Appl. Comput. Math., V.xx, N.xx, 20xx, pp.xx-xx

LIKERT-SCALE FUZZY UNCERTAINTY FROM A TRADITIONAL DECISION MAKING VIEWPOINT: HOW SYMMETRY HELPS EXPLAIN HUMAN DECISION MAKING (INCLUDING SEEMINGLY IRRATIONAL BEHAVIOR)

J. LORKOWSKI AND V. KREINOVICH[1]

Abstract. One of the main methods for eliciting the values of the membership function $\mu(x)$ is to use the Likert scales, i.e., to ask the user to mark his or her degree of certainty by an appropriate mark k on a scale from 0 to n and take $\mu(x) = k/n$. In this paper, we show how to describe this process in terms of the traditional decision making, and we onclude that the resulting membership degrees incorporate both probability and utility information. It is therefore not surprising that fuzzy techniques often work better than probabilistic techniques (which only take into account the probability of different outcomes). We also show how symmetry helps explain human decision making, including seemingly irrational behavior.

Keywords: Decision Theory, Bounded Rationality, Fuzzy Uncertainty, Irrational Behavior.

AMS Subject Classification: 03B52, 03E72, 91B06

1. Formulation of the Problem: Need to Reconcile Theoretical Decision Making with Empirical Decision Making and with Fuzzy Techniques

Decisions are important. One of the main objectives of science and engineering is to help people make decisions.

For example, we try to predict weather, so that people will be able to dress properly (and take an umbrella if needed), and so that if a hurricane is coming, people can evacuate. We analyze quantum effects in semi-conductors so that engineers can design better computer chips. We analyze diseases so that medical doctors can help select the best treatment, etc.

In complex situations, people need help in making their decisions. In simple situations, an average person can easily make a decision. For example, if the weather forecast predicts rain, one should take an umbrella, otherwise one should not.

In more complex situations, however, even when we know all the possible consequences of each action, it is not easy to make a decision. For example, in medicine, many treatments come with side effects: a surgery can sometimes result in a patient's death, immune system suppression can result in a infectious disease, etc. In such situations, it is not easy to compare different actions, and even skilled experts would appreciate computer-based help.

To help people make decisions, we need to analyze how people make decisions. One of the difficulties in designing computer-based systems which would help people make decisions is that to make such systems successful, we need to know what exactly people want when they make decisions. Often, people cannot explain in precise terms why exactly they have selected this or that alternative.

In such situations, we need to analyze how people actually make decisions, and then try to come up with formal descriptions which fit the observed behavior.

[1]Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA, e-mails: lorkowski@computer.org, vladik@utep.edu
Manuscript received xx.

Experiments start with decision making under full information. To analyze how people make decisions, researchers start with the simplest situations, in which we have the full information about the situation:

- we know all possible outcomes $o_1, \ldots, o_n$ of all possible actions;
- we know the exact value u_i (e.g., monetary) of each outcome o_i; and
- for each action a and to each outcome i, we know the probability $p_i(a)$ of this outcome.

Seemingly reasonable behavior. The outcome of each action a is not deterministic. For the same action, we may get different outcomes u_i with different probabilities $p_i(a)$. However, usually similar situations are repeated again and again.

If we repeat a similar situation several times, then the average expected gain of selecting an action a becomes close to the mathematical expectation of the gain, i.e., to the value

$$u(a) \stackrel{\text{def}}{=} \sum_{i=1}^{n} p_i(a) \cdot u_i.$$

Thus, we expect that a decision maker selects the action a for which this expected value $u(a)$ is the largest.

This idea is behind the traditional decision making theory; see, e.g., [6, 7, 9, 15, 23]. The traditional decision making theory has been effectively used in business and in other decision areas.

How people actually make decisions is somewhat different. In their famous experiments, the Nobelist Daniel Kahneman and his co-author Amos Tversky found out that people often deviate from the ideal (rational) behavior; see, e.g., [8] and references therein.

First problem: how can we explain this difference? How can we reconcile theoretical decision making and empirical decision making?

Fuzzy uncertainty: a usual description. In addition to applying traditional decision theory, another very successful way of making decisions under uncertainty is to use techniques based on fuzzy logic and fuzzy uncertainty.

Fuzzy logic (see, e.g., [10, 21, 29]) has been designed to describe imprecise ("fuzzy") natural language properties like "big", "small", etc. In contrast to "crisp" properties like $x \le 10$ which are either true or false, experts are not 100% sure whether a given value x is big or small. To describe such properties P, fuzzy logic proposes to assign, to each possible value x, a degree $\mu_P(x)$ to which the value x satisfies this property:

- the degree $\mu_P(x) = 1$ means that we are absolutely sure that the value x satisfies the property P;
- the degree $\mu_P(x) = 0$ means that we are absolutely sure that the value x does not satisfy the property P; and
- intermediate degrees $0 < \mu_P(x) < 1$ mean that we have *some* confidence that x satisfies the property P but we also have a certain degree of confidence that the value x does not satisfy this property.

How do we elicit the degree $\mu_P(x)$ from the expert? One of the usual ways is to use a *Likert scale*, i.e., to ask the expert to mark his or her degree of confidence that the value x satisfies the property P by one of the labels $0, 1, \ldots, n$ on a scale from 0 to n. If an expert marks m on a scale from 0 to n, then we take the ratio m/n as the desired degree $\mu_P(x)$. For example, if an expert marks her confidence by a value 7 on a scale from 0 to 10, then we take $\mu_P(x) = 7/10$.

For a fixed scale from 0 to n, we only get $n+1$ values this way: 0, $1/n$, $2/n$, $\ldots$, $(n-1)/n = 1 - 1/n$, and 1. If we want a more detailed description of the expert's uncertainty, we can use a more detailed scale, with a larger value n.

Second problem: how can we reconcile traditional decision making theory with fuzzy techniques? The traditional decision theory describes rational human behavior, it has

many practical applications. On the other hand, fuzzy techniques are also very successful in many application problems, in particular, in control and in decision making (see, e.g., [10, 21]).

It is therefore desirable to combine these two techniques, so that we would able to capitalize on the successes of both types of techniques. To enhance this combination, it is desirable to be able to describe both techniques in the same terms. In particular, it is desirable to describe fuzzy uncertainty in terms of traditional decision making.

What we do in this paper. In our opinion, one of the main reasons why we still do not have a convincing reconciliation between the traditional decision making theory and fuzzy techniques is that there is a lot of confusion and misunderstanding about such basic notions of traditional decision theory as utility, subjective probability, etc. – just like many decision making researchers have misunderstandings about fuzzy techniques.

Because of this, we start, in Section 2, with providing a brief overview of the traditional decision theory and its main concepts. A short Section 3 describes the main idea behind the symmetry approach. In Section 4, we describe the Likert scale techniques in these terms, and we show that, when we apply the idea of symmetry to this description, then the traditional decision making theory leads exactly to formulas normally used in fuzzy logic. In Section 5, we show that similar ideas of symmetry can explain another example of seemingly irrational behavior – so-called compromise effect.

In Section 6, a similar explanation is provided for empirical decision weights discovered by Kahneman and Tversky; this explanation is closely related to our interpretation of fuzzy techniques. Finally, in Section 7, we return to the Likert-scale fuzzy uncertainty. Instead of simply using the decisions made by experts (as we did in Section 4), this time, we use the traditional decision making theory to analyze why people select this or that mark on a Likert scale. Our conclusion is that the resulting membership degrees incorporate both probability and utility information. It is therefore not surprising that fuzzy techniques often work better than probabilistic techniques – which only take into account the probability of different outcomes.

Comment. Some results from this paper were first presented at major fuzzy conferences [12, 13, 14]; several other results appear in this paper for the first time.

2. Traditional Decision Theory and Its Main Concepts: A Brief Overview

Main assumption behind the traditional decision theory. Traditional approach to decision making is based on an assumption that for each two alternatives A' and A'', a user can tell:

- whether the first alternative is better for him/her; we will denote this by $A'' < A'$;
- or the second alternative is better; we will denote this by $A' < A''$;
- or the two given alternatives are of equal value to the user; we will denote this by $A' = A''$.

Towards a numerical description of preferences: the notion of utility. Under the above assumption, we can form a natural numerical scale for describing preferences. Namely, let us select a very bad alternative A_0 and a very good alternative A_1. Then, most other alternatives are better than A_0 but worse than A_1.

For every probability $p \in [0, 1]$, we can form a lottery $L(p)$ in which we get A_1 with probability p and A_0 with probability $1 - p$.

- When $p = 0$, this lottery coincides with the alternative A_0: $L(0) = A_0$.
- When $p = 1$, this lottery coincides with the alternative A_1: $L(1) = A_1$.

For values p between 0 and 1, the lottery is better than A_0 and worse than A_1. The larger the probability p of the positive outcome increases, the better the result:

$$p' < p'' \text{ implies } L(p') < L(p'').$$

Thus, we have a continuous scale of alternatives $L(p)$ that monotonically goes from $L(0) = A_0$ to $L(1) = A_1$. We will use this scale to gauge the attractiveness of each alternative A.

