experiment

In this experiment, we want to test Transitivity and Order of Operations (see Section 2).

4.1 Presentation

All participants were shown a picture of a person⁴ on a computer screen (representing the object x in our empirical relational structure $\langle \mathcal{F}, \succ_x, \cup, \cap \rangle$)

⁴ Unfortunately, not having the written permission of this person to publish the picture, it cannot be presented here.

and consequently presented with N binary choices of the form: the person on this picture belongs more to: ‘the set of people who are tall’ – ‘the set of people who are old and who have grey hair’ (the alternatives representing the elements of the set \mathcal{F} in our empirical relational structure). A choice was made by pointing one of the two alternatives and left-clicking the mouse. To continue the experiment, no other option was available (forced choice).

So the linguistic terms (tall, old, grey hair,...) used in the alternatives were considered as labels of a fuzzy set. A combination of up to 3 of these linguistic terms (fuzzy sets) was possible. The assemblage of these combined fuzzy sets was achieved through the use of the linguistic terms ‘and’ and ‘or’, which were considered as representing the fuzzy ‘intersection’ and ‘union’ respectively. Of course, other representations for these operators were possible. Our choice was in favor of ‘and’ and ‘or’ because of their pervasiveness in the fuzzy literature and in many applications (e.g. Bilgiç and Türksen, 1995; Schwartz, 2000). We are aware of the criticisms raised by some researchers against the modelling of the linguistic ‘and’ by an intersection and of the linguistic ‘or’ by a union (e.g. Tamburini, De Santis and Zamuner, 2002; Türksen, 2006). But since all experiments aiming at verifying that ‘and’ can be represented by an intersection and ‘or’ by a union have tested the representation itself and not the axioms underlying it (e.g. Zimmermann and Zysno, 1980), we do not know today why the linguistic connectives cannot be adequately modelled by the set-theoretic ones. That is why we think it is important to carry out some experiments testing which axioms underlying the intersection (resp. union), modelled by min (resp. max), are satisfied by ‘and’ (resp. ‘or’).

The length N of the sequence of binary choices is based on all pairwise combinations of the n alternatives and was chosen as the maximum comfortable sequence of binary choices for the participants (typically we chose $n=23$, so we obtained a length of 253 pairwise comparisons). These 23 alternatives can be found in table 1.

Of these 23 alternatives, the first 9 alternatives are single fuzzy sets and have been used for testing Transitivity. The following 6 (10–15 in table) alternatives consist of three unions and three intersections of the same fuzzy sets, and have been used for testing Order of Operations. The remaining 8 (16–23 in table) alternatives are randomly generated combinations of fuzzy sets and have been used as distractors. In order to construct alternatives 10–15, we tried not to use fuzzy sets 1–9. The reason for this is that we want to avoid that a participant, when presented with a pair, say ‘adult and thick’ vs. ‘adult or thick’, be influenced by his previous response on the pair ‘adult’ vs. ‘thick’. This increases the chance that the responses be independent. Yet, two of the alternatives 10–14 involve one of the alternatives 1–9: ‘thin’, because it is difficult to find many different linguistic expressions that are relevant for describing a given person.

1	middle age	13	red hair and long hair
2	grey hair	14	thin and round face
3	adult	15	small and young adult
4	slender	16	very thick and grey hair and square face
5	large	17	very thick or black hair or adult
6	black hair	18	middle length or juvenile or blond hair
7	middle long hair	19	black hair and middle length and old
8	thick	20	very small and grey hair and adult
9	thin	21	very thick and small and brown hair
10	red hair or long hair	22	small and brown hair and adult
11	thin or round face	23	very small and brown hair and adult
12	small or young adult		

Table 1

Alternatives used in the experiment (translated from Dutch).

4.2 Procedure

The experiment was conducted in a session normally lasting no more than one hour. Participants could determine the pace of the presentations. A new binary choice was shown only after having answered a previous one. The pairwise alternatives presentation we used was based on a design which is optimal in two respects: (i) the alternatives are optimally spaced across the sequence (in the sense that we have about the same number of pairs of alternatives in between two pairs of alternatives that include the same alternative) and (ii) the left-right position of each alternative is altered as often as possible on subsequent presentations.

