Towards qualitative approaches to multi-stage decision making

Régis Sabbadin, Hélène Fargier and Jérôme Lang
IRIT - Université Paul Sabatier - 31062 Toulouse Cedex (France)
e-mail: {sabbadin, fargier, lang} @irit.fr

Abstract

In this paper we propose a generalisation to multi-stage decision making of Dubois and Prade's qualitative decision theory. Our framework is a qualitative, possibilistic counterpart to Markov decision processes, and the computation of an optimal policy is done in a way similar to dynamic programing. We first study in detail the case where uncertainty about the results of actions is represented by possibility distributions and goals are described in a non-fuzzy way by a subset of the set of final states. Then we extend our framework to the case where goalness is defined fuzzily, by a qualitative utility function on the set of final states.

1 Introduction

For a few years, there has been a growing interest in the Artificial Intelligence community towards the foundations and computational methods of decision making under uncertainty. This is especially relevant for applications to planning, where a suitable sequence of decisions is to be found, starting from a description of the initial world, of the available decisions and their effects, and of the goals to reach. Several authors have thus proposed to integrate some parts of decision theory into the planning paradigm; but up to now, they have focussed on "classical" models for decision making, based on *Markov decision processes* (where actions are stochastic and the satisfaction of agents expressed by a numerical, additive utility function), and its computational counterpart, *dynamic programming.* However, transition probabilities for representing the effects of actions are not always available, especially in AI applications where uncertainty is often ordinal, qualitative.

The same remark applies to utilities: it is often more adequate to represent preference over states simply with an ordering relation rather than with additive utilities. Recently, several authors have advocated this qualitative view of decision making and have proposed qualitative versions of decision theory, together with suitable logical languages for expressing preferences, namely, Boutilier [5], Tan and Pearl [21], Dubois and Prade [13] [14]. The latter propose a qualitative utility theory based on possibility theory, where preferences and uncertainty are both qualitative. Our purpose is to extend Dubois and Prade's possibilistic framework for qualitative decision theory so as to enable multiple-stage decision making [10] [17].

In order to have a synthetic view on problems and approaches pertaining to decision making under uncertainty, we may consider the following taxonomy, which gives various classes of problems (from elementary to more complex ones) when the different criteria vary. To the very general class of problems we consider, we give the generic name of "generalized Markov decision processes" (GMDP for short), since we always make the Markovian assumption that the past of the system cannot influence the choice of the policy at a given time point.

- *Temporal structure of the decision stages.* There may be only one decision stage, or there may be an ordered set of time points (stages) where decisions are to be made; this set may be either finite (*finite horizon decision making*) or discrete infinite (*infinite horizon decision making*).

- *Available knowledge of the initial state.* This knowledge may be precise (thus described by one possible initial state), probabilistic (probability distribution on the set of possible states), or possibilistic (possibility distribution on the set of possible states). At this point we recall that a possibility distribution on a set of states S is a mapping $\pi : S \rightarrow [0, 1]$, where $\pi(s)$ measures to what extent s is likely to be the actual state, ranging from 1 (completely possible) to 0 (impossible). A possibility distribution is generally assumed to be normalized (and we will make this assumption throughout the paper), *i.e.* $\exists s$ such that $\pi(s) = 1$. When π takes its value in $\{0, 1\}$, the possibility distribution is said to be *crisp* and is equivalent to a classical set. Thus, a description of the initial state by a crisp possibility distribution comes down to specifying a set of possible initial states.

- *Knowledge on the effects of actions.* Actions may be *deterministic*, meaning that for a given state s and an action a allowable in s, there is only one possible subsequent state $Result(s, a) \in X$, where X is a set of *consequences*; they are *nondeterministic* iff for each state and each action there is a *set* of possible subsequent states; they are *stochastic* (resp. *possibilistic*) when their effects are described by probability (resp. possibility) distributions. These probability (resp. possibility) distributions are denoted by $pr(s'|s, a)$ (resp. $\pi(s'|s, a)$) where s is a state and a an action, these quantities being respectively the probability (resp. possibility) of reaching the state s' from state s when action a is performed. For each fixed state s and each fixed action a, $pr(.|s, a)$ (resp. $\pi(.|s, a)$) is a probability (resp. possibility) distribution, namely, $\sum_{s'} pr(s'|s, a) = 1$ (resp. $\max_{s'} pr(s'|s, a) = 1$). From the possibilistic case we recover the nondeterministic case by allowing only crisp possibility distributions: $\pi(s'|s, a) = 1$ if and only if the transition from s to s' when a is performed is a possible transition, $\pi(s'|s, a) = 0$ otherwise. From the possibilistic case we recover the nondeterministic case by allowing only crisp possibility distributions: $\pi(s'|s, a) = 1$ if and only if the transition from s to s' when a is performed is a possible transition, $\pi(s'|s, a) = 0$ otherwise.

- *Description of the goals.* In the case of finite-horizon decision making, the final state reached is often of primary importance for the global satisfaction of the agent (it may even be the only criterion). The agent may have to achieve a crisp, non-flexible goal, *i.e.*, to reach one of the *goal states*; the notion of goal is sometimes defined in a more flexible way, by a utility function or more generally a function on an arbitrary ordered satisfaction scale. Sometimes, several heterogeneous quantities are needed to evaluate the quality of a consequence (such as cost and time); this gives rise to multicriteria decision making.

- *Role of intermediate states and actions in the global satisfaction of the agent.* Intermediate states may also receive a utility degree, or more generally a satisfaction degree, which is taken into account when computing the global satisfaction attached to a path (a succession of states). The same remark applies to actions which may have a cost which also has to be taken into account. The satisfaction degrees and costs of the different states and actions may be aggregated additively as in classical utility theory, or qualitatively by the minimum, or by other operators. The utilities of the different states reached may be weighted by a discounting factor, especially in the case of infinite horizon decision making [19].

- *Choice criterion for the policy.* An optimal policy consists in attaching to each reachable state the *best* action, following a criterion which has to be defined. In classical decision theory, this criterion consists in maximizing the expected utility. As to qualitative approaches, a possible criterion (which is used by Dubois and Prade [13] [14]) consists in making an assumption of commensurability between the uncertainty and the satisfaction scales and then maximizing a "pessimistic" qualitative utility (see Section 3) – but alternative methods are possible, including partial ordering relations.

- *Observability.* A GMDP is *fully observable* if the state of the world is known at each step of the process, *non-observable* iff no further knowledge can be gathered about the state of the world after each action is performed, and *partially observable* if the agent may only have an incomplete knowledge of the state of the world. In the latter case, some tests (or knowledge-gathering actions) may be available to the agent, who uses their results to maintain his/her beliefs about the current state (represented by a set of states, or a probability distribution, or a possibility distribution...).

We name the different models we are going to discuss according to four of the abovementioned criteria, namely, the temporal structure of the decision stages (respectively 1, N and ∞ for one / a finite number / an infinite discrete number of decision stages), the uncertainty model for the results of actions (D for deterministic, ND for nondeterministic, pr for stochastic, π for possibilistic), the description of the goals (G for crispness, $+$ for additive utilities, min for qualitative utilities combined by min), and the choice criterion (for instance $\bar{u}$ for maximum expected utility; others criteria are considered later on). Full observability is assumed everywhere (and therefore omitted). With these notations, the "standard" approach to multistage decision under uncertainty, namely, fully observable Markov decision processes, correspond to the 4-uple $\langle N, pr, +, \bar{u} \rangle$.

When both uncertainty and goalness are represented by means of fuzzy sets (over a universe of states and consequences, respectively), the quality of a decision may be evaluated, in the general case, by a fuzzy qualitative utility degree $\tilde{u}(a)$, where $\mu_{\tilde{u}(a)}(\alpha)$ (for $0 \leq \alpha \leq 1$) represents the possibility that decision a leads to a consequence whose utility degree is α. Before giving some details about the general case, let us focus on particular cases. In the following, S is the universe of all possible initial states, X the universe of all possible consequences, and a decision a is a mapping from S to X.[1] . Uncertainty about the initial state is represented by a normalized possibility distribution

$$\pi : S \to [0,1]$$

while goalness is represented by a qualitative utility function

$$u : X \to [0,1]$$

where $u(x) = 0$ (resp. $u(x) = 1$) means that x is completely unsatisfactory (resp. satisfactory).

The fuzzy utility degree evaluating the goodness of the decision $a : S \to X$ is defined exactly the same way as Zadeh's compatibility degree between a fuzzy statement and an uncertain state of facts [26] [12]:

$$\mu_{\tilde{u}(a)}(\alpha) = \sup_{\substack{s \in S \\ u(a(s))=\alpha}} \pi(s)$$

In the general case, this fuzzy utility is any fuzzy number on the utility scale $[0,1]$. Let us look briefly at particular cases obtained by adding some restrictive hypotheses on the nature of the uncertainty on the initial state and/or the expression of the goals.

- *no uncertainty, binary goals.*
  In this case the initial state s_0 is known with a complete precision, and goalness is defined by a partition of the set of consequences into goal states ($G \subseteq X$) and non-goal states ($\bar{G}$). When evaluating the quality of a decision a, there are only two possible cases: either $a(s_0) \in G$, and in this case $\mu_{\tilde{u}(a)}(1) = 1$ and $\forall \alpha \neq 1, \mu_{\tilde{u}(a)}(\alpha) = 0$ (a is a *good* decision) or $a(s_0) \notin G$, and in this case $\mu_{\tilde{u}(a)}(0) = 1$ and $\forall \alpha \neq 0, \mu_{\tilde{u}(a)}(\alpha) = 0$ (a is a *bad* decision). Now, the ranking of available decisions is obvious:

[1] This assumes that decisions are deterministic, the only source of uncertainty bearing on the initial state (1); however, if the initial state is precisely known and actions have possibilistic effects (2), this can be rewritten equivalently in the previous framework (1): to a given action a and an initial state s_0 whose possible effects are described by the possibility distribution $\pi(s_i|s_0, a), i = 1...n$, we associate an abstract uncertain initial state described by the possibility distribution $\pi_0 : \pi_0(s_0^i) = \pi(s_i|s_0, a)$ and the (deterministic) effect of action a on state s_0^i is s_i; this is then easily generalized without difficulty to the case of several nondeterministic actions.

good decisions are preferred to bad decisions, two good (resp. bad) decisions being equivalent.

- *qualitative uncertainty, binary goals.*
  Now we have a nonempty set of possible initial states $S^* \subseteq S$, and a set of goal states $G \subseteq X$. There are now three different cases for a given decision a: let $a(S^*) = \{a(s), s \in S^*\}$, then either $a(S^*) \subseteq G$, which means that whatever the initial state, the decision a is guaranteed to lead to a goal state (a is a good decision), either $a(S^*) \cap G = \emptyset$, which means that whatever the initial state, the decision a is guaranteed to lead to a non-goal state (a is a bad decision), or $a(S^*) \cap G \neq \emptyset$, $a(S^*) \cap \bar{G} \neq \emptyset$, which means that a is ambiguous (it is completely possible that it leads to a goal state and also completely possible that it leads to a non-goal state). This last case corresponds to the following fuzzy utility: $\mu_{\tilde{u}(a)}(0) = \mu_{\tilde{u}(a)}(1) = 1$, $\mu_{\tilde{u}(a)}(\alpha) = 0 \quad \forall \alpha \in (0,1)$. The ranking of available decisions is still obvious: a good decision is preferred to an ambiguous one, an ambiguous one to a bad one, two good (resp. ambiguous, bad) decisions being equivalent.