Due to the above monotonicity, when p increases, we first have $L(p) < A$, then we have $L(p) > A$, and there is a threshold separating values p for which $L(p) < A$ from the values p for which $L(p) > A$. This threshold value is called the *utility* of the alternative A:

$$u(A) \stackrel{\text{def}}{=} \sup\{p : L(p) < A\} = \inf\{p : L(p) > A\}.$$

Then, for every $\varepsilon > 0$, we have

$$L(u(A) - \varepsilon) < A < L(u(A) + \varepsilon).$$

We will describe such (almost) equivalence by $\equiv$, i.e., we will write that $A \equiv L(u(A))$.

How to elicit the utility from a user: a fast iterative process. Initially, we know the values $\underline{u} = 0$ and $\overline{u} = 1$ such that $A \equiv L(u(A))$ for some $u(A) \in [\underline{u}, \overline{u}]$.

On each stage of this iterative process, once we know values $\underline{u}$ and $\overline{u}$ for which $u(A) \in [\underline{u}, \overline{u}]$, we compute the midpoint u_{mid} of the interval $[\underline{u}, \overline{u}]$ and ask the user to compare A with the lottery $L(u_{\text{mid}})$ corresponding to this midpoint. There are two possible outcomes of this comparison: $A \le L(u_{\text{mid}})$ and $L(u_{\text{mid}}) \le A$.

- In the first case, the comparison $A \le L(u_{\text{mid}})$ means that $u(A) \le u_{\text{mid}}$, so we can conclude that $u \in [\underline{u}, u_{\text{mid}}]$.
- In the second case, the comparison $L(u_{\text{mid}}) \le A$ means that $u_{\text{mid}} \le u(A)$, so we can conclude that $u \in [u_{\text{mid}}, \overline{u}]$.

In both cases, after an iteration, we decrease the width of the interval $[\underline{u}, \overline{u}]$ by half. So, after k iterations, we get an interval of width 2^{-k} which contains $u(A)$ – i.e., we get $u(A)$ with accuracy 2^{-k}.

How to make a decision based on utility values. Suppose that we have found the utilities $u(A')$, $u(A'')$, ..., of the alternatives A', A'', ... Which of these alternatives should we choose?

By definition of utility, we have:

- $A \equiv L(u(A))$ for every alternative A, and
- $L(p') < L(p'')$ if and only if $p' < p''$.

We can thus conclude that A' is preferable to A'' if and only if $u(A') > u(A'')$. In other words, we should always select an alternative with the largest possible value of utility. So, to find the best solution, we must solve the corresponding optimization problem.

Before we go further: caution. We are *not* claiming that people estimate probabilities when they make decisions: we know they often don't. Our claim is that when people make *definite* and *consistent* choices, these choices *can* be described by probabilities. (Similarly, a falling rock does not solve equations but follows Newton's equations $ma = m\dfrac{d^2x}{dt^2} = -mg$.) In practice, decisions are often *not* definite (uncertain) and *not* consistent.

How to estimate utility of an action. For each action, we usually know possible outcomes $S_1, \ldots, S_n$. We can often estimate the probabilities $p_1, \ldots, p_n$ of these outcomes.

By definition of utility, each situation S_i is equivalent to a lottery $L(u(S_i))$ in which we get:

- A_1 with probability $u(S_i)$ and
- A_0 with the remaining probability $1 - u(S_i)$.

Thus, the original action is equivalent to a complex lottery in which:

- first, we select one of the situations S_i with probability p_i: $P(S_i) = p_i$;
- then, depending on S_i, we get A_1 with probability $P(A_1 \,|\, S_i) = u(S_i)$ and A_0 with probability $1 - u(S_i)$.

The probability of getting A_1 in this complex lottery is:

$$P(A_1) = \sum_{i=1}^{n} P(A_1 \,|\, S_i) \cdot P(S_i) = \sum_{i=1}^{n} u(S_i) \cdot p_i.$$

In this complex lottery, we get:

- A_1 with probability $u = \sum\limits_{i=1}^{n} p_i \cdot u(S_i)$, and
- A_0 with probability $1 - u$.

So, the utility of the complex action is equal to the sum u.

From the mathematical viewpoint, the sum defining u coincides with the expected value of the utility of an outcome. Thus, selecting the action with the largest utility means that we should select the action with the largest value of expected utility $u = \sum p_i \cdot u(S_i)$.

Subjective probabilities. In practice, we often do not know the probabilities p_i of different outcomes. How can we gauge our subjective impressions about these probabilities?

For each event E, a natural way to estimate its subjective probability is to fix a prize (e.g., \$1) and compare:

- a lottery ℓ_E in which we get the fixed prize if the event E occurs and 0 is it does not occur, with
- a lottery $\ell(p)$ in which we get the same amount with probability p.

Here, similarly to the utility case, we get a value $ps(E)$ for which, for every $\varepsilon > 0$:

$$\ell(ps(E) - \varepsilon) < \ell_E < \ell(ps(E) + \varepsilon).$$

Then, the utility of an action with possible outcomes $S_1, \ldots, S_n$ is equal to $u = \sum\limits_{i=1}^{n} ps(E_i) \cdot u(S_i)$.

Auxiliary issue: almost-uniqueness of utility. The above definition of utility u depends on the selection of two fixed alternatives A_0 and A_1. What if we use different alternatives A_0' and A_1'? How will the new utility u' be related to the original utility u?

By definition of utility, every alternative A is equivalent to a lottery $L(u(A))$ in which we get A_1 with probability $u(A)$ and A_0 with probability $1 - u(A)$. For simplicity, let us assume that $A_0' < A_0 < A_1 < A_1'$. Then, for the utility u', we get $A_0 \equiv L'(u'(A_0))$ and $A_1 \equiv L'(u'(A_1))$. So, the alternative A is equivalent to a complex lottery in which:

- we select A_1 with probability $u(A)$ and A_0 with probability $1 - u(A)$;
- depending on which of the two alternatives A_i we get, we get A_1' with probability $u'(A_i)$ and A_0' with probability $1 - u'(A_i)$.

In this complex lottery, we get A_1' with probability $u'(A) = u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0)$. Thus, the utility $u'(A)$ is related with the utility $u(A)$ by a linear transformation $u' = a \cdot u + b$, with $a > 0$.

Traditional approach summarized. We assume that

- we know possible actions, and
- we know the exact consequences of each action.

Then, we should select an action with the largest value of expected utility.

3. Symmetry Idea: A Brief Description

General idea. In many, what is often helpful is a *symmetry approach*.

The main idea behind this approach is that if the situation is invariant with respect to some natural symmetries, then it is reasonable to select an action which is also invariant with respect to all these symmetries.

Laplace Indeterminacy Principle as an example of the symmetry approach. Let us assume that we have n possible alternatives A_1, A_2, ..., A_n, and we have no information about the probabilities of different alternatives. To make decisions, we need to assign subjective probabilities $ps(A_i)$ to these alternatives – in such a way that $\sum_{i=1}^{n} ps(A_i) = 1$.

In this case, since we have no reason to prefer one alternative to another, we can conclude that this situation has natural symmetries: namely, permutations, such as a permutation that simply swaps two alternatives A_i and A_j.

The symmetry approach implies that the subjective probabilities $ps(A_i)$ should not change under these perturbations. In particular, from the fact that the subjective probabilities should not change after the swap between A_i and A_j, we can conclude that $ps(A_i) = ps(A_j)$ for every two alternatives A_i and A_j.

Since all n subjective probabilities are equal to each other, and their sum is equal to 1, thus each of these probabilities is equal to $ps(A_i) = \dfrac{1}{n}$. This conclusion is known (and used) for many centuries. It was first explicitly formulated by P.-S. Laplace and is thus known as Laplace Indeterminacy Principle.

Thus, this known principle is indeed an example of the symmetry approach.

The symmetry approach has been helpful in dealing with uncertainty. There have been many applications of this approach. In particular, it has been shown that for many empirically successful techniques related to neural networks, fuzzy logic, and interval computations, their empirical success can be explained by the fact that these techniques can be deduced from the appropriate symmetries; see, e.g., [18]. In particular, this explains the use of a sigmoid activation function $s(z) = \dfrac{1}{1+\exp(-z)}$ in neural networks, the use of the most efficient t-norms and t-conorms in fuzzy logic, etc.

4. How to Transform Labels from a Likert Scale into Uncertainty Degrees: Fuzzy Technique Reconciled with the Traditional Decision Making Approach

Transforming labels on a Likert scale into uncertainty degrees: formulation of the problem. A Likert scale is a sorted list of labels. Let us denote the number of labels on this list by n. In these terms, e.g., a Likert scale for uncertainty is a list of n labels sorted in the increased order of certainty:

- the first label corresponds to the smallest certainty,
- the second label corresponds to the second smallest certainty,
- etc.,
- until we reach the last label which corresponds to the largest certainty.

According to the traditional decision making theory approach, we need to describe these levels of certainty by numerical degrees (what the traditional decision making theory calls *subjective probabilities*). In other words, for each i from 1 to n, we need to assign, to each i-th label, a value p_i in such a way that labels corresponding to higher certainty should get larger numbers: $p_1 < p_2 < \ldots < p_n$.

How can we do it?