4.3 Equipment

The picture (stimulus) and the alternatives were presented by the use of a Java program⁵ and as such could be presented in any (Java supporting) browser. Normally the experiments were conducted on a normal Pentium 4 pc, running Linux. Presentation-media was a 17inch monitor, placed at 50 cm distance in front of the participants.

⁵ This program was written by M. Covents of the Testpracticum of Ghent University, whom we owe due thanks.

4.4 Participants

Eight people participated in this experiment, of which 4 were man and 4 women. Some of them were MA-students and the rest was staff of Ghent University. The students received a compensation of 10 euro per session.

4.5 Analysis - Transitivity

Because we have no arguments against or in favour of Transitivity, we used in our statistical analysis a uniform prior: $P[\zeta_T \in X] = P[\zeta_T \notin X] = 1/2$.

For this analysis we used the first 9 alternatives from table 1. So we needed a monte carlo simulation to conduct the above presented statistical analysis (see section 3.4). A drawback of simulation methods is that it is often not straightforward to decide when it is safe to terminate them. To have an idea of the convergence of our calculations we decided to sample—for each separate analysis—a number of parallel, independent chains and plot the 90%-posterior band of our statistic of interest (see below) as well as the average of the same statistic computed over the parallel generated relations at each point t of the sequence of length T ⁶.

Because of reasons of convergence we did not conduct a single Monte Carlo—for each analysis—in order to compute expression (5), but carried out two Monte Carlo simulations, one for the product of the numerators of the second and the third fraction of the right hand side of expression (5), (see expression (6)) and one for the product of the denominators of the same fractions, (see expression (7)). Sometimes the number of generated relations needed in order to achieve convergence was different for expression (6) and expression (7).

$$\frac{\#X}{k} \times \sum_{i=1}^k P[\gamma_O \mid \zeta_T = R_i] \quad (6)$$

$$\frac{3^{n(n-1)/2} - \#X}{l} \times \sum_{i=1}^l P[\gamma_O \mid \zeta_T = S_i] \quad (7)$$

⁶ Initially also the Gelman and Rubin statistic (Gelman and Rubin, 1992) was computed as a convergence diagnostic, but since this measure indicated convergence even when there was no graphical evidence for it we discarded this diagnosis from our analyses.

subject	$\sum_{R \in X} P[\succ_O \mid \succ_{T= R}]$	$\sum_{R \notin X} P[\succ_O \mid \succ_{T= R}]$	Ratio
1.	18.08596	8356106	45837.86
2.	18.8051	8154478	48838.94
3.	17.92734	9078993	41818.16
4.	1.309643	6511686	4259.38
5.	17.17977	6371685	57101.80
6.	1.402265	7855661	3780.37
7.	18.33732	8498536	45696.02
8.	0.3898225	6612845	1248.43

Table 2

Results Monte Carlo analysis for Transitivity.

4.6 Results - Transitivity

When we look at table 2 we can find in the second column the numerator of the third fraction of the right hand side of expression (3), which is estimated by expression (6). In the third column we can find the denominator of the third fraction of the right hand side of expression (3), which is estimated by expression (7). In the last column of this table one can find the result for the posterior odds computed for each of the subjects (i.e. the ratio of the second and the third column multiplied by $\#\bar{X}/\#X$, see expression 3, with $\#X = 7087261$ ⁷ and $\#\bar{X} = 3^{n(n-1)/2} - 7087261$). It is clear that the evidence obtained from each subject speaks clearly in favor of transitivity, because the odds are much larger than 1. A graphical display of the convergence of n_t parallel sequences of Monte Carlo estimations for expressions (6) and (7), are printed here for two of our participants (i.e. subj. 6 and subj. 8). In fig.1 convergence plots are shown for expression (6) for subject 6 (left, with $t_n = 1000$ and $T = 1.0 \times 10^5$) and subject 8 (right, with $t_n = 300$ and $T = 1.0 \times 10^5$). The full line indicates the running average of $P[\succ_O \mid \succ_{T= R}]$, i.e. $\sum_{i=1}^k P[\succ_O \mid \succ_{T= R_i}]/k$, while the dashed lines denote the running empirical percentiles (.05 and .95), from bottom to top. The full lines becoming stationary indicates convergence for our estimators.