- *possibilistic uncertainty, binary goals.*
  Now the knowledge about the initial state is described by a normalized possibility distribution π, and goalness by a set of goal states $G \subseteq X$. The quality of a decision a will be evaluated by 2 numbers, namely the possibility and the necessity that a leads to a goal state:

$$\Pi(Good(a)) = sup_{s \in S | a(s) \in G} \pi(s)$$
$$N(Good(a)) = inf_{s \in S | a(s) \in \bar{G}} 1 - \pi(s)$$

  This corresponds to the following fuzzy utility: $\mu_{\tilde{u}(a)}(0) = 1 - N(Good(a))$, $\mu_{\tilde{u}(a)}) = \Pi(Good(a))$, and $\mu_{\tilde{u}(a)}(\alpha) = 0 \quad \forall \alpha \in (0,1)$. These degrees are standard possibility and necessity degrees, and verify thus $N(Good(a)) > 0 \Rightarrow \Pi(Good(a)) = 1$. They evaluate the extent to which the set of possible consequences of a, respectively, has a non-empty intersection with the goals, and is included in the set of goals. Obviously, when the possibility distribution π is crisp we recover the previous case (qualitative uncertainty, binary goals) where the three possible evaluations correspond respectively to $\Pi(Good(a)) = N(Good(a)) = 1$, $\Pi(Good(a)) = N(Good(a)) = 0$, and $\Pi(Good(a)) = 1$, $N(Good(a)) = 0$. The ranking of available decisions is still easy: a is at least as good as a' iff $\Pi(Good(a)) \geq \Pi(Good(a'))$ and $N(Good(a)) \geq$

$N(Good(a'))$; this is a complete ranking, because $N(Good(a)) > 0 \Rightarrow \Pi(Good(a)) = 1$ holds for all decisions.

- *no uncertainty, fuzzy goals.*
  Here the knowledge about the initial state is represented by a single state s_0 and the goalness by a fuzzy subset $\tilde{G}$ of the set of consequences $\mu_{\tilde{G}} : X \to [0, 1]$. The membership degree $\mu_{\tilde{G}}(x)$ of x to the fuzzy set of goals $\tilde{G}$ represents the qualitative utility resulting from the obtention of consequence x (in particular, $\mu_{\tilde{G}}(x) = 0$ (resp. $= 1$) means that x is completely undesirable (resp. desirable)). Note that $\mu_{\tilde{G}}(x)$ does not have to be normalized, since it may be the case that no consequence is fully satisfactory. Now, the quality of a decision a is evaluated by a single value, namely the goalness degree of the consequence obtained by applying a to the only possible initial state, i.e., $\mu_{\tilde{G}}(a(s_0))$. This corresponds to a crisp qualitative utility, namely, $\mu_{\tilde{u}(a)}(\mu_{\tilde{G}}(a(s_0))) = 1$, $\mu_{\tilde{u}(a)}(\alpha) = 0 \quad \forall \alpha \neq \mu_{\tilde{G}}(a(s_0))$. The ranking of available decisions is still easy: a is at least as good as a' iff $\mu_{\tilde{G}}(a(s_0)) \leq \mu_{\tilde{G}}(a'(s_0))$.

- *qualitative uncertainty, fuzzy goals.*
  The knowledge about the initial state is represented by $S^* \subseteq S$ and the goalness by $\mu_{\tilde{G}} : X \to [0, 1]$. The evaluation of the quality of a is evaluated by a nonempty set of possible qualitative utilities, namely, $\{\mu_{\tilde{G}}(a(s)) | s \in S^*\}$.

- the general case (*possibilistic uncertainty, fuzzy goals*). Now the quality of a decision is evaluated by the fuzzy goalness degree

$$\mu_{\tilde{G}(a)}(\alpha) = \sup_{\substack{s \in S \\ u(a(s)) = \alpha}} \pi(s)$$

While the ranking of available decisions can be done in an obvious way in the first four cases, this is far less obvious for the last two ones. Comparing two fuzzy quantities can be done in several different ways. We first eliminate partial orderings (such as $\tilde{\alpha} \geq \tilde{\beta}$ iff $\tilde{\max}(\tilde{\alpha}, \tilde{\beta}) = \tilde{\alpha}$) since they have an unsufficiently discriminating power among decisions. We also eliminate "quantitative" defuzzification methods such as averaging, since they are not in the spirit of our qualitative modelling of uncertainty and flexibility of decision processes. We are left with two remaining possibilities:

- using Dubois and Prade's *comparison indices* introduced in [11]. The use of these indices assumes that the fuzzy quantities involved are *fuzzy intervals*[2], which is far from being guaranteed! (in particular, this is never the case if the decision space is discrete, unless all possible consequences of a given decision have the same satisfaction degree). However it is possible to take the convex closure of the involved fuzzy quantities, which gives fuzzy intervals, and then to compute the comparison indices. These four indices, measuring to what extent $\tilde{\alpha}$ is greater than $\tilde{\beta}$, are recalled below:

- $\Pi_{\tilde{\alpha}}([\tilde{\beta}, +\infty)) = \sup_{\substack{u,v \\ u \geq v}} \min(\mu_{\tilde{\alpha}}(u), \mu_{\tilde{\beta}}(v))$ measures to what extent the least possible values of $\tilde{\beta}$ are smaller or equal to the greatest possible values of $\tilde{\alpha}$.
- $\Pi_{\tilde{\alpha}}(]\tilde{\beta}, +\infty)) = \sup_u \inf_{\substack{v \\ v \geq u}} \min(\mu_{\tilde{\alpha}}(u), 1 - \mu_{\tilde{\beta}}(v))$ measures to what extent the greatest possible values of $\tilde{\beta}$ are smaller or equal to the greatest possible values of $\tilde{\alpha}$.
- $N_{\tilde{\alpha}}([\tilde{\beta}, +\infty)) = \inf_u \sup_{\substack{v \\ v \leq u}} \max(1 - \mu_{\tilde{\alpha}}(u), \mu_{\tilde{\beta}}(v))$ measures to what extent the least possible values of $\tilde{\beta}$ are smaller or equal to the least possible values of $\tilde{\alpha}$.
- $N_{\tilde{\alpha}}(]\tilde{\beta}, +\infty)) = 1 - \sup_{u \leq v} \min(\mu_{\tilde{\alpha}}(u), \mu_{\tilde{\beta}}(v))$ measures to what extent the least possible values of $\tilde{\alpha}$ are greater to the greatest possible values of $\tilde{\beta}$.

These indices can then be used in order to rank n fuzzy intervals, by computing for each of them a possibility and a necessity degree of dominance. Details can be found in [11].

- Making a *qualitative commensurability assumption* between the uncertainty scale and the satisfaction scale, leading to the computation of the possibility and the necessity degree of the fuzzy event "the consequence resulting from decision s is a goal". This principle is the core of Dubois and Prade's possibilistic decision theory [13] [14]. Let S and A the set of possible states and available actions, respectively. The possible consequences of action a from s_0 is described by the possibility distribution $\pi(.|a, s_0)$ on X. The qualitative utility is a mapping u from X to $[0, 1]$. Then the qualitative value of a is measured by two

[2] A fuzzy quantity is a fuzzy subset of a real scale, here $[0, 1]$; a fuzzy interval is a convex fuzzy quantity, or equivalently, a fuzzy quantity whose all α-cuts are intervals; a fuzzy number is a unimodal fuzzy interval, i.e., with a single value having a membership degree of 1.

qualitative utility functions (which plays the same role as expected utility in standard decision theory), defined by

$$u^{PES}(s_0, a) = \min_{s \in S} \max(1 - \pi(s|s_0, a), u(s))$$

$$u^{OPT}(s_0, a) = \max_{s \in S} \min(\pi(s|s_0, a), u(s))$$

These two quantities are respectively the necessity and the possibility of a fuzzy event, namely, it can be viewed as a degree of inclusion (resp. non-empty intersection) of the fuzzy set of more or less possible situations in (resp. with) the fuzzy set of preferred outcomes [14]. $u^{PES}(s_0, a)$ is thus a pessimistic criterion, while $u^{OPT}(s_0, a)$ is an optimistic one.

Note that in the abovementioned particular cases, $u^{PES}(s_0, a)$ and $u^{OPT}(s_0, a)$ generalize the quantities proposed for evaluating decisions. Namely:
- no uncertainty, crisp goals: $u^{OPT}(s_0, a) = u^{PES}(s_0, a) = 1$ if a is good, $= 0$ if a is bad.
- qualitative uncertainty, crisp goals: idem, plus $u^{OPT}(s_0, a) = 1$ and $u^{PES}(s_0, a) = 0$ if a is ambiguous.
- possibilistic uncertainty, crisp goals: $u^{OPT}(s_0, a) = \Pi(Good(a))$, $u^{PES}(s_0, a) = N(Good(a))$.
- no uncertainty, fuzzy goals: $u^{OPT}(s_0, a) = u^{PES}(s_0, a) = \mu_{\tilde{G}}(a(s_0))$.
- qualitative uncertainty, fuzzy goals:

$$u^{OPT}(s_0, a) = \max_{s \in S^*} \mu_{\tilde{G}}(a(s_0))$$

$$u^{PES}(s_0, a) = \min_{s \in S^*} \mu_{\tilde{G}}(a(s_0))$$

This deserves further comments. $u^{OPT}(s_0, a)$ is the utility of the best possible outcome when performing a, and $u^{PES}(s_0, a)$ is the utility of the worst possible outcome, also known in decision theory as the *Wald index* (see [13]).

Let us give some more details about Dubois and Prade's framework for qualitative decision theory. First, the framework can be defined with a higher level of generality, by allowing for the use of utility values not necessarily in $[0, 1]$ but in any completely ordered lattice L, where $u(s) = 1_L$ and $u(s) = 0_L$ respectively mean complete satisfaction and dissatisfaction, equipped with an order reversing function n from L to L satisfying $n(0_L) =$

1_L and $n(1_L) = 0_L$ (when $L = [0, 1]$ the prototypical order reversing function is $n(x) = 1 - x$).

The use of the indices $u^{PES}(s_0, a)$ and $u^{OPT}(s_0, a)$ to rank decisions is similar to the use of the necessity and possibility of pattern matching in fuzzy databases [15]. Both couples of indices are actually the necessity and the possibility of a fuzzy event given an uncertain initial fact modelled by a possibility distribution; in the case of qualitative decision, this fuzzy event is "a goal state is reached from the uncertainly known initial state when action a is performed"; while in the case of a fuzzy database, the fuzzy event is "the uncertain data concerning a given object satisfy the flexible request".

