Nonmonotonic desires - A possibility theory viewpoint

Didier Dubois, Emiliano Lorini, and Henri Prade

IRIT, CNRS & Université Paul Sabatier, France
dubois@irit.fr, emiliano.lorini@irit.fr, prade@irit.fr

Abstract. If an agent desires that φ and desires that ψ, this agent often also desires that φ and ψ hold at the same time ($\varphi \wedge \psi$). However, there are cases where $\varphi \wedge \psi$ may be found less satisfactory for the agent than each of φ or ψ alone. This paper is a first attempt at modeling such nonmonotonic desires. The approach is developed in the setting of possibility theory, since it has been recently pointed out that guaranteed (or strong) possibility measures are a good candidate for modeling graded desires. Although nonmonotonic reasoning has been studied extensively for knowledge, and that preferential nonmonotonic consequence relations can be faithfully represented in the possibilistic setting, nonmonotonic desires appear to require a different approach.

1 Introduction

In a recent work [11], we have advocated the idea that if an agent is satisfied with $\varphi \vee \psi$, it should be also the case that the agent is satisfied with φ and that the agent is satisfied with ψ. This claim is justified if by "the agent is satisfied with φ" we mean: "in all situations where φ holds". This has led us to propose the modeling of desires by means of guaranteed possibility measures Δ that obey axiom $\Delta(\varphi \vee \psi) = \min(\Delta(\varphi), \Delta(\psi))$.

This approach leads us to consider that φ is desirable for the agent, it is still the case in any context χ, i.e. saying that having φ true is satisfactory forces us to consider that having $\chi \wedge \varphi$ true is satisfactory as well, whatever χ. Even if it may be generally the case, it may happen that when an agent specifies that it has the desire to have φ true, it may not specify that in the abnormal / rare case where χ is true, it rather desires having $\neg\varphi$ true. Another worth considering option is when, in the context where χ is true, it becomes indifferent which of φ or $\neg\varphi$ is true. This means that although the agent generally desires to have φ true, there may exist some particular circumstances where this desire no longer exists, or even that an opposite desire may take place.

Such a concern sounds like a nonmonotonic reasoning issue [7], where a set of explicit desires may express some apparent logical contradiction ("I definitely desire φ, but not necessarily in any circumstances", in particular "I definitely desire $\neg\varphi$ if χ occurs"). It is not clear that nonmonotonic reasoning approaches, which have been developed for handling default knowledge in the presence of incomplete information, still apply to computing a preference rank-ordering between situations. The paper intends to discuss this situation, and to provide an approach suitable for handling non monotonic desires, a problem that has been mentioned by philosophers [21], but not apparently considered for its specific features in artificial intelligence until now.

The paper is organized as follows. We first provide a short background on the possibility theory framework and its suitability for representing desires. We then recall how reasoning with default knowledge in the presence of incomplete information can be properly handled in the possibilistic reasoning setting. Then we examine the problems arising in the treatment of nonmonotonic desires, propose and discuss an approach.

2 Background on possibility theory and the modeling of desires

Let π be a mapping from a set of worlds W to $[0, 1]$ that rank-orders them. Note that this encompasses the particular case where π reduces to the characteristic function of a subset $E \subseteq W$. The possibility distribution π may represent a plausibility ordering (E the set of situations considered not impossible) when modeling epistemic uncertainty, or a preference ordering (E is then the subset of satisfactory worlds) when modeling preferences. Let us recall the complete system of the 4 set functions underlying possibility theory [13] and their characteristic properties:

i) The *(weak) possibility measure* (or potential possibility) $\Pi(A) = \max_{w \in A} \pi(w)$ evaluates to what extent there is a world in A that is possible. When π reduces to E, $\Pi(A) = 1$ if $A \cap E \neq \emptyset$, which expresses the consistency of the event A with E, and $\Pi(A) = 0$ otherwise. Possibility measures are characterized by the following decomposability property: $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$.

ii) The dual (*strong* or or actual) *necessity measure* $N(A) = \min_{w \notin A} 1 - \pi(w) = 1 - \Pi(\overline{A})$ evaluates to what extent it is certain (necessarily true) that all possible worlds are in A. When π reduces to E, $N(A) = 1$ if $E \subseteq A$, which expresses that E entails event A (when E represents evidence), and $N(A) = 0$ otherwise. The duality of N w. r. t. Π expresses that A is all the more certain as the opposite event $\overline{A}$ is impossible. Necessity measures are characterized by the following decomposability property: $N(A \cap B) = \min(N(A), N(B))$.