Convergence plots for expression (7) for subject 6 (left, with $t_n = 3000$ and

⁷ The total number of weak orders W_m on m objects is given by $W_m = w(m, 1) + \dots + w(m, m)$ with $w(m, k) = k!S(m, k)$, $S(m, k)$ being a Stirling number of the second kind. The quickest way to compute W_m is by use of the recursion $w(m, k) = k(w(m-1, k) + w(m-1, k-1))$ with $w(1, 1) = 1$, $w(i, 0) = 0$ and $w_{i < j}(i, j) = 0$ for all $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, m$. A detailed description of this method can be found in Maassen and Bezembinder (2002)

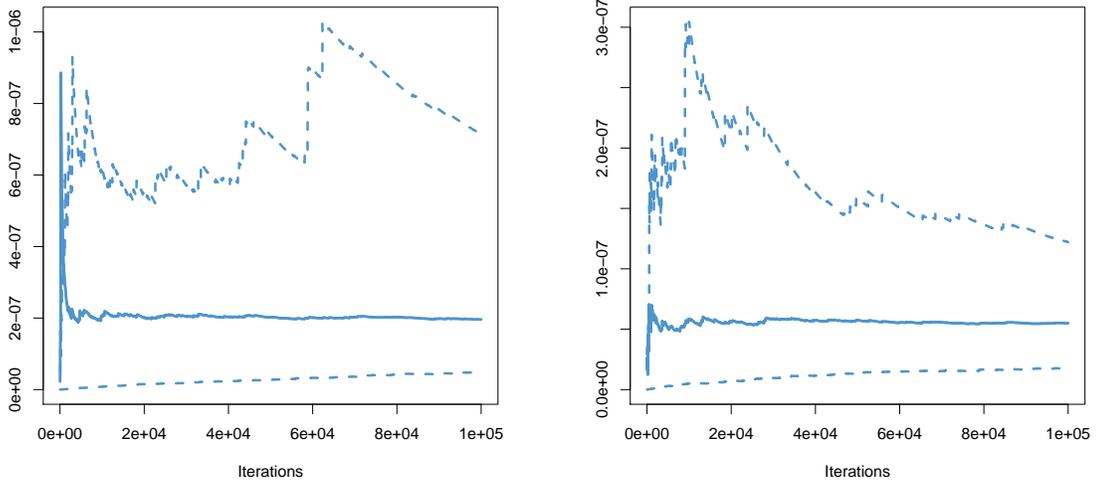


Fig. 1. Convergence plots for expression (6). Left: subject 6; Right: subject 8.

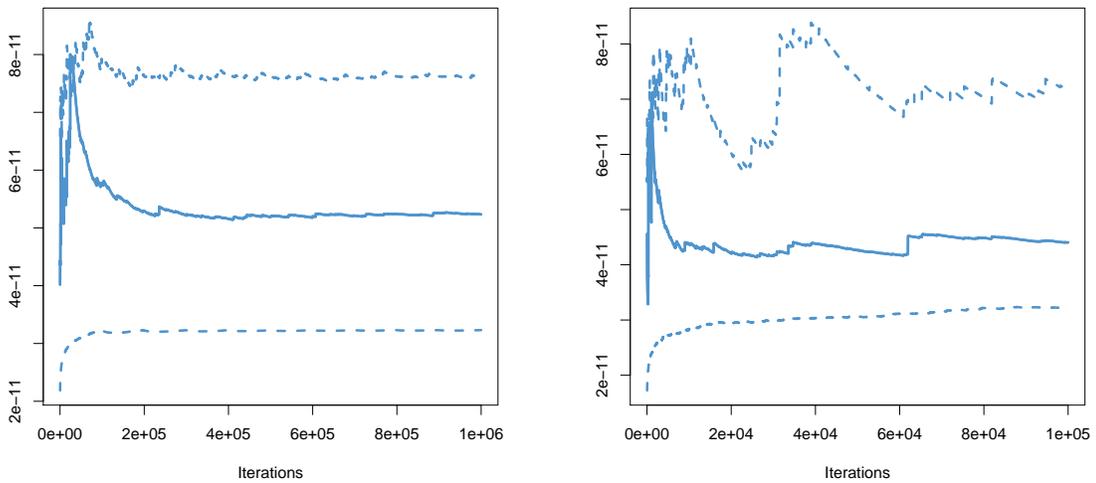


Fig. 2. Convergence plots for expression (7). Left: subject 6; Right: subject 8.

