RANDOM SETS UNIFY, EXPLAIN, AND AID KNOWN UNCERTAINTY METHODS IN EXPERT SYSTEMS

VLADIK KREINOVICH*

Abstract. Numerous formalisms have been proposed for representing and processing uncertainty in expert systems. Several of these formalisms are somewhat *ad hoc*, in the sense that some of their formulas seem to have been chosen rather arbitrarily.

In this paper, we show that random sets provide a natural general framework for describing uncertainty, a framework in which many existing formalisms appear as particular cases. This interpretation of known formalisms (e.g., of fuzzy logic) in terms of random sets enables us to justify many "ad hoc" formulas. In some cases, when several alternative formulas have been proposed, random sets help to choose the best ones (in some reasonable sense).

One of the main objectives of expert systems is not only to *describe* the current state of the world, but also to provide us with reasonable *actions*. The simplest case is when we have the *exact* objective function. In this case, random sets can help in choosing the proper method of "fuzzy optimization."

As a test case, we describe the problem of choosing the best tests in technical diagnostics. For this problem, feasible algorithms are possible.

In many real-life situations, instead of an *exact* objective function, we have several participants with *different* objective functions, and we must somehow reconcile their (often conflicting) interests. Sometimes, standard approaches of game theory are not working. We show that in such situations, random sets present a working alternative. This is one of the cases when particular cases of random sets (such as fuzzy sets) are not sufficient, and general random sets are needed.

Key words. Random Sets, Fuzzy Logic, Expert Systems, Maximum Entropy, Fuzzy Optimization, Technical Diagnostics, Cooperative Games, von Neumann-Morgenstern Solution.

AMS(MOS) subject classifications. 60D05, 03B52, 04A72, 90C70, 90D12, 94C12, 94A17, 68T35

1. Main Problems: Too many different formalisms describe uncertainty, and some of these formalisms are not well justified. Many different formalisms have been proposed for representing and processing uncertainty in expert systems (see, e.g., [30]); some of these formalisms are not well justified. There are two problems here:

1.1. First Problem: Justification. First of all, we are not sure whether all these formalisms are truly adequate for describing uncertainty. The fact that these formalisms have survived means that they have been successful in application and thus, that they do represent (exactly or approximately) some features of expert reasoning. However, several of the

* Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968. This work was partially supported by NASA grant No. NAG 9-757. The author is greatly thankful to O.N. Bondareva, P. Fishburn, O.M. Kosheleva, H.T. Nguyen, P. Suppes, and L.A. Zadeh for stimulating discussions, and to J. Goutsias and to the anonymous referees for valuable suggestions.

formalisms are somewhat *ad hoc* in the sense that some of their formulas (such as combination formulas in Dempster-Shafer formalism or in fuzzy logic) seem to have been chosen rather arbitrarily. It may, in principle, turn out that these formulas are only approximations to other formulas that describe human reasoning much better. It is therefore desirable to either *explain* the original formulas (from some reasonable first principles), or to *find* other formulas (close to the original ones) that can be thus explained.

1.2. Second Problem: Combination. In many cases, different pieces of knowledge are represented in different formalisms. To process all knowledge, it is therefore desirable to combine (*unify*) these formalisms.

1.3. What we are planning to do. In this paper, we will show that both problems can be solved (to a large extent) in the framework of random sets: Namely, random sets provide a natural general framework for describing uncertainty, a framework of which many existing formalisms appear as particular cases. This interpretation of known formalisms (like fuzzy logic) in terms of random sets enables us to justify many "ad hoc" formulas.

In this paper, due to size limitations, we will mainly talk about the first (*justification*) problem. As a result of our analysis, we will conclude that most existing formalisms can be justified within the same framework: of *random sets.*

The existence of such a unified framework opens possibilities for solving the second (*combination*) problem. In this paper, we will show how this general random set framework can be helpful in solving two *specific* classes of problems: *decision making* (on the example of technical diagnostics) and *conflict resolution* (on the example of cooperative games). In contrast to these (well-defined and algorithmically analyzed) *specific* solutions, a *general* random-set-based description of the possible combined formalisms has only been developed *in principle* and requires further research before it can be practically used.

2. Random Sets: Solution to the main problems.

2.1. Uncertainty naturally leads to random sets.

2.1.1. In rare cases of complete knowledge, we know the exact state of the system. In rare cases, we know the *exact* state of a system, i.e., in more mathematical terms, we know exactly which element s of the set of all states A describes the current state of the system.

In the majority of the situations, however, we only have a *partial* knowledge about the system, i.e., we have *uncertainty.*

2.1.2. If all statements about the system are precise, then such incomplete knowledge describes a set of possible states. Usually, only a *part* of the knowledge is formulated in precise terms. In some situations, however, *all* the knowledge is formulated in precise terms. In such situations, the knowledge consists of one or several *precise* statements $E_1(s), \ldots, E_n(s)$ about the (unknown) state s.

For example, if we are describing the temperature s in El Paso at this very moment, a possible knowledge base may consist of a single statement $E_1(s)$, according to which this temperature s is in the nineties, meaning that it is between 90 and 100 Fahrenheit.

We can describe such a knowledge by the set S of all states $s \in A$ for which all these statements $E_1(s), \ldots, E_n(s)$ are true; e.g., in the above example of a single temperature statement, S is the set of all the weather states s in which the temperature is exactly in between 90 and 100.

2.1.3. Vague statements naturally lead to random sets of possible states. Ideally, it would be great if all our knowledge was precise. In reality, an important part of knowledge consists of expert statements, and experts often cannot express their knowledge in precise terms. Instead, they make statements like "it will be pretty hot tomorrow," or, even worse, "it will most probably be pretty hot tomorrow." Such *vague* statements definitely carry some information. The question is: how can we formalize such vague statements? If a statement $P(s)$ about the state s is precise, then, for every state s, we can definitely tell whether this state does or does not satisfy this property $P(s)$. As a result, a precise statement characterizes a subset S of the set of all states A. In contrast to precise statements, for vague statements $P(s)$, we are often not sure whether a given state does or does not satisfy this "vague" property: e.g., intuitively, there is no strict border between "pretty hot" and "not pretty hot": there are some intermediate values about which we are not sure whether they are "pretty hot" or not. In other words, a vague statement means that an expert is not sure which set S of possible values is described by this statement. For example, the majority of the people understand the term "pretty hot" pretty well; so, we can ask different people to describe exactly what this "pretty hot" means. The problem is that different people will interpret the same statement by different subsets $S \subseteq A$ (not to mention that some people will have trouble choosing any formalization at all).

Summarizing, a vague statement is best represented not by a *single* subset S of possible states, but by a *class* $\mathcal{S}$ of possible sets $S \subseteq A$. Some sets from this class are more probable, some are less probable. It is natural to describe this "probability" by a number from the interval $[0, 1]$. For example, we can ask several people to interpret a vague statement like "pretty hot" in precise terms, and for each set S (e.g., for the set $[80, 95]$), take the fraction of those people who interpret "pretty hot" as belonging to this particular set S, as the probability $p(S)$ of this set (so that, e.g., if 10% of the people interpret "pretty hot" as "belonging to $[80, 95]$," we take $p([80, 95]) = 0.1$). Another possibility is to ask a single "interpreter" to provide us with the "subjective" probabilities $p(S)$ that describe how probable each set $S \in \mathcal{S}$ is. In both cases, we get a probability measure on a class of sets, i.e., a *random set*.

Comment. Our approach seems to indicate that random sets are a *reasonable* general description of uncertainty. However, this does not mean that random set are the *only* possible general description of uncertainty. Another reasonable possibilities could include *interval-valued* probability measures, in which instead of a single probability of an event, we get its lower and upper probabilities; *Bayesian*-type approach, in which in addition to the *interval* of possible values of probability, we have (second-order) *probabilities* of different probability values; etc.