iii) The *strong* (or actual, or *"guaranteed"*) possibility measure $\Delta(A) = \min_{w \in A} \pi(w)$ evaluates to what extent *all* situations in A are possible. When π reduces to E, $\Delta(A) = 1$ if $A \subseteq E$, and $\Delta(A) = 0$ otherwise. Strong possibility measures are characterized by the following property: $\Delta(A \cup B) = \min(\Delta(A), \Delta(B))$.

iv) The dual *(weak)* (or potential) necessity measure $\nabla(A) = \max_{w \notin A} 1 - \pi(w) = 1 - \Delta(\overline{A})$ evaluates to what extent there is a situation outside A that is impossible. When π reduces to E, $\nabla(A) = 1$ if $A \cup E \neq U$, and $\nabla(A) = 0$ otherwise. Weak necessity measures are characterized by property: $\nabla(A \cap B) = \max(\nabla(A), \nabla(B))$.

Δ, ∇ are decreasing set functions, while the (weak) possibility and (strong) necessity measures are increasing. A modal logic counterpart of these 4 modalities has been proposed in the *binary*-valued case (things are possible or impossible) [9]. There is a close link between Spohn functions and (weak) possibility / (strong) necessity measures [12].

2.1 Possibility theory as basis for a logical theory of desires

The possibility and necessity operators Π and N have a clear epistemic meaning both in the frameworks of possibility theory, and of Spohn's uncertainty theory [22] (also referred to as 'κ calculus', or as 'rank-based system' and 'qualitative probabilities'). Differently from the operators Π and N, the operators Δ and ∇ are less employed to model epistemic attitudes. In fact while the epistemic use of N models the idea of "knowing at least", the one of Δ accounts for the idea of "knowing at most". Combining both modalities provides a representation of the idea of "only knowing" [1, 15]. In epistemic logic, the notion of strong belief put forward by Battigalli and Siniscalchi [2], corresponds to the inequality $\Delta(A) > \Pi(\overline{A})$ (which provably implies $N(A) > N(\overline{A}) = 0$).

Here we advocate the idea that Δ and ∇ can be viewed as operators modeling motivational mental attitudes such as goals or desires.[1] In particular, we claim that Δ can be used to model the notion of *desire*, whereas ∇ can be used to model the notion of *potential* desire. [2] According to the philosophical theory of motivation based on Hume [18], a desire can be conceived as an agent's motivational attitude which consists in an anticipatory mental representation of a pleasant (or desirable) state of affairs (representational dimension of desires) that motivates the agent to achieve it (motivational dimension of desires). In this perspective, the motivational dimension of an agent's desire is realized through its representational dimension. For example when an agent desires to be at the Japanese restaurant eating sushi, he imagines himself eating sushi at the Japanese restaurant and this representation gives him pleasure. This pleasant representation motivates him to go to the Japanese restaurant in order to eat sushi.

Intuitively speaking, with the term *potential* desire, we refer to a weaker form of motivational attitude. We assume that an agent considers a given property φ potentially desirable if φ does not conflict with the agent's current desires. In this sense, φ is potentially desirable if it is not incompatible with the agent's current desires. Following ideas presented in [11], let us explain why the operator Δ is a good candidate for modeling the concept of desire and why ∇ is a good candidate for modeling the idea of potential desire.

Mental states We define an agent's mental state as a tuple $M = (E, D)$ where:

- $E \subseteq W$ is a *non-empty* subset of the set of all worlds, and
- $D \subset W$ is a proper subset of the set of all worlds.

The set E defines the set of worlds not ruled out by the agent (i.e., the maximal set of worlds that the agent considers possible), whereas D is the set of desirable worlds for the agent. Let $\mathcal{M}$ denote the set of all mental states. We here assume for every mental state M there exists a world with a minimal degree of desirability 0 corresponding to indifference (this is why $D \neq W$). This type of normality constraint for guaranteed possibility distributions is usually assumed in possibility theory. More generally, a graded mental state is a pair $\widetilde{M} = (\pi, \delta)$ where:

- $\pi : W \to L$ is a normal possibility distribution over the set of all worlds, and
- $\delta : W \to L$ is a function mapping every world w to its desirability (or pleasantness) degree in L, with $\delta(w) = 0$ for some $w \in W$.