$T = 1.0 \times 10^6$) and subject 8 (right, with $t_n = 300$ and $T = 1.0 \times 10^5$) can be found in fig. 2. Again, the full line indicates $\sum_{i=1}^l P[\gamma_O | \sum_{T=R_i}] / l$, while the dashed lines denote the running empirical percentiles (.05 and .95), from bottom to top. It is clear that again convergence is reached for our estimators.

4.7 Analysis - Order of Operations

We decided to test Order of Operations and Transitivity simultaneously, for the following reason.

When we ask a subject whether a person on a picture belongs more to the set ‘middle age’ or ‘grey hair’, we can expect that the subject answers in function of his/her perception of the picture and of the labels used to represent the fuzzy sets. But when we ask a subject whether a person on a picture belongs more to the set $A \cup B$ or $A \cap B$ (whatever A and B are), the subject will probably notice that the elementary fuzzy sets (A and B) are the same in both alternatives and that the only difference is the operation (\cup or \cap). The subject may then be tempted to use some kind of logical reasoning instead of using his/her perception or knowledge. There are thus chances that the answers of a subject satisfy Order of Operations even if \succsim_T does not satisfy the condition. Besides, if we present all pairs of alternatives formed by alternatives 10–15 (see Table 1), only three pairs out of 15 are relevant for Order of Operations: (10, 13), (11, 14) and (12, 15). So, a test of Order of Operations only would be based on very little data and would not be very reliable.

If we simultaneously test Transitivity and Order of Operations, then we use not just the three above-mentioned pairs (which are not very reliable) but all pairs. Indeed, if we have $(A \cup B) \succ_O C$ and $C \succ_O (A \cap B)$, then, assuming Transitivity, we have some indirect evidence in favour of Order of Operations. Conversely, if we have $(A \cap B) \succ_O C$ and $C \succ_O (A \cup B)$, then, assuming Transitivity, we have some indirect evidence against Order of Operations.

Since we previously tested Transitivity for the 8 subjects and since they were all found to be transitive, the outcome of the simultaneous test of Order of Operations and Transitivity can be used for deciding whether the subjects satisfy Order of Operations. The two competing hypotheses in this part of the paper are thus ‘Transitivity and Order of Operations’ against ‘no Transitivity or no Order of Operations’.

The same procedure as in the test for Transitivity was used, except for the fact that we now generated weak order relations that were in agreement with Order of Operations (and with the same number of relevant unions and intersections on the same single fuzzy sets, i.e. three unions and three intersections and three unique pairs of fuzzy sets) on one hand and random relations that were not transitive or not in agreement with Order of Operations on the other.

4.8 Results - Order of Operations

The results of this analysis are shown in table 3⁸. Only for subject 4, 6 and 7 the results are in agreement with Order of Operations and Transitivity. For

⁸ In this case $\#X = 787$, which was approximated by generating random weak order relations and testing for Order of Operations, while $\#\bar{X} = 4683 - 787$, which, in this analysis, is the total number of weak orders on 6 objects minus $\#X$.

Subject	$\sum_{R \in X} P[\gamma_O \zeta_T = R]$	$\sum_{R \notin X} P[\gamma_O \zeta_T = R]$	Ratio
1.	0.0297	6.7869	0.022
2.	0.0304	0.9961	0.151
3.	0.4757	5.1674	0.456
4.	0.0617	0.0615	4.969
5.	0.4988	5.4990	0.449
6.	0.7355	0.5340	62.44
7.	1.9025	4.0389	2.332
8.	0.5140	5.3717	0.416

Table 3

Results Monte Carlo analysis for Order of Operations.

the other subjects, Transitivity or Order of Operations does not hold. Since all subjects were previously found to be transitive (Section 4.6), we thus conclude that subjects 1, 2, 3, 5 and 8 do not satisfy Order of Operations. Notice that most Bayes factors are close to 1. This indicates that our conclusions should be considered with some scepticism.