In this paper, we selected random sets mainly for one *pragmatic* reason: because the theory of random sets is, currently, the most developed and thus, reduction to random sets is, currently, most useful. It may happen, however, that further progress will show that interval-valued or second order probabilities provide an even better general description of uncertainty.

2.2. If random sets are so natural, why do we need other formalisms for uncertainty. Since random sets are such a natural formalism for describing uncertainty, why not use them? The main reason why generic random sets are rarely used (and other formalisms are used more frequently) becomes clear if we analyze how much computer memory and computer time we need to process general random sets.

Indeed, for a system with n possible states, there exist 2^n possible sets $S \subseteq A$. Hence, to describe a generic random set, we need $2^n - 1$ real numbers $p(S) \geq 0$ corresponding to different sets S (we need $2^n - 1$ and not 2^n because we have a relation $\sum p(S) = 1$)[1]. For sets A of moderate and realistic size (e.g., for $n \approx 300$), this number $2^n - 1$ exceeds the number of particles in the known Universe; thus, it is impossible to store all this information. For the same reason, it is even less possible to process it.

Thus, for practical use in expert systems, we must use *partial* information about these probabilities. Let us show that three formalisms for uncertainty — statistics, fuzzy, and Dempster-Shafer — can indeed be interpreted as such partial information. (This fact has been known for some time, but unfortunately many researchers in uncertainty are still unaware of it; so, without going into technical detail, we will briefly explain how these formalisms can be interpreted in the random set framework.)

[1] It is quite possible that some experts believe the knowledge to be inconsistent, and thus, $p(\emptyset) > 0$. If we assume that all the experts believe the knowledge to be consistent, then we get $p(\emptyset) = 0$ and thus, only $2^n - 2$ real numbers are needed to describe a random set.

2.3. Three main uncertainty formalisms as particular cases of random set framework.

2.3.1. Standard Statistical Approach.

Description. In the standard statistical approach, we describe uncertainty by assigning a probability $P(s)$ to each *event* s. (The sum of these probabilities must add up to 1).

Interpretation in terms of random sets. This description is clearly a particular case of a random set, in which only one-element sets have non-zero probability $p(\{s\}) = P(s)$, and all other sets have probability 0. For this description, we need $n-1$ numbers instead of $2^n - 1$; thus, this formalism is quite feasible.

Limitations. The statistical description is indeed a *particular* case not only in the mathematical sense, but also from the viewpoint of common sense reasoning: Indeed, e.g., if only two states are possible, and we know nothing about the probability of each of them, then it is natural to assign equal probability to both; thus, $P(s_1) = P(s_2)$ and hence, $P(s_1) = P(s_2) = 0.5$. Thence, within the standard statistical approach, we are unable to distinguish between the situation in which we know nothing about the probabilities, and the situations like tossing a coin, in which we are absolutely sure that the probability of each state is exactly 0.5.

2.3.2. Fuzzy Formalism.

Description. Statistical approach has shown that if we store only one real number per state, we get a computationally feasible formalism. However, the way this is done in the standard statistical approach is too restrictive. It is therefore desirable to store some partial information about the set without restricting the underlying probability measure $p(S)$ defined on the class of sets S.

A natural way to do this is to assign, to each such probability measure, the values

$$\mu(s) = P(s \in S) = P(\{S \mid s \in S\}) = \sum_{S \ni s} p(S).$$

These values belong to the interval $[0,1]$ and do not necessarily add up to 1.

In fuzzy formalism (see, e.g., [10,24]), numbers $\mu(s) \in [0,1]$, $s \in A$ (that do not necessarily add up to 1), form a *membership function* of a *fuzzy set.*

Interpretation in terms of random sets. Hung T. Nguyen has shown [22] that every membership function can be thus interpreted, and that, moreover, standard (initially *ad hoc*) operations with fuzzy numbers can be interpreted in this manner.

For example, the standard fuzzy interpretation of the vague property "pretty hot" is as follows: we take, say, $A = [60, 130]$, and assign to every value $s \in A$, a number $\mu(s)$ that describes to what extent experts believe that s is pretty hot.

The random set interpretation is that with different probabilities, "pretty hot" can mean belonging to different subsets of the set $[60, 130]$, and, for example, $\mu(80)$ is the total probability that 80 is pretty hot.

Limitations. In this interpretation, a fuzzy set contains only a *partial* information about uncertainty; this information contains the probabilities $\mu(s)$ that different states s are possible, but it does not contain, e.g., probabilities that a pair (s, s') is possible. It may happen that the possibility of a state s (i.e., the fact that $s \in S$) renders some other states impossible. E.g., intuitively, if two values $s < s'$ correspond to "pretty hot," then any temperature in between s and s' must also correspond to "pretty hot"; thus, the probability measure $p(S)$ should be only located on convex sets S. However, "fine" information of this type cannot be captured by a fuzzy description.

2.3.3. Dempster-Shafer Formalism.

Description and interpretation in terms of random sets. This formalism (see, e.g., [32]) is the closest to random sets: each expert's opinion is actually represented by the values ("masses") $m(S) \geq 0$ assigned to sets $S \subseteq A$ that add up to 1 and thus form a random set. The only difference from the general random set framework is that instead of representing the entire knowledge as a random set, this formalism represents each expert's opinion as a random set, with somewhat *ad hoc* combination rules.

Limitations. Since this formalism is so close to the general random set framework, it suffers (although somewhat less) from the same computational complexity problems as the general formalism; for details and for methods to overcome this complexity, see. e.g., [15].

2.4. Random set interpretation explains seemingly ad hoc operations from different uncertainty formalisms. Operations from standard statistics are usually pretty well justified. The Dempster-Shafer formalism is also mainly developed along statistical lines and uses *mostly* well-justified operations[2]. Since both formalisms are explicitly formulated in terms of probabilities, these justifications, of course, are quite in line with the random set interpretation.

The only formalism in which operations are mainly *ad hoc* and need justification is fuzzy formalism. Let us show that its basic operations can be naturally justified by the random set interpretation.

[2] It should be mentioned that Dempster's rules is not *completely* justified along statistical/probabilistic lines because it performs some *ad hoc* conditioning.

The basic operations of fuzzy formalism are operations of *fuzzy logic* that describe *combinations* of different parts of knowledge. For example, if we know the membership functions $\mu_1(s)$ and $\mu_2(s)$ that describe two "vague" statements $E_1(s)$ and $E_2(s)$, what membership function $\mu(s)$ corresponds to their conjunction $E_1(s)\&E_2(s)$? In other words, how to interpret "and" and "or" in fuzzy logic?

2.4.1. "and"- and "or"-Operations of Fuzzy Logic: Extremal Approach. Let us first show how the simplest "and" and "or" operations $f_\&(a,b) = \min(a,b)$ and $f_\vee(a,b) = \max(a,b)$ can be explained (in this description, we follow [22] and [2]):

Definition 2.1. Assume that a (crisp) set A is given. This set will be called a *Universum.*

- By a *random set*, we mean a probability measure P on a class 2^A of all subsets S of A.
- By a *fuzzy set* C, we mean a map (*membership function*) $\mu : A \to [0,1]$.
- We say that a random set P *represents* a fuzzy set C with a membership function $\mu(s)$, if for every $s \in A$, $P(s \in S) = \mu(s)$.

If we have *two* membership functions, this means that we actually have *two* unknown sets. To describe this uncertainty in probabilistic terms, we therefore need a probability measure on a set $2^A \times 2^A$ of all *pairs* of sets.

Definition 2.2. By a *random pair of sets*, we mean a probability measure P on a class $2^A \times 2^A$ of all pairs (S_1, S_2) of subsets of A. We say that a random pair P *represents* a pair of fuzzy sets (C_1, C_2) with membership functions $\mu_1(s)$ and $\mu_2(s)$ if for every $s \in A$: $P(s \in S_1) = \mu_1(s)$ and $P(s \in S_2) = \mu_2(s)$.