Now, how can we use the optimistic and pessimistic indices to rank decisions? Following Dubois and Prade, we give priority to the pessimistic index, which generalizes the well-known Wald index. The best action in s_0 is then the action a maximizing $u^{PES}(s_0, a)$. It is also possible to use the optimistic index to refine the ordering among decisions (see Section 4).

Using our notations, Dubois and Prade's framework corresponds to the 4-uple $\langle 1, \pi, min, u^{PES} \rangle$ where u^{PES} is the pessimistic utility with a commensurability assumption. Note that a similar pessimistic criterion has been proposed by Whalen [22], in terms of "disutility".

Dubois and Prade's qualitative decision theory only applies to the single-step decision case. In this paper we aim to generalize Dubois and Prade's framework to multistage decision, first (Section 2) assuming full observability, possibilistic uncertainty and crisp goals, and then (Section 3) assuming full observability, possibilistic uncertainty and qualitative utility with the abovementioned commensurability assumption. An alternative multistage generalization of qualitative decision theory has been proposed by Da Costa Pereira [8], in which the environment is non-observable – thus the policiea are unconditional sequences of actions.

We will first consider the use of the multi-stage generalisation of the pessimistic index, and we will then discuss about the possible refinement by an optimistic index (which is not as obvious as it may appear at a first glance). We will propose sketches of algorithms, in the spirit of dynamic programming. We will relate our work with some other approaches to qualitative multi-stage decision making, and we will briefly give some hints on how to further generalize our framework in several ways.

2 Multi-stage decision making with possibilistic uncertainty and crisp goals

2.1 States, policies and trajectories

Following Puterman [19], $T = \{1, 2, ..., N\}$ denotes the finite set of time points (or stages) at which decisions are to be made[3]. The set of possible states at stage $t \in T$ is denoted by S_t[4]. The initial state is s_1; since the last action takes place at time N, the last state obtained (*i.e.*, the final state, about which goals are expressed) is s_{N+1}.

Under our assumption of full observability, at stage t, the decision maker observes the system in state $s \in S_t$ and chooses an action a from the set of allowable actions at t in state s, denoted by $A_{s,t}$[5]. Since the effects of the actions are ill-known, the agent's knowledge about the subsequent state is described by a transition possibility function π_t : $\pi_t(s'|s, a)$ is the possibility that the state reached at stage $t + 1$ is $s' \in S_{t+1}$ knowing that the state obtained at stage t was s and that action $a \in A_{s,t}$ has been performed at stage t.

A *policy* consists in a collection, for each t, of a decision rule d_t mapping every state s from S_t to an action $d_t(s)$ in $A_{s,t}$. The set of allowable decision rules at stage t is denoted by D_t and called the decision set at t. It is the set of all the mappings from S_t to $A_{s,t}$. A *partial policy* $d_{t \to N} = \{d_t, d_{t+1}, ..., d_N\}$ specifies the sequence of decision rules to be used by the decision maker from the stage $t \in T$ to the end of the planning horizon. The set of all allowable partial policies from t to N will be denoted by $D_{t \to N}$. A (full) policy $d = \{d_1, d_{t+1}, ..., d_N\}$ completely specifies the sequence of decision rules to be taken from the beginning[6].

A *(full) trajectory* τ is a sequence of states obtained from stage 1 to stage $N + 1$. Thus, $\tau = (s_1, ..., s_{N+1})$; we will sometimes make use of the notation $s_i = \tau(i)$. The set of all conceivable trajectories $S_1 \times ... \times S_{N+1}$ will be denoted $TRAJ$. Given two stages t_{min} and t_{max} such that $t_{min} < t_{max}$,

[3] Infinite horizon GMDP are not considered in this paper.

[4] It often happens that the set of possible states does not vary over time, *i.e.*, $S_t = S_{t'}$ $\forall t, t'$ but we prefer to keep the subscript in all cases, to make it easier to distinguish between two identical states obtained at two different time points.

[5] It may happen that $A_{s,t}$ is independent from s, from t or both.

[6] In the Markov Decision Processes literature, policies are usually denoted by π; unfortunately, in the fuzzy set literature, possibility distributions are usually denoted by π as well. We had to make a choice, which has been guided by the belief that readers of this journal are more familiar with fuzzy sets than with MDPs.

a *partial trajectory* from t_{min} to t_{max}, denoted by $\tau_{t_{min} \to t_{max}}$ is a sequence of states $(s_{t_{min}}, s_{t_{min}+1}, ..., s_{t_{max}})$ obtained from stage t_{min} to stage t_{max}; the set of all conceivable trajectories from t_{min} to t_{max}, *i.e.*, $S_{t_{min}} \times ... \times S_{t_{max}}$, will be denoted by $TRAJ_{t_{min} \to t_{max}}$.

Lastly, in Section 2, the set of goal states is defined by a (crisp) subset G of S_{N+1}. This definition does not allow for a gradation of goals: states in G are good states (any two states in G being equally good) while states in $S_{N+1} \setminus G$ are bad states (any two of these being equally bad).

The rest of Section 2 is organized as follows. First we focus on the simple case with a single decision stage (thus considering only stages N and $N+1$); we will define, for each state s in S_N, the possibility and the necessity that a given action a performed in s leads to a goal state, which will lead us to the definition of an optimal action for each state. Then, we will switch to the general case; we will see that a given policy induces in a natural way a possibility distribution on the set of trajectories, which will lead us to the definition of an optimal policy. Then, we will see how it is possible to compute an optimal policy recursively, in a way much similar to dynamic programming. This backwards algorithm will be based on the computation, for each t from N downto 1, for each $s \in S_t$ and for each $a \in A_{s,t}$, of the two quantities $\Pi(Good_t(s,a))$ and $\mathsf{N}(Good_t(s,a))$ measuring respectively the possibility and the necessity that performing a in state s will eventually lead to a goal state *provided that an optimal policy is performed from stage* $t+1$ *on*.

2.2 The single-stage case

We first consider a state s at stage N, and an action $a \in A_{s,N}$ to be performed in s. Since the subsequent state is described by the possibilistic transition function $\pi(.|s,a)$, it is possible to compute the possibility and the necessity of the event "the subsequent state will be a goal state", denoted by $Good_N(s,a)$:

Definition 1
$\Pi(Good_N(s,a)) = \max_{s' \in G} \pi_N(s'|s,a)$;
$\mathsf{N}(Good_N(s,a)) = \min_{s' \notin G}(1 - \pi_N(s'|s,a)$

$\Pi(Good_N(s,a))$ and $\mathsf{N}(Good_N(s,a))$ are respectively the possibility and the necessity of the (crisp) subset G of S_{N+1} induced by the possibility distribution $\pi(.|s,a)$. Thus, $\Pi(Good_N(s,a))$ and $\mathsf{N}(Good_N(s,a))$ are standard possibility and necessity degrees, which satisfy the following property: $\mathsf{N}(Good_N(s,a)) > 0 \Rightarrow \Pi(Good_N(s,a)) = 1$ (1)

A cautious, pessimistic approach consists in preferring an action which maximizes the necessity to reach a goal state. While this criterion seems rather natural, it is however often not discriminating enough since there may often be no action at all leading to a goal state with some strictly positive certainty. The idea is then to discriminate further among the actions using $\Pi(Good_N(s,a))$, which is an optimistic index. This leads us to the following ranking over actions, where $a \geq_s a'$ (respectively $a >_s a'$) reads "a is at least as good as (resp. better than) a' in state s " :

Definition 2

1. *$a >_s a'$ if and only if one of these two conditions holds:*
   - $\mathsf{N}(Good_N(s,a)) > \mathsf{N}(Good_N(s,a'))$
   - $\mathsf{N}(Good_N(s,a)) = \mathsf{N}(Good_N(s,a'))$ *and* $\Pi(Good_N(s,a)) > \Pi(Good_N(s,a'))$*;*
2. *$a \geq_s a'$ if and only if $a' >_s a$ does not hold;*
3. *$a \sim_s a'$ if and only if $a \geq_s a'$ and $a' \geq_s a$.*

Obviously, the following properties hold:

- $\geq_s$ is a complete preorder;
- $a \geq_s a'$ iff $\mathsf{N}(Good_N(s,a)) \geq \mathsf{N}(Good_N(s,a'))$ and $\Pi(Good_N(s,a)) \geq \Pi(Good_N(s,a'))$;
- $a \sim_s a'$ iff $\mathsf{N}(Good_N(s,a)) = \mathsf{N}(Good_N(s,a'))$ and $\Pi(Good_N(s,a)) = \Pi(Good_N(s,a'))$

An action a will then be *optimal for s* iff there is no action a' such that $a' >_s a$. The set of optimal actions for s (at stage N) will be denoted by $A^*_{s,N}$. Due to (1), optimal actions can be characterized more intuitively by the following property: in case there is an action a leading from s to a goal state with some positive certainty, *i.e.*, $\mathsf{N}(Good_N(s,a)) > 0$, then $A^*_{s,N}$ is the set of actions for which $\mathsf{N}(Good_N(s,a))$ is maximal; otherwise, if it is the case that $\mathsf{N}(Good_N(s,a)) = 0$ for each action a, then $A^*_{s,N}$ is the set of actions for which $\Pi(Good_N(s,a))$ is maximal.

Now, for a given state s at stage N, an optimal policy will assign an arbitrary action in $A^*_{s,N}$. Then, we can define the possibility and the necessity that a goal state can be reached from s, being the corresponding possibility and necessity degrees obtained for an optimal action for s.

Definition 3
Let a^* be an arbitrary action in $A^*_{s,N}$; then
$\Pi(Good_N(s)) = \Pi(Good_N(s, a^*))$
$\mathsf{N}(Good_N(s)) = \mathsf{N}(Good_N(s, a^*))$

This definition is well-founded since $\Pi(Good_N(s, a^*))$ and $\mathsf{N}(Good_N(s, a^*))$ are constant for all actions a^* in $A^*_{s,N}$. Due to (1), it can be checked easily that the definition is equivalent to $\Pi(Good_N(s)) = \max_{a \in A_{s,N}} \Pi(Good_N(s, a))$ and $\mathsf{N}(Good_N(s)) = \max_{a \in A_{s,N}} \mathsf{N}(Good_N(s, a))$. In other terms, $\Pi(Good_N(s))$ (resp. $\mathsf{N}(Good_N(s))$) is the possibility (resp. the necessity) of the event "there is a policy which leads from s to a goal state", or equivalently, "performing an optimal action in s leads to a goal state". Knowing that for any s, $\Pi(Good_N(s))$ and $\mathsf{N}(Good_N(s))$ are equal to $\Pi(Good_N(s, a))$ and $\mathsf{N}(Good_N(s, a))$ for a given a (namely, an optimal action for s), it follows that $\Pi(Good_N(s))$ and $\mathsf{N}(Good_N(s))$ are standard possibility and necessity degrees, *i.e.*,
$\mathsf{N}(Good_N(s)) > 0 \Rightarrow \Pi(Good_N(s)) = 1.$

2.3 The multistage case: optimal policies

We are now going to generalize the notion of optimal policy to the multistage case. For this we first need to define the possibility (resp. the necessity) that a given policy leads from an initial state s to a goal state.