5 Conclusion

We have introduced a method that can be used for the empirical testing of any measurement-theoretic model and, in particular, of the measurement-theoretic foundations of fuzzy set theory. We applied this method to test Transitivity and Order of Operations.

The analysis of our data clearly showed that the observed relations from our experiment were in agreement with Transitivity. This does not mean that all subjects satisfy Transitivity in all circumstances. In order to draw such a conclusion, we would need much more subjects and experiments not only with pictures of people but also in other contexts. Besides, drawing such a conclusion is not our goal since differences across subjects and contexts can exist. The conclusion that we draw is just that Transitivity is a plausible condition. It is thus worthwhile studying models relying on Transitivity. But a prudent real-life application of such a model with an expert would require testing Transitivity with this expert.

The analysis of our data leads to the rejection of Order of Operations for 5 subjects out of 8. This could lead to the conclusion that a large part of the population violates Order of Operations while the other part satisfies it. Nevertheless, given the fact that the Bayes factors are close to 1 and given

the small number of alternatives in the experiment, we think more research is needed.

Since the above tested axioms are only a small part of the big heap of conditions formulated by the authors working on the measurement-theoretic foundations of fuzzy set theory (e.g. Bollman-Sdorra et al., 1993; Bilgiç and Türksen, 1995; Marchant, 2004a,b), many other parts of this pile are still waiting for an empirical validation, an undertaking we plan to do in the near future.

Let us note a drawback of our method, in that it is computationally very intensive. For example, in the case of subject 6, for the test of Transitivity (relation over 9 objects), in order to conduct the Monte Carlo analysis for expression (7) (see fig. 2, top), it took about 150 hours (program written in R, running on a Pentium 4 pc).

Another thought of consideration should be directed towards the choice of our priors, which in our case are chosen in accordance with Laplace’s principle. A reasonable choice we think, because this reflects the knowledge we have about our hypotheses before considering the data, however, we cannot say this is the ‘best’ possible knowledge. So maybe we are overlooking additional prior information, the basis on which we can construct different, but maybe better priors.

A Monte Carlo Method

In this section a description is given of a Monte Carlo procedure developed in order to estimate $\sum_{R \in X} P[\succ_O \mid \succ_T = R]$ (see expression 3, the numerator of the second fraction of the right hand term).

- (1) Initialise 3 vectors, named ‘pEstim’, ‘sumEstim’ and ‘Estim’, each of length t_n with 0 components.
- (2) Initialise a vector named ‘Average’ with length T .
- (3) Initialise counter t . Set $t = 0$.
- (4) $t \leftarrow t + 1$
- (5) Generate t_n weak order relations according to the procedure described in Maassen and Bezembinder (2002).
- (6) For each of the generated relations, compute $P[\succ_O \mid \succ_T = R_i]$ (expression (4)). Place the results in pEstim.
- (7) $\text{sumEstim} \leftarrow \text{sumEstim} + \text{pEstim}$
- (8) $\text{Estim} \leftarrow \text{sumEstim}/N$
- (9) $\text{Average}[t] \leftarrow \text{mean}(\text{Estim})$
- (10) If $t < T$, go to 4, else exit.

In this case $\text{Average}[T] \times \#X$ gives us the estimate of $\sum_{R \in X} P[\succ_O \mid \succ_{T=R}]$, the results of which can be found in the second column of table (2). To approximate the denominator term $\sum_{R \notin X} P[\succ_O \mid \succ_{T=R}]$, a similar procedure can be used except that, this time, we generate complete but intransitive relations and we multiply $\text{Average}[T]$ with $(3^{n(n-1)/2} - \#X)$.