Technical comment: from exact real numbers to computer-representable real numbers. One of the main objectives of assigning numerical values to different degrees is the same objective that motivated fuzzy logic in the first place:

- we want computers to help us solve the corresponding decision problems, and
- computers are not very good in dealing with labels on a scale; their natural language is the language of numbers.

From this viewpoint, it makes sense to consider not all theoretically possible exact real numbers, but only computer-representable real numbers.

In a computer, real numbers from the interval $[0, 1]$ are usually represented by the first d digits of their binary expansion. Thus, computer-representable numbers are 0, $h \stackrel{\text{def}}{=} 2^{-d}$, $2h$, $3h$, ..., until we reach the value $2^d \cdot h = 1$.

In our analysis, we will assume that the "machine unit" $h > 0$ is fixed, and we will this assume that only multiples of this machine units are possible values of all n probabilities p_i.

Illustrative example. When $h = 0.1$, each probability p_i takes 11 possible values: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0.

Comment. In the modern computers, the value h is extremely small; thus, whenever necessary, we can assume that $h \approx 0$ – i.e., use limit case of $h \to 0$ instead of the actual small "machine unit" h.

Possible combinations of probabilities. For each h, we consider all possible combinations of probabilities $p_1 < \ldots < p_n$ in which all the numbers p_i are proportional to the selected step h, i.e., all possible combinations of values $(k_1 \cdot h, \ldots, k_n \cdot h)$ with $k_1 < \ldots < k_n$.

Example. When $n = 2$ and $h = 0.1$, we consider all possible combinations of values $(k_1 \cdot h, k_2 \cdot h)$ with $k_1 < k_2$:

- For $k_1 = 0$ and $p_1 = 0$, we have 10 possible combinations $(0, 0.1)$, $(0, 0.2)$, ..., $(0, 1)$.
- For $k_1 = 1$ and $p_1 = 0.1$, we have 9 possible combinations $(0.1, 0.2)$, $(0.1, 0.3)$, ..., $(0.1, 1)$.
- ...
- Finally, for $k_1 = 9$ and $p_1 = 0.9$, we have only one possible combination $(0.9, 1)$.

How to assign probabilities: an approach based on the traditional decision theory. As we have describes in Section 2, in the traditional decision theory, subjective probabilities are defined, in effect, as utilities of the corresponding lotteries.

For each i, for different possible combinations $(p_1, \ldots, p_n)$, we get different value of this utility p_i. According to the traditional decision theory, the proper utility value P_i to assign to the i-th label should be the *expected utility*, i.e., the weighted combination of the values p_i assigned by different combinations, with the probability of each combination as a weight.

So, to find the desired values P_i, we need to assign probabilities to different combinations $(p_1, \ldots, p_n)$.

How to assign probabilities to different combinations: let us use symmetry. Since we have no reason to believe that some combinations $(p_1, \ldots, p_n)$ are more probable and some are less probable, it is thus reasonable to assume that all these combinations are equally probable – this is exactly what the symmetry idea (as described in Section 3) prescribes in such cases.

Let us use this assumption to estimate P_i.

Example. For $n = 2$ and $h = 0.1$, we thus estimate P_1 by taking an arithmetic average of the values p_1 corresponding to all possible pairs. Specifically, we average:

- ten values $p_1 = 0$ corresponding to ten pairs $(0, 0.1)$, ..., $(0, 1)$;
- none values $p_1 = 0.1$ corresponding to nine pairs $(0.1, 0.2)$, ..., $(0.1, 1)$;
- ...
- and a single value $p_1 = 0.9$ corresponding to the single pair $(0.9, 1)$.

As a result, we get the value

$$P_1 = \frac{10 \cdot 0.0 + 0 \cdot 0.1 + \ldots + 1 \cdot 0.9}{10 + 9 + \ldots + 1} = \frac{16.5}{55} = 0.3.$$

Similarly, to get the value p_2, we average:

- a single value $p_2 = 0.1$ corresponding to the single pair $(0, 0.1)$;

- two values $p_2 = 0.2$ corresponding to two pairs $(0, 0.2)$ and $(0.1, 0.2)$;
- ...
- ten values $p_2 = 1.0$ corresponding to ten pairs $(0, 1)$, ..., $(0.9, 1)$.

As a result, we get the value

$$P_2 = \frac{1 \cdot 0.1 + 2 \cdot 0.2 + \ldots + 10 \cdot 1.0}{1 + 2 + \ldots + 10} = \frac{37.5}{55} = 0.7.$$

Estimating P_i: towards a precise formulation of the problem. The probability p_i of each label can take any of the equidistant values 0, h, $2h$, $3h$, ..., with equal probability. In the limit $h \to 0$, the resulting probability distribution tends to the uniform distribution on the interval $[0, 1]$.

In this limit $h \to 0$, we get the following problem:

- we start with n independent random variable $v_1, \ldots, v_n$ which are uniformly distributed on the interval $[0, 1]$;
- we then need to find, for each i, the conditional expected value $E[v_i \,|\, v_1 < \ldots < v_n]$ of each variable v_i under the condition that the values v_i are sorted in increasing order.

Estimating P_i: solving the problem. Conditional expected values are usually more difficult to compute than unconditional ones. So, to solve our problem, let us reduce our problem to the problem of computing the unconditional expectation.

Let us consider n independent random variables each of which is uniformly distributed on the interval $[0, 1]$. One can easily check that for any two such variables v_i and p_j, the probability that they are equal to each other is 0. Thus, without losing generality, we can safely assume that all n random values are different. Therefore, the whole range $[0, 1]^n$ is divided into $n!$ sub-ranges corresponding to different orders between v_i. Each sub-range can be reduced to the sub-range corresponding to $v_1 < \ldots < v_n$ by an appropriate permutation in which v_1 is swapped with the smallest $v_{(1)}$ of n values, v_2 is swapped with the second smallest $v_{(2)}$, etc.

Thus, the conditional expected value of v_i is equal to the (unconditional) expected value of the i-th value $v_{(i)}$ in the increasing order. This value $v_{(i)}$ is known as an *order statistic*, and for uniform distributions, the expected values of all order statistics are known (see, e.g., [1, 2, 5]): $P_i = \dfrac{i}{n+1}$.

Resulting formula. If all we know is that our degree of certainty is expressed by i-th label on an n-label Likert scale, then we should assign, to this case, the degree of certainty $p_i = \dfrac{i}{n+1}$.

Relation with fuzzy logic. For the scale from 0 to m, the total number of labels is $n = m+1$, and each label j is the $(j+1)$-st in increasing order. Thus, we should assign, to this label j, the degree $p_j = \dfrac{j+1}{m+2}$.

This value is very close to the usual assignment $p_j = \dfrac{j}{m}$, especially when m is large. Thus, the usual fuzzy assignment can indeed be justified by the traditional decision theory approach.

5. Symmetry Approach Helps Explain Seemingly Irrational Human Behavior: A Case Study

Customers make decisions. A customer shopping for an item usually has several choices. Some of these choices have better quality, leading to more possibilities, etc. – but are, on the other hand, more expensive. For example, a customer shopping for a photo camera has plenty of choices ranging from the cheapest one, whose photos are good, to very professional cameras enabling the user to make highest-quality photos, even under complex circumstances. A person planning to spend a night at a different city has a choice from the cheapest motels which provide

a place to sleep to luxurious hotels providing all kinds of comfort, etc. A customer selects one of the alternatives by taking into account the additional advantages of more expensive choices versus the need to pay more money for these choices.

It is important to understand customer's decisions. Whether we are motivated by a noble goal of providing alternatives which are the best for the customers – or whether a company wants to make more money by providing what is wanted by the customers – it is important to understand how customers make decisions.

Experimental studies. In many real-life situations, customers face numerous choices. As usual in science, a good way to understand complex phenomena is to start by analyzing the simplest cases. In line with this reasoning, researchers provided customers with two alternatives and recorded which of these two alternatives a customer selected. In many particular cases, these experiments helped better understand the customer's selections – and sometimes even predict customer selections.

At first glance, it seems like such pair-wise comparisons are all we need to know: if a customer faces several choices a_1, a_2, ..., a_n, then a customer will select an alternative a_i if and only if this alternative is better in pair-wise comparisons that all other possible choices. To confirm this common-sense idea, in the 1990s, several researchers asked the customers to select one of the three randomly selected alternatives.

What was expected. The experimenters expected that since the three alternatives were selected at random, a customers would:

- sometimes select the cheapest of the three alternative (of lowest quality of all three),
- sometimes select the intermediate alternative (or intermediate quality), and
- sometimes select the most expensive of the three alternatives (of highest quality of all three).

What was observed. Contrary to the expectations, the experimenters observed that in the overwhelming majority of cases, customers selected the intermediate alternative; see, e.g., [24, 25, 28]. In all these cases, the customer selected an alternative which provided a compromise between the quality and cost; because of this, this phenomenon was named *compromise effect*.

Why is this irrational? At first glance, selecting the middle alternative is reasonable. However, it is not.

For example, let us assume that we have four alternative $a_1 < a_2 < a_3 < a_4$ ordered in the increasing order of price and at the same time, increasing order of quality. Then:

- if we present the user with three choices $a_1 < a_2 < a_3$, in most cases, the user will select the middle choice a_2; this means, in particular, that, to the user, a_2 better than the alternative a_3;
- on the other hand, if we present the user with three other choices $a_2 < a_3 < a_4$, in most cases, the same user will select the middle choice a_3; but this means that, to the user, the alternative a_3 better than the alternative a_2.

