Embedding Probabilities, Utilities and Decisions in a Generalization of Abstract Dialectical Frameworks

Atefeh Keshavarzi Zafarghandi   A.KESHAVARZI.ZAFARGHANDI@RUG.NL
Rineke Verbrugge   L.V.VERBRUGGE@RUG.NL
Bart Verheij   BART.VERHEIJ@RUG.NL
Bernoulli Institute of Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands

Abstract

Life is made up of a long list of decisions. In each of them there exists quite a number of choices and most of the decisions are effected by uncertainties and preferences, from choosing a healthy lunch and nice clothes to choosing a profession and a field of study. Uncertainties can be modeled by probabilities and preferences by utilities. A rational decision maker prefers to make a decision with the least regret or the most satisfaction. The principle of maximum expected utility can be helpful in this issue. Expected utility deals with problems in which agents make a decision under conditions in which probabilities of states play a role in the choice, as well as the utilities of outcomes.

Argumentation formalisms could be an option to model these problems and to pick one or several alternatives.

In this paper, a new argument-based framework, numerical abstract dialectical frameworks (nADFs for short), is introduced to do so. First, the semantics of this formalism, which is a generalization of abstract dialectical frameworks (ADFs for short), based on many-valued interpretations are introduced, including preferred, grounded, complete and model-based semantics. Second, it is shown how nADFs are expressive enough to formalize standard decision problems. It is shown that the different types of semantics of an nADF that is associated with a decision problem all coincide and have the standard meaning. In this way, it is shown how the nADF semantics can be used to choose the best set of decisions.

Keywords: argumentation, abstract dialectical frameworks, utility, probability, decision problem, expected utility theory

1. Introduction

During life, people are faced with a long series of decisions. A good decision may lead to a cure for a disease, to an investment in a proper project by a business person, to a judgment in a crime case, and to a fair debate. Definitely, different decisions that are made by an agent yield different consequences. At the moment of decision making, we are usually not certain of what is the consequence of our deci-

sions, but we may know the set of possible consequences that our decision can lead to. That is, we usually make decisions under uncertainty. The uncertainty mostly arises because of external factors that are out of control of agents, which are called *state* such as the probability of needing to undergo emergency surgery.

Assume that Maryam wants to travel abroad. She wants to decide whether or not to buy an international health insurance by spending 100 euros. The decision depends on some factors. Here, the external factor is the probability of having to undergo emergency surgery abroad. For example, if Maryam had a heart attack recently, the need for health insurance abroad is higher than for healthy people. Maryam's decision leads to either: 1) buying an international insurance for 100 euros and needing it when she is abroad; 2) losing 100 euros because of buying an international health insurance and not needing it; 3) needing an emergency surgery without any insurance, that is, she has to spend at least $10,000$ euros; 4) not buying international health insurance and not needing it, that is, spending nothing. Another factor with crucial importance in making decisions is the preferences that Maryam has on different consequences, which are called *outcomes*. Maryam prefers not to spend any money for an insurance and not to undergo emergency surgery to other outcomes, however, in the case that she needs emergency surgery abroad, she prefers to spend 100 euros rather than at least $10,000$ euros.

Maryam can choose among actions (buying an international health insurance or not), but she does not have any control over the states (having to undergo emergency surgery abroad or not). However, the probability of occurrence of each state has an effect on her decision. Actually, if a state of the world can be affected by an agent, it is not a state in the sense of decision theory. An agent has control but not belief over actions, however, over states she/he has belief but no control.

A theory concerned with making the best decision under uncertainty is called expected utility theory [32, 33, 25, 12, 18, 21, 24]. The expected utility of each decision or action is the weighted average of utilities of the possible outcomes, where *utility* is a numerical measure of preference of outcomes from an agent point of view, representing the agent's

desire. These utilities are weighted by the probability of the state that leads to that outcome for a specific action.

Although there exist many formalisms, solvers and automated methods in decision theory such as influence diagrams [19, 22, 27], because of the importance of decision making in human life and wide variety of decision problems, new approaches of modeling and evaluating them are required.

Argumentation is a reasoning model that can help to select one or several alternative actions, or explain an already adopted decision. Several efforts have been put into the study and definition of argumentation formalisms within which the values or preferences of agents are of crucial importance for everyday reasoning [1, 3, 4, 5, 15, 20, 29, 31]. One might wonder whether an argumentation formalism can be considered for modeling and solving decision problems. Motivated by this question, we will here introduce an argumentation formalism to represent problems in which both the probability of states and utility over outcomes play a role in making decisions. Then in our future work, we generalize abstract dialectical framework (ADF) solvers to numerical abstract dialectical frameworks (nADFs) to make a decision automatically.

The main goal of this paper is to investigate how an argumentation formalism can accommodate a decision problem. We model scenarios with utility, using a formalism of argumentation that will allow us to compute the maximum expected utility of a problem with the help of semantics of that argumentation framework. To this end, we introduce *numerical Abstract Dialectical Frameworks* (nADFs for short), which are a generalization of abstract dialectical frameworks, introduced first in [8] and then revised in [9], as a generalization of Dung's argumentation frameworks (AFs) [14]. An nADF shows how the structure of arguments can be constructed from a given knowledge base and how arguments interact with each other. In argumentation formalisms like AFs and ADFs the area that deals with evaluating arguments is called semantics. Semantics are criteria used to select subsets of available arguments that satisfy desirable properties. We follow the same way in our work to choose the best action in the nADF which is constructed based on a decision problem. We do not claim whether our results will make decision theory computationally more efficient. The reasons why we combined decision theory with ADFs are as follows:

- Argumentation theory can shed light on the process of decision making, from modeling to evaluating a problem. ADFs are expressive formalisms in that area.

- Decision theory uses the well-known tools of probabilities and utilities, of which the relation with argumentation theory are still to be well-understood.

In nADFs as well as ADFs, each argument is associated with an acceptance condition. However, in contrast with ADFs, the language used to define acceptance conditions of nADFs is a variation of propositional logic allowing numerical calculation.

This paper is organized as follows. In Section 2 we summarize the relevant background. In particular, we provide a short reminder on decision problems, expected utility theory and ADFs. In Section 3, the structures of numerical abstract argumentation frameworks, which are generalization of ADFs, are introduced. Semantics of nADFs are defined based on many-valued interpretation on rational numbers of the unit interval. In Section 4, we investigate how nADFs can be used to model decision problems, that is, how an nADF can be constructed from a given decision problem. Then, we show that in the constructed nADF all semantics collapse to the same set of interpretation. Moreover, it is shown how this unique interpretation can be used to choose the best action. Finally, in Section 5 we will summarize and conclude the presented results and refer to the open questions we would like to address next. Moreover, we compare nADFs with two argumentation formalisms, ADFs and weighted abstract dialectical frameworks [11], which form a generalization of ADFs.