Definition 4
Let s_t be a state from S_t, $d_{t\to N}$ a policy from $D_{t\to N}$ and $\tau_{t+1\to N+1} = (s_{t+1}, ..., s_{N+1})$ a trajectory from $TRAJ_{t+1\to N+1}$. The possibility that trajectory $\tau_{t+1\to N+1}$ results from performing $d_{t\to N}$ from s_t on is defined by

$$\pi(\tau_{t+1\to N+1}|s_t, d_{t\to N}) = \min_{i=t...N} \pi(s_{i+1}|s_i, d_i(s_i))$$

This definition deserves some comments. First, $\pi(.|s_t, d_{t\to N})$ is a normalized possibility distribution on $TRAJ_{t+1\to N+1}$, because all transition possibility functions are normalized, and consequently, any trajectory composed only of elementary transitions with possibility 1 is itself of possibility 1. Then, defining a possibility distribution on trajectories from transition possibilities (*i.e.*, from possibilities on elementary transitions) is equivalent to defining a joint possibility distribution on a Cartesian product of sets, to each of which is attached a possibility distribution. Indeed, a trajectory can be seen as a tuple of elementary transitions (s_i, s_{i+1}) which are furthermore non-interactive, because of the Markov-like assumption that the transition possibility distribution at stage t only depends on s_t and the action performed, and not on the history of the system (namely, the previous transitions). The most usual choice is the minimum [12]; the intuition which lays behind it is that a trajectory is exactly as possible as the less possible of its elementary transitions. For the sake of simplicity, this choice will not be questioned again (except in Section 4); it is important noticing any T-norm would be a valuable choice[7], and furthermore the results and algorithms in the rest of the paper generalize easily.

It remains now to define the possibility and the necessity that a given policy applied from an initial state leads to a goal state. It is induced from the possibility distribution on trajectories by considering the possibility and the necessity measures of the set of *good trajectories*:

Definition 5
A trajectory $\tau_{t\to N+1}$ is good *iff its last state $\tau(N+1)$ is in G. $GoodTraj_{t\to N+1} \subseteq TRAJ_{t\to N+1}$ denotes the set of all good trajectories from t to $N+1$.*

Definition 6

- $\Pi(Good_t(s_t, d_{t\to N})) = \Pi(GoodTraj_{t\to N+1} | s_t, d_{t\to N})$
- $\mathsf{N}(Good_t(s_t, d_{t\to N})) = \mathsf{N}(GoodTraj_{t\to N+1} | s_t, d_{t\to N})$

Using the expression of the possibility of a trajectory given a policy, the complete expression of these degrees is

$$\begin{aligned}
\Pi(Good_t(s_t, d_{t\to N})) &= \max_{\tau_{t\to N+1} \in GoodTraj_{t\to N+1}} \pi(\tau_{t\to N+1} | s_t, d_{t\to N}) \\
&= \max_{\substack{(s_{t+1},\dots,s_{N+1}) \in TRAJ_{t\to N+1} \\ s_{N+1} \in G}} \min_{i=t\dots N} \pi(s_{i+1} | s_i, d_i(s_i)) \\
\mathsf{N}(Good_t(s_t, d_{t\to N})) &= \min_{\tau_{t\to N+1} \in TRAJ_{t\to N+1} \setminus GoodTraj_{t\to N+1}} 1 - \pi(\tau_{t\to N+1} | s_t, d_{t\to N}) \\
&= \min_{\substack{(s_{t+1},\dots,s_{N+1}) \in TRAJ_{t\to N+1} \\ s_{N+1} \in S_{N+1} \setminus G}} 1 - \min_{i=t\dots N} \pi(s_{i+1} | s_i, d_i(s_i))
\end{aligned}$$

This enables us now to rank policies and define optimal ones.

Definition 7
Let two policies $d_{t\to N}$ and $d'_{t\to N}$, and $s \in S_t$.

[7] A T-norm $*$ is a mapping from $[0,1] \times [0,1]$ to $[0,1]$ satisfying commutativity, associativity, monotonicity and having 1 as neutral element. The most usual T-norms are the minimum, the product, and the Lukasiewicz T-norm $a, b \mapsto \max(0, a+b-1)$.

$d_{t\to N} >_s d'_{t\to N}$ *if and only if one of these two conditions holds:*
- $\mathsf{N}(Good_t(s, d_{t\to N})) > \mathsf{N}(Good_t(s, d'_{t\to N}))$
- $\Pi(Good_t(s, d_{t\to N})) > \Pi(Good_t(s, d'_{t\to N}))$

The relations $\geq_s$ and $\sim_s$ are defined in a similar way as they were defined in the single-stage case. Likewise, $\Pi(Good_t(s, d_{t\to N}))$ and $\mathsf{N}(Good_t(s, d_{t\to N}))$ are standard possibility and necessity measures.

Definition 8
$d_{t\to N}$ *is an optimal policy for* $s \in S_t$ *if and only if there is no policy* $d'_{t\to N}$ *such that* $d'_{t\to N} >_s d_{t\to N}$.

2.4 Backwards computation of an optimal policy

By backwards induction, we are now going to compute, for every stage t, for every $s \in S_t$ and every $a \in A_{s,t}$:

(i) the possibility and the necessity degrees that performing a in s, followed by an optimal policy for stages $t+1$ to N, will eventually lead to a goal state; this event will be denoted by $Good_t(s, a)$; and

(ii) the possibility $\Pi(Good_t(s))$ and the necessity $\mathsf{N}(Good_t(s))$ that a goal state can be reached from s following an optimal policy from stage t to stage N.

Definition 9
$\Pi(Good_t(s, a)) = \max_{s'\in S_{t+1}} \min(\pi_t(s'|s, a), \Pi(Good_{t+1}(s')))$
$\mathsf{N}(Good_t(s, a)) = \min_{s'\in S_{t+1}} \max(1 - \pi_t(s'|s, a), \mathsf{N}(Good_{t+1}(s')))$

Hence we can compare actions with respect to a given state $s \in S_t$ in the same way as done for states in S_N. Namely,

Definition 10
$a >_s a'$ *if and only if one of these two conditions holds:*
- $\mathsf{N}(Good_t(s, a)) > \mathsf{N}(Good_t(s, a'))$
- $\mathsf{N}(Good_t(s, a)) = \mathsf{N}(Good_t(s, a'))$ *and* $\Pi(Good_t(s, a)) > \Pi(Good_t(s, a'))$

An action a is optimal for s iff there is no action a' such that $a' >_s a$. The set of optimal actions for s (at stage t) will be denoted by $A^*_{s,t}$. *An optimal policy assigns to each stage s an optimal action .*

Lastly, we define $\Pi(Good_t(s))$ and $\mathsf{N}(Good_t(s))$, which are meant to measuring to what extent *applying an optimal policy* from s will lead possibly (resp. certainly) to a goal state at stage $N+1$. They are actually the possibility and necessity measures of the event "there is a policy which leads from s to a goal state".

Definition 11
Let a^* be an arbitrary action in $A^*_{s,t}$. Then,
$\Pi(Good_t(s)) = \Pi(Good_t(s, a^*))$
$\mathsf{N}(Good_t(s)) = \mathsf{N}(Good_t(s, a^*))$

Proposition 1 $\Pi(Good_t(s,a))$, $\mathsf{N}(Good_t(s,a))$, $\Pi(Good_t(s))$ *and* $\mathsf{N}(Good_t(s))$ *are standard possibility and necessity degrees.*

We show now that this backward computation of a policy always gives an optimal policy.

Proposition 2

- $\Pi(Good_t(s)) = \max_{d_{t\to N}\in D_{t\to N}} \Pi(Good_t(s, d_{t\to N}))$
- $\mathsf{N}(Good_t(s)) = \max_{d_{t\to N}\in D_{t\to N}} \mathsf{N}(Good_t(s, d_{t\to N}))$

Corollary 1 *Any policy computed by backwards induction is optimal.*

This comes directly from the previous result and the definition of an optimal policy.

Thus, our backwards algorithm is sound. Note that there are optimal policies that cannot be computed by the previous backwards induction. As an example, consider the following problem with $N = 2$ and $G = \{s_6\}$:

On Figure 2 we show four possible policies (the assigned action figures below each state, and transition edges are labelled by the corresponding possibility degreees).

In the four cases, the most plausible trajectory leading to a bad state is (s_1, s_2, s_4). The possibility of this trajectory when respectively d, d', d'' or d''' is applied is respectively 0.8, 0.7, 0.7 and 1. Therefore,
$\mathsf{N}(Good_1(s, d_{1\to 3})) = 0.2;$
$\mathsf{N}(Good_1(s, d'_{1\to 3})) = \mathsf{N}(Good_1(s, d''_{1\to 3})) = 0.3$
$\mathsf{N}(Good_1(s, d'''_{1\to 3})) = 0$ (and $\Pi(Good_1(s, d'''_{1\to 3})) = 1$)).
d' and d'' are optimal trajectories.
Note however that d'' cannot be obtained by the backward computation,

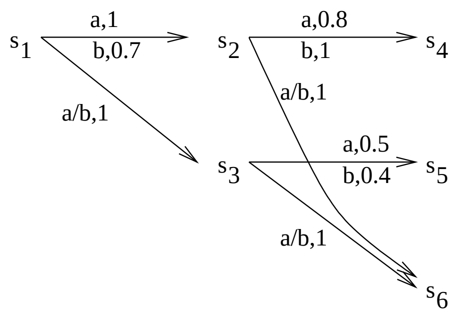

Figure 1: An action model

because of the subpolicy of d'' from stage 2 to 3, assigning a suboptimal action to s_3. This shows that subpolicies of an optimal policy may be suboptimal, which is a consequence of the use of the idempotent operator min for computing the possibility of a trajectory.

Now, using the above results, an optimal policy can be computed by a possibilistic variant of dynamic programming [2]; it computes the policy backwards (from later stages to earlier ones). The correctness of the algorithm comes straightforwardly from the recursive definition of an optimal policy.