If in a pair-wise comparison, a_2 is better, then the second choice is wrong. If in a pair-wise comparison, the alternative a_3 is better, then the first choice is wrong. In both cases, one of the two choices is irrational.

This is not just an experimental curiosity, customers' decisions have been manipulated this way. At first glance, the above phenomena may seem like one of optical illusions or logical paradoxes: interesting but not that critically important. Actually, it is serious and important, since, according to anecdotal evidence, many companies have tried to use this phenomenon to manipulate the customer's choices: to make the customer buy a more expensive product.

For example, if there are two possible types of a certain product, a company can make sure that most customers select the most expensive type – simply by offering, as the third option, an even more expensive type of the same product.

Manipulation possibility has been exaggerated. Recent research shows that manipulation is not very easy: the compromise effect only happens when a customer has no additional information – and no time (or no desire) to collect such information. In situations when customers were given access to additional information, they selected – as expected from rational folks – one of the three alternatives with almost equal frequency, and their pairwise selections, in most cases, did not depend on the presence of any other alternatives; see, e.g., [27].

Compromise effect: mystery remains. The new experiment shows that the compromise effect is not as critical and not as wide-spread as it was previously believed. However, in situation when decisions need to be made under major uncertainty, this effect is clearly present – and its seemingly counterintuitive, inconsistent nature is puzzling.

How can we explain such a seemingly irrational behavior?

It is possible to find a rational explanation for such a behavior. In this section, we use symmetry to show that it is possible to find a rational explanation for such a behavior.

Comment. This section focuses on *one* specific example of a seemingly irrational behavior. We should emphasize, however, that there are *many* well-known examples of such behavior; see, e.g., [8] and references therein. These examples cover both seemingly irrational individual choices and seemingly irrational group choices.

In some cases, a seemingly irrational behavior can be explained in rational terms – sometimes, by using fuzzy techniques; see, e.g., [12, 13]. In particular, many seeming paradoxes related to group decision making – paradoxes related to Arrow's impossibility result – can be explained if instead of simply recording which participant prefers which alternative, we also take into account the degree to which to each participants prefer one alternative to another (see, e.g., [17]) – in line with Zadeh's idea pioneered in [3, 4, 30, 31, 32, 33].

It should be also mentioned that many instances of human behaviors are indeed irrational – in the sense that humans sometimes select actions which are detrimental to the decision maker's own interests. There are many examples of such actions, from irrational decision making in simple economic situations to self-damaging behavior related to drug and alcohol addiction.

For purposes of this explanation, we concentrate only on *one* specific case of seemingly irrational human behavior.

Description of the situation. We have three alternative a, a' and a'':

- the alternative a is the cheapest – and is, correspondingly, of the lowest quality among the give there alternatives;
- the alternative a' is intermediate in terms of price – and is, correspondingly, intermediate in terms of quality;
- finally, the alternative a'' is the most expensive – and is, correspondingly, of the highest quality among the give there alternatives.

What do we know about the utility of each alternative. The utility of each alternatives comes from two factors:

- the first factor comes from the quality: the higher the quality, the better – i.e., larger the corresponding component u_1 of the utility;
- the second factor comes from price: the lower the price, the better for the user – i.e., the larger the corresponding component u_2 of the utility.

In the first experiments which established the compromise effect, the users did not have enough time and/or information to find the corresponding utility values u_i, u_i', and u_i'' corresponding to different alternatives. Also, we do not know how, for each alternative, the corresponding

components u_1 and u_2 are combined into a single utility value characterizing this alternative – we do not even know which of the two components is more important.

Since we do not know how utility components are combined, a reasonable way to represent each alternative is by assigning to it a pair consisting of the two component utilities:

- to the alternative a, we assign the pair of values (u_1, u_2);
- to the alternative a', we assign the pair of values (u_1', u_2'); and
- to the alternative a'', we assign the pair (u_1'', u_2'').

We do not know the actual values of the component utilities, all we know is the relative order of the corresponding values: namely, we know that $u_1 < u_1' < u_1''$ and $u_2'' < u_2' < u_2$. Since we do not know the actual values of each utility component, the only we know about each of these values is whether this value is:

- the lowest of the three values; we will denote such a value by L;
- the intermediate (median) value; we will denote such a value by M; and
- the highest of the three values; we will denote such a value by H.

In these terms, we have:

- for the first utility component, $u_1 = L$, $u_1' = M$, and $u_1'' = H$;
- for the second utility component: $u_2 = H$, $u_2' = M$, and $u_2'' = L$.

In these terms, the above description of each alternative by the corresponding pair of utility values takes the following form:

- the alternative a is characterized by the pair (L, H);
- the alternative a' is characterized by the pair (M, M); and
- the alternative a'' is characterized by the pair (H, L).

Natural transformations and natural symmetries. As we have mentioned, we do not know a priori which of the utility components is more important. As a result, it is reasonable to treat both components equally. So, swapping the two components is a reasonable transformation, in the sense that we should select the same of three alternatives before and after swap:

- if we are selecting an alternative based on the pairs (L, H), (M, M), and (H, L),
- then we should select the exact same alternative if the pairs were swapped, i.e., if:
  - the alternative a was characterized by the pair (H, L);
  - the alternative a' was characterized by the pair (M, M); and
  - the alternative a'' was characterized by the pair (L, H).

Similarly, there is no reason to a priori prefer one alternative versus the other. So, the selection should not depend on which of the alternatives we mark as a, which we mark as a', and which we mark as a''. In other words, any permutation of the three alternatives is a reasonable transformation. For example, if, in our case, we select an alternative a which is characterized by the pair (L, H), then, after we swap a and a'' and get the choice of the following three alternatives:

- the alternative a which is characterized by the pair (H, L);
- the alternative a' is characterized by the pair (M, M); and
- the alternative a'' is characterized by the pair (L, H),

then we should select the same alternative – which is now denoted by a''.

What can be conclude based on these symmetries. Now, we can observe the following: that if we *both* swap u_1 and u_2 *and* swap a and a'', then you get the exact same characterization of all alternatives:

- the alternative a is still characterized by the pair (L, H);
- the alternative a' is still characterized by the pair (M, M); and
- the alternative a'' is still characterized by the pair (H, L).

The only difference is that:

- now, a indicates an alternative which was previously denoted by a'', and
- a'' now denotes the alternative which was previously denoted by a.

As we have mentioned, it is reasonable to conclude that:

- if in the original triple selection, we select the alternative a,
- then in the new selection – which is based on the exact same pairs of utility values – we should also select an alternative denoted by a.

But this "new" alternative a is nothing else but the old a''. So, we conclude that:

- if we selected a,
- then we should have selected a different alternative a'' in the original problem.

This is clearly a contradiction:

- we started by assuming that, to the user a was better than a'' (because otherwise a would not have been selected in the first place), and
- we ended up concluding that to the same user, the original alternative a'' is better than a.

This contradiction shows that, under the symmetry approach, we cannot prefer a.

Similarly:

- if in the original problem, we preferred an alternative a'',
- then this would mean that in the new problem, we should still select an alternative which marked by a''.

But this "new" a'' is nothing else but the old a. So, this means that:

- if we originally selected a'',
- then we should have selected a different alternative a in the original problem.

This is also a contradiction:

- we started by assuming that, to the user a'' was better than a (because otherwise a'' would not have been selected in the first place), and
- we ended up concluding that to the same user, the original alternative a is better than a''. This contradiction shows that, under the symmetry approach, we cannot prefer a''.

We thus conclude that out of the three alternatives a, a', and a'':

- we cannot select a, and
- we cannot select a''.

This leaves us only once choice: to select the intermediate alternative a'. This is exactly the compromise effect that we planned to explain.

Conclusion. Experiments show when people are presented with three choices $a < a' < a''$ of increasing price and increasing quality, and they do not have detailed information about these choices, then in the overwhelming majority of cases, they select the intermediate alternative a'.

This "compromise effect" is, at first glance, irrational: selecting a' means that, to the user, a' is better than a'', but in a similar situation when the user is presented with $a' < a'' < a'''$, the same principle would indicate that the user will select a'' – meaning that a'' is better than a'.

Somewhat surprisingly, a natural symmetry approach explains this seemingly irrational behavior.

6. How to Explain Kahmenan and Tversky's Empirical Decision Weights

How people actually make decisions. Based on the arguments of the traditional decision making theory, we expect that a decision maker selects the action a for which this expected value $u(a)$ is the largest.

However, in their famous experiments, the Nobelist Daniel Kahneman and his co-author Amos Tversky found out that a much more accurate description of human decision making can be obtained if we assume that, instead of maximizing the expected gain, people maximize a

weighted gain, with weights determined by the corresponding probabilities; see, e.g., [8] and references therein.

In other words, people select the action a for which the weighted gain

$$w(a) \stackrel{\text{def}}{=} \sum_i w_i(a) \cdot u_i$$

attains the largest possible value, where $w_i(a) = f(p_i(a))$ for an appropriate function $f(x)$.