We are interested in the membership functions $\mu_{C_1 \cap C_2}(s)$ and $\mu_{C_1 \cup C_2}(s)$. It is natural to interpret these numbers as $P(s \in S_1 \cap S_2)$ and $P(s \in S_1 \cup S_2)$. The problem is that these numbers are *not* uniquely defined by μ_1 and μ_2. So, instead of a single value, we get a whole class of possible values. However, this class has a very natural bound:

Proposition 2.1. Let C_1 and C_2 be fuzzy sets with membership functions $\mu_1(s)$ and $\mu_2(s)$. Then, the following is true:

- For every random pair P that represents (C_1, C_2), and for every $s \in A$, $P(s \in S_1 \cap S_2) \le \min(\mu_1(s), \mu_2(s))$.
- There exists a random pair P that represents (C_1, C_2) and for which for every $s \in A$, $P(s \in S_1 \cap S_2) = \min(\mu_1(s), \mu_2(s))$.

So, min is an *upper bound* of possible values of probability. This min is exactly the operation originally proposed by Zadeh to describe "and" in fuzzy logic.

Similarly, for union, max turns out to be the *lower bound* [2].

Comment. One of the main original reasons for choosing the operations $\min(a,b)$ and $\max(a,b)$ as analogues of "and" and "or" is that these are the only "and" and "or" operations that satisfy the intuitively clear requirement that for every statement E, both $E\&E$ and $E \vee E$ have the same degree of belief as E itself. However, this reason does not mean that min and max are the *only* reasonable "and" and "or" operations: in the following text, we will see that other requirements (also seemingly reasonable) lead to different pairs of "and" and "or" operations.

2.4.2. "and"- and "or"-Operations of Fuzzy Logic: Maximum Entropy Approach. In the previous section, we considered *all possible* probability distributions consistent with the given membership function. As a result, for each desired quantity (such as $P(s \in S_1 \cap S_2)$), we get an *interval* of possible values.

In some situations, this interval may be too wide (close to $[0,1]$), which makes the results of this "extremal" approach rather meaningless. It is therefore necessary to consider only *some* probability measures, ideally, the measures that are (in some reasonable sense) "most probable."

The (reasonably known) formalization of this idea leads to the so-called *maximum entropy (MaxEnt) method* in which we choose the distribution for which the entropy $-\sum p(S) \cdot \log_2(p(S))$ is the largest possible (this method was originally proposed in [9]; for detailed derivation of this method, see, e.g., [4,19,7]).

In this approach, it is natural to take the value $f_\&(\mu_1(s), \mu_2(s))$ as the membership function $\mu_{C_1 \cap C_2}(s)$ corresponding to the intersection $C_1 \cap C_2$, where for each a and b, the value $f_\&(a,b)$ is defined as follows: As a Universum A, we take a 2-element set $A = \{E_1, E_2\}$; for each set $S \subseteq A$, we define $p(S)$ as a probability that all statements $E_i \in S$ are true and all statements $E_i \notin S$ are false: e.g., $p(\emptyset) = P(\neg E_1 \& \neg E_2)$ and $p(\{E_1\}) = P(E_1 \& \neg E_2)$, where $\neg E$ denotes the false of statement E. We consider only distributions for which $P(E_1) = a$ for $P(E_2) = b$. In terms of $p(S)$, this means that $p(\{E_1\}) + p(\{E_1, E_2\}) = a$ and $p(\{E_2\}) + p(\{E_1, E_2\}) = b$. Among all probability distributions $p(S)$ with this property, we choose the one p_{ME} with the maximal entropy. The value $P_{ME}(E_1 \& E_2)$ corresponding to this distribution p_{ME} is taken as the desired value of $f_\&(a,b)$. Similarly, we can define $f_\vee(a,b)$ as $P_{ME}(E_1 \vee E_2)$ for the MaxEnt distribution $p_{ME}(S)$. We will call the resulting operations $f_\&(a,b)$ and $f_\vee(a,b)$ *MaxEnt operations.*

To find $f_\&(a,b)$ and $f_\vee(a,b)$, we thus have to solve a conditional nonlinear optimization problem with four unknowns $p(S)$ (for $S = \emptyset$, $\{E_1\}$, $\{E_2\}$, and $\{E_1, E_2\}$). This problem has the following explicit solution (see, e.g., [18]):

Proposition 2.2. For $\&$ and $\vee$, the MaxEnt operations are $f_\&(a,b) = a \cdot b$ and $f_\vee(a,b) = a + b - a \cdot b$.

These operations have also been originally proposed by Zadeh. It is worth mentioning that both the pair of operations $(\min(a,b), \max(a,b))$ coming from the extremal approach and this pair of MaxEnt operations are in good accordance with the general group-theoretic approach [14,16,17] for describing operations $f_\&$ and $f_\vee$ which can be optimal w.r.t. different criteria.

In [18], a general description of MaxEnt operations is given. For example, for implication $A \to B$, the corresponding MaxEnt operation is $f_\to(a,b) = 1 - a + a \cdot b$, which coincides with the result of a step-by-step application of MaxEnt operations $f_\&$ and $f_\vee$ to the formula $B \vee \neg A$ (which is a representation of $A \to B$ in terms of &, $\vee$, and $\neg$).

For equivalence $A \equiv B$, the corresponding MaxEnt operation is $f_\equiv(a,b) = 1 - a - b + 2a \cdot b$. Unlike $f_\to$, the resulting expression *cannot* be obtained by a step-by-step application of $f_\&$, $f_\vee$, and $f\neg$ to any propositional formula.

Similar formulas can be obtained for logical operations with three, four, etc., arguments. In particular, the MaxEnt analogues of $f_\&(a,b,\ldots,c)$ and $f_\vee(a,b,\ldots,c)$ turn out to be equal to the results of consequent applications of the corresponding *binary* operations, i.e., to $a \cdot b \cdot \ldots \cdot c$ and to $f_\vee(\ldots(f_\vee(a,b),\ldots),c)$.

2.4.3. Extremal and MaxEnt approaches are not always applicable. At first glance, MaxEnt seems to be a reasonable approach to choosing "and" and "or" operations. It has indeed lead to successful expert systems (see, e.g., Pearl [26]). Unfortunately, the resulting operations are not always the ones we need.

For example, if we are interested in *fuzzy control* applications (see, e.g., [23,10,24]), then it is natural to choose "and" and "or" operations for which the resulting *control* is the most stable or the most smooth. Let us describe this example in some detail.

In some control problems (e.g., in *tracking* a spaceship), we are interested in the most *stable* control, i.e., in the control that would, after the initial deviation, return us within the prescribed distance of the original trajectory in the shortest possible time. We can take this time as the measure of stability.

In some other problems, e.g., in *docking* a spaceship to the Space Station, we are not that worried about the time, but we want the controlled trajectory to be as smooth as possible, i.e., we want the derivative dx/dt to be as small as possible. In such problems, we can take, e.g., the "squared average" of the derivative $\int (dx/dt)^2\, dt$ as the numerical characteristic of smoothness that we want to minimize.

The simplest possible controlled system is a 1D system, i.e., a system characterized by a single parameter x. For 1D systems, fuzzy control rules take the form $A_1(x) \to B_1(u), \ldots, A_n(x) \to B_n(u)$, where $A_i(x)$ and $B_i(u)$ are fuzzy predicates characterized by their membership functions $\mu_{Ai}(x)$

and $\mu_{Bi}(u)$. For these rules, the standard fuzzy control methodology prescribes the use of a control value $\bar{u} = (\int u \cdot \mu(u)\,du)/(\int \mu(u)\,du)$, where $\mu(u) = f_\vee(r_1, \ldots, r_n)$ and $r_i = f_\&(\mu_{Ai}(x), \mu_{Bi}(u))$.