Algorithm 1:

$d = \emptyset$; {the policy}
for $s \in S_{N+1}$ **loop** { initialisation}
  if $s \in G$ **then** {good final state}
    $\mathsf{N}(Good_N(s)) = 1$; $\Pi(Good_N(s)) = 1$;
  else {bad final state}
    $\mathsf{N}(Good_N(s)) = 0$; $\Pi(Good_N(s)) = 0$;
  endif;
end loop;
for $t := N$ downto 1 **loop1** {backward computing}
  for $s \in S_t$ **loop2**
    for $a \in A_{s,t}$ **loop3**
      compute $\mathsf{N}(Good_t(s,a)) = \min_{s' \in S_{t+1}} \max(1 - \pi_t(s'|s,a), \mathsf{N}(Good_{t+1}(s')))$

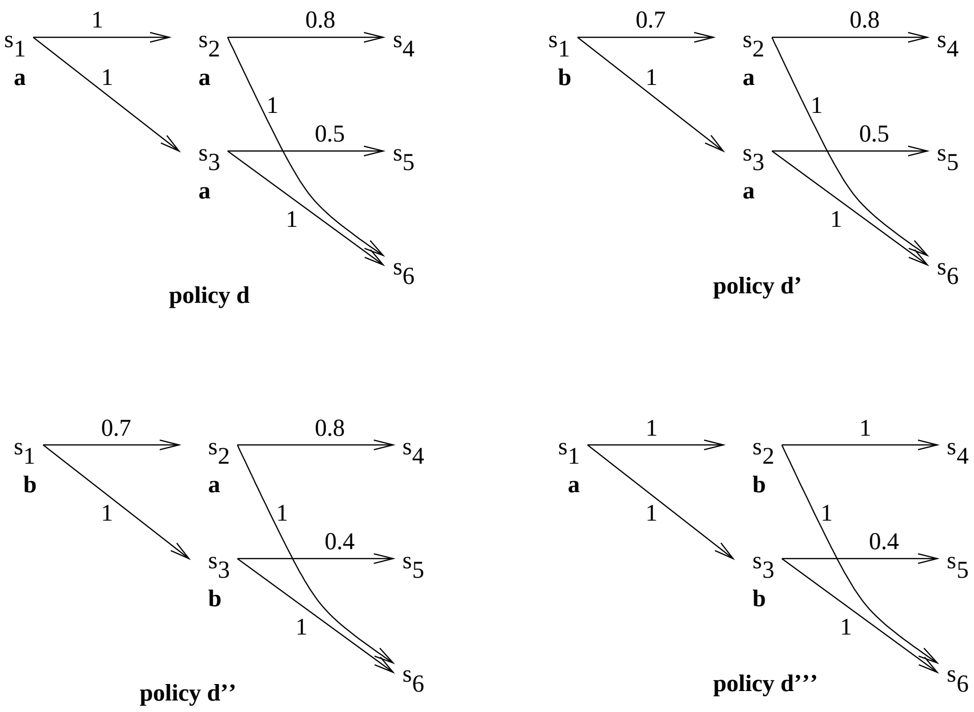

Figure 2: Four policies

compute $\Pi(Good_t(s,a)) = \max_{s' \in S_{t+1}} \min(\pi_t(s'|s,a), \Pi(Good_{t+1}(s')))$

{ choose the best action }

end loop3;

$d_t(s) = a^*$ { a^* maximizes $[\mathsf{N}(Good_t(s,a)), \Pi(Good_t(s,a))]$}

$\mathsf{N}(Good_t(s))$ $=\mathsf{N}(Good_t(s,a^*))$

$\Pi(Good_t(s)) = \Pi(Good_t(s,a^*))$

end loop2;

add the decision rule d_t to the policy d.

end loop1;

Interestingly, the use of qualitative operators (min and max) – instead of + and product for classical MDP – gives us more opportunities to avoid unnecessary computations; thus, a given action may be detected to be sub-optimal w.r.t. a given state before the whole computation of $\Pi(Good_t(s,a))$ and $\mathsf{N}(Good_t(s,a))$ is completed (*). This leads to the following improved version of the algorithm:

Algorithm 1':

$d = \emptyset$; {the policy}
for $s \in S_{N+1}$ **loop** {initialisation}
    if $s \in G$ **then** {good final state}
      $\mathsf{N}(Good_N(s)) = 1$; $\Pi(Good_N(s)) = 1$;
    else {bad final state}
      $\mathsf{N}(Good_N(s)) = 0$; $\Pi(Good_N(s)) = 0$;
    endif;
end loop;
for $t := N$ downto 1 **loop1** {backward computing}
    for $s \in S_t$ **loop2**
      $n_{opt} = 0$; $pi_{opt} = 0$;
      for $a \in A_{s,t}$ **loop3** {compute N and Π }
          $n = 1$; $pi = 0$;
          for $s' \in S_{t+1}$ **loop4**
              if $pi_{opt} < 1$ **then**
                $pi = \max(pi, \min(\pi_t(s'|s,a), \Pi(Good_{t+1}(s')))$;
                if $pi > pi_{opt}$ **then**
                  $pi_{opt} = pi$; $a_{opt} = a$;
                { else, pi_{opt} cannot be improved, and $\Pi(Good_t(s,a))$ is useless}
                end if;
              end if;
              $n = \min(n, \max(1 - \pi_t(s'|s,a), \mathsf{N}(Good_{t+1}(s'))))$;
              if $n < n_{opt}$ **then** EXIT loop4 (*) **endif** ;
              {in this case a is sub-optimal, since n can only decrease}
          end loop4;
          if $n > n_{opt}$ **then**
              $n_{opt} = n$; $a_{opt} = a$;
          end if;
      end loop3;
      $\Pi(Good_t(s)) = pi_{opt}$; $\mathsf{N}(Good_t(s)) = n_{opt}$; $d(s) = a_{opt}$;
    end loop2;
end loop1;

As for traditional dynamic programming, the complexity of this algorithm[8]

[8] assuming for simplifying the notations that $S_t = S_{t'} = S$ for any t, t' and that $A_{s,t} = A_{s',t'}$ for any s, s', t, t'

is in $o(N.|S|^2|A|)$.

3 Taking flexible goals into account

3.1 The fuzzy set of good trajectories

Up to now, we have taken into account only two types of final states : those which satisfy the goals, and those which do not. This dichotomy is inadequate to model many real decision problems, where the actor only expresses preference or indifference between states of the world: goals are flexible. To this purpose we assign to states a qualitative utility degree (as introduced in Section 1).

In many problems, only final states (at stage $N+1$) are assigned a qualitative utility. This utility is interpreted as a *goalness* degree, i.e., $u(s) = \mu_G(s)$ where μ_G is the membership function of the fuzzy set of goal states. Thus μ_G is a mapping from S_{N+1} to $[0,1]$ (not necessarily normalized), $\mu_G(s)$ being the degree to which s is an admissible final state.

Now, more generally, qualitative utilities are assigned not only to final states but to intermediary states as well[9].

The global utility of a trajectory is then defined from the utilities of the states it contains:

Definition 12 *Given a qualitative utility function $u_t : S_t \to [0,1]$ for each t, the global utility of a trajectory $\tau = (s_1, \ldots, s_{N+1})$ is*

$$u(\tau) = \min_{i=1}^{N+1} u_t(s_i)$$

Note that we could have chosen another triangular norm than min to aggregate elementary utilities into a global one.

The qualitative utility u on trajectories induces a fuzzy set of good trajectories, such that $\mu_{GoodTraj}(\tau) = u(\tau)$. This fuzzy set is not necessarily normalized, since there may be stages where no state is completely satisfactory.

[9]We might as well assign qualitative utilities to *actions*. These utility degrees would them be meant to be "qualitative anti-costs", with the convention that the higher $u(a)$, the *cheaper* a – $u(a) = 1$ meaning that a is free and $u(a) = 0$ that a costs so much that a policy containing a is not admissible at all. However, without loss of generality, and for the sake of brevity, we omit them.

3.2 Optimal policies according to the pessimistic utility

Now, from the possibility distribution and the qualitative utility function defined on trajectories, we are able to compute the pessimistic and the optimistic counterparts to expected utility, as discussed in Section 1. We start by the pessimistic index.

Definition 13 *The pessimistic utility associated to a policy $d_{t\to N}$ is defined by*

$$u^{PES}(d_{t\to N}|s_t) = \min_{\tau \in TRAJ_{t\to N+1}} \max(1 - \pi(\tau|s_t, d_{t\to N}), u(\tau))$$

$d_{t\to N}$ is u^{PES}-optimal iff it maximizes u^{PES}.

The intuition underlying the definition of pessimistic utility is that a policy is all the better as the most plausible trajectories all have a high utility. This assumes commensurability between the uncertainty and the utility scale [14].

Note that, as in Section 2, there may be optimal policies containing suboptimal subpolicies.

Two particular cases are noticeable. First, if the transition possibilities $\pi_N(s'|(s,a))$ only involve 0 or 1 possibility levels, we recognize here the usual Maximin Criterion of decision theory (also known as Wald criterion), which chooses the policy maximizing the utility of the *worst* possible trajectory. Secondly, if u only uses 0 or 1 utility levels, the set of good trajectories becomes crisp and $u^{PES}(d_{t\to N}|s)$ is nothing but $N(Good_t(s, d_{t\to N}))$ defined in Section 2.

As in Section 2, we show that a u^{PES}-optimal policy can be computed by backwards induction.

Definition 14

- $\forall s \in S_{N+1}$, $u^{PES}_{N+1}(s) = u_{N+1}(s)$
- $\forall s \in S_t$, $\forall a \in A_{s,t}$, $u^{PES}_t(s,a) = \min_{s' \in S_{t+1}} \max(1 - \pi_t(s'|(s,a)), u^{PES}_{t+1}(s'))$
- $\forall s \in S_t$, $u^{PES}_t(s) = \min(u_t(s), \max_{a \in A_{s,t}} u^{PES}_t(s,a)$

Moreover, we say that a is at least as good as a' in state $s \in S_t$, denoted by $a \geq^{PES}_s a'$, if and only if $u^{PES}_t(s,a) \geq u^{PES}_t(s,a')$. Now, we define $A^{*PES}_{s,t}$ as the set of optimal actions for s, i.e. the set of all actions a in $A_{s,t}$

maximizing $u_t^{PES}(s,a)$. We get easily that $\forall s \in S_t$, $u_t^{PES}(s) = u_t^{PES}(s,a^*)$ for an arbitrary $a \in A_{s,t}^{*PES}$.

A backward computed policy then consists in applying Definition 14 by choosing, at each stage t and for each state $s \in S_t$, an action a in $a \in A_{s,t}^{*PES}$.

$u_t^{PES}(s)$ is actually the necessity of the fuzzy event "there is a policy which leads from s to a good trajectory". In other terms :

Proposition 3 $u_t^{PES}(s) = Sup_{d \in D_t} u_t^{PES}(s, d_t(s))$

Its proof is omitted because it is very similar to the proof of Proposition 2.

Corollary 2 *Any policy computed by backward induction following Definition 14 is u^{PES}-optimal.*

Hence, as in the case of binary utilities on the final states, policies maximizing the necessity to eventually lead to a goal state may be computed backwards, assigning to each state s in stage t, an action in $A_{s,t}^{*PES}$.

3.3 Optimal policies according to the optimistic utility

In a similar way, we can rank policies according to the optimistic utility index already discussed in Section 1. According to an *optimistic* point of view, a policy is considered all the better as the fuzzy set of possible trajectories having a high utility is not empty.

Definition 15 *The optimistic utility associated to a policy $d_{t\to N}$ is defined by*

$$u^{OPT}(d_{t\to N}|s_t) = \max_{\tau \in TRAJ_{t\to N+1}} \min(\pi(\tau|s_t, d_{t\to N}), u(\tau))$$

$d_{t\to N}$ is u^{OPT}-optimal iff it maximizes u^{OPT}.

The intuition underlying the definition of the optimistic utility is that a policy is all the better as the (fuzzy) set of possible trajectories having a high utility is not empty. If the transition possibilities $\pi_N(s'|(s,a))$ only involve 0 or 1 possibility levels, we recognize maximax criterions. If u only uses 0 or 1 utility levels, then $u^{OPT}(d_{t\to N|s}|s) = \Pi(Good_t(s, d_{t\to N})$. This criterion u_{OPT} is a generalisation to multi-stage decision of the index $E(\pi) = max_{x\in X} min(\pi(x), u(x))$ first proposed by Yager [23] and later used by Kacprzyk [18] (see also Conclusion).