This empirical transformation $f(x)$ from probabilities to weights takes the following form:

probability	0	1	2	5	10	20	50	80	90	95	98	99	100
weight	0	5.5	8.1	13.2	18.6	26.1	42.1	60.1	71.2	79.3	87.1	91.2	100

How can we explain this empirical transformation? There are qualitative explanations for this phenomenon, but not the quantitative one.

So, we propose a quantitative explanation – related to the fuzzy logic ideas.

Main idea. The main idea behind our explanation is based on the fact that when people make decisions, they do not estimate probabilities as numbers from the interval $[0, 1]$ and do not process them. If a person is asked about the probability of a certain event, in many cases, the answer will not come as an exact number, it will most probably come as an imprecise ("fuzzy") word, like "low", "high", "medium", etc.; see, e.g., [10, 21, 29].

In other words, instead of using all *infinitely* many possible real numbers from the interval $[0, 1]$, we only use *finitely* many possible values – i.e., in effect, we estimate the probability on a finite Likert scale. The reason for this discretization is that if the two probability values are close to each other, intuitively, we do not feel the difference. For example, there is a clear different between 10% chances of rain or 50% chances of rain, but we do not think that anyone can feel the difference between 50% and 51% chances. So, the discrete scale is formed by probabilities which are distinguishable from each other. Let us show how this idea can be formalized.

Comment. In this formalization, we will follow ideas first outlined in [19].

How to formalize when probabilities are distinguishable. Probability of an event is estimated, from observations, as the frequency with which this event occurs. For example, if out of 100 days of observation, rain occurred in 40 of these days, then we estimate the probability of rain as 40%. In general, if out of n observations, the event was observed in m of them, we estimate the probability as the ratio $\dfrac{m}{n}$.

This ratio is, in general, different from the actual (unknown) probability. For example, if we take a fair coin, for which the probability of head is exactly 50%, and flip it 100 times, we may get 50 heads, but we may also get 47 heads, 52 heads, etc.

It is known (see, e.g., [26]), that the expected value of the frequency is equal to p, and that the standard deviation of this frequency is equal to $\sigma = \sqrt{\dfrac{p \cdot (1-p)}{n}}$. It is also known that, due to the Central Limit Theorem, for large n, the distribution of frequency is very close to the normal distribution (with the corresponding mean p and standard deviation σ).

For normal distribution, we know that with a high certainty all the values are located within 2 to 3 standard deviations from the mean, i.e., in our case, within the interval

$$(p - k_0 \cdot \sigma, p + k_0 \cdot \sigma),$$

where $k_0 = 2$ or $k_0 = 3$: for example, for $k_0 = 3$, this is true with confidence 99.9%. We can thus say that the two values of probability p and p' are (definitely) distinguishable if the corresponding intervals of possible values of frequency do not intersect – and thus, we can distinguish between these two probabilities just by observing the corresponding frequencies.

In precise terms, the probabilities $p < p'$ are distinguishable if

$$(p - k_0 \cdot \sigma, p + k_0 \cdot \sigma) \cap (p' - k_0 \cdot \sigma', p + k_0 \cdot \sigma') = \emptyset,$$

where

$$\sigma' \stackrel{\text{def}}{=} \sqrt{\frac{p' \cdot (1-p')}{n}},$$

i.e., if $p' - k_0 \cdot \sigma' \geq p + k_0 \cdot \sigma$. The smaller p', the smaller the difference $p' - k_0 \cdot \sigma'$. Thus, for a given probability p, the next distinguishable value p' is the one for which

$$p' - k_0 \cdot \sigma' = p + k_0 \cdot \sigma.$$

When n is large, these value p and p' are close to each other; therefore, $\sigma' \approx \sigma$. Substituting an approximate value σ instead of σ' into the above equality, we conclude that

$$p' \approx p + 2k_0 \cdot \sigma = p + 2k_0 \cdot \frac{p \cdot (1-p)}{n}.$$

If the value p corresponds to the i-th level out of m – i.e., in fuzzy terms, corresponds to the truth value $\frac{i}{m}$, then the next value p' corresponds to the $(i+1)$-st level, i.e., to the truth value $\frac{i+1}{m}$.

Let $g(p)$ denote the fuzzy truth value corresponding to the probability p. Then, $g(p) = \frac{i}{m}$ and $g(p') = \frac{i+1}{m}$. Since the values p and p' are close, the difference $p' - p$ is small, and therefore, we can expand the expression $g(p') = g(p + (p' - p))$ in Taylor series and keep only linear terms in this expansion: $g(p') \approx g(p) + (p' - p) \cdot g'(p)$, where $g'(p) = \frac{dg}{dp}$ denotes the derivative of the function $g(p)$. Thus,

$$g(p') - g(p) = \frac{1}{m} = (p' - p) \cdot g'(p).$$

Substituting the known expression for $p' - p$ into this formula, we conclude that

$$\frac{1}{m} = 2k_0 \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \cdot g'(p).$$

This can be rewritten as

$$g'(p) \cdot \sqrt{p \cdot (1-p)} = \text{const}$$

for some constant, and thus,

$$g'(p) = \text{const} \cdot \frac{1}{\sqrt{p \cdot (1-p)}}.$$

Integrating this expression and taking into account that $p = 0$ corresponds to the lowest 0-th level – i.e., that $g(0) = 0$ – we conclude that

$$g(p) = \text{const} \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1-q)}}.$$

This integral can be easily computed if introduce a new variable t for which $q = \sin^2(t)$. In this case,

$$dq = 2 \cdot \sin(t) \cdot \cos(t) \cdot dt,$$

$1 - p = 1 - \sin^2(t) = \cos^2(t)$ and therefore,

$$\sqrt{p \cdot (1-p)} = \sqrt{\sin^2(t) \cdot \cos^2(t)} = \sin(t) \cdot \cos(t).$$

The lower bound $q = 0$ corresponds to $t = 0$ and the upper bound $q = p$ corresponds to the value t_0 for which $\sin^2(t_0) = p$ – i.e., $\sin(t_0) = \sqrt{p}$ and $t_0 = \arcsin\left(\sqrt{p}\right)$. Therefore,

$$g(p) = \text{const} \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1-q)}} = \text{const} \cdot \int_0^{t_0} \frac{2 \cdot \sin(t) \cdot \cos(t) \cdot dt}{\sin(t) \cdot \cos(t)} = \int_0^{t_0} 2 \cdot dt = 2 \cdot \text{const} \cdot t_0.$$

We know how t_0 depends on p, so we get

$$g(p) = 2 \cdot \text{const} \cdot \arcsin\left(\sqrt{p}\,\right).$$

We can determine the constant from the condition that the largest possible probability value $p = 1$ should correspond to the right-most point $g(p) = 1$. From the condition that $g(1) = 1$, taking into account that

$$\arcsin\left(\sqrt{1}\right) = \arcsin(1) = \frac{\pi}{2},$$

we conclude that

$$g(p) = \frac{2}{\pi} \cdot \arcsin\left(\sqrt{p}\,\right). \tag{6.1}$$

Description of the resulting discretization. For a scale from 0 to some number m, the value $g(m)$ is equal to the ratio $\frac{i}{m}$. So, $i = m \cdot g(p)$.

Thus, the desired discretization means that to each probability p, we assign the value $i \approx m \cdot g(p)$ on the scale from 0 to m, where $g(p)$ is described by the above formula.

How do we select weights? If we need to select finitely many weights from the interval $[0, 1]$, then it is natural to select weights which are equally distributed on this interval, i.e., weights

$$0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}, 1. \tag{6.2}$$

This is how it is done in fuzzy logic, this is what, according to Section 4, follows from the traditional decision making theory.

How to assign weights to probabilities: idea. We have m a finite list of distinguishable probabilities $0 = p_0 < p_1 < \ldots < p_m = 1$. These probabilities correspond to degree

$$g(p_i) = \frac{i}{m}, \tag{6.3}$$

where $g(p)$ is determined by the formula (6.1). We need to assign, to each of these probabilities, an appropriate weight from the above list (6.2).

The larger the probability, the more weight we should assign to the corresponding outcome. Thus, we arrive at the following assignment of weights to probabilities:

- to the value $p_0 = 0$, we assign the smallest possible weight $w_0 = 0$;
- to the next value p_1, we assign the next weight
$$w_1 = \frac{1}{m};$$
- to the next value p_2, we assign the next weight
$$w_2 = \frac{2}{m};$$
- $\ldots$
- to the value p_{m-1}, we assign the weight
$$w_{m-1} = \frac{m-1}{m};$$
- finally, to the value $p_m = 1$, we assign the weight $w_m = 1$.

In general, to the value p_i, we assign the weight

$$w_i = \frac{i}{m}.$$

By comparing this assignment with the formula (6.3), we conclude that to each value p_i, we assign the value $w_i = g(p_i)$.

How to assign weights to probabilities: result. Our arguments show that to each probability $p \in [0,1]$, we assign the weight $g(p)$, where the function $g(p)$ is determined by the formula (1).