[1] We use the term 'motivational' mental attitude (e.g., a desire, a goal or an intention) in order to distinguish it from an 'epistemic' mental attitude such as knowledge or belief.

[2] Here, the word *potential* does not refer to the idea that φ would be desired by the agent as a consequence of his mental state, but the agent has not enough deductive power to become aware of it. It is more the idea that the agent has no reason not to desire φ. Another possible term is *desire admissibility* or *desire compatibility*.

– L is a bounded chain acting as a qualitative scale for possibility and desirability, that make these notions commensurate.

Let us stress the point that while $\delta(w) = 1$ expresses complete desirability, $\delta(w) = 0$ expresses indifference, rather than repulsion. The condition $\delta(w) = 0$ for some $w \in W$ accounts for the claim that desire presupposes not everything is desired.

Modeling desire using $\varDelta$ function We here assume that in order to determine how much a proposition φ is desirable an agent takes into consideration the worst situation in which φ is true. In other words, $\varDelta(||\varphi||) = \alpha$ means that the agent desires all situations where φ holds to at least level α. Thus, denoting by $||\varphi||$ the set of situations where φ is true, for all graded mental states $\tilde{M} = (\pi, \delta)$ and for all propositions φ, we can interpret $\varDelta(||\varphi||) = \min_{u \in ||\varphi||} \delta(u)$ as the extent to which the agent desires φ to be true. Let us justify the following two properties for desires:

$$\varDelta(||\varphi \vee \psi||) = \min(\varDelta(||\varphi||), \varDelta(||\psi||))$$

and

$$\varDelta(||\varphi \wedge \psi||) \geq \max(\varDelta(||\varphi||), \varDelta(||\psi||)).$$

According to the first property, an agent desires φ to be true with a given strength α and desires ψ to be true with a given strength β if and only if the agent desires $\varphi \vee \psi$ to be true with strength equal to $\min(\alpha, \beta)$. Note that the copula "and" in "desiring φ and desiring ψ" corresponds to performing the disjunction of the propositions, as we must perform the set-union of all situations where φ is true and all those where ψ is true (these propositions are not viewed as events referring to a single real world, but as collections of situations). Clearly, in the case of representing uncertainty, this property would not make any sense because the plausibility of $\varphi \vee \psi$ should be clearly *at least* equal to the maximum of the plausibilities of φ and ψ. For the notion of desires, it seems intuitively satisfactory to have the opposite, namely the level of desire of $\varphi \vee \psi$ should be *at most* equal to the minimum of the desire levels of φ and ψ. Indeed, we only deal with here with "*positive*"[3] desires (i.e., desires to reach something with a given strength).

Under this proviso, the level of desire of $\varphi \wedge \psi$ cannot be less than the maximum of the levels of desire of φ and ψ. Here $\varphi \wedge \psi$ does not refer to the disjunction "desiring φ or desiring ψ", but to the idea of desiring φ and ψ simultaneously. According to the second property, the joint occurrence of two desired events φ and ψ is more desirable than the single occurrence of one of the two events. This is the reason why in the right side of the equality we have the $\max$. The latter property does not make any sense in the case of epistemic attitudes like beliefs, as the joint occurrence of two events φ and ψ is epistemically less plausible than the occurrence of a single event. On the contrary the opposite inequality makes perfect sense for motivational attitudes like desires.

By way of example, suppose Peter wishes to go to the cinema in the evening with strength α (i.e., $\varDelta(||goToCinema||) = \alpha$) and, at the same time, he wishes to

[3] The distinction between positive and negative desires is a classical one in psychology. Negative desires correspond to state of affairs the agent wants to avoid with a given strength, and then desires the opposite to be true. However, we do not develop this bipolar view here.

spend the evening with his friend with strength β (i.e., $\Delta(||stayWithFriend||) = \beta$). Then, according to the preceding property, Peter wishes to to go the cinema with his friend with strength at least $\max\{\alpha, \beta\}$ (i.e., $\Delta(||goToCinema \wedge stayWithFriend||) \geq \max\{\alpha, \beta\}$). This is a reasonable conclusion because the situation in which Peter achieves his two desires is (for Peter) at least as pleasant as the situation in which he achieves only one desire. A similar intuition can be found in [8] about the min-decomposability of disjunctive desires, where however it is emphasized that it corresponds to a pessimistic view.