As for u^{PES}, we show that a u^{OPT}-optimal policy can be computed by backward induction.

Definition 16

- $\forall s \in S_{N+1},\ u^{OPT}_{N+1}(s) = u_{N+1}(s)$
- $\forall s \in S_t,\ \forall a \in A_{s,t},\ u^{OPT}_t(s,a) = \max_{s' \in S_{t+1}} \min(\pi_t(s'|(s,a)), u^{OPT}_{t+1}(s'))$
- $\forall s \in S_t,\ u^{OPT}_t(s) = \min(u_t(s), \max_{a \in A_{s,t}} u^{PES}_t(s,a)$

$\geq^{OPT}_s$ is defined in a similar way as $\geq^{PES}_s$. $A^{*OPT}_{s,t}$ is the set of all actions a in $A_{s,t}$ maximizing $u^{OPT}_t(s,a)$. We get easily that $\forall s \in S_t$, $u^{OPT}_t(s) = u^{OPT}_t(s,a^*)$ for an arbitrary $a \in A^{*OPT}_{s,t}$.

A backward computed policy then consists in applying Definition 16 by choosing, at each stage t and for each state $s \in S_t$, an action a in $a \in A^{*OPT}_{s,t}$.

$u^{OPT}_t(s)$ is actually the possibility of the fuzzy event "there is a policy which leads from s to a good trajectory":

Proposition 4 $u^{OPT}_t(s) = Sup_{d \in D_t} u^{OPT}_t(s, d_t(s))$

Corollary 3 *Any policy computed by backward induction following Definition 16 is u^{OPT}-optimal.*

3.4 On flexible and non flexible goals

In the two previous sections, we made a distinction between the cases where the goals posted at stage $N+1$ were flexible or not. We will show here that the latter case can be reduced to the former, supposing that the utility of state s_{N+1} is respectively a degree of *necessity* or *possibility* to reach a (binary) goal state at stage $N+2$. This leads to consider an additional decision stage, leading from S_{N+1} to S_{N+2} where only binary goal states are posted, and only one action is available.

However, the transition possibilities from stage S_{N+1} to stage S_{N+2} will be different depending on whether we wish to compute backwards *pessimistic* or *optimistic* utilities.

Suppose that we have a utility function u on S_{N+1}. If we define a new problem with an additionnal stage $N+2$ by :

- $S_{N+2} = \{g, \bar{g}\}$
- $Goal_{N+2} = \{g\}$
- $\forall s \in S_{N+1}, A_{s,N+1} = \{a^*\}$
- $\forall s \in S_{N+1}, \pi_{N+1}(g|s,a^*) = 1,\ \pi_{N+1}(\bar{g}|s,a^*) = 1 - u(s)$.

Then we can prove the following proposition.

Proposition 5 $\forall s \in S_{N+1}, u(s) = \mathsf{N}(Good_{N+1}(s))$.

Proof :

According to the definition of $\mathsf{N}(Good_{N+1}(s))$, we have :
$\mathsf{N}(Good_{N+1}(s)) = sup_{a \in A_{s,N+1}} \min_{s' \notin Goal_{N+2}} 1 - \pi_{N+1}(s'|(s,a))$
So, $\mathsf{N}(Good_{N+1}(s)) = 1 - (1 - u(s)) = u(s)\Box$

In the same way as before, extending the horizon of the problem by :

- $S_{N+2} = \{g, \bar{g}\}$
- $Goal_{N+2} = \{g\}$
- $\forall s \in S_{N+1}, A_{s,N+1} = \{a^*\}$
- $\forall s \in S_{N+1}, \pi_{N+1}(g|s, a^*) = u(s), \pi_{N+1}(\bar{g}|s, a^*) = 1.$

We can prove :

Proposition 6 $\forall s \in S_{N+1}, u(s) = \Pi(Good_{N+1}(s))$.

Proof :

According to the definition of $\Pi(Good_{N+1}(s))$, we have :
$\Pi(Good_{N+1}(s)) = sup_{a \in A_{s,N+1}} \max_{s' \in Goal_{N+2}} \pi_{N+1}(s'|s, a)$
So, $\Pi(Good_{N+1}(s)) = u(s)\Box$

Then $u(s)$ can be obtained either as a necessity degree, or as a possibility degree, considering an additional stage $N+2$, and a unique available decision at stage $N+1$, which will either be a, or a'.

It is important to notice that pessimistic and optimistic utilities are not dual necessity and possibility degrees in so far as $\Pi(Good_{N+1}(s)) < 1$ does not imply $\mathsf{N}(Good_t(s)) = 0$.

3.5 Refining the pessimistic criterion by an optimistic one

There are many reasons for preferring to use the pessimistic utility index rather than the optimistic one. Indeed, the former is more cautious, more reliable, since it takes account of all cases, focusing on the worst ones, while the optimistic index takes account of one state only (the best one). However, using only u^{PES} for ranking policies may have a weak discrimination power, because the set of u^{PES}-optimal policies may be too large. We may then think of refining u^{PES} by u^{OPT}, in a way recalling how we ranked policies in Section 2.

Definition 17

1. $d_{t\to N} \geq_s d'_{t\to N}$ *iff one of these two conditions holds:*
   - $u_t^{PES}(s, d_{t\to N}) > u_t^{PES}(s, d_{t\to N})$
   - $u_t^{PES}(s, d_{t\to N}) = u_t^{PES}(s, d_{t\to N})$ *and* $u_t^{OPT}(s, d_{t\to N}) > u_t^{OPT}(s, d_{t\to N})$
2. $d_{t\to N}$ *is optimal iff there is no* $d'_{t\to N}$ *such that* $d'_{t\to N} >_s d_{t\to N}$

It is then easy to show that $d_{t\to N}$ is optimal if and only if it is u^{PES}-optimal *and* it maximizes u^{OPT} among the set of all u^{PES}-optimal policies.

Since $u_t^{PES}(s, d_{t\to N})$ and $u_t^{OPT}(s, d_{t\to N})$ are respectively the necessity and the possibility of a fuzzy event, generally they do *not* verify $u_t^{PES}(s, d_{t\to N}) > 0 \Rightarrow u_t^{OPT}(s, d_{t\to N}) = 1$ (but only the weaker relationship $u_t^{OPT}(s, d_{t\to N}) \geq u_t^{OPT}(s, d_{t\to N})$). As a consequence, optimal actions with respect to u^{PES} are not necessarily optimal with respect to u^{OPT}. For instance, let $T = 1$, $S_1 = \{s_0\}$, $S_2 = \{s_1, s_2, s_3\}$, $A_{s_0,1} = \{a, b, c\}$ with $u_1(s_0) = 1$ and

$$\left|\begin{array}{cccc} \pi(s_1|s_0,a) = 1 & \pi(s_2|s_0,b) = 0 & \pi(s_3|s_0,a) = 0.2 & u_2(s_1) = 1 \\ \pi(s_1|s_0,b) = 0.7 & \pi(s_2|s_0,b) = 1 & \pi(s_3|s_0,b) = 0.1 & u_2(s_2) = 0.6 \\ \pi(s_1|s_0,c) = 1 & \pi(s_2|s_0,b) = 0 & \pi(s_3|s_0,b) = 0 & u_2(s_3) = 0.2 \end{array}\right|$$

The policies d, d' and d'' assigning respectively a,b and c to s_0 have the following indices: $\left|\begin{array}{ll} u_1^{PES}(d) = 0.2 & u_1^{OPT}(d) = 1 \\ u_1^{PES}(d') = 0.6 & u_1^{OPT}(d') = 0.7 \\ u_1^{PES}(d'') = 0.6 & u_1^{OPT}(d'') = 0.6 \end{array}\right|$

Thus, d' and d'' are u^{PES}-optimal; *the optimal policy according to the refined criterion* is d'. d'' is u^{OPT}-optimal.

Now, we sho that an optimal policy for the refined criterion can be computed again by backwards induction. $u_t^{PES}(s,a)$ and $u_t^{PES}(s)$ have been defined in Section 3.2.

Definition 18

- $\forall s \in S_{N+1}$, $u_{N+1}^{*OPT}(s) = u_{N+1}(s)$
- $\forall s \in S_t$, $\forall a \in A_{s,t}$, $u_t^{*OPT}(s,a) = Sup_{s' \in S_{t+1}} min(\pi_t(s'|s,a), u_{t+1}^{*OPT}(s'))$
- $a >_s a'$ *iff* $u_t^{PES}(s,a) > u_t^{PES}(s,a')$ *or* $(u_t^{PES}(s,a) = u_t^{PES}(s,a')$ *and* $u_t^{*OPT}(s,a) > u_t^{*OPT}(s,a'))$
- $A^*_{s,t} = \{a \in A_{s,t}|$ *there is no* $a' \in A_{s,t}$ *such that* $a' >_s a\}$
- $\forall s \in S_t$, $u_t^{*OPT}(s) = \min(u_t(s), u_t^{*OPT}(s,a^*))$ *for an arbitrary* $a^* \in A^*_{s,t}$

Thus, $u_t^{*OPT}(s)$ (resp $u_t^{PES}(s)$) stands for the possibility (resp. the necessity) to reach a goal state when *applying an optimal policy*. Note that we have the inequalities, for all T and s: $u_t^{OPT}(s) \geq u_t^{*OPT}(s) \geq u_t^{PES}(s)$ and that generally, these inequalities are strict. In particuler, u^{*OPT} should not be confused with u^{OPT}: while $u_t^{OPT}(s)$ is the possibility of the fuzzy event "a good trajectory results from t when applying an optimal policy *according to the pure optimistic criterion*", $u_t^{OPT}(s)$ is the the possibility of the fuzzy event "a good trajectory results from t when applying an optimal policy *according to the refined criterion*".

Proposition 7 $u_t^{*OPT}(s) = Sup_{d \in D_t | d \ is \ u^{PES}-optimal} u_t^{OPT}(s, d_t(s))$

Corollary 4 *Any policy computed by backward induction following Definition 18 is optimal for the refined criterion.*

In the light of this backward induction computation, the following variant of dynamic programming computes an optimal policy for the refined criterion. As in Section 2, it is possible to avoid unnecessary computations (we omit the details for the sake of brevity).

algorithm 2;

{ initialization }

$d = \emptyset$ {the policy}
for s $\in S_{N+1}$ **loop**
    $u_{N+1}^{*OPT}(s) = u(s)$;
    $u_{N+1}^{PES}(s) = u(s)$;
end loop;

{ Backward computing }

for $t := N$ downto 1 **loop1**
    for $s \in S_t$ **loop2**
        for $a \in A_{s,t}$ **loop3**
            $Compute(u_t^{PES}(s, a))$
            $Compute(u_t^{*OPT}(s, a))$
        end loop3;
        $d_t(s) = a^*$ { a^* maximizes $u_t^{PES}(s, a)$ and then $u_t^{*OPT}(s, a)$ }

$u_t^{PES}(s) = \min(u_t(s), u_t^{PES}(s, a^*))$
$u_t^{*OPT}(s) = \min(u_t(s), u_t^{*OPT}(s, a^*))$
end loop2;
Add d_t to the policy d
end loop1;

An alternative way of refining the pessimistic ordering would consist in replacing the minimum operator in the computation of the possibility of a trajectory from the possibility degrees of its elementary transitions by a lexicographic minimum. This would consist in storing, for each trajectory, not only the minimum $\pi(s_{i+1}|s_i, d_i(s_i)$ but all of them, ranked increasingly. Then, two trajectories are compared by comparing first their lowest components, and in case of equality, their second lowest components, etc. Details are omitted.