2. Background

In this section we summarized, decision problems, expected utility theory and abstract dialectical frameworks.

2.1. Decision Problems

Decision making under uncertainty infuses the life of every decision maker, which can be an individual, an organization or a society. To say that a decision is made by a decision maker, called an *agent*, means an action among the set of *actions A* is chosen to be done. Uncertainty in decision making means an available action may lead to the set of *outcomes O*. The outcome of each decision is also influenced by some external factors which are called *states S*.

Following the example introduced in the introduction, Maryam can choose whether to buy a health insurance. The consequences of her decision depend on whether she gets emergency surgery abroad. That is, Maryam's decision depends on the probability of getting an emergency surgery. Beyond the probability of states, Maryam's decision depends on her preferences on the consequences. For instance, she prefers not to buy a health insurance and not to get a surgery to other consequences. However, she prefers to spend 100 euros to buy a health insurance rather than to spend at least 10,000 euros to get emergency surgery. The basic model of decision under uncertainty is a table or matrix in which the columns are labeled with states and the rows are labeled with actions and the consequence of picking an action in each state is an outcome, as depicted in Figure 1. The notation $o_1 >_p o_2$ means an agent strictly

$$
\begin{array}{ccccc}
\text{state} \rightarrow & s_1 & s_2 & \cdots & s_n \\
\text{act} \downarrow & & & & \\
a_1 & o_{11} & o_{12} & \cdots & o_{1n} \\
\vdots & & & & \\
a_m & o_{m1} & o_{m2} & \cdots & o_{mn}
\end{array}
$$

Figure 1: The table of a decision-making problem

prefers o_1 to o_2, $o_1 \sim_p o_2$ means o_1 and o_2 are equally preferred by an agent or an agent is indifferent between o_1 and o_2, and $o_1 \succeq_p o_2$ means o_1 is preferred at least as much as is o_2. The preference relation $\succeq_p$ over the set of outcomes is called *rational* iff it is transitive and complete. The technical name for the value of a possible outcome is *utility*. In [6, 28], utility is interpreted as a measure of pleasure or happiness. Contemporary decision theorists typically interpret utility as a measure of preference [32, 33, 26]. That is, it is not the case that an agent prefers outcome o_1 over o_2 because o_1 generates a higher utility than o_2. But for an agent, o_1 has a higher utility than o_2 because she/he prefers o_1 to o_2.

Definition 1 *Given $\succeq_p$ a rational order over the finite set of outcomes O. A function $u : O \rightarrow \mathbb{R}$ is called a utility function that represents $\succeq_p$ if, for every two outcomes o_1 and o_2, $u(o_1) \geq u(o_2)$ iff $o_1 \succeq_p o_2$.*

In Cantor's result characterizing dense order, dating from around 1895, it is presented that a binary relation $\succeq_p$ over a finite set can be represented by a real-valued function u if and only if $\succeq_p$ is a rational order. Note that in the current paper, utility functions are defined over $[0,1] \cap \mathbb{Q}$, in which $\mathbb{Q}$ denotes the set of rational numbers. A decision problem is formally defined in Definition 2.

Definition 2 *A decision problem is a tuple (A, S, O, p, u) where:*

- *A is a finite set of actions that can be chosen by an agent;*

- *S is a finite set of states;*

- *O is a finite set of outcomes;*

- *p is a probability function on states, $p : S \rightarrow [0,1]$ such that $\Sigma_s p(s) = 1$;*

- *u is a utility function on outcomes, $u : O \rightarrow [0,1] \cap \mathbb{Q}$.*

The criterion that deals with the analysis of situations where individuals must make a decision without knowing which outcomes may result from that decision (act) is called expected utility, which was first introduced by Daniel Bernoulli in his work on a paradox of probability [2]. Expected utility theory (EUT) states that a decision maker chooses among actions A under uncertainty by comparing the expected utility [18, 21, 33] of each action computed as the sum of the utilities of outcomes which are weighted by states respective probability. EUT is a standard theory of individual choice under uncertainty. The expected utility theory says the higher the expected utility of an action is, the better to be chosen. The expected utility of each act $a \in A$ depends on two features of the problem: The value of each outcome $o \in O$ forms an agent's standpoint, the utility of an outcome, $u(o)$; and the probability of each state s conditional on action a and outcome o is represented as $p(s|a,o)$.

Definition 3 *Let A be a set of acts that could be chosen by an agent, S a set of states, and O a set of outcomes. The expected utility of $a \in A$ is defined as:*

$$EU(a) = \Sigma_{o \in O} p(s|a,o) u(o)$$

In expected utility theory, probability can be interpreted as subjective estimate by the individual or as objectively obtained from relevant (past) data. The former is a measure of individual degrees of belief as described in [23, 25]. However, probability can be interpreted as an objective chance as in [33, 32]. In the current work, $p(s|a,o)$ is the probability of state s that, when combined with the act a, leads to the outcome o. The principle of maximum expected utility (MEU) says that a rational agent should choose the action that belongs to the set of actions with maximum expected utility. An action a belongs to the set of *maximum expected utility* if for each $a' \in A$, $EU(a) \geq EU(a')$.

2.2. Abstract Dialectical Frameworks

An abstract dialectical framework (ADF) is a directed graph in which nodes represent statements, positions or arguments. Links indicate relations among nodes that can be beyond simple attack. The conditions under which a node n is accepted are indicated by an acceptance condition C_n attached to the argument, which is a function from a subset of $par(n)$ to one of the truth value $\mathbf{t}$ or $\mathbf{f}$, where $par(n) = \{a \mid (a,n) \in L\}$. ADFs are defined formally as follows.

Definition 4 *[8, 9, 10] An abstract dialectical framework (ADF) is a tuple $D = (N, L, C)$ where:*

- *N is a finite set of nodes (arguments, statements, positions);*

- *$L \subseteq N \times N$ is a set of links;*

- *$C = \{C_n\}_{n \in N}$ is a collection of total functions $C_n : (par(n) \rightarrow \{\mathbf{t}, \mathbf{f}\}) \rightarrow \{\mathbf{t}, \mathbf{f}\}$.*

Acceptance conditions in finite cases can be equivalently represented as propositional formulas using atoms from $par(n)$, that is, C is $\{\varphi_n\}_{n \in N}$, a collection of propositional

formulas φ_n. Thus, φ_n represent a truth function C_n as a propositional formula i.e., whenever $v : par(n) \to \{\mathbf{t}, \mathbf{f}\}$ is a truth function, then $C_n(v)$ is the evaluation of φ_n under v.