The resulting control $\bar{u}(x)$, and therefore, the smoothness and stability of the resulting trajectory $x(t)$, depend on the choice of "and" and "or" operations. It turns out that thus defined control is the most *stable* if we use $f_\&(a,b) = \min(a,b)$ and $f_\vee(a,b) = \min(a+b,1)$, while the *smoothest* control is attained for $f_\&(a,b) = ab$ and $f_\vee(a,b) = \max(a,b)$ (for proofs, see, e.g., [17,31]). None of these pairs can be obtained by using extremal or MaxEnt approaches (some of these operations actually correspond to the *minimal* entropy [28,29]).

3. Random sets help in choosing appropriate actions. One of the main objectives of expert systems is not only to *describe* the current state of the world, but also to provide us with reasonable *actions*. Let us show that in choosing an appropriate action, random sets are also very helpful.

3.1. Random Sets and Fuzzy Optimization. The simplest case of choosing an action is when we have the *exact* objective function. Let us describe such a situation in some detail. We have a set A of all actions that are (in principle) possible (i.e., about which we have no reasons to believe that these actions are impossible). For each action $s \in A$, we know exactly the resulting gain $f(s)$. We also know that our knowledge is only partial; in particular, as we learn more about the system, some actions from the set A which we now consider possible may turn out to be actually impossible. In other words, not all actions from the set A may be actually possible.

If we know exactly which actions are possible and which actions are not, then we know the set C of possible actions, and the problem of choosing the best action becomes a conditional optimization problem: to find $s \in A$ for which $f(s) \to \max$ under the condition that $s \in C$. In many real-life situations, we only vaguely know which actions are possible and which are not, so the set C of possible actions is a "vague" set. If we formalize C as a fuzzy set (with a membership function $\mu(s)$), we get a problem of optimizing $f(s)$ under the fuzzy condition $s \in C$. This informal problem is called *fuzzy optimization*. Since we are not sure what the conditions are, the answer will also be vague. In other words, the answer that we are looking for is not a single state s, but a *fuzzy set* $\mu_D(s)$ (that becomes a one-element crisp set when C is a crisp set with a unique maximum of $f(s)$).

There are several heuristic methods of defining what it means for a fuzzy set $\mu_D(s)$ to be the solution of the fuzzy optimization problem. Most of these methods are not well justified. To get a well justified method, let us use the random set interpretation of fuzzy sets. In this interpretation, $\mu(s)$ is interpreted as the probability that $s \in S$ for a random set S, and $\mu_D(s)$ is the probability that a conditional maximum of the function f on a set S is attained at s, i.e., the probability

$$P(s \in S \,\&\, f(s) = \max_{s' \in S} f(s')).$$

Proposition 3.1. [2] Let C be a fuzzy set with a membership function $\mu(s)$ and let $f : A \to \mathbb{R}$ be a function. Then, the following is true:

- For every random set $p(S)$ that represents the fuzzy set C, and for every $s \in A$,

$$P(s \in S \,\&\, f(s) = \max_{s' \in S} f(s')) \le \mu_{DP}(s),$$

  where

$$\mu_{DP}(s) = \min(\mu(s), 1 \;-\; \sup_{s' :\, f(s') > f(s)} \mu(s')).$$

- For every $s \in A$, there exists a random set $p(S)$ that represents C and for which

$$P(s \in S \,\&\, f(s) = \max_{s' \in S} f(s')) \le \mu_{DP}(s).$$

Just like Proposition 2.1 justified the choice of min and max as $\&-$ and $\vee-$operations, this result justifies the use of $\mu_{DP}(x)$ as a membership function that describes the result of fuzzy optimization.

3.2. Random sets are helpful in intelligent control. A slightly more complicated case is when the objective function is *not* exactly known. An important class of such situations is given by *intelligent control* situations [23]. Traditional expert systems produce a list of possible actions, with "degree of certainty" attached to each. For example, if at any given moment of time, control is characterized by a single real-valued variable, the output of an expert system consists of a number $\mu(u)$ assigned to every possible value u of control; in other words, this output is a membership function. In order to use this output in automatic control, we must select a single control value u; this selection of a single value from a fuzzy set is called a *defuzzification*.

The most widely used defuzzification method (called "centroid") chooses the value $\bar{u} = (\int u \cdot \mu(u)\, du)/(\int \mu(u)\, du)$. Centroid defuzzification is a very successful method, but it is usually only heuristically justified. It turns out that the random-set interpretation of fuzzy sets (described above), together with the MaxEnt approach, leads exactly to centroid defuzzification; for precise definitions and proofs, see [18].

4. Case Study: Technical Diagnostics.

4.1. Real-life problem: Which tests should we choose. A typical problem of technical diagnostics is as follows: the system does not work, and we need to know which components malfunction. Since a system can have lots of components, and it is very difficult (or even impossible) to check them one by one, usually, some tests are undertaken to narrow down the set of possibly malfunctioning components.

Each of the tests brings us some additional information, but costs money. The question is: within a given budget C, how can we choose the set of test that will either solve our problem or at least, that would bring us, on average, the maximum information.

This problem and methods of solving it was partially described in [12,6,27].

To select the tests, we must estimate the information gain of each test. To be able to do that, we must first describe what we mean by *information*.

Intuitively, an information means a *decrease in uncertainty*. Thus, in order to define the notion of information, we will first formalize the notion of *uncertainty*.

4.2. To estimate the information gain of each test, we must estimate the amount of uncertainty in our knowledge. The amount of uncertainty in our knowledge is usually defined as the average number of binary ("yes"-"no") questions that we need to ask to gain the complete knowledge of the system, i.e., in our case, to determine the set F of faulty components. The only thing we know about this set is that it is non-empty, because otherwise, the system would not be faulty and we would not have to do any testing in the first place.

In the ideal situation, when for each non-empty subset F of the set of all components $\{1,\ldots,n\}$, we know the probability $p(F)$ that components from F and only these components are faulty, then we can determine this uncertainty as the *entropy* $-\sum p(F)\log_2(p(F))$. In reality, we only have a *partial* information about these probabilities, i.e., we only know a *set* $\mathcal{P}$ of possible distributions.

There exist many different probability distributions that are consistent with our information, and these distributions have different entropy. Therefore, it may happen that in say, 5 questions (in average) we will get the answer, or it may happen that we will only get the complete picture in 10 questions. The only way to guarantee that after a certain (average) amount of questions, we will get the complete information, is to take the *maximum* of possible entropies.

In mathematical terms, we thus define the *uncertainty* U of a situation characterized by the set $\mathcal{P}$ of probability measures as its *worst-case* entropy, i.e., as the *maximum* possible value of entropy for all distributions $\{p(F)\}\in \mathcal{P}$.

In particular, in technical diagnostics, the number of components n is usually large. As a result, the number 2^n is so astronomical that it is impossible to know $2^n - 1$ numbers corresponding to all possible non-empty sets F. What we usually know instead is the probability p_i of each component i to be faulty: $p_i = \sum\{p(F) \mid i \in F\}$. Since it is quite possible that two or more component are faulty at the same time, the sum of these n probabilities is usually greater than 1.

Example. Let us consider a simple example, in which a system consists of two components and each of them can be faulty with probability 20%; we will also assume that the failures of these two components are independent events.

In this example, we get:

- a non-faulty situation with probability $80\% \cdot 80\% = 0.64$;
- both components faulty with probability $20\% \cdot 20\% = 0.04$;
- the first component only faulty with probability $20\% \cdot 80\% = 0.16$; and
- the second component only faulty also with probability 0.16.

Totally, the probability of a fault is $1 - 0.64 = 0.36$. Hence:

- the probability $p(\{1,2\})$ that both components are faulty can be computed as $p(\{1,2\}) = 0.04/0.36 = 1/9$;
- the probability that the first component only is faulty is $p(\{1\}) = 0.16/0.36 = 4/9$, and
- the probability that the second component only is faulty is $p(\{2\}) = 4/9$.