References

- Bilgiç, T., Türksen, I.B., 1995. Measurement-theoretic justification of connectives in fuzzy set theory. *Fuzzy Sets and Systems*. 76, 289–308.
- Bilgiç, T., Türksen, I.B., 1997. Measurement-theoretic frameworks for fuzzy set theory. In: Martin, T., Ralescu, A. (Eds.), *Fuzzy logic in artificial intelligence: towards intelligent systems*. Springer, Berlin, pp. 252–265.
- Bilgiç, T., Türksen, I.B., 2000. Measurement of membership functions: theoretical and empirical work. In: Dubois, D., Prade, H. (Eds.), *Fundamentals of fuzzy sets. The Handbook of Fuzzy Sets*. Vol. 7. Kluwer Academic Press, Dordrecht, pp. 195–232.
- Bollman-Sdorra, P., Wong, S.K.M., Yao, Y.Y., 1993. A measure-theoretic axiomatization of fuzzy sets. *Fuzzy Sets and Systems*. 60, 295–307.
- Carbone, E., 1997. Investigation of stochastic preference theory using experimental data. *Economics Letters* 57, 305–311.
- Fishburn, P. C., 1991. Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty* 4, 113–134.
- Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*. 7, 457–511.
- Harless, D. W., Camerer, C. F., 1994. The predictive utility of generalized expected utility theories. *Econometrica* 62, 1251–1290.
- Iverson, G., Falmagne, J.-C., 1985. Statistical issues in measurement. *Mathematical Social Sciences*. 10, 131–153.
- Karabatsos, G., 2001. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*. 2, 389–423.
- Karabatsos, G., 2005. The exchangeable multinomial model as an approach to testing deterministic axioms of choice and measurement. *Journal of Mathematical Psychology*. 49, 51–69.
- Karabatsos, G., Sheu, C.-F., 2001. Testing measurement theory axioms with Markov Chain Monte Carlo. Paper presented at the 34th Annual Meeting of the Society for Mathematical Psychology. Brown University, Providence, Rhode Island.
- Karabatsos, G., Ullrich, J.R., 2002. Enumerating and testing conjoint measurement models. *Mathematical Social Sciences*. 43, 485–504.
- Krantz, D.H., Luce, R.D., Suppes, P., Tversky, A., 1971. *Foundations of Measurement*. Vol. 1. Academic Press, New York.

- Luce, R.D., 2000. Utility of gains and losses: measurement theoretical and experimental approaches. Lawrence Erlbaum, Mahwah, NJ.
- Maassen, H., Bezembinder, T., 2002. Generating random weak orders and the probability of a Condorcet winner. *Social Choice and Welfare*. 19, 517–532.
- Marchant, T., 2004a. The measurement of membership by comparison. *Fuzzy Sets and Systems*. 148, 157–177.
- Marchant, T., 2004b. The measurement of membership by subjective ratio estimation. *Fuzzy Sets and Systems*. 148, 179–199.
- Myung, J.I., Karabatsos, G., Iverson, G., 2005. A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*. 49, 205–225.
- Norwich, A.M., Türksen, I.B., 1982. The fundamental measurement of fuzziness. In: Yager, R. (Ed.), *Fuzzy sets and possibility theory: Recent developments*. Pergamon, Oxford.
- Roberts, F.S., 1979. *Measurement theory: with applications to decision making, utility, and the social sciences*. Addison-Wesley Pub, Co.
- Schwartz, D.G., 2000. Layman's probability theory: A calculus for reasoning with linguistic likelihood. *Information Sciences*. 126, 71–82.
- Tamburini, F., De Santis, C., Zamuner, E., 2002. Identifying phrasal connectives in Italian. Using quantitative methods, 45–64. In: Nuccorini, S. (Ed.), *Phrases and Phraseology - Data and Descriptions*. Peter Lang.
- Türksen, I.B., 1991. Measurement of membership functions and their acquisition. *Fuzzy Sets and Systems*. 40, 5–38.
- Türksen, I.B., 2006. Breakdown Of Classical Equivalences In Fuzzy Theory. *Proceedings of 5th International Symposium on Intelligent Manufacturing Systems*, May 29-31, 2006, Sakarya University. 1–8.
- Tversky, A., 1969. Intransitivity of preferences. *Psychological Review*. 76, 31–48.
- Zadeh, L.A., 1965. Fuzzy sets. *Information and Control* 8, 338–353.
- Zimmermann, H. - J. and Zysno, P., 1980. Latent connectives in human decision making. *Fuzzy Sets and Systems*. 4, 37–51.