Definition 5 *Let $D = (N, L, C)$ be an ADF, a three-valued interpretation v is a function that maps each argument to either true ($\mathbf{t}$), false ($\mathbf{f}$) or undecided ($\mathbf{u}$). It is called a two-valued interpretation (or a total interpretation) if all arguments are mapped to either $\mathbf{t}$ or $\mathbf{f}$.*

The truth values $\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ are partially ordered by information ordering $\leq_i$ such that $\mathbf{u} <_i \mathbf{t}$ and $\mathbf{u} <_i \mathbf{f}$ and no other pair in $<_i$. The pair $(\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}, \leq_i)$ is a complete meet-semilattice with the meet operator $\sqcap$, such that, $\mathbf{t} \sqcap \mathbf{t} = \mathbf{t}$, $\mathbf{f} \sqcap \mathbf{f} = \mathbf{f}$, and returns $\mathbf{u}$ otherwise. The meet of two interpretations v and w is then defined as $(v \sqcap w)(n) = v(n) \sqcap w(n)$ for all $n \in N$. Let $\mathcal{V}_3$ be the set of all three-valued interpretations of D. The ordered pair $(\mathcal{V}_3, \leq_i)$ is a partially ordered set (poset) in which an interpretation $w \in \mathcal{V}_3$ is at least as informative as another interpretation $v \in \mathcal{V}_3$, denoted by $v \leq_i w$, if $v(n) \leq_i w(n)$ for each $n \in N$. The set of all total interpretations that extends v is denoted by $[v]_2 = \{w \in \mathcal{V}_2 \mid v \leq_i w\}$ such that $\mathcal{V}_2$ is a set of all total interpretations. When the logical language which is used in acceptance conditions is a language of propositional logic, for any interpretation v we have $C_n(v) = v(\varphi_n)$. That is, the acceptance condition C_n evaluates v just as partial evaluation of φ_n under v.

Semantics for ADFs are introduced based on *characteristic operator* Γ_D over three-valued interpretations, which maps interpretations to interpretations. For an ADF D, given an interpretation v (for D) Γ_D is defined as follows:

$$\Gamma_D(v) = v' \text{ such that } v'(n) = \prod \{C_n(w) \mid w \in [v]_2\}$$

That is, given an interpretation v, for each argument n, the operator returns the meet of all total extension of v on n.

It is shown in [9] that the semantics of ADFs are generalization of AFs semantics. Different semantics reflect different types of point of view about the acceptance of arguments. Semantics of ADFs are based on the concept of admissibility. In ADFs an admissible interpretation does not consist of any unjustifiable information.

In particular, an interpretation v for a given ADF D is called admissible iff $v \leq_i \Gamma_D(v)$; it is complete iff $v = \Gamma_D(v)$; it is grounded iff is the $\leq_i$-least fixed-point of Γ_D; it is preferred iff it $\leq_i$-maximal admissible (resp. complete); it is two-valued model iff it is two-valued and $\forall n \in N : v(n) = v(\varphi_n)$.

The intuition of semantics which are defined in ADFs are as follows. An interpretation is called preferred if it represents maximum information about arguments without losing admissibility. An interpretation is complete if it exactly contains justifiable information. In addition, an interpretation is grounded if it collects all the information that is beyond any doubt. Further, an interpretation is two-valued model if it exactly consists of justifiable information

and does not contains any argument with undecided truth value.

3. Numerical Abstract Dialectical Frameworks

In many argumentation situations it is natural to assume n-valued acceptance degree of arguments, for $n > 3$. For instance, if one wants to investigate in a given semantics (preferred, complete, ...) whether the probability of a state is below/above of some threshold in an interpretation.

In this section we introduce a modification of ADFs called numerical abstract dialectical frameworks (nADFs). nADFs enhance ADFs by allowing numerical acceptance conditions of arguments and arithmetical computations among them. The logic used to define the acceptance conditions of arguments in nADFs is a variation of propositional logic, defined in Definition 6.

Definition 6 *This logic contains:*

- *a countably infinite number of variables: $x_1, x_2, \ldots$;*

- *a countably infinite number of constants which are called propositional atoms: $a, b, s, \ldots$;*

- *truth constants: $\top$ and $\bot$;*

- *the connectives of propositional logic: $\wedge, \vee, \neg$;*

- *binary function symbols: $\oplus$ and $\otimes$;*

- *a binary predicate symbol $\geq$ that takes entities in the domain of discourse as input while outputs are either 1 (True), 0 (False), or $\mathbf{u}$ (unknown or undecided).*

The set of terms $\{t_1, t_2, \ldots\}$ is inductively defined by the following rules:

- *any variable and any propositional atom is a term;*

- *applying of each binary function of the language on two terms t_1 and t_2 also results in a term, for instance, $t_1 \otimes t_2$;*

- *nothing else is a term.*

The set of formulas is inductively defined by the following rules:

- *any formula of propositional logic is also a formula;*

- *for arbitrary terms t_1 and t_2, $t_1 \geq t_2$ is a formula;*

- *nothing else is a formula.*

Note that interpretations of the connectives and function symbols of Definition 6 are given below in Section 3.1. An nADF is introduced in Definition 7.

Definition 7 *Let V be $[0, 1] \cap \mathbb{Q}$. An nADF is a tuple $U = (N, L, C, i)$ in which the following hold:*

- *N is a finite set of nodes;*

- $L \subseteq N \times N$ *is a set of links;*

- $C = \{C_n\}_{n \in N}$ *is a collection of total functions called acceptance conditions over V, that is, $C_n : (par(n) \to V) \to V$;*

- *i is a function called input function , $i : N' \to V$ where $N' \subseteq N$.*

Note that this definition is a generalization of Definition 4 of ADFs. In the current work, the C_n correspond to formulas of the language introduced in Definition 6 indicated by φ_n. Note that the set of links L is also implicitly determined by the acceptance conditions.