4 Example

A robot is moving in a room in which it entered by the top-left square. Its objective is entirely satisfied if it finishes in the down-right square and partly if it finishes in one of the neighbor squares. The state-space, starting square and the utility function on the objective states are depicted in Figure 3.

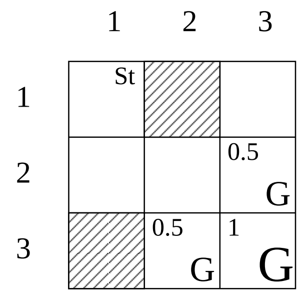

Figure 3: State space and utility function.

The available actions are to move (T)op, (D)own, (L)eft, (R)ight or to (S)tay in place. If the robot chooses to stay, it will *certainly* remain in the same square. If it goes T, D, L or R it will (entirely) possibly reach the desired square ($\pi = 1$) if it is free but there will be some possibilities that it reaches a neighbor square, as depicted in Figure 4 for the action R. The other transition possibility functions are of course symetric to these.

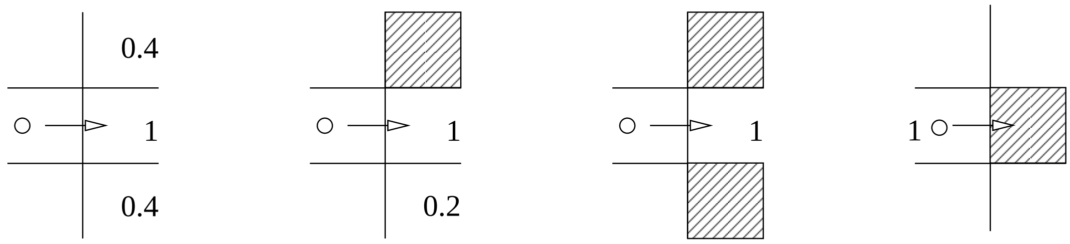

Figure 4: Transition possibilities for moving right.

Now, suppose that the horizon of the problem is 5, that is goals are set at step 6. Figure 3 resumes the utility function u, so it also resumes u_6^{PES} : $u_6^{PES}(s_{33}) = 1$, $u_6^{PES}(s_{32}) = u_6^{PES}(s_{23}) = 0.5$ and $u_6^{PES}(s) = 0$ for every other s. Let us now compute the optimal actions for the states in S_5. For every action a and state s, we have $u_5^{PES}(s, a) = min_{s' \in S_6} max(1 - \pi(s'|s, a), u_6^{PES}(s'))$ (π does not depend on the step) and $u_5^{PES}(s) = max_{a \in \{T,D,L,R,S\}} u_5^{PES}(s, a)$.

Figure 5 Resumes the utility of each state in S_5 as well as the optimal action for each state with a non-null pessimistic utility. The optimal action is unique, except for state s_{33} for which D and R would be optimal actions as well.

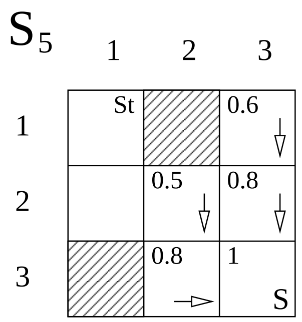

Figure 5: .

Now we can iterate the process and get an optimal policy. The iterated process is described in Figure 6. Note that after 4 iterations, the utility of each state and the associated optimal action do not change anymore.

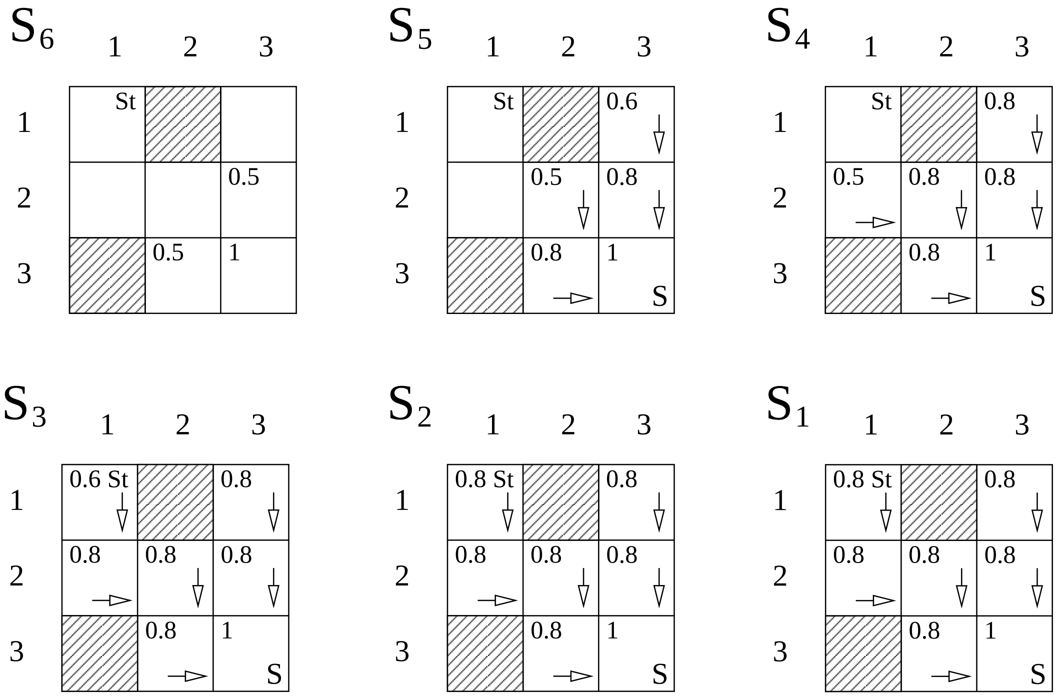

Figure 6: Pessimistic optimal policy computation.

5 Related work and conclusion

The main contribution of this article consisted in extending Dubois and Prade's possibilistic decision theory from the single-stage to the multiple-stage case. We have proposed two successive extensions (the latter generalising the former), corresponding, in our notations, to the 4-uples $\langle N, \pi, G$, maximise $(\mathsf{N}, \Pi)\rangle$ and $\langle N, \pi, min$, maximise u^{PES} and then $u^{*OPT})\rangle$. We have sketched algorithms for these extensions, in the spirit of dynamic programming.

An alternative framework for possibilistic multistage decision making has been intensively developed by Da Costa et al.[6] [7] [8]. Their work share with ours the use of possibility distributions for representing uncertain effects of actions. The main difference between both approaches relies on observability: while our approach assumes that the environment is fully observable, theirs assume non-observability. Consequently, rather than computing policies, they compute unconditional plans (or sequences of actions) which maximize the necessity or the possibility to reach a goal state.

They use a STRIPS-like representation of possibilistic actions, which avoid an explicit enumeration of states as we do; they developed an algorithm for computing optimal plans which is in the spirit of traditional AI planning, while our algorithms are in the spirit of dynamic programming. Lastly, their approach is generalized in order to handle flexible goals in approximatively the same way of us, thus leading to an alternative extension to multistage decision theory of Dubois and Prade's qualitative decision theory [8] in the case of non-observability.

Apart from Dubois and Prade's possibilistic approach to one-stage decision making, some authors have considered more or less qualitative approaches to decision, either one-stage or multi-stage. Yager [23] proposed to use the optimistic criterion defined by

$$u^{OPT}(s_0, a) = max_{s \in S} min(\pi_{a,s_0}(s), u(s))$$

which is the optimistic counterpart of Dubois and Prade's u^{PES}. As observed by Dubois and Prade [14], this criterion can be overoptimistic. More general criteria for one-stage decision making, recovering optimistic and pessimistic criteria as particular cases, have been proposed by Bolaños, Lamata and Moral [4] and more recently by Yager [24], Yager and Lamata [25]. Extending these general frameworks to multi-stage decision is worth considering for further research.

Several fuzzy extensions of dynamic programming have been proposed, a review of which is in [16]. The seminal one is Bellman and Zadeh's [3] which assume a qualitative utility function (where intermediate utilities of different states of a both are aggregated by min) and transition functions which are either deterministic, either stochastic – these two approaches corresponding thus to $\langle N, D, min, u \rangle$ and to $\langle N, pr, min, \bar{u} \rangle$, respectively. This last approach can be seen as a semi-qualitative, semi-quantitative approach to decision making – the selection criterion for the optimal policy consists in maximizing the expected qualitative utility. On the contrary, our framework is fully qualitative (both on the uncertainty and the utility side). Other significant work about using possibility theory for multi-stage decision making has been done by Kacprzyk [18] and also Baldwin and Pilsworth [1]. Kacprzyk [18] defines an optimal policy by maximising the aforementioned optimistic criterion u^{OPT} at each stage – his approach corresponds thus to the 4-uple $\langle N, \pi, min, u^{OPT} \rangle$; he computes it with a branch and bound algorithm; however, as noticed by Dubois and Prade [14], maximizing u^{OPT} is practically not reasonable; to make the argument simpler, suppose that all transition possibilities are binary ($\forall s, s, s', \pi(s'|(s,a)) \in \{0,1\}$), then maximizing u^{OPT} comes down to assume at each stage that the best possible outcome occurs (in other words, the agent takes its desires for reality). The author also proposes extensions of his approach to the cases where termination time is fuzzy, or infinite. The alternative approach of Baldwin and Pilsworth [1], similar to dynamic programming, also consists in maximising an optimistic criterion, but the search is performed over a set of fuzzy states and the maximisation over a set of fuzzy decisions, which, as noticed by Kacprzyk [18], makes the approach practically unreasonable due to its computational complexity.

Further work would consist first in implementing the proposed algorithms, taking account of the suggested techniques (and possibly others) for avoiding unnecessary computations. Next, we think of generalising our framework in several directions:

- Considering the contribution of intermediate states and actions performed to the global utility function: the most natural way to do it in a purely qualitative way consists in combining these qualitative utilities by the minimum, for instance for u^{PES} we get

  $$u_t^{PES}(s,a) = \min(u(s), u(a), \min_{s' \in S_{t+1}} \max(1 - \pi(s'|(s,a)), u_{t+1}^{PES}(s'))$$

  and $u_t^{PES}(s) = \max_{a \in A_{s,t}} u_t^{PES}(s,a)$. Integrating these intermediate utilities in the framework described in Section 3 does not present any particular difficulty.