From the normality constraint of δ, we can deduce the following inference rule:

Proposition 1. *For every $M \in \mathcal{M}$, if $\Delta(||\varphi||) > 0$ then $\Delta(||\neg\varphi||) = 0$.*

This means that if an agent desires φ to be true — i.e., with some strength $\alpha > 0$ — then he does not desire φ to be false. In other words, an agent's desires must be consistent.

The operator Δ satisfies the following additional property:

Proposition 2. *For every $M \in \mathcal{M}$, if $||\varphi|| = \emptyset$ then $\Delta(||\varphi||) = 1$.*

i.e. in the absence of actual situations where φ is true, the property φ is desirable by default.

Modeling potential desire using ∇ As pointed out above, we claim that the operator ∇ allows us to capture a concept of potential desire (or non-incompatibility with desire): $\nabla(||\varphi||)$ represents the extent to which an agent considers φ a potentially desirable property or, alternatively, the extent to which the property φ is not incompatible with the agent's desires. An interesting situation is when the property φ is *maximally* potentially desirable for the agent (i.e., $\nabla(||\varphi||) = 1$). This is the same thing as saying that the agent does not desire φ to be false (i.e., $\Delta(||\neg\varphi||) = 0$). Intuitively, this means that φ is totally potentially desirable in as much as the level of desire for $\neg\varphi$ is 0. In particular, given a graded mental state $M = (\pi, \delta)$, let $D = \{w \in W : \delta(w) > 0\}$ be the set of somewhat satisfactory or desirable worlds in M. Then, we have $\nabla(||\varphi||) = 1$ if and only if $\overline{D} \cap ||\neg\varphi|| \neq \emptyset$, i.e., $\neg\varphi$ is consistent with what is not desirable, represented by the set $\overline{D}$.

Another interesting situation is when the property φ is *maximally* desirable for the agent (i.e., $\Delta(||\varphi||) = 1$). This is the same thing as saying that $\neg\varphi$ is not at all potentially desirable for the agent (i.e., $\nabla(||\neg\varphi||) = 0$). It is worth noting that if an agent desires φ to be true, then φ should be *maximally* potentially desirable. This property is expressed by the following valid inference rule which follows straightforwardly from the previous one and from the definition of $\nabla(||\varphi||)$ as $1 - \Delta(||\neg\varphi||)$:

Proposition 3. *For every $\tilde{M}$, if $\Delta(||\varphi||) > 0$ then $\nabla(||\varphi||) = 1$.*

Let us now consider the case in which the agent does not desire φ (i.e., $\Delta(||\varphi||) = 0$). In this case two different situations are possible: either $\Delta(||\neg\varphi||) = 0$ and φ is *fully* compatible with the agent's desires (i.e., $\nabla(||\varphi||) = 1$), or $\Delta(||\neg\varphi||) > 0$ and then φ is not *fully* compatible with the agent's desires (i.e., $\nabla(||\varphi||) < 1$).

Some valid inference rules for desires The following is a valid inference rule for Δ-based logic, see [9] for the proof:

Proposition 4. *For every $M \in \mathcal{M}$, if $\Delta(||\varphi \wedge \psi||) \geq \alpha$ and $\Delta(||\neg\varphi \wedge \chi||) \geq \beta$ then $\Delta(||\psi \wedge \chi||) \geq \min(\alpha, \beta)$.*

Therefore, if we interpret Δ as a desire operator, we have that if an agent desires $\varphi \wedge \psi$ with strength at least α and desires $\neg\varphi \wedge \chi$ with strength at least β, then he desires $\psi \wedge \chi$ with strength at least $\min(\alpha, \beta)$. This seems a reasonable property of desires. By way of example, suppose Peter is planning what to do in the weekend. He has two concomitant desires. On the one hand, Peter desires to go to the contemporary art museum on saturday afternoon and to have dinner at a japanese restaurant on saturday evening with strength at least α. On the other hand, Peter desires not to go the contemporary art museum on saturday afternoon but to go to the sea on sunday morning with strength at least β. Then, it is reasonable to conclude that Peter desires to have dinner at a japanese restaurant on saturday evening and to go to the sea on sunday morning with strength at least $\min(\alpha, \beta)$.