An nADF, just like an ADF, is a directed graph in which nodes indicate arguments or statements and links represent relations between statements. Each node n has an attached formula, denoted by φ_n, of the logical language introduced in Definition 6, which is a language of propositional logic with new binary functions: $\oplus$ used for the plus function and $\otimes$ used for the times function, plus a binary relation $\geq$ used for the preference relation. In Definition 7, i is a partial function on nodes; however, $i(n)$ does not appear in the acceptance conditions. It is used to indicate the input value of n and $i(n)$ is called input value of n. Input function i is used in the computation of semantics of nADFs. In our setting, $i(n)$ will be used to represent the probabilities of states and the utilities of outcomes. In general, if $i(n)$ is defined in an nADF, this does not mean that the degree of acceptance of φ_n is $i(n)$ or the initial value of n is $i(n)$, but the input value that is considered for n is $i(n)$. For instance, an atom n can be used to show the number of heart-beats per minute and $i(n)$ indicates the normal number of heart-beats per minute. The input value of normal heart-beats $i(n)$ can be compared with a person's number of heart-beats n or can be used in an equation to decide whether a person's heart beats normally, but it does not mean that n is assigned the value $i(n)$. Example 1 is an abstract nADF with three arguments. Their values in an interpretation are computed in Example 3. In Section 4 we give a concrete example in terms of decision making.

Example 1 *Let $U = (\{a,b,c\}, L, \{\varphi_a, \varphi_b, \varphi_c\}, \{i(b) = 1/5, i(c) = 4/5\})$ be an nADF in which $\varphi_a = a$, $\varphi_b = b \vee \neg a$, $\varphi_c = (a \otimes c) \geq b$, depicted in Figure 2. In this nADF, function i is defined on b and c and this means that the input value of b is $1/5$ and the input value of c is $4/5$. The acceptance condition of a says that the degree of acceptance of a depends only on a. The acceptance condition of b says the degree of acceptance of b depends on the degree of acceptance of b and $\neg a$. The acceptance condition of c is composed from the predicate $\geq$ on terms; $a \otimes c$ and b.*

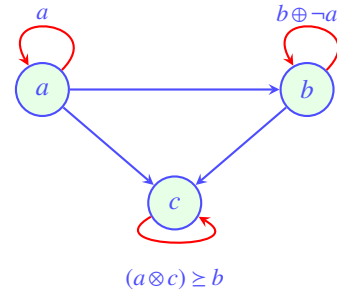nADFs are also used to answer queries, for instance, in Example 1 an nADF can be used to clarify for which amount

Figure 2: nADF of Example1

of a the acceptance condition of c is 1 (true). The computation of acceptance degrees of nodes is introduced in Section 3.1.

3.1. Semantics of nADFs

Semantics of nADFs indicate the degree of acceptance of each argument and they are introduced based on many-valued interpretation given below.

Definition 8 *A many-valued interpretation v for an nADF U is a function mapping each argument to a rational number in the unit interval or to undecided (**u**), namely $v : N \to V_{\mathbf{u}}$ such that $V_{\mathbf{u}} = ([0,1] \cap \mathbb{Q}) \cup \{\mathbf{u}\}$.*

The definition is a generalization of three-valued interpretation of ADFs (Definition 5). That is, an interpretation assigns a rational number between 0 and 1 or **u** to the nodes of an nADF. The intuition of 0, 1 and **u** is that an argument is false (rejected), true (accepted) or unknown (undecided), respectively. Any number between 0 and 1 assigned to an argument shows the degree of acceptance. Interpretations can be extended to assign a degree of acceptance to each acceptance condition. The evaluation of the acceptance condition of each argument n under a given interpretation v is a partial evaluation of φ_n under i-correction **v** introduced in Definition 9.

Definition 9 *Let $U = (N, L, C, i)$ be an nADF and let v be a many-valued interpretation. The i-correction of v under U denoted by **v** is defined such that $\mathbf{v}(n) = i(n)$ if i is defined on $n \in N$ in U and $\mathbf{v}(n) = v(n)$ otherwise.*

The evaluation of non-standard connectives, functions and the predicate $\geq$ under the i-correction of a given interpretation v in a given nADF U is as follows.

Definition 10 *Given an nADF $U = (N, L, C, i)$, let v be a many-valued interpretation. The partial evaluation of acceptance conditions under v is defined inductively as follows, in which **v** is the i-correction of v under U, lowercase a, b are propositional atoms, uppercase A, B are formulas and t_1, t_2 are terms.*

$$\mathbf{v}(A \wedge B) := min\{\mathbf{v}(A), \mathbf{v}(B)\},$$
$$\mathbf{v}(A \vee B) := max\{\mathbf{v}(A), \mathbf{v}(B)\},$$
$$\mathbf{v}(\neg A) := 1 - \mathbf{v}(A),$$
$$\mathbf{v}(a \otimes b) := \mathbf{v}(a) \times \mathbf{v}(b),$$
$$\mathbf{v}(a \oplus b) := \mathbf{v}(a) + \mathbf{v}(b),$$
$$\mathbf{v}(t_1 \geq t_2) := \begin{cases} 1 & \text{if } \mathbf{v}(t_1), \mathbf{v}(t_2) \in \mathbb{Q} \text{ and } \mathbf{v}(t_1) \geq \mathbf{v}(t_2), \\ 0 & \text{if } \mathbf{v}(t_1), \mathbf{v}(t_2) \in \mathbb{Q} \text{ and } \mathbf{v}(t_1) < \mathbf{v}(t_2), \\ \mathbf{u} & \text{if either } \mathbf{v}(t_1) \text{ or } \mathbf{v}(t_2) \text{ is undecided.} \end{cases}$$

Here, multiplication $\times$ on rational numbers of the unit interval is the standard multiplication. Moreover, $\mathbf{u} \times 0 = 0 \times \mathbf{u} = 0$ and $\mathbf{u} \times n = n \times \mathbf{u} = \mathbf{u}$ for $n \neq 0$. Also, $+$ and $-$ on rational numbers of the unit interval are the standard addition and subtraction, respectively, such that $n - \mathbf{u} = \mathbf{u} - n = \mathbf{u} + n = n + \mathbf{u} = \mathbf{u}$ for $n \in \mathbb{Q} \cap [0,1]$. Finally, $\mathbf{v}(A \vee B)$ and $\mathbf{v}(A \wedge B)$ are $\mathbf{u}$ if either $\mathbf{v}(A)$ or $\mathbf{v}(B)$ is $\mathbf{u}$.