Comparing our weights with empirical weights: first try. Let us compare the probabilities p_i, Kahneman's empirical weights $\widetilde{w}_i$, and the weight $w_i = g(p_i)$ computed by using the formula (1):

p_i	0	1	2	5	10	20	50	80	90	95	98	99	100
$\widetilde{w}_i$	0	5.5	8.1	13.2	18.6	26.1	42.1	60.1	71.2	79.3	87.1	91.2	100
$w_i = g(p_i)$	0	6.4	9.0	14.4	20.5	29.5	50.0	70.5	79.5	85.6	91.0	93.6	100

The estimates $w_i = g(p_i)$ are closer to the observed weights $\widetilde{w}_i$ than the original probabilities, but the relation does not seem very impressive.

We will show that the fit is much better than it seems at first glance. At first glance, the above direct comparison between the observed weights $\widetilde{w}_i$ and the estimated weights $w_i = g(p_i)$ seems to make perfect sense. However, let us look deeper.

The weights come from the fact that users maximize the weighted gain $w(a) = \sum w_i(a) \cdot u_i$. It is easy to observe that if we multiply all the weights by the same positive constant $\lambda > 0$, i.e., consider the weights $w'_i(a) = \lambda \cdot w_i(a)$, then for each action, the resulting value of the weighted gain will also increase by the same factor:

$$w'(a) = \sum w'_i(a) \cdot u_i = \sum \lambda \cdot w_i(a) \cdot u_i = \lambda \cdot \sum w_i(a) \cdot u_i = \lambda \cdot w_i(a).$$

The relation between the weighted gains of two actions a and a' does not change if we simply multiply both gains by a positive constant:

- if $w_i(a) < w_i(a')$, then, multiplying both sides of this inequality by λ, we get
$$w'_i(a) < w'_i(a');$$
- if $w_i(a) = w_i(a')$, then, multiplying both sides of this equality by λ, we get $w'_i(a) = w'_i(a')$;
- if $w_i(a) > w_i(a')$, then, multiplying both sides of this inequality by λ, we get
$$w'_i(a) > w'_i(a').$$

All we observe is which of the two actions a person selects. Since multiplying all the weights by a constant does not change the selection, this means that based on the selection, we cannot uniquely determine the weights: an empirical selection which is consistent with the weights w_i is equally consistent with the weights $w'_i = \lambda \cdot w_i$.

This fact can be use to *normalize* the empirical weights, i.e., to multiply them by a constant so as to satisfy some additional condition.

In [8], to normalize the weights, the authors use the requirement that the weight corresponding to probability 1 should be equal to 1. Since for $p = 1$, the estimated weight $g(1)$ is also equal to 1, we get a perfect match for $p = 1$, but a rather lousy match for probabilities intermediate between 0 and 1.

Instead of this normalization, we can select λ so as to get the best match "on average".

How to improve the fit: details. A natural idea is to select λ from the Least Squares method, i.e., select λ for which the relative mean squares difference

$$\sum_i \left(\frac{\lambda \cdot w_i - \widetilde{w}_i}{w_i} \right)^2$$

is the smallest possible. Differentiating this expression with respect to λ and equating the derivative to 0, we conclude that

$$\sum_i \left(\lambda - \frac{\widetilde{w}_i}{w_i} \right) = 0,$$

i.e., that

$$\lambda = \frac{1}{m} \cdot \sum_i \frac{\widetilde{w}_i}{w_i}.$$

Resulting match. For the above values, this formula leads to $\lambda = 0.910$.

Result. The resulting values $w_i' = \lambda \cdot w_i$ are much closer to the empirical weights $\widetilde{x}_i$:

p_i	0	1	2	5	10	20	50	80	90	95	98	99	100
$\widetilde{w}_i$	0	5.5	8.1	13.2	18.6	26.1	42.1	60.1	71.2	79.3	87.1	91.2	100
$w_i' = \lambda \cdot g(p_i)$	0	5.8	8.2	13.1	18.7	26.8	45.5	64.2	72.3	77.9	82.8	87.4	91.0

For most probabilities p_i, the difference between the fuzzy-motivated weights w_i' and the empirical weights $\widetilde{w}_i$ is so small that it is below the accuracy with which the empirical weights can be obtained from the experiment.

Thus, fuzzy-related ideas indeed explain Kahneman and Tversky's empirical decision weights.

7. Likert-Scale Fuzzy Uncertainty from a Traditional Decision Making Viewpoint (cont-d): Advantages of Fuzzy Approach

How do we select a mark on a Likert scale? In Section 4, we simply used the labels marked by people on a Likert scale. But how do people select which labels to mark? To understand this, let us recall how this marking is done. Suppose that we have a Likert scale with $n + 1$ labels 0, 1, 2, ..., n, ranging from the smallest to the largest.

Then, if the actual value of the quantity x is very small, we mark label 0. At some point, we change to label 1; let us mark this threshold point by x_1. When we continue increasing x, we first have values marked by label 1, but eventually reach a new threshold after which values will be marked by label 2; let us denote this threshold by x_2, etc. As a result, we divide the range $[\underline{X}, \overline{X}]$ of the original variable into $n + 1$ intervals $[x_0, x_1]$, $[x_1, x_2]$, ..., $[x_{n-1}, x_n]$, $[x_n, x_{n+1}]$, where $x_0 = \underline{X}$ and $x_{n+1} = \overline{X}$:

- values from the first interval $[x_0, x_1]$ are marked with label 0;
- values from the second interval $[x_1, x_2]$ are marked with label 1;
- ...
- values from the n-th interval $[x_{n-1}, x_n]$ are marked with label $n - 1$;
- values from the $(n + 1)$-st interval $[x_n, x_{n+1}]$ are marked with label n.

Then, when we need to make a decision, we base this decision only on the label, i.e., only on the interval to which x belongs. In other words, we make n different decisions depending on whether x belongs to the interval $[x_0, x_1]$, to the interval $[x_1, x_2]$, ..., or to the interval $[x_n, x_{n+1}]$.

Decisions based on the Likert discretization are imperfect. Ideally, we should take into account the exact value of the variable x. When we use Likert scale, we only take into account an interval containing x and thus, we do not take into account part of the original information. Since we only use part of the original information about x, the resulting decision may not be as good as the decision based on the ideal complete knowledge.

For example, an ideal office air conditioner should be able to maintain the exact temperature at which a person feels comfortable. People are different, their temperature preferences are different, so an ideal air conditioner should be able to maintain any temperature value x within a certain range $[\underline{X}, \overline{X}]$. In practice, some air conditioners only have a finite number of settings. For example, if we have setting corresponding to 65, 70, 75, and 80 degrees, then a person who prefers 72 degrees will probably select the 70 setting or the 75 setting. In both cases, this person will be somewhat less comfortable than if there was a possibility of an ideal 72 degrees setting.

How do we select a Likert scale: main idea. According to the general ideas of traditional (utility-based) approach to decision making, we should select a Likert scale for which the expected utility is the largest.

To estimate the utility of decisions based on each scale, we will take into account the just-mentioned fact that decisions based on the Likert discretization are imperfect. In utility terms, this means that the utility of the Likert-based decisions is, in general, smaller than the utility of the ideal decision.

Which decision should we choose within each label? In the ideal situation, if we could use the exact value of the quantity x, then for each value x, we would select an optimal decision $d(x)$, a decision which maximizes the person's utility.

If we only know the label k, i.e., if we only know that the actual value x belongs to the $(k+1)$-st interval $[x_k, x_{k+1}]$, then we have to make a decision based only on this information. In other words, we have to select one of the possible values $\widetilde{x}_k \in [x_k, x_{k+1}]$, and then, for all x from this interval, use the decision $d(\widetilde{x}_k)$ based on this value.

Which value $\widetilde{x}_k$ should we choose: idea. According to the traditional approach to decision making, we should select a value for which the expected utility is the largest.

Which value $\widetilde{x}_k$ should we choose: towards a precise formulation of the problem. To find this expected utility, we need to know two things:

- we need to know the probability of different values of x; these probabilities can be described, e.g., by the probability density function $\rho(x)$;
- we also need to know, for each pair of values x' and x, what is the utility $u(x', x)$ of using a decision $d(x')$ in the situation in which the actual value is x.

In these terms, the expected utility of selecting a value $\widetilde{x}_k$ can be described as

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx. \tag{7.1}$$

Thus, for each interval $[x_k, x_{k+1}]$, we need to select a decision $d(\widetilde{x}_k)$ corresponding to the value $\widetilde{x}_k$ for which the expression (7.1) attains its largest possible value. The resulting expected utility is equal to

$$\max_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx. \tag{7.2}$$

How to select the best Likert scale: general formulation of the problem. The actual value x can belong to any of the $n+1$ intervals $[x_k, x_{k+1}]$. Thus, to find the overall expected utility, we need to add the values (7.2) corresponding to all these intervals. In other words, we need to select the values $x_1, \ldots, x_n$ for which the following expression attains its largest possible value:

$$\sum_{k=0}^{n} \max_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx. \tag{7.3}$$

Equivalent reformulation in terms of disutility. In the ideal case, for each value x, we should use a decision $d(x)$ corresponding to this value x, and gain utility $u(x, x)$. In practice, we have to use decisions $d(x')$ corresponding to a slightly different value, and thus, get slightly worse utility values $u(x', x)$. The corresponding decrease in utility $U(x', x) \stackrel{\text{def}}{=} u(x, x) - u(x', x)$ is usually called *disutility*. In terms of disutility, the function $u(x', x)$ has the form

$$u(x', x) = u(x, x) - U(x', x),$$

and thus, the optimized expression (7.1) takes the form

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x, x)\, dx - \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx.$$

The first integral does not depend on $\widetilde{x}_k$; thus, the expression (7.1) attains its maximum if and only if the second integral attains its minimum. The resulting maximum (7.2) thus takes the form

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x,x)\,dx - \min_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\,dx. \tag{7.4}$$

Thus, the expression (7.3) takes the form

$$\sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x,x)\,dx - \sum_{k=0}^{n} \min_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\,dx.$$

The first sum does not depend on selecting the thresholds. Thus, to maximize utility, we should select the values $x_1, \ldots, x_n$ for which the second sum attains its smallest possible value:

$$\sum_{k=0}^{n} \min_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\,dx \to \min. \tag{7.5}$$

Let is recall that are interested in the membership function. For a general Likert scale, we have a complex optimization problem (7.5). However, we are not interested in general Likert scales per se, what we are interested in is the use of Likert scales to elicit the values of the membership function $\mu(x)$.