3 Nonmonotonic reasoning with incomplete knowledge

Nonmonotonic reasoning has been extensively studied in AI in relation with the problem of reasoning with rules having exceptions under incomplete information [17], or for dealing with the frame problem in dynamic worlds [7]. In the following, we recall the possibilistic approach [6], which has been proved [5] to provide a faithful representation of the postulate-based approach proposed by Kraus, Lehmann and Magidor [19], and completed in [20].

A default rule "if φ then ψ, generally", denoted by $\varphi \rightsquigarrow \psi$, is then understood formally as the constraint

$$\Pi(\varphi \wedge \psi) > \Pi(\varphi \wedge \neg\psi) \tag{1}$$

on a possibility measure Π describing the semantics of the available knowledge. It expresses that in the context where φ is true, there exist situations where having ψ true is strictly more likely than any situation where ψ is false in the same context.

This constraint can be shown to be equivalent to $N(\psi \mid \varphi) > 0$, when $\Pi(\psi \mid \varphi)$ is defined as the greatest solution of the min-based equation

$$\Pi(\varphi \wedge \psi) = \min(\Pi(\psi \mid \varphi), \Pi(\varphi)) \tag{2}$$

and the duality $N(\psi \mid \varphi) = 1 - \Pi(\neg\psi \mid \varphi)$ holds. There also exists a product-based definition instead of (2). But only the min-based conditioning is used in the following.

Let us consider the following classical example with default rules $d1$: "birds fly", $d2$: "penguins do not fly", $d3$: "penguins are birds", symbolically written

$$d1 : b \rightsquigarrow f;\, d2 : p \rightsquigarrow \neg f;\, d3 : p \rightsquigarrow b.$$

The set of three defaults is thus represented by the following set C of constraints:

$$b \wedge f >_\Pi b \wedge \neg f; p \wedge \neg f >_\Pi p \wedge f; p \wedge b >_\Pi p \wedge \neg b.$$

Let Ω be the finite set of interpretations of the considered propositional language, generated by b, f, p in the example. If this language is made of the literals $\varphi_1, \ldots, p_n$, these interpretations correspond to the possible worlds (i.e., the completely described situations) where the conjunctions $*p_1 \wedge \ldots \wedge *p_n$ are true, where $*$ stands for the presence of the negation sign $\neg$ or its absence. In our example, $\Omega = \{\omega_0 : \neg b \wedge \neg f \wedge \neg p$, $\omega_1 : \neg b \wedge \neg f \wedge p, \omega_2 : \neg b \wedge f \wedge \neg p, \omega_3 : \neg b \wedge f \wedge p, \omega_4 : b \wedge \neg f \wedge \neg p, \omega_5 : b \wedge \neg f \wedge p$, $\omega_6 : b \wedge f \wedge \neg p, \omega_7 : b \wedge f \wedge p\}$. Any interpretation ω thus corresponds to a particular proposition. A possibility distribution π is a mapping from Ω to $[0, 1]$. It induces a ranking of Ω according to the level of normality of each situation. Let $>_\Pi$ denote a ranking of Ω, such that $\omega >_\Pi \omega'$ iff $\pi(\omega) > \pi(\omega')$ on Ω. π is indeed the restriction to propositions describing complete situations of a possibility measure Π defined by $\Pi(\varphi) = \max_{\omega \models \varphi} \pi(\omega)$, where $\omega \models \varphi$ means ω is an interpretation which makes φ true. For instance $\Pi(b \wedge f) = \max(\pi(\omega_6), \pi(\omega_7))$.

Then the set of constraints $\mathcal{C}$ on interpretations is:

$$C1 : \max(\pi(\omega_6), \pi(\omega_7)) > \max(\pi(\omega_4), \pi(\omega_5)),$$

$$C2 : \max(\pi(\omega_5), \pi(\omega_1)) > \max(\pi(\omega_3), \pi(\omega_7)),$$

$$C3 : \max(\pi(\omega_5), \pi(\omega_7)) > \max(\pi(\omega_1), \pi(\omega_3)).$$

Any finite consistent set of constraints of the form $\varphi_i \wedge \psi_i >_\Pi p_i \wedge \neg \psi_i$, representing a set of defaults $\varphi_i \rightsquigarrow \psi_i$ induces a partially defined ranking $>_\Pi$ on Ω, that can be completed according to the principle of minimal specificity, e.g. [5]. This principle assigns to each world ω the highest possibility level (in forming a well-ordered partition of Ω) without violating the constraints. This defines a unique $>_\Pi$.