The set of all many-valued interpretations over N is denoted by $\mathcal{V}$, i.e. $\mathcal{V} = \{v \mid v : N \rightarrow V_\mathbf{u}\}$. Interpretations can be ordered by the ordering $<_i$ which assigns a greater value to the rational numbers of the unit interval than to $\mathbf{u}$, that is, $\mathbf{u} <_i x$ for $x \in [0,1] \cap \mathbb{Q}$. The reflexive, transitive closure of $<_i$ is $\leq_i$ i.e. $\mathbf{u} \leq_i \mathbf{u}$ and $x \leq_i x$ for each $x \in [0,1] \cap \mathbb{Q}$. Thus,

$$v_1 \leq_i v_2 \text{ iff for each } n \in N, \ v_1(n) \leq_i v_2(n)$$

Note that in the current work we assume that all rational numbers are incomparable via $\leq_i$. That is, for each $x, y \in [0,1] \cap \mathbb{Q}$, neither $x <_i y$ nor $y <_i x$, which is a proper generalization of the information ordering defined in Section 2.2 for ADFs.

Definition 11 *Let $\mathcal{V}$ be the set of all many-valued interpretations and let v_1 and v_2 be two interpretations of $\mathcal{V}$. v_2 is called an extension of v_1 if $v_1 \leq_i v_2$. v_1 and v_2 are called incomparable and it is denoted by $v_1 \not\sim v_2$ if neither $v_1 \leq_i v_2$ nor $v_2 \leq_i v_1$.*

The least interpretation, which is called *trivial interpretation*, is the one that maps all arguments to undecided, which is denoted by $v_u : N \rightarrow \{\mathbf{u}\}$.

Example 2 *Let $v = \{a \mapsto \mathbf{u}, s \mapsto \mathbf{u}, o \mapsto 1/3\}, v_1 = \{a \mapsto \mathbf{u}, s \mapsto 1/10, o \mapsto 1/3\}$ and $v_2 = \{a \mapsto \mathbf{u}, s \mapsto 1/10, o \mapsto 1/2\}$ be three interpretations of $\mathcal{V}$. Since v and v_1 are equivalent on an argument which is assigned to a rational number by v and $v(s) <_i v_1(s)$, v_1 is an extension of v. However, v_2 and v are incomparable $v \not\sim v_2$, because neither $v(o) \leq_i v_2(o)$ nor $v_2(o) \leq_i v(o)$.*

Definition 12 *Let $v \in \mathcal{V}$ be a many-valued interpretation. Then an extension w of v is called a completion of v if it is a total interpretation. The set of completions of v is denoted by $[v]_c$.*

Semantics of nADFs, similarly to semantics of ADFs defined in Section 2.2, are defined based on a *characteristic*

operator Γ_U on many-valued interpretations which are ordered by ordering $\leq_i$, which shows that nADFs are appropriate generalization of ADFs. The operator Γ_U transforms interpretations of nADFs into others, $\Gamma_U : \mathcal{V} \rightarrow \mathcal{V}$. The operator takes a many-valued interpretation v as an input and returns a many-valued interpretation $\Gamma_U(v)$. For a given nADF $U = (N, L, C, i)$, the characteristic operator Γ_U on an argument n for the given interpretation v is a meet of completions of v on n.

Definition 13 *Let $U = (N, L, C, i)$ be an nADF, let v be an interpretation and let φ_n be an acceptance condition of n. The operator $\Gamma_U(v)$ yields a new interpretation:*

$$\Gamma_U(v) : N \rightarrow V_\mathbf{u} \quad with \quad n \mapsto \bigsqcap \{w(\varphi_n) \mid w \in [v]_c\}.$$

Some of the different types of semantics of nADFs are given below. These are the same as the semantics of standard ADFs when interpretations are three-valued and input function i does not define on any argument.

The intuition of defining semantics of nADFs is exactly the same as the intuition of semantics of ADFs, presented in Section 2.2.

Definition 14 *Let $U = (N, L, C, i)$ be an nADF, let v be an interpretation and let $\mathbf{v}$ be the i-correction of v under U. An interpretation v is:*

- *admissible in U iff $v \leq_i \Gamma_U(\mathbf{v})$;*

- *complete in U iff $v = \Gamma_U(\mathbf{v})$;*

- *grounded in U iff v is the $\leq_i$-least fixed point of Γ_U;*

- *preferred in U iff v is $\leq_i$-maximal admissible;*

- *model in U iff $v = \Gamma_U(\mathbf{v})$ and $\forall n \in N, \ v(n) \neq \mathbf{u}$;*

Note that in Definition 14, Γ_U is applied to $\mathbf{v}$, the i-correction of v under U. The sets of adm(U), com(U), grd(U), prf(U) and mod(U) denote the set of all admissible interpretations, complete interpretations, the unique grounded interpretation, preferred interpretations and model of U, respectively.

Example 3 *Continuing Example 1, let $v = \{a \mapsto 0, b \mapsto u, c \mapsto u\}$. The i-correction of v under U is $\mathbf{v} = \{a \mapsto 0, b \mapsto 1/5, c \mapsto 4/5\}$ since i is defined on b and c. Since none of the arguments of $\mathbf{v}$ assign to $\mathbf{u}$, $[\mathbf{v}]_c = \{\mathbf{v}\}$. Therefore, $\mathbf{v}(\varphi_b) = \mathbf{v}(b \vee \neg a) = max\{\mathbf{v}(b), \mathbf{v}(\neg a)\} = max\{i(b), 1 - \mathbf{v}(a)\} = max\{1/5, 1\} = 1$. That is, $\Gamma_U(\mathbf{v})(b) = 1$. In the same way, since $\mathbf{v}(a \otimes c) = \mathbf{v}(a) \times i(c) = 0$ and $\mathbf{v}(b) = 1/5$, $\mathbf{v}(a \otimes c) < \mathbf{v}(b)$ and $\Gamma_U(\mathbf{v})(c) = \mathbf{v}(\varphi_c) = 0$. Since $\Gamma_U(\mathbf{v}) = \{a \mapsto 0, b \mapsto 1, c \mapsto 0\}$ and $v \leq_i \Gamma_U(\mathbf{v})$, v is an admissible interpretation of U. In addition, $\Gamma_U(\mathbf{v})$ is a preferred interpretation, a complete interpretation, and a model of U. However, the revision of the trivial interpretation v_u under Γ_U is v_u, which is the unique grounded interpretation and a complete interpretation of U as well but not a preferred interpretation.*

251

4. Embedding of Decision Problems in nADFs

In this section we investigate how the standard decision problems introduced in Definition 2 can be embedded in nADFs. Since there are three main different types of arguments, namely act, state and outcome, in a decision problem, different symbols are used for distinct types of them. Circles are used to show act nodes; diamonds are clarifying state nodes; and boxes illustrate outcome statements, depicted in Figure 3.