Among all faulty cases, the first component fails in $p_1 = 1/9 + 4/9 = 5/9$ cases, and the second component fails also in $1/9 + 4/9 = 5/9$ cases. Here, $p_1 + p_2 = 10/9 > 1$.

This inequality is in perfect accordance with the probability theory, because when we add p_1 and p_2, we thus add some events *twice*: namely, those events in which both components fail.

We arrive at the following problem:

Definition 4.1. By an *uncertainty estimation problem for diagnostics*, we mean the following problem: *We know* the probabilities $p_1, \ldots, p_n$, and *we want to find* the maximum U of the expression $-\sum p(F)\log(p(F))$ (where F runs over all possible non-empty subsets $F \subseteq \{1, \ldots, n\}$) under the conditions that $\sum p(F) = 1$ and $\sum\{p(F) \mid i \in F\} = p_i$ for all $i = 1, \ldots, n$.

The following algorithm solves this problem:

Proposition 4.1. The following algorithm solves the uncertainty estimation problem for diagnostics:

1. Find a real number α from the equation
$$\prod_{i=1}^{n}(1+\alpha(1-p_i)) = (1+\alpha)^{n-1}$$
by using a bisection method; start with an interval $[0, 1/\prod(1-p_i)]$.
2. Compute the value
$$U = -\sum_{i=1}^{n} p_i \cdot \log\left(\frac{p_i}{1+\alpha(1-p_i)}\right) - \log_2(\alpha)\cdot\left(\sum_{i=1}^{n} p_i - 1\right).$$

(For a proof, see [12,6,27].)

4.3. The problem of test selection: To select the tests, we must estimate the information gain of each test. Now that we know how to estimate the *uncertainty* of each situation, let us describe how to estimate the *information gain* of each test t.

After each test, we usually decrease the uncertainty. For every possible outcome r of each test t, we usually know the probability $p(r)$ of this outcome, and we know the (conditional) probabilities of different components to be faulty under the condition that the test t resulted in r. In this case, for each outcome r, we can estimate the *conditional* information gain of the test t as the *difference* $I_t(r) = U - U(r)$ between the original uncertainty U and the uncertainty $U(r)$ after the test. It is, therefore, natural to define the information gain I_t of the test t as the *average* value of this conditional gain: $I_t = \sum_r p(r)\cdot I_t(r)$.

Since we know how to solve the problem of estimating uncertainty (and thus, how to estimate U and $U(r)$ for all possible outcomes r), we can, therefore, easily estimate the information gain I_t of each test t.

4.4. How to select a sequence of tests.. Usually, a single test is not sufficient for diagnostics, so, we need to select not a *single* test, but a *sequence* of tests. Tests are usually independent, so, the total amount of information that we get from a set S of tests is equal to the sum of the amounts of informations that we gain from each one of them. Hence, if we denote the total number of tests by T, the cost of the t^{th} test by c_t, and the amount of information gained by using the t^{th} test by I_t, we can reformulate the problem of choosing the optimal set of tests as follows: among all sets $S \subseteq \{1,\ldots,T\}$ for which
$$\sum_{t\in S} c_t \le C,$$
find a set for which
$$\sum_{t\in S} I_t \to \max.$$

As soon as we know the values c_t and I_t that correspond to each test, this problem becomes a *knapsack problem* well known in computer science. Although in general, this problem formulated *exactly* is computationally intractable (NP-hard) [5], there exist many efficient heuristic algorithms for solving this problem *approximately* [5,3].

From the practical viewpoint, finding an "almost optimal" solution is OK. This algorithm was actually applied to real-life technical diagnostics in manufacturing (textile coloring) [12,6] and in civil engineering (hoisting crane) [27].

5. Random sets help in describing conflicting interests. In Section 3, we considered the situations in which we either know the exact objective function $f(s)$ or at least in which we know *intuitively* which action is best. In many real-life situations, instead of a precise idea of which actions are best, we have several participants with different objective functions, and we must somehow reconcile their (often conflicting) interests.

We will show that sometimes, standard approaches of game theory (mathematical discipline specifically designed to solve conflict situations) are not working, and that in such situations, random sets present a working alternative.

5.1. Traditional Approach to Conflict Resolution. Game theory was invented by von Neumann and Morgenstern [21] (see also [25,8]) to assist in conflict resolution, i.e., to help several participants (*players*), with different goals, to come to an agreement.

5.1.1. To resolve conflicts, we must adopt some standards of behavior. A normal way to resolve a conflict situation with many players is that first, several players find a compromise between their goals, so they form a *coalition*; these coalitions merge, split, etc., until we get a coalition that is sufficiently strong to impose its decision on the others (this is, e.g., the way a multi-party parliament usually works).

The main problem with this coalition formation is that sometimes it goes on and on and never seems to stop: indeed, when a powerful coalition is formed, outsiders can often ruin it by promising more to some of its minor players; thus, a new coalition is formed, etc. This long process frequently happens in multi-party parliaments.

How to stop this seemingly endless process? In real economic life, not all outputs and not all coalitions that are mathematically possible are considered: there exist legal restrictions (like anti-trust law) and ethical restrictions (like "good business practice") that represent the ideas of social justice, social acceptance, etc. Von Neumann and Morgenstern called these restrictions the "standard of behavior". So, in real-life conflict situations, we look not for an arbitrary outcome, but only for the outcome that belongs to some *a priori* fixed set S of "socially acceptable" outcomes, i.e., outcomes that are in accordance with the existing "standard of behavior". This set

S is called a *solution*, or *NM-solution*.

For this standard of behavior to work, we must require two things:

- First, as soon as we have achieved a "socially acceptable" outcome (i.e., outcome x from the set S), no new coalition can force a change in this decision (as long as we stay inside the social norm, i.e., inside the state S).
- Second, if some coalition proposes an outcome that is not socially acceptable, then there must always exist a coalition powerful enough to enforce a return to the social norm.

In this framework, conflict resolution consists of two stages:

- first, the society selects a "standard of behavior" (i.e., a NM solution S);
- second, the players negotiate a compromise solution within the selected set S.

Let us describe Neumann-Morgenstern's formalization of the idea of "standard of behavior."

5.1.2. The notion of a "standard of behavior" is traditionally formalized as a NM solution.

General case. Let us denote the total number of players by n. For simplicity, we will identify players with their ordinal numbers, and thus, identify the set of all players with a set $N = \{1, \ldots, n\}$. In these terms, a *coalition* is simply a subset $C \subseteq N$ of the set N of all players.

Let us denote the set of all possible outcomes by X.

To formalize the notion of an NM-solution, we need to describe the enforcement abilities of different coalitions. The negotiating power of each coalition C can be described by its ability, given an outcome x, to change it to some other outcome y. We will denote this ability by $x <_C y$. In these terms, the above requirements on a "standard of behavior" S can be formalized as follows:

Definition 5.1. By a *conflict situation*, we mean a triple $(N, X, \{<_C\}_{C \subseteq N})$, where:

- N is a finite set whose elements are called *players* or *participants*;
- X is a set whose elements are called *outcomes*;
- $<_C$ for every *coalition* C (i.e., for every subset $C \subseteq N$), is a binary relation on the set X.

A set $S \subseteq X$ is called a *NM-solution* if it satisfies the following two conditions:

1. If $x, y \in C$, then for every coalition C, $x \not<_C y$.
2. If $x \notin S$, then there exists a coalition C and an outcome $y \in S$ for which $x <_C y$.

Comment. One can easily see that the definition of an NM-solution depends only on the *union*

$$< \; = \bigcup_C <_C$$

of binary relations $<_C$ that correspond to different coalitions. Thus, we can reformulate this definition as follows:

Definition 5.2. Let $(N, X, \{<_C\})$ be a conflict situation. We say that an outcome y *dominates* an outcome x (and denote it by $x < y$) if there exists a coalition C for which $x <_C y$.