As we have mentioned earlier, in an n-valued scale:

- the smallest label 0 corresponds to the value $\mu(x) = 0/n$,
- the next label 1 corresponds to the value $\mu(x) = 1/n$,
- ...
- the last label n corresponds to the value $\mu(x) = n/n = 1$.

Thus, for each n:

- values from the interval $[x_0, x_1]$ correspond to the value $\mu(x) = 0/n$;
- values from the interval $[x_1, x_2]$ correspond to the value $\mu(x) = 1/n$;
- ...
- values from the interval $[x_n, x_{n+1}]$ correspond to the value $\mu(x) = n/n = 1$.

The actual value of the membership function $\mu(x)$ corresponds to the limit $n \to \infty$, i.e., in effect, to very large values of n. Thus, in our analysis, we will assume that the number n of labels is huge – and thus, that the width of each of $n+1$ intervals $[x_k, x_{k+1}]$ is very small.

Let us take into account that each interval is narrow. Let us use the fact that each interval is narrow to simplify the expression $U(x', x)$ and thus, the optimized expression (7.5).

In the expression $U(x', x)$, both values x' and x belong to the same narrow interval and thus, the difference $\Delta x \stackrel{\text{def}}{=} x' - x$ is small. Thus, we can expand the expression $U(x', x) = U(x + \Delta x, x)$ into Taylor series in Δx, and keep only the first non-zero term in this expansion. In general, we have

$$U(x + \Delta, x) = U_0(x) + U_1 \cdot \Delta x + U_2(x) \cdot \Delta x^2 + \ldots,$$

where

$$U_0(x) = U(x,x), \quad U_1(x) = \frac{\partial U(x + \Delta x, x)}{\partial(\Delta x)},$$

$$U_2(x) = \frac{1}{2} \cdot \frac{\partial^2 U(x + \Delta x, x)}{\partial^2(\Delta x)}. \tag{7.7}$$

Here, by definition of disutility, we get $U_0(x) = U(x,x) = u(x,x) - u(x,x) = 0$. Since the utility is the largest (and thus, disutility is the smallest) when $x' = x$, i.e., when $\Delta x = 0$, the derivative $U_1(x)$ is also equal to 0 – since the derivative of each (differentiable) function is equal to 0 when

this function attains its minimum. Thus, the first non-trivial term corresponds to the second derivative:

$$U(x+\Delta x, x) \approx U_2(x)\cdot \Delta x^2,$$

i.e., in other words, that

$$U(\widetilde{x}_k, x) \approx U_2(x)\cdot (\widetilde{x}_k - x)^2.$$

Substituting this expression into the expression

$$\int_{x_k}^{x_{k+1}} \rho(x)\cdot U(\widetilde{x}_k, x)\,dx$$

that needs to be minimized if we want to find the optimal $\widetilde{x}_k$, we conclude that we need to minimize the integral

$$\int_{x_k}^{x_{k+1}} \rho(x)\cdot U_2(x)\cdot (\widetilde{x}_k - x)^2\,dx. \tag{7.8}$$

This new integral is easy to minimize: if we differentiate this expression with respect to the unknown $\widetilde{x}_k$ and equate the derivative to 0, we conclude that

$$\int_{x_k}^{x_{k+1}} \rho(x)\cdot U_2(x)\cdot (\widetilde{x}_k - x)\,dx = 0,$$

i.e., that

$$\widetilde{x}_k \cdot \int_{x_k}^{x_{k+1}} \rho(x)\cdot U_2(x)\,dx = \int_{x_k}^{x_{k+1}} x\cdot \rho(x)\cdot U_2(x)\,dx,$$

and thus, that

$$\widetilde{x}_k = \frac{\int_{x_k}^{x_{k+1}} x\cdot \rho(x)\cdot U_2(x)\,dx}{\int_{x_k}^{x_{k+1}} \rho(x)\cdot U_2(x)\,dx}. \tag{7.9}$$

This expression can also be simplified if we take into account that the intervals are narrow. Specifically, if we denote the midpoint of the interval $[x_k, x_{k+1}]$ by $\overline{x}_k \stackrel{\text{def}}{=} \dfrac{x_k + x_{k+1}}{2}$, and denote $\Delta x \stackrel{\text{def}}{=} x - \overline{x}_k$, then we have $x = \overline{x}_k + \Delta x$. Expanding the corresponding expressions into Taylor series in terms of a small value Δx and keeping only main terms in this expansion, we get

$$\rho(x) = \rho(\overline{x}_k + \Delta x) = \rho(\overline{x}_k) + \rho'(\overline{x}_k)\cdot \Delta x \approx \rho(\overline{x}_k),$$

where $f'(x)$ denoted the derivative of a function $f(x)$, and

$$U_2(x) = U_2(\overline{x}_k + \Delta x) = U_2(\overline{x}_k) + U_2'(\overline{x}_k)\cdot \Delta x \approx U_2(\overline{x}_k).$$

Substituting these expressions into the formula (7.9), we conclude that

$$\widetilde{x}_k = \frac{\rho(\overline{x}_k)\cdot U_2(\overline{x}_k)\cdot \int_{x_k}^{x_{k+1}} x\,dx}{\rho(\overline{x}_k)\cdot U_2(\overline{x}_k)\cdot \int_{x_k}^{x_{k+1}} dx} = \frac{\int_{x_k}^{x_{k+1}} x\,dx}{\int_{x_k}^{x_{k+1}} dx} = \frac{\frac{1}{2}\cdot (x_{k+1}^2 - x_k^2)}{x_{k+1} - x_k} = \frac{x_{k+1} + x_k}{2} = \overline{x}_k.$$

Substituting this midpoint value $\widetilde{x}_k = \overline{x}_k$ into the integral (7.8) and taking into account that on the k-th interval, we have $\rho(x) \approx \rho(\overline{x}_k)$ and $U_2(x) \approx U_2(\overline{x}_k)$, we conclude that the integral (8) takes the form

$$\int_{x_k}^{x_{k+1}} \rho(\overline{x}_k)\cdot U_2(\overline{x}_k)\cdot (\overline{x}_k - x)^2\,dx = \rho(\overline{x}_k)\cdot U_2(\overline{x}_k)\cdot \int_{x_k}^{x_{k+1}} (\overline{x}_k - x)^2\,dx. \tag{7.8a}$$

When x goes from x_k to x_{k+1}, the difference $\Delta x = x - \overline{x}_k$ between the value x and the interval's midpoint $\overline{x}_k$ ranges from $-\Delta_k$ to Δ_k, where Δ_k is the interval's half-width:

$$\Delta_k \stackrel{\text{def}}{=} \frac{x_{k+1} - x_k}{2}.$$

In terms of the new variable Δx, the integral in the right-hand side of (7.8a) has the form

$$\int_{x_k}^{x_{k+1}} (\overline{x}_k - x)^2\,dx = \int_{-\Delta_k}^{\Delta_k} (\Delta x)^2\,d(\Delta x) = \frac{2}{3}\cdot\Delta_k^3.$$

Thus, the integral (7.8) takes the form

$$\frac{2}{3}\cdot\rho(\overline{x}_k)\cdot U_2(\overline{x}_k)\cdot\Delta_k^3.$$

The problem (7.5) of selecting the Likert scale thus becomes the problem of minimizing the sum (7.5) of such expressions (7.8), i.e., of the sum

$$\frac{2}{3}\cdot\sum_{k=0}^{n}\rho(\overline{x}_k)\cdot U_2(\overline{x}_k)\cdot\Delta_k^3. \tag{7.10}$$

Here, $\overline{x}_{k+1} = x_{k+1} + \Delta_{k+1} = (\overline{x}_k + \Delta_k) + \Delta_{k+1} \approx \overline{x}_k + 2\Delta_k$, so $\Delta_k = (1/2)\cdot\Delta\overline{x}_k$, where $\Delta\overline{x}_k \stackrel{\text{def}}{=} \overline{x}_{k+1} - \overline{x}_k$. Thus, (7.10) takes the form

$$\frac{1}{3}\cdot\sum_{k=0}^{n}\rho(\overline{x}_k)\cdot U_2(\overline{x}_k)\cdot\Delta_k^2\cdot\Delta\overline{x}_k. \tag{7.11}$$