Definition 15 *A decision problem $D = (A, S, O, p, u)$ can be modeled by nADF $U_D = (N, L, C, i)$ as follows:*

- $N = A \cup S \cup O$;
- $\varphi_s = s$ for $s \in S$;
  $\varphi_o = o$ for $o \in O$;
  $\varphi_{a_i} = \bigotimes_{i \neq k} (\bigoplus_j (s_j \otimes o_{ij}) \geq \bigoplus_j (s_j \otimes o_{kj}))$ for $a_i \in A$;
- $i(s) = p(s)$ for $s \in S$ and $i(o) = u(o)$ for $o \in O$.

Self-loops in a graph of a decision problem can be utilized as a guess whether or not to accept an argument or to which extent to accept an argument. For instance, self-loops on state nodes are used to show that the degree of acceptance of each state node depends on the probability of occurrence of that state. This notion leads to name this type of links as self-dependent links; the set of self-dependent links is denoted by R_d. Self-dependent links are reflexive relations that can be defined on N. In nADFs, any other link which is not self-dependent is called an event link; the set of event links is denoted by R_e. In the nADF depicted in Figure 3, $(s_1, s_1) \in R_d$ and $(s_1, a_1) \in R_e$. Since both the probability of occurrence of states and the utility of outcomes play a role in choosing actions, there are event links from states and outcomes to actions. The degree of acceptance of states only depends on the probability of occurrence of that state. Therefore, the acceptance condition of each state node $s \in S$ is $\varphi_s = s$. Similarly, the degree of the acceptance of outcomes only depends on the utility of that outcome from an agent point of view, that is, $\varphi_o = o$ for each $o \in O$. Thus, in each nADF there exists a self-dependent link on each state and outcome node. The acceptance condition of each action is defined in a way that the best action can be chosen via semantics. That is, $\varphi_{a_i} = \bigotimes_{i \neq k} (\bigoplus_j (s_j \otimes o_{ij}) \geq \bigoplus_j (s_j \otimes o_{kj}))$ for $a_i \in A$. To model decision problems by nADFs, function i on each state node s is equivalent with $p(s)$ and on each outcome node o is $u(o)$. In the current work, we assume that in a decision problem an agent is aware of the probability of states and her/his utility of each outcome. That is, the values of these functions are part of an input of the decision problem.

Example 4 *Continuing the example introduced in Section 1 in which Maryam wants to decide whether to buy the international health insurance the following propositional atoms are used to model this knowledge base.*

a_1: *Maryam buys the international health insurance.*

a_2: *Maryam does not buy any international health insurance.*

s_1: *Maryam gets emergency surgery when she is abroad.*

s_2: *Maryam does not get emergency surgery when she is abroad.*

o_{11}: *Maryam gets emergency surgery when she is abroad and it is paid by the health insurance company.*

o_{12}: *Maryam buys the international health insurance but she does not use it.*

o_{21}: *Maryam gets emergency surgery when she is abroad and she has to pay by herself.*

o_{22}: *Maryam does not buy the international health insurance and she does not need it.*

We assume that p is a probability function on states, that is, $p(s_1)$ shows the probability of Maryam gets emergency surgery when she is abroad and $p(s_2)$ indicates the probability of Maryam does not get emergency surgery when she is abroad. Assume that $p(s_1) = 1/10$ and $p(s_2) = 9/10$. Maryam's preference order on outcomes is as follow: $o_{22} >_p o_{12} >_p o_{11} >_p o_{21}$. The utility function u which keeps the same order, from Maryam's point of view, is: $u(o_{22}) = 7/8, u(o_{12}) = 5/8, u(o_{11}) = 1/2, u(o_{21}) = 3/8$. Therefore, this problem is modeled by decision problem $D = (\{a_1, a_2\}, \{s_1, s_2\}, \{o_{11}, o_{12}, o_{21}, o_{22}\}, \{p(s_1), p(s_2)\}, \{u(o_{11}), u(o_{12}), u(o_{21})\})$.
The corresponding nADF of D is $U_D = (N, L, C, i)$ where $N = \{a_1, a_2, s_1, s_2, o_{11}, o_{12}, o_{21}, o_{22}\}$, $C = \{\varphi_{a_1} = \bigoplus_j (s_j \otimes o_{1j}) \geq \bigoplus_j (s_j \otimes o_{2j}), \varphi_{a_2} = \bigoplus_j (s_j \otimes o_{2j}) \geq \bigoplus_j (s_j \otimes o_{1j}), \varphi_{s_1} = s_1, \varphi_{s_2} = s_2, \varphi_{o_{11}} = o_{11}, \varphi_{o_{12}} = o_{12}, \varphi_{o_{21}} = o_{21}, \varphi_{o_{22}} = o_{22}\}$ and $\{i(s_1) = p(s_1), i(s_2) = p(s_2), i(o_{11}) = u(o_{11}), i(o_{12}) = u(o_{12}), i(o_{21}) = u(o_{21}), i(o_{22}) = u(o_{22})\}$), depicted in Figure 3.

The notation X_x^v is used to show the set of arguments of X which are assigned to x by v such that $x \in X$ and X can be either A, S or O. For instance, in Example 2 that $v = \{a \mapsto \mathbf{u}, s \mapsto \mathbf{u}, o \mapsto 1/3\}$, $A_1^v = \{\}$ and $S_{1/10}^v = \{s\}$.