Definition 5.3. A subset $S \subseteq X$ of the set of all outcomes is called a *NM-solution* if it satisfies the following two conditions:

1. if x and y are elements of S, then y cannot dominate x;
2. if x doesn't belong to S, then there exists an outcome y belonging to S that dominates x ($x < y$).

Important particular case: Cooperative games. The most thoroughly analyzed conflict situations are so-called *cooperative games*, in which cooperation is, in principle, profitable to all players. An outcome is usually described by the gains $x_1 \geq 0, \ldots, x_n \geq 0$ (called *utilities*) of all the players. In these terms, each outcome $x \in X$ is an $n-$dimensional vector $(x_1, \ldots, x_n)$ called a *payoff vector*. The total amount of gains of all the players is bounded by the maximal amount of money that the players can gain by cooperating; this amount is usually denoted by $v(N)$. In these terms, the set of all possible outcomes is the set of all vectors x_i for which $\sum x_i \leq v(N)$.

For cooperative games, the "enforcing" binary relation $<_C$ is usually described as follows: For every coalition C, we can determine the largest possible amount of money $v(C)$ that this coalition can gain in the hypothetical situation when all its members work together and all the others work against them. The function v that assigns the value $v(C)$ to each coalition C is called a *characteristic function* of the game.

The values $v(C)$ that correspond to different coalitions C must satisfy the following natural requirement: If two disjoint coalitions C and C' join forces, they can gain at least the same amount of money as when they acted separately. Thus, $v(C \cup C') \geq v(C) + v(C')$.

In terms of a characteristic function, a coalition can force the transition from x to y if the following two conditions hold:

- first, when C *can* gain for itself this new amount of money, i.e., when the total amount of money gained by the coalition C in the outcome y does not exceed $v(C)$;
- second, when all players from the coalition C gain by going from x to y ($x_i < y_i$ for all $i \in C$), and are thus interested in this transition.

Let us describe such conflict situations formally:

Definition 5.4. Let n be a positive integer and $N = \{1, \ldots, n\}$. By a *cooperative game*, we mean a function $v : 2^N \to [0, \infty)$ for which $v(C \cup C') \geq v(C) + v(C')$ for disjoint C and C'. For each cooperative game, we can define the conflict situation $(N, X, \{<_C\})$ as follows:

- X is the set of all $n-$dimensional vectors $x = (x_1, \ldots, x_n)$ for which $x_i \geq 0$ and $\sum x_i = v(N)$.
- $x <_C y$ if $\sum\{y_i \mid i \in C\} \leq v(C)$ and $x_i < y_i$ for all $i \in C$.

5.2. Sometimes, this traditional approach to conflict resolution does not work. Von Neumann and Morgenstern have shown that NM-solutions exist for many reasonable cooperative games, and have conjectured that such a solution exists for every cooperative game. It turned out, however, that there exist games without NM-solutions (see [20,25]).

At first glance, it seems that for such conflict situations, no "standard of behavior" is possible, and thus, endless coalition re-formation is inevitable. We will show, however, that in such situations, not the original idea of "standard of behavior" is inconsistent, but only its deterministic formalization, and that in a more realistic *random-set* formalization, a "standard of behavior" always exists.

5.3. A more realistic formalization of the "standard of behavior" requires random sets. Von Neumann-Morgenstern's formalization of the notion of the "standard of behavior" (described above) was based on the simplifying assumption that this notion is deterministic, i.e., that about every possible outcome $x \in X$, either everyone agrees that x is socially acceptable, or everyone agrees that x is not socially acceptable. In reality, there are many "gray zone" situations, in which different lawyers and experts have different opinions. Thus, the actual societal "standard of behavior" is best described not by a single set S, but by a *class* $\mathcal{S}$ of sets that express the views of different experts.

Some opinions (and sets S) are more frequent, some are rarer. Thus, to adequately describe the actual standard of behavior, we must know not only this class $\mathcal{S}$, but also the *frequencies* (*probabilities*) $p(S)$ of different sets S from this class. In other words, a more adequate formalization of the "standard of behavior" is not a *set* $S \subseteq X$, but a *random set* $p(S)$.

Since S is a *random* set, we *cannot* anymore demand that the resulting outcome x is socially acceptable for all the experts; what we *can* demand is that this outcome should be socially acceptable for the *majority* of them, or, better yet, for an *overwhelming majority*, i.e., that $P(\{S \mid x \in S\}) > 1 - \varepsilon$ for some (small) $\varepsilon > 0$.

Similarly, we can reformulate the definition of a NM-solution. The first condition of the original definition was that if y dominates x, then it is impossible that both outcomes x and y are socially acceptable. A natural "random" analogue of this requirement is as follows: if y dominates x, then the overwhelming majority of experts believe that x and y cannot be both socially acceptable, i.e., $P(\{S\,|\,x \in S \,\&\, y \in S\}) < \varepsilon$.

The second condition was that if x is not socially acceptable, then we can enforce socially acceptable y, i.e., if $x \notin S$, then $\exists y(y \in S \,\&\, x < y)$. We would like to formulate a "random" analogue of this notion as requiring a similar property for "almost all" elements of S's complement, but unfortunately, the set S can be non-measurable [25] and the probability measure on its complement can be difficult to define. To overcome this difficulty, let us reformulate the second condition in terms of *all* x, not only $x \notin S$. This can be easily done: the second condition means that for every x there exists a $y \in S$ that is either equal to x or dominates x. This reformulated condition can be easily modified for the case when S is a random set: for every $x \in X$, according to the overwhelming majority of experts, either x is already socially acceptable, or there exists another outcome y that is socially acceptable and that dominates x:

$$P(\{S\,|\,\exists y \in S(x < y \vee x = y\}) > 1 - \varepsilon.$$

Thus, we arrive at the definition described in the following section.

5.4. Random sets help in conflict resolution.

5.4.1. Definitions and the Main Result.

Definition 5.5. Let $(N, X, \{<_C\})$ be a conflict situation, and let $\varepsilon \in (0,1)$ be a real number. A random set $p(S)$, $S \subseteq X$, is called a *(random) $\varepsilon-$solution* if it satisfies the following two conditions:

1. if $x < y$ then $P(\{S\,|\,x \in S \,\&\, y \in S\}) < \varepsilon$;
2. for every $x \in X$, $P(\{S\,|\,\exists y \in S(x < y \vee x = y\}) > 1 - \varepsilon$.

Comment. If p is a *degenerate* random set, i.e., $p(S_0) = 1$ for some set $S_0 \subseteq X$, then p is an $\varepsilon-$solution if and only if S_0 is a NM-solution.

Proof. If $p(S)$ is degenerate, then all the probabilities are either 0 or 1; so the inequality $P(\{S\,|\,x \in S \,\&\, y \in S\}) < \varepsilon$ means that it is simply impossible that x and y both belong to S, and the fact that $P(\{S\,|\,\exists y \in S(x < y \vee x = y\}) > 1 - \varepsilon$ means that such a y really exists. □

Proposition 5.1. [13] For every cooperative game and for every $\varepsilon > 0$, there exists an $\varepsilon-$solution.

Before we prove this result about cooperative games, let us show that a similar result is true not only for cooperative games, but also for arbitrary conflict situations that satisfy some natural requirements.

5.4.2. This result holds not only for cooperative games, but for arbitrary natural conflict situations. To describe these "naturality" requirements, let us recall the definition of a *core* as the set of all non-dominated outcomes. It is easy to see that every outcome from the core belongs to every NM-solution.

Definition 5.6. We say that a conflict situation is *natural* if for every outcome x that doesn't belong to the core, there exists infinitely many different outcomes y_n that dominate x.

Motivation. If x doesn't belong to the core, this means that some coalition C can force the change from x to some other outcome y. We can then take an arbitrary probability $p \in [0,1]$ and then, with probability p undertake this "forcing" and with probability $1-p$ leave the outcome as is. The resulting outcomes (it is natural to denote them by $p*y+(1-p)*x$) are different for different values of p, and they all dominate x, because C can always force x into each of them. So, in natural situations, we really have infinitely many $y_n > x$.