In terms of the membership function, we have $\mu(\overline{x}_k) = k/n$ and $\mu(\overline{x}_{k+1}) = (k+1)/n$. Since the half-width Δ_k is small, we have

$$\frac{1}{n} = \mu(\overline{x}_{k+1}) - \mu(\overline{x}_k) = \mu(\overline{x}_k + 2\Delta_k) - \mu(\overline{x}_k) \approx \mu'(\overline{x}_k)\cdot 2\Delta_k,$$

thus, $\Delta_k \approx \dfrac{1}{2n}\cdot\dfrac{1}{\mu'(\overline{x}_k)}$. Substituting this expression into (7.11), we get the expression $\dfrac{1}{3\cdot(2n)^2}\cdot I$, where

$$I = \sum_{k=0}^{n}\frac{\rho(\overline{x}_k)\cdot U_2(\overline{x}_k)}{(\mu'(\overline{x}_k))^2}\cdot\Delta\overline{x}_k. \tag{7.12}$$

The expression I is an integral sum, so when $n\to\infty$, this expression tends to the corresponding integral

$$I = \int\frac{\rho(x)\cdot U_2(x)}{(\mu'(x))^2}\,dx. \tag{7.11}$$

Minimizing (7.5) is equivalent to minimizing I. With respect to the derivative $d(x) \stackrel{\text{def}}{=} \mu'(x)$, we need to minimize the objective function

$$I = \int\frac{\rho(x)\cdot U_2(x)}{d^2(x)}\,dx \tag{7.12}$$

under the constraint that

$$\int_{\underline{X}}^{\overline{X}} d(x)\,dx = \mu(\overline{X}) - \mu(\underline{X}) = 1 - 0 = 1. \tag{7.13}$$

By using the Lagrange multiplier method, we can reduce this constraint optimization problem to the unconstrained problem of minimizing the functional

$$I = \int\frac{\rho(x)\cdot U_2(x)}{d^2(x)}\,dx + \lambda\cdot\int d(x)\,dx, \tag{7.14}$$

for an appropriate Lagrange multiplier λ. Differentiating (7.14) with respect to $d(x)$ and equating the derivative to 0, we conclude that $-2\cdot\dfrac{\rho(x)\cdot U_2(x)}{d^3(x)} + \lambda = 0$, i.e., that $d(x) = c\cdot(\rho(x)\cdot U_2(x))^{1/3}$ for some constant c. Thus, $\mu(x) = \int_{\underline{X}}^{x} d(t)\,dt = c\cdot\int_{\underline{X}}^{x}(\rho(t)\cdot U_2(t))^{1/3}\,dt$. The constant c must be determined by the condition that $\mu(\overline{X}) = 1$. Thus, we arrive at the following formula (7.15).

Resulting formula. The membership function $\mu(x)$ obtained by using Likert-scale elicitation is equal to

$$\mu(x) = \frac{\int_{\underline{X}}^{x} (\rho(t) \cdot U_2(t))^{1/3}\, dt}{\int_{\underline{X}}^{\overline{X}} (\rho(t) \cdot U_2(t))^{1/3}\, dt}, \tag{7.15}$$

where $\rho(x)$ is the probability density describing the probabilities of different values of x, $U_2(x) \stackrel{\text{def}}{=} \dfrac{1}{2} \cdot \dfrac{\partial^2 U(x + \Delta x, x)}{\partial^2 (\Delta x)}$, $U(x', x) \stackrel{\text{def}}{=} u(x, x) - u(x', x)$, and $u(x', x)$ is the utility of using a decision $d(x')$ corresponding to the value x' in the situation in which the actual value is x.

Comment. The above formula only applies to membership functions like "large" whose values monotonically increase with x. It is easy to write a similar formula for membership functions like "small" which decrease with x. For membership functions like "approximately 0" which first increase and then decrease, we need to separately apply these formula to both increasing and decreasing parts.

Conclusion. The resulting membership degrees incorporate both probability and utility information. This fact *explains why fuzzy techniques often work better than probabilistic techniques* – because the probability techniques only take into account the probability of different outcomes.

8. Acknowledgment

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

The authors are thankful to Dr. Fikret Aliev for his support and encouragement, and to the participants of the Joint World Congress of the International Fuzzy Systems Association and Annual Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS'2013 (Edmonton, Canada, June 24–28, 2013), 4th World Conference on Soft Computing (Berkeley, California, May 25–27, 2014), and 2014 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2014 (Boston, Massachusetts, June 24–26, 2014) for valuable discussions.

References

[1] M. Ahsanullah, V. B. Nevzorov, and M. Shakil, *An Introduction to Order Statistics*, Atlantis Press, Paris, 2013.
[2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*, Society of Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, 2008.
[3] R. E. Bellman and L. A. Zadeh, "Decision making in a fuzzy environment", *Management Science*, 1970, Vol. 17, No. 4, pp. B141–B164.
[4] M. D. Cohen, G. Huber, R. L. Keeney, A. H. Levis, L. L. Lopes, A. P. Sage, S. Sen, A. B. Whinston, R. L. Winkler, D. von Winterfeldt, and L. A. Zadeh, "Research needs and the phenomena of decision making and operations", *IEEE Transactions on Systems, Man, and Cybernetics*, 1985, Vol. 15, No. 6, pp. 764–775.
[5] H. A. David and H. N. Nagaraja, *Order Statistics*, Wiley, New York, 2003.
[6] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
[7] P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.
[8] D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus, and Giroux, New York, 2011.
[9] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives*, John Wiley and Sons, New York, 1976.
[10] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
[11] V. Kreinovich, "Decision Making under Interval Uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163-193.
[12] J. Lorkowski and V. Kreinovich, "Likert-scale fuzzy uncertainty from a traditional decision making viewpoint: it incorporates both subjective probabilities and utility information", *Proceedings of the Joint World Congress of the International Fuzzy Systems Association and Annual Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS'2013*, Edmonton, Canada, June 24–28, 2013, pp. 525–530.

[13] J. Lorkowski and V. Kreinovich, "Fuzzy Logic Ideas Can Help in Explaining Kahneman and Tversky's Empirical Decision Weights", *Proceedings of the 4th World Conference on Soft Computing*, Berkeley, California, May 25–27, 2014, pp. 285–289.
[14] J. Lorkowski and V. Kreinovich, "Interval and Symmetry Approaches to Uncertainty – Pioneered by Wiener – Help Explain Seemingly Irrational Human Behavior: A Case Study", *Proceedings of the 2014 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2014*, Boston, Massachusetts, June 24–26, 2014, to appear.
[15] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
[16] J. March, "Bounded Rationality, Ambiguity, and the Engineering of Choice," *The Bell Journal of Economics*, Vol. 9, No. 2, 1978, pp. 587–608.
[17] H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision Making Beyond Arrow's 'Impossibility Theorem', With the Analysis of Effects of Collusion and Mutual Attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.
[18] H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.
[19] H. T. Nguyen, V. Kreinovich, and B. Lea, "How to combine probabilistic and fuzzy uncertainties in fuzzy control", *Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, December 2–4, 1992, pp. 117–121.
[20] H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
[21] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
[22] S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.
[23] H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.
[24] D. Redelmeier and E. Shafir, "Medical decision mading in situations that offer multiple alternatives," *Journal of the American Medical Association*, 1995, Vol. 273, No. 4, pp. 302–305.
[25] E. Shafir, I. Simonson, and A. Tversky, "Reason-based choice," *Cognition*, 1993, Vol. 49, pp. 11–36.
[26] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2011.
[27] I. Simonson and E. Rosen, *Absolute Value: What Really Influences Customers in the Age of (Nearly) Perfect Information*, HarperBusiness, New York, 2014.
[28] A. Tversky and I. Simonson, "Context-dependent preferences," *Management Science*, 1993, Vol. 39, No. 10, pp. 1179–1189.
[29] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.
[30] L. A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes", *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, Vol. 3, No. 1, pp. 28–44.
[31] L. A. Zadeh. "Precisiated Natural Language-Toward a Radical Enlargement of the Role of Natural Languages in Information Processing, Decision and Control", In: L. Wang, S. K. Halgamuge, and X. Yao (Eds.), *Proceedings of the 1st International Conference on Fuzzy Systems and Knowledge Discovery FSDK'02: Computational Intelligence for the E-Age*, Singapore, November 18–22, 2002, Vol. 1, pp. 1–3.
[32] L. A. Zadeh, "Computing with Words and Perceptions - A Paradigm Shift in Computing and Decision Analysis and Machine Intelligence", In: A. Wani, K. J. Cios, and K. Hafeez (Eds.), *Proceedings of the 2003 International Conference on Machine Learning and Applications ICMLA'2003*, Los Angeles, California, June 23–24, 2003, pp. 3–5.
[33] L. A. Zadeh, "A New Direction in Decision Analysis-Perception-based Decisions", In: A. L. Ralescu (Ed.), *Proceedings of the Fourteenth Midwest Artificial Intelligence and Cognitive Sciences Conference MAICS'2003*, Cincinnati, Ohio, April 12–13, 2003, pp. 1–2.