Example 5 *Continuing Example 4, let $v = \{a_1 \mapsto 1, a_2 \mapsto \mathbf{u}, s_1 \mapsto \mathbf{u}, s_2 \mapsto 1/5, o_{11} = \mathbf{u}, o_{12} = 5/8, o_{21} = \mathbf{u}, o_{22} = \mathbf{u}\}$. Intuitively, interpretation v want to investigate whether it is reasonable for an agent pick action a_1 when she/he only knows the probability of occurring of s_2 and utility of output o_{12}, as input values. Particularly, in the current example, Maryam wants to decide whether it is reasonable for her to buy the international health insurance when she assumed*
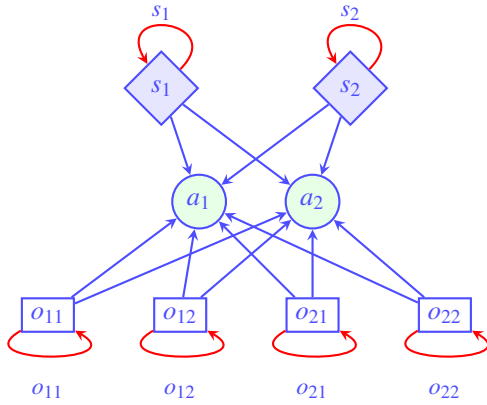
Figure 3: nADF of whether to buy international health insurance, used in Example 4

that the probability of not getting emergency surgery is 1/5 and her utility of buying the international health insurance and not using it is 5/8. To do so, we compute the revise of $\mathbf{v}$ *by* Γ_U, $v_1 = \Gamma_U(\mathbf{v}) = \{a_1 \mapsto 0, a_2 \mapsto 1, s_1 \mapsto 1/10, s_2 \mapsto 9/10, o_{11} = 1/2, o_{12} = 5/8, o_{21} = 3/8, o_{22} = 7/8\}$. *Since* v *and* v_1 *are incomparable on* a_1 *and* s_2, $v \nleq_i v_1$. *That is,* v *is not an admissible interpretation of* U_D. *That is, based on this piece of information that is presented in* v, *it is not reasonable that Maryam pick* a_1. *However,* v_1 *is the unique complete interpretation of* U_D. *That is, if the information of Maryam increases to* v_1 *about the probabilities of states and utilities of outcome, then choosing* a_2, *is a feasible choice for Maryam.*

The constructive proof of the uniqueness of the model in an nADF which is constructed based on a decision problem is investigated by Proposition 16.

Proposition 16 *Let* $D = (A, S, O, p, u)$ *be a decision problem and let* $U_D = (N, L, C, i)$ *be the corresponding nADF. Let* v *be an arbitrary interpretation of* U_D *and let* $\mathbf{v}$ *be the i-correction of* v *under* U_D. *The least fixed point of* Γ_{U_D} *on* $\mathbf{v}$ *is a model of* U_D.

Proof Let v be an arbitrary interpretation on U_D and let $\mathbf{v}$ be the *i*-correction of v under U_D. By the definition of acceptance conditions of states and outcomes nodes and the definition of Γ_{U_D}, $v_1 = \Gamma_{U_D}(\mathbf{v})$ assigns each state node to its probability, each outcome node to its utility, and after computation each action to either 1 or 0. Moreover, the value of actions, states and outcomes nodes do not change by iteration of this operator on v_1. That is, v_1 is the least fixed point of Γ_{U_D}. If $v \leq_i v_1$, then v_1 is the least fixed point of Γ_{U_D} and v is an admissible interpretation. However, if v and v_1 are incomparable, then v_1 is the least fixed point of Γ_{U_D} and v is not an admissible interpretation. Therefore, in all cases v_1 is the least fixed point of Γ_{U_D}, $v_1 = \Gamma_{U_D}(v_1)$. Since $v_1(n) \neq \mathbf{u}$ for each $n \in N$, v_1 is a model of U_D. ∎

Corollary 17 *Assume that a decision problem* $D = (A, S, O, p, u)$ *is modeled by nADF* $U_D = (N, L, C, i)$. *All semantics of* U_D *coincide.*

Proof Let v be an arbitrary interpretation and let $\mathbf{v}$ be the *i*-correction of v under U_D. By Proposition 16 the least fixed point of Γ_{U_D} on $\mathbf{v}$ is a model of U_D and by the Definition 14 it is a preferred interpretation. By the Definition 14, each grounded interpretation is a complete interpretation. It is enough to show that this complete interpretation is unique. Thus, all semantics of U_D are equivalent. Toward a contradiction, assume that $|com(U_D)| > 1$, then by the definition $\sqcap com(U_D)$ is the least fixed point of Γ_{U_D} that cannot be a model of U_D. This is a contradiction by the assumption that the least fixed point of Γ_{U_D} is a model of U_D. ∎

Theorem 18 investigates how semantics of nADFs can be used to choose the set of the best actions of decision problems of an agent.

Theorem 18 *Let* $D = (A, S, O, p, u)$ *be a decision problem, let* $U_D = (N, L, C, i)$ *be the corresponding nADF, and let* v *be the grounded interpretation of* U_D, *which is also the unique preferred interpretation, complete interpretation and model of* U_D. *Then the set* A_1^v *of actions evaluated as 1 in the grounded interpretation* v *equals the set of actions with maximal expected utility in the decision problem* D.

Proof It is enough to show that A_1^v is non-empty. Toward a contradiction, assume that A_1^v is the empty set. Since v is the grounded interpretation by Corollary 17 v is a model of U_D, as well. If A_1^v is empty, then all $a \in A$ are mapped to 0 by v. That is, for each i there exists k such that $v(\bigoplus_j(s_j \otimes o_{ij}) \succeq \bigoplus_j(s_j \otimes o_{kj})) = 0$. That is, for each a_i there exists a_k such that the expected utility of a_i is less than the expected utility of a_k. It is a contradiction by the assumption that the number of actions are finite, therefore, A_1^v is non-empty. If $a_i \in A_1^v$, by the definition of φ_{a_i} the expected utility of a_i is greater or equal with any other actions. That is, a_i is the best action to do. ∎

5. Related Works and Conclusion

In the paper, argumentation is formally connected to decision making, by developing a formal connection between argumentation formalisms and EUT. This is significant for argumentation since the general issue how argumentation relates to the standard setting of EUT is not fully understood. This paper provides a step in that understanding. The result is significant for decision making since an argumentation perspective provides insight in how to defend different positions, which remains unaddressed in theories of decision making. Generally, it is relevant to study the bridging of qualitative and quantitative theories (here: ADF as a theory of argumentation and EUT as a theory of decision making).

This paper proposes an argumentation formalism, numerical abstract dialectical frameworks (nADFs), that can model standard decision problems. In [7, 16, 17, 30], other formalisms for modeling a decision problem are presented. The ability of doing arithmetical calculation makes nADFs an applicable formalism in decision-making problems, for example, in the medical domain. Our proposal specifically generalizes abstract dialectical frameworks ADFs to allow the modeling of standard decision problems. ADFs are special cases of nADFs in which formulas are limited to the standard language of propositional logic, i is empty, and the semantics is defined based on three-valued interpretations.