Proposition 5.2. Conflict situations that correspond to cooperative games are natural.

Proof. The proof follows the informal argument given as a motivation of the naturality notion: if $x \in X$ and x is not in the core, this means that $x <_C y$ for some y and for some coalition C. But one can easily check that if $x <_C y$, then $x <_C p \cdot y + (1-p) \cdot x$ for all $p \in (0,1]$. Therefore, $x < p \cdot y + (1-p) \cdot x$ for all such p. □

Proposition 5.3. For every natural conflict situation and for every $\varepsilon > 0$, there exists an $\varepsilon-$solution.

Proof. Due to the definition of an $\varepsilon-$solution, every outcome from the core belongs to the solution with probability 1. As for other outcomes, some of them may belong to the solution, some of them may not. Let's consider the simplest possible random set with this property: this random set contains all the points from the core with probability 1, and as for all the other points, the probability that it contains each of them is one and the same (say α), and the events corresponding to different points belonging or not to this random set are independent. Formally: $P(\{S \mid x \in S\}) = \alpha$, $P(\{S \mid x \notin S\}) = 1-\alpha$, and

$$P(\{S \mid x_1 \in S, \ldots, x_k \in S, y_1 \notin S, \ldots, y_m \notin S\}) = \alpha^k \cdot (1-\alpha)^m.$$

All the values $\chi_S(x)$ of the (random) *characteristic function* of the random set S for x outside the core are independent, and this function thus describes a *white noise*.

We want to prove that for an appropriate value of α, this random set is an $\varepsilon-$solution. Let's prove the first condition first.

If $x < y$, then x cannot belong to the core, so either y belongs to the core, or both do not. If y belongs to the core, then y belongs to S with probability 1, so $P(\{S \mid x \in S \,\&\, y \in S\}) = P(\{S \mid x \in S\}) = \alpha$. Therefore, for $\alpha < \varepsilon$, we have the desired property.

If neither x, nor y belong to the core, then $P(\{S \mid x \in S \,\&\, y \in S\}) = \alpha^2$. If $\alpha < \varepsilon$, then automatically $\alpha^2 < \alpha < \varepsilon$, and thus, the first condition is satisfied.

Let's now check the second condition. If x belongs to the core then $x \in S$ with probability 1, so, we can take $y = x$. If x does not belong to the core then, according to the definition of a natural conflict situation there exist infinitely many different outcomes y_i such that $y_i > x$. If at least one of these outcomes y_i belongs to the core, then $y = y_i$ belongs to S and dominates x with probability 1. To complete the proof, it is thus sufficient to consider the case when all outcomes y_i are from outside the core.

In this case, if y_i belongs to S for some i, then for this S, we get an element $y = y_i \in S$ that dominates x. Hence, for every integer m, the desired probability P that such a dominating element $y \in S$ exists is greater than or equal to the probability P_m that one of the elements $y_1, \ldots, y_m$ belongs to S. But $P_m = 1 - P(\{S \mid y_1 \notin S, \ldots, y_m \notin S\}) = 1 - (1-\alpha)^m$, so $P \geq 1 - (1-\alpha)^m$ for all m. By choosing a sufficiently large m, we can conclude that $P > 1 - \varepsilon$.

So, for an arbitrary $\alpha < \varepsilon$, the "white noise" random set defined above is an $\varepsilon-$solution. □

5.4.3. Discussion: Introduction of random sets is in complete accordance with the history of game theory. To our viewpoint, the main idea of this section is in complete accordance with the von Neumann-Morgenstern approach to game theory. Indeed, where did they start? With zero-sum games, where if we limit ourselves to deterministic strategies, then not every game has a stable solution. Then, they noticed that in real life, when people do not know what to do, they sometimes flip coins and choose an action at random. So, they showed that if we add such "randomized" strategies, then *every zero-sum game has a stable solution.*

Similarly, when it turned out that in some games, no set of outcomes can be called socially acceptable, we noticed that in reality, whether an outcome is socially acceptable or not is sometimes not known. If we allow such "randomized" standards of behavior, then we also arrive at the conclusion that *every cooperative game has a solution.*

5.5. Fuzzy solutions are not sufficient for conflict resolution. Is it reasonable to use such a complicated formalism as generic random sets? Maybe, it is sufficient to use other, simpler formalisms for expressing experts' uncertainty like fuzzy logic. In this section, we will show that fuzzy solutions are not sufficient, and general random sets are indeed necessary.

Historical comment. A definition of a fuzzy NM-solution was first proposed in [1]; the result that fuzzy solutions are not sufficient was first announced in [11].

5.5.1. Fuzzy NM-Solution: Motivations. We want to describe the "vague" "standard of behavior" S, i.e., the standard of behavior in which for some outcomes $x \in X$, we are not sure whether this outcome is acceptable or not. In fuzzy logic formalism, our degree of belief in a statement E about which we may be unsure is described by a number $t(E)$ from the interval $[0,1]$: 0 corresponds to "absolutely false," 1 corresponds to "absolutely true," and intermediate values describe uncertainty. Therefore, within this formalism, a "vague" set S can be described by assigning, for every $x \in X$, a number from the interval $[0,1]$ that describe our degree of belief that this particular outcome x is acceptable. This number is usually denoted by $\mu(x)$, and the resulting function $\mu : X \to [0,1]$ describes a *fuzzy set.*

We would like to find a fuzzy set S for which the two properties describing NM solution are, to a large extent, true. These two properties are:

1. if $x < y$, then x and y cannot both be elements of S, i.e., $\neg(x \in S \,\&\, y \in S)$;
2. for every $x \in X$, we have $\exists y(y \in S \,\&\, (x < y \vee x = y))$.

In order to interpret these conditions for the case when S is a fuzzy set, and thus, when the statements $x \in S$ and $y \in S$ can have degree of belief different from 0 or 1, we must define logical operations for intermediate truth values (i.e., for values $\in (0,1)$).

Quantifiers $\forall x E(x)$ and $\exists x E(x)$ mean, correspondingly, infinite "and" $E(x_1) \,\&\, E(x_2) \ldots$ and infinite "or" $E(x_1) \vee E(x_2) \vee \ldots$. For most "and" and "or" operations, e.g.,, for $f_{\&}(a,b) = a \cdot b$ or $f_{\vee}(a,b) = a + b - a \cdot b$, an infinite repetition of "and" leads to a meaningless value 0, and a meaningless repetition of "or" leads to a meaningless value 1. It can be shown (see, e.g., [2]) that under certain reasonable conditions, the only "and" and "or" operations that do not always lead to these meaningless values after an infinite iteration are $\min(a,b)$ and $\max(a,b)$. Thus, as a degree of belief $t(E\&F)$ in the conjunction $E\&F$, we will take $t(E\&F) = \min(t(E), t(F))$ and, similarly, we will take $t(E \vee F) = \max(t(E), t(F))$. For negation, we will take the standard operation $t(\neg E) = 1 - t(E)$.

Thus, we can define

$$t(\forall x E(x)) = \min(t(E(x_1)), t(E(x_2)), \ldots) = \inf_x t(E(x))$$

and, similarly,

$$t(\exists x E(x)) = \max(t(E(x_1)), t(E(x_2)), \ldots) = \sup_x t(E(x)).$$

Thus, to get the degree of belief of a complex logical statement that uses propositional connectives and quantifiers, we must replace & by min, $\vee$ by max, $\neg$ by $t \to 1 - t$, $\forall$ by inf, and $\exists$ by sup.

As a result, for both conditions that define a NM-solution, we will get a degree of belief t that describes to what extent this condition is satisfied.