- Retracting the commensurability assumption between possibility degrees and utility degrees. This would enable us to consider mixed qualitative/quantitative frameworks, with, for instance, additive utilities and possibilistic transition functions – the latter would probably lead us to rank fuzzy numbers in order to define optimal policies.

- Retracting the assumption of full observability. To this purpose we may adapt some methodologies and results from partially observable Markov decision processes. In this case we would have to consider the presence of information-gathering actions in policies. Now, some work has been done in the field of possibilistic planning [6] [7] [8], where, under the assumption of no observability, one looks for an unconditional sequence of actions leading to a goal state with a maximal necessity degree. This has lead to the implementation of a possibilistic planner.

Another direction for further research would consist in assessing automatically the possibility distributions describing the uncertain effects of actions, from a set of qualitative, nonmonotonic rules. Namely, it has been shown in [9] that from a set of hard and defeasible dynamic laws describing the effects of actions, it is possible to build a possibility distribution taking account of the specificity of the rules; namely, an effect given by a more specific rule will get a higher possibility than an effect given by a less specific one. See also [8] for such a use of qualitative default rules in the context of possibilistic planning.

References

[1] J.F. Baldwin and B.W. Pilsworth. Dynamic programming for fuzzy systems with fuzzy environment. *J. Math. Anal. & Appl.* 85, 1-23, 1982.

[2] R.A. Bellman. *Dynamic Programming.* Princeton University Press, 1957.

[3] R.E. Bellman and L.A. Zadeh. Decision making in a fuzzy environment. *Management Science*, 17, B-141-B-164, 1970.

[4] M.J. Bolaños, M.T. Lamata and S. Moral. Decision making problems in a general environment. *Fuzzy Sets and Systems* 25, 135-144, 1988.

[5] Craig Boutilier. Towards a logic for qualitative decision theory. *Proceedings of the 4th Intern. Conf. on the Principles of Knowledge Representation and Reasoning (KR'94)*, 75-86, Morgan Kaufmann, 1994.

[6] Célia Da Costa Pereira, Frédérick Garcia, Jérôme Lang and Roger Martin-Clouaire. Planning with graded nondeterministic actions: a possibilistic approach. *Int. Journal of Intelligent Systems*, 12:935-962, 1997.

[7] Célia Da Costa Pereira, Frédérick Garcia, Jérôme Lang and Roger Martin-Clouaire. Possibilistic planning: representation and complexity. In *Recent Advances in AI Planning* (Sam Steel, Rachid Alami, eds.), Lectures Notes in Artificial Intelligence, Spriger Verlag, 143-155, 1997.

[8] Célia Da Costa Pereira. *Planification d'actions en environnement incertain : une approche fondée sur la théorie des possibilités.* PhD, Université Paul Sabatier, Toulouse, France, May 1998. In French.

[9] Didier Dubois, Florence Dupin de Saint-Cyr and Henri Prade. Updating, transition constraints and possibilistic Markov chains. *Advances in Intelligent Systems - IPMU'94 (B. Bouchon-Meunier, R. R. Yager, L.A. Zadeh, eds.)*, LNCS 945, Springer-Verlag, 263-272.

[10] D. Dubois, H. Fargier, J. Lang, H. Prade, R. Sabbadin, Qualitative decision theory and multistage decision making-A possibilistic approach. In *Proceedings of the European Workshop on Fuzzy Decision Analysis for Management, Planning and Optimization (EFDAN'96)*, Dortmund, 1996,

[11] Didier Dubois and Henri Prade. Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences* 30, 1983, 183-224.

[12] Didier Dubois and Henri Prade. *Possibility theory.* Plenum Press, 1988.

[13] Didier Dubois and Henri Prade. Possibility theory as a basis for qualitative decision theory. *Proceedings of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI"95)*, Montréal, Aug. 1995, 1924-1930.

[14] Didier Dubois and Henri Prade. Towards possibilistic decision theory. *Preprints of the 3rd Fuzzy Logic in AI Workshop*, Montreal, August 95. Also in Tech. Report IRIT, Toulouse (1995).

[15] Didier Dubois, Henri Prade and Claudette Testemale. Weighted fuzzy pattern matching. *Fuzzy Sets and Systems* 28 (1988), 313-331.

[16] Augustine O. Esogbue and Janusz Kacprzyk. Fuzzy Dynamic Programming. In *Fuzzy Sets in Decision Analysis, Operations Research and Statistics* (R. Slowinski, editor), Kluwer Academic Publishers, pp. 281-307, 1998.

[17] H. Fargier, J. Lang, R. Sabbadin. Towards qualitaive approaches to multi-stage decision making. In *Proceedings of the Sixth Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'96)*, Granada, July 1-5, 1996.

[18] Janucz Kacprzyk. *Multi-stage decision making under fuzziness.* Verlag TÜV Rheinland, 1983.

[19] Martin L. Puterman. *Markov Decision Processes.* John Wiley and Sons, New York, 1994.

[20] Martin L. Puterman. Dynamic programming. *Encyclopedia of Physical Science and Technology*, Vol. 4 (1987), 438-463.

[21] Sek-Wah Tan and Judea Pearl. Qualitative decision theory. *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI'94)*, MIT Press, 1994, 928-933.

[22] Thomas Whalen. Decision making under uncertainty with various assumptions about available information. *IEEE Trans. on Systems, Man and Cybernetics*, 14, 888-900 (1984).

[23] Ronald R. Yager. Possibilistic decision making. *IEEE Trans. on Systems, Man and Cybernetics* 9 (1979), 388-392.

[24] Ronald R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning* 12, 237-261, 1995.

[25] Ronald R. Yager and Maria T. Lamata. Decision making under uncertainty with nonnumeric payoffs. Tech. Report MII-1525, Iona College, New Rochelle, New York, 1995.

[26] Lotfi A. Zadeh. A theory of approximate reasoning. *Machine Intelligence*, Vol. 9, edited by J.E. Hayes, D. Michie, L.I. Mikulich, Elsevier, NY, 149-154.

Appendix

Proof of Proposition 1:
We prove by backward induction, that $\forall t, \Pi(Good_t(s)) < 1 \Rightarrow \mathsf{N}(Good_t(s)) = 0$. We have already proved that this is true for $t = N$, we now show that if this is true for $t+1$, it is true for t :
Let $s \in S_t$ and assume that $\Pi(Good_t(s)) < 1$. By definition of $\Pi(Good_t(s))$, this is equivalent to $\max_{a \in A_{s,t}} \Pi(Good_t(s,a)) < 1$, i.e., $\forall a \in A_{s,t}, \Pi(Good_t(s,a)) < 1$, which, using the definition of $\Pi(Good_t(s,a))$, is equivalent to :

$$\forall a \in A_{s,t}, \forall s' \in S_{t+1}, \pi_t(s'|s,a) = 1 \Rightarrow \Pi(Good_{t+1}(s')) < 1 (1)$$

Now, the induction hypothesis and (1) entail that $\forall a \in A_{s,t}, \forall s' \in S_{t+1}, \pi_t(s'|s,a) = 1 \Rightarrow \mathsf{N}(Good_{t+1}(s')) = 0$.
Then, let $a \in A_{s,t}$ and $s' \in S_{t+1}$ such that $\pi_t(s'|s,a) = 1$ (the existence of such a state is guaranteed by the normalisation of the possibility distribution π_t), then we have $\mathsf{N}(Good_{t+1}(s')) = 0$, and thus $\mathsf{N}(Good_t(s,a)) = \min_{s' \in S_{t+1}} \max(1 - \pi_t(s'|s,a), \mathsf{N}(Good_{t+1}(s'))) = 0$.

This is true $\forall a \in A_{s,t}$, hence we conclude that $\mathsf{N}(Good_t(s)) = 0$. $\square$

Proof of Proposition 2:
First, notice that
$\Pi(Good_t(s)) = \max_{a \in A_{s,t}} \Pi(Good_t(s,a)) = \max_{d_t \in D_t} \Pi(Good_t(s, d_t(s)))$, and
$\mathsf{N}(Good_t(s)) = \max_{a \in A_{s,t}} \mathsf{N}(Good_t(s,a)) = \max_{d_t \in D_t} \mathsf{N}(Good_t(s, d_t(s)))$.

- Let us prove by backwards induction that
  $\Pi(Good_t(s) = \max_{d \in D_{t \to N}} \max_{\substack{(s_{t+1},\ldots,s_{N+1}) \in TRAJ_{t+1 \to N+1} \\ s_{N+1} \in G}} \min_{j=t\ldots N} \pi(s_{j+1}|s_j, d(s_j))$

  It is obvious for $t = N$ (from the definition of $\Pi(Good_N(s)))$. Let us prove that if the result is true for stages $t+1 \in 2 \ldots N+1$ it is true for stage t :
  $\Pi(Good_t(s)) = \max_{d_t \in D_t} \max_{s_{t+1} \in S_{t+1}} \min(\pi_t(s_{t+1}|s, d_t(s)), \Pi(Good_{t+1}(s_{t+1})))$.
  If we suppose that the result is true at stage $t+1$, we get :
  $\Pi(Good_t(s)) = \max_{d_t \in D_t} \max_{s_{t+1} \in S_{t+1}} \min(\pi_t(s_{t+1}|s, d_t(s)), \max_{d \in D_{t+1 \to N}}$
  $\max_{\substack{s_{t+2},\ldots,s_{N+1} \in S_{t+2} \times,\ldots,\times S_{N+1}, \\ s_{N+1} \in G}} \min_{j=t+1\ldots N} \pi(s_{j+1}|s_j, d(s_j))$
  knowing (from the Markovian assumption) that $\pi_t(s_{t+1}|s, d_t(s))$ does not depend on further stages, we get :
  $\Pi(Good_t(s)) =$
  $\max_{d_t \in D_t} \max_{s_{t+1} \in S_{t+1}} \max_{d \in D_{t+1 \to N}} \max_{\substack{s_{t+2},\ldots,s_{N+1} \in S_{t+2} \times,\ldots,\times S_{N+1} \\ s_{N+1} \in G}} \min_{j=t\ldots N} \pi(s_{j+1}|s_j, d(s_j))$
  We can swap $\max_{s_{t+1} \in S_{t+1}}$ and $\max_{d \in D_{t+1 \to N}}$:
  $\Pi(Good_t(s)) =$
  $\max_{d_t \in D_t} \max_{d \in D_{t+1 \to N}} \max_{s_{t+1} \in S_{t+1}} \max_{\substack{s_{t+2},\ldots,s_{N+1} \in S_{t+2} \times,\ldots,\times S_{N+1} \\ s_{N+1} \in G}} \min_{j=t\ldots N} \pi(s_{j+1}|s_j, d(s_j))$
  And from this comes the result at stage t.

- It can be proved similarly that
  $\mathsf{N}(Good_t(s)) = \max_{d \in D_{t \to N}} \min_{\substack{(s_{t+1},\ldots,s_{N+1}) \in TRAJ_{t+1 \to N+1} \\ s_{N+1} \notin G}} \max_{j=t\ldots N}(1 - \pi(s_{j+1}|s_j, d(s_j))$.

$\square$