Semantics of nADFs are defined based on many-valued interpretations, similarly to weighted abstract dialectical frameworks wADFs [11], which are also generalizations of ADFs. A weighted ADF is a tuple $(N, L, C, V, \leq_i)$ in which V indicates the set of truth values of arguments and $\leq_i$ is an ordering on V. That is, semantics of wADFs are also defined based on many-valued interpretation. To do calculation in nADFs, the set of truth values is fixed to the rational numbers of the unit interval and the information ordering is a generalization of the standard information ordering defined in ADFs. The language which is used in the acceptance conditions of nADFs is a variation of the language of propositional logic, with two new function symbols $\otimes$ and $\oplus$ and a predicate $\succeq$, and the partial function i in nADFs. These additions empower the formalism to represent arithmetical calculations. In general, nADFs are not a special case of wADFs. However, if in an nADF the formulas are restricted to propositional logic and the input function is empty, then it can also be viewed as a wADF in which the set of truth values V is $[0, 1] \cap \mathbb{Q}$ and $\leq_i$ is a standard generalization of information ordering in ADFs.

It is constructively proven in [13] that in each acyclic ADF, all semantics coincide. In the current work, it is shown that in each nADF that formalizes a decision problem, all semantics coincide, as well. In Section 4 it is shown how an nADF can be constructed for a decision problem for a single-agent system to choose the best action. As to future work, it can be investigated whether nADFs can be used for modeling decision problems in multi-agent systems. In addition, t would be interesting to investigate whether nADFs are powerful enough to answer queries, for instance, "for which probabilities of needing an emergency surgery Maryam will decide to buy an insurance?" Where the answer can be an interval of probabilities. Moreover, the computational complexity of decision problems in nADFs can be studied. Finally, it can be interesting to study simulation experiments that show the effectiveness of nADFs modeling decision problems.

References

[1] Leila Amgoud and Henri Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4):413–436, 2009.

[2] Kenneth J. Arrow. The use of unbounded utility functions in expected-utility maximization: Response. *The Quarterly Journal of Economics*, 88(1):136–138, 1974.

[3] Katie Atkinson and Trevor J. M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10-15):855–874, 2007.

[4] Katie Atkinson and Trevor J. M. Bench-Capon. Taking account of the actions of others in value-based reasoning. *Artificial Intelligence*, 254:1–20, 2018.

[5] Trevor J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[6] J. Bentham. *An Introduction to the Principles of Morals and Legislation*. Doubleday, Originally published in 1789, Garden City, 1961.

[7] Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1-2):63–101, 1997.

[8] Gerhard Brewka and Stefan Woltran. Abstract dialectical frameworks. In *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)*, pages 102–111, 2010.

[9] Gerhard Brewka, Hannes Strass, Stefan Ellmauthaler, Johannes Peter Wallner, and Stefan Woltran. Abstract dialectical frameworks revisited. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI 2013)*, pages 803–809, 2013.

[10] Gerhard Brewka, Stefan Ellmauthaler, Hannes Strass, Johannes Peter Wallner, and Stefan Woltran. Abstract dialectical frameworks. An overview. *IFCoLog Journal of Logics and their Applications (FLAP)*, 4(8), 2017.

[11] Gerhard Brewka, Hannes Strass, Johannes Peter Wallner, and Stefan Woltran. Weighted abstract dialectical frameworks. In *Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, pages 1779–1786, 2018.

[12] Rachael Briggs. Normative theories of rational choice: Expected utility. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition, 2017. https://plato.stanford.edu/archives/spr2017/entries/rationality-normative-utility/.

[13] Martin Diller, Atefeh Keshavarzi Zafarghandi, Thomas Linsbichler, and Stefan Woltran. Investigating subclasses of abstract dialectical frameworks. In *Computational Models of Argument: Proceedings of (COMMA 2018)*, pages 61–72, Amsterdam, 2018. IOS Press.

[14] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[15] Phan Minh Dung and Phan Minh Thang. Towards (probabilistic) argumentation for jury-based dispute resolution. In *Computational Models of Argument: Proceedings of (COMMA 2010)*, pages 171–182, Amsterdam, 2010. IOS Press.

[16] Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. Assumption-based argumentation. In Guillermo Ricardo Simari and Iyad Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 199–218. Springer, Berlin, 2009.

[17] Xiuyi Fan and Francesca Toni. Assumption-based argumentation dialogues. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 198–203, 2011.

[18] Itzhak Gilboa. *Theory of Decision under Uncertainty*. Cambridge University Press, New York, NY, 2009.

[19] Ronald A Howard and James E Matheson. Influence diagrams. *Decision Analysis*, 2(3):127–143, 2005.

[20] Anthony Hunter and Matthias Thimm. Probabilistic argument graphs for argumentation lotteries. In *Computational Models of Argument: Proceedings of (COMMA 2014)*, pages 313–324, Amsterdam, 2014. IOS Press.

[21] Philippe Mongin. Expected utility theory. In John B. Davis, D. Wade Hands, and Uskali Maki, editors, *The Handbook of Economic Methodology*, pages 171–178. Edward Elgar Publishing, Cheltenham, 1998.

[22] Scott M. Olmsted. *On Representing and Solving Decision Problems*. PhD thesis, Stanford University, US, 1985.

[23] Frank P. Ramsey. Truth and probability. In Horacio Arlo-Costa, Vincent F. Hendricks, and Johan van Benthem, editors, *Readings in Formal Epistemology*, pages 21–45. Springer, Berlin, 2016. Written in 1926, first published in 1931.

[24] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, New Jersey, NJ, 3rd edition, 2009.

[25] L. Savage. *The Foundations of Statistics*. Wiley, New York, NY, 1954.

[26] Amartya Sen. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, pages 317–344, 1977.

[27] Ross D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, 1986.

[28] Henry Sidgwick. *The Methods of Ethics*. Hackett Publishing, London, 1981. First Edition 1874.

[29] Bart Verheij. Arguments for ethical systems design. In *Legal Knowledge and Information Systems (JURIX 2016)*, pages 101–110, Amsterdam, 2016. IOS Press.

[30] Bart Verheij. Formalizing value-guided argumentation for ethical systems design. *Artificial Intelligence and Law*, 24(4):387–407, 2016.

[31] Charlotte S. Vlek, Henry Prakken, Silja Renooij, and Bart Verheij. A method for explaining Bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law*, 24:285–324, 2016.

[32] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 2nd edition, 1947.

[33] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 2007. 60th anniversary edition, with an introduction by Harold W. Kuhn and an afterword by Ariel Rubinstein.