We can, in principle, require that the desired fuzzy set S satisfies these conditions with degree of belief at least 1, or at least 0.99, or at least 0.9, or a least t_0 for some "cut-off" value t_0. The choice of this "cut-off" t_0 is reasonably arbitrary; the only thing that we want to guarantee is that our belief that S is a solution should exceed our degree of belief that S is not a NM-solution. In other words, we would like to guarantee that $t > 1 - t$; this inequality is equivalent to $t > 1/2$. Thus, we can express the "cut-off" degree of belief as $t_0 = 1 - \varepsilon$ for some $\varepsilon \in [0, 1/2)$.

So, we arrive at the following definition:

5.5.2. Fuzzy NM-Solution: Definition and the Main Result.

Definition 5.7. Let $(N, X, \{<_C\})$ be a conflict situation, and let $\varepsilon \in [0, 1/2)$ be a real number. A fuzzy set $S \subseteq X$ with a membership function $\mu : X \to [0, 1]$ is called a *fuzzy $\varepsilon-$solution* iff the following two conditions hold:

1. if $x < y$, then $1 - \min(\mu(x), \mu(y)) \geq 1 - \varepsilon$;
2. for every $x \in X$, $\sup\{\mu(y) \,|\, x < y \text{ or } y = x\} \geq 1 - \varepsilon$.

The following result shows that this notion does not help when there is no NM-solution:

Proposition 5.4. A conflict situation has a fuzzy $\varepsilon-$solution iff it has a (crisp) NM-solution.

Proof. It is easy to check that if S is a NM-solution, then its characteristic function is a fuzzy $\varepsilon-$solution.

Vice versa, let a fuzzy set S with a membership function $\mu(x)$ be a fuzzy $\varepsilon-$solution. Let us show that the set $S_0 = \{x \,|\, \mu(x) > \varepsilon\}$ is then an NM-solution. Indeed, from 1., it follows that if $x < y$, then both x and y cannot belong to S_0: otherwise both values $\mu(x)$ and $\mu(y)$ would be $> \varepsilon$, and thus, their minimum would also be $> \varepsilon$, and $1-$ this minimum would be $< 1 - \varepsilon$.

From the second condition of Definition 5.7, it follows that no matter what small $\delta > 0$ we take, for every x, there exists a y for which $\mu(y) < 1 - \varepsilon - \delta$, and either $x < y$ or $y = x$.

But $\varepsilon < 1/2$, so $1 - \varepsilon - \delta > \varepsilon$ for sufficiently small δ; so, if we take y_0 that corresponds to such δ, we will get a $y \in S_0$ for which $x < y$ or $y = x$.

So, the set S_0 satisfies both conditions of the NM-solution and thus, it is a NM-solution. □

REFERENCES

[1] O.N. Bondareva and O.M. Kosheleva, *Axiomatics of core and von Neumann-Morgenstern solution and the fuzzy choice*, Proc. USSR National conference on optimization and its applications, Dushanbe, 1986, pp. 40–41 (in Russian).

[2] B. Bouchon-Meunier, V. Kreinovich, A. Lokshin, and H.T. Nguyen, *On the formulation of optimization under elastic constraints (with control in mind)*, Fuzzy Sets and Systems, vol. 81 (1996), pp. 5–29.

[3] Th.H. Cormen, Ch.L. Leiserson, and R.L. Rivest, *Introduction to algorithms*, MIT Press, Cambridge, MA, 1990.

[4] G.J. Erickson and C.R. Smith (eds.), *Maximum-entropy and Bayesian methods in science and engineering*, Kluwer Acad. Publishers, 1988.

[5] M.R. Garey and D.S. Johnson, *Computers and intractability: A guide to the theory of NP-completeness*, W.F. Freeman, San Francisco, 1979.

[6] R.I. Freidzon*et al.*, *Hard problems: Formalizing creative intelligent activity (new directions)*, Proceedings of the Conference on Semiotic aspects of Formalizing Intelligent Activity, Borzhomi–88, Moscow, 1988, pp. 407–408 (in Russian).

[7] K. Hanson and R. Silver, Eds., *Maximum Entropy and Bayesian Methods, Santa Fe, New Mexico, 1995*, Kluwer Academic Publishers, Dordrecht, Boston, 1996.

[8] J.C. Harshanyi, *An equilibrium-point interpretation of the von Neumann-Morgenstern solution and a proposed alternative definition*, In: *John von Neumann and modern economics*, Claredon Press, Oxford, 1989, pp. 162–190.

[9] E.T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. 1957, vol. 106, pp. 620–630.

[10] G.J. Klir and Bo Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, NJ, 1995.

[11] O.M. Kosheleva and V.Ya. Kreinovich, *Computational complexity of game-theoretic problems*, Technical report, Informatika center, Leningrad, 1989 (in Russian).

[12] V.Ya. Kreinovich, *Entropy estimates in case of a priori uncertainty as an approach to solving hard problems*, Proceedings of the IX National Conference on Mathematical Logic, Mathematical Institute, Leningrad, 1988, p. 80 (in Russian).

[13] O.M. Kosheleva and V.Ya. Kreinovich, *What to do if there exist no von Neumann-Morgenstern solutions*, University of Texas at El Paso, Department of Computer Science, Technical Report No. UTEP-CS-90-3, 1990.

[14] V. Kreinovich, *Group-theoretic approach to intractable problems*, In: Lecture Notes in Computer Science, Springer-Verlag, Berlin, 1990, vol. 417, pp. 112–121.

[15] V. Kreinovich *et al.*, *Monte-Carlo methods make Dempster-Shafer formalism feasible*, in [32], pp. 175–191.

[16] V. Kreinovich and S. Kumar, *Optimal choice of &- and $\vee$-operations for expert values*, Proceedings of the 3rd University of New Brunswick Artificial Intelligence Workshop, Fredericton, N.B., Canada, 1990, pp. 169–178.

[17] V. Kreinovich *et al.*, *What non-linearity to choose? Mathematical foundations of fuzzy control*, Proceedings of the 1992 International Conference on Fuzzy Systems and Intelligent Control, Louisville, KY, 1992, pp. 349–412.

[18] V. Kreinovich, H.T. Nguyen, and E.A. Walker, *Maximum entropy (MaxEnt) method in expert systems and intelligent control: New possibilities and limitations*, In: [7].

[19] V. Kreinovich, *Maximum entropy and interval computations*, Reliable Computing, vol. 2 (1996), pp. 63–79.

[20] W.F. Lucas, *The proof that a game may not have a solution*, Trans. Amer. Math. Soc., 1969, vol. 136, pp. 219–229.

[21] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton University Press, Princeton, NJ, 1944.

[22] H.T. Nguyen, *Some mathematical tools for linguistic probabilities*, Fuzzy Sets and Systems, vol. 2 (1979), pp. 53–65.

[23] H.T. Nguyen *et al.*, *Theoretical aspects of fuzzy control*, J. Wiley, N.Y., 1995.

[24] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 1996.

[25] G. Owen, *Game theory*, Academic Press, N.Y., 1982.

[26] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo, CA, 1988.

[27] D. Rajendran, *Application of discrete optimization techniques to the diagnostics of industrial systems*, University of Texas at El Paso, Department of Mechanical and Industrial Engineering, Master Thesis, 1991.

[28] A. Ramer and V. Kreinovich, *Maximum entropy approach to fuzzy control*, Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems, College Station, December 2–4, 1992, pp. 113–117.

[29] A. Ramer and V. Kreinovich, *Maximum entropy approach to fuzzy control*, Information Sciences, vol. 81 (1994), pp. 235–260.

[30] G. Shafer and J. Pearl (eds.), *Readings in Uncertain Reasoning*, Morgan Kaufmann, San Mateo, CA, 1990.

[31] M.H. Smith and V. Kreinovich, *Optimal strategy of switching reasoning methods in fuzzy control*, in [23], pp. 117–146.

[32] R.R. Yager, J. Kacprzyk, and M. Pedrizzi (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*, Wiley, N.Y., 1994.