The well ordered partition of Ω which is obtained in the example is

$$\{\omega_0, \omega_2, \omega_6\} >_\Pi \{\omega_4, \omega_5\} >_\Pi \{\omega_1, \omega_3, \omega_7\}.$$

Let $E_1, \ldots, E_m$ be the obtained partition in the general case. A numerical counterpart to $>_\Pi$ can be defined by $\pi(\omega) = \frac{m+1-i}{m}$ if $\omega \in E_i, i = 1, \ldots, m$. In our example we have $m = 3$ and $\pi(\omega_0) = \pi(\omega_2) = \pi(\omega_6) = 1$; $\pi(\omega_4) = \pi(\omega_5) = 2/3$; $\pi(\omega_1) = \pi(\omega_3) = \pi(\omega_7) = 1/3$. Note that it is purely a matter of convenience to use a numerical scale, and any other numerical counterpart such that $\pi(\omega) > \pi(\omega')$ iff $\omega >_\Pi \omega'$ will work as well. Namely the range of π is used as an ordinal scale.

From this possibility distribution π, we can compute for any proposition φ its necessity degree $N(\varphi)$. For instance,

$$\begin{aligned} N(\neg p \vee \neg f) &= \min\{1 - \pi(\omega) | \omega \models p \wedge f\} \\ &= \min(1 - \pi(\omega_3), 1 - \pi(\omega_7)) = 2/3, \end{aligned}$$

while $N(\neg b \vee f) = \min\{1 - \pi(\omega) \mid \omega \models b \wedge \neg f\} = \min(1 - \pi(\omega_4), 1 - \pi(\omega_5)) = 1/3$

and $N(\neg p \vee b) = \min(1 - \pi(\omega_1), 1 - \pi(\omega_3)) = 2/3$.

The default rule-base can then be encoded in possibilistic logic [10]. The method consists in turning each default $\varphi_i \rightsquigarrow \psi_i$ into a possibilistic clause $(\neg \varphi_i \vee \psi_i, N(\neg \varphi_i \vee \psi_i))$, where N is computed from the greatest possibility distribution π induced by the set of constraints corresponding to the default knowledge base, as already explained. Then we apply the possibilistic inference machinery for reasoning with the defaults together with the available factual knowledge. In our example, we obtain the possibilistic logic base $K = \{(\neg p \vee \neg f, 2/3), (\neg p \vee b, 2/3), (\neg b \vee f, 1/3)\}$. This encodes the generic knowledge embedded in the default rules. Suppose that all we know about the factual situation under consideration is that "Tweety" is a bird, which is encoded by $(b, 1)$. Then we apply the possibilistic logic resolution rule [10]

$$(\neg a \vee b, \alpha), (a \vee c, \beta) \vdash (b \vee c, \min(\alpha, \beta)).$$

Then, we can check that $K \cup \{(b, 1)\} \vdash (f, 1/3)$, i.e., we conclude that if all we know about "Tweety" is that it is a bird, then it flies. If we are said that "Tweety" is in fact a penguin, encoded by $(p, 1)$, then $K \cup \{(b, 1)\} \cup \{(p, 1)\} \vdash (\perp, 1/3)$, which means that K augmented with the available factual information is now inconsistent (at level 1/3).

However, the conclusions which can be obtained with a certainty level strictly greater than the level of inconsistency are safe (the level of inconsistency of a possibilistic logic base is the greatest weight with which $\perp$ can be derived from the base, applying the resolution rule repeatedly). Namely, here, we have $K \cup \{(b, 1)\} \cup \{(p, 1)\} \vdash (\neg f, 2/3)$. Thus, knowing that "Tweety" is a penguin, we now conclude that it does not fly (since $2/3 >$ level of inconsistency$(K \cup \{(b, 1)\} \cup \{(p, 1)\}) = 1/3$). Roughly speaking, the most specific rules w.r.t. a given context remain above the level of inconsistency.

4 Nonmonotonic desires

As recalled in Section 2, expressing that φ true is desired can be arguably represented by $\Delta(\varphi) > 0$. Still $\Delta(\varphi) > 0$ means that *any* situation where φ is true is somewhat desirable, which may appear quite strong, since even if φ true is indeed desired in general, there may exist quite particular situations where this is no longer the case. Thus, one may need to make the modeling nonmonotonic.

This raises the question of defining conditioning for the Δ function, an issue that has been only briefly discussed in [3].

Conditional desires Conditioning can be also defined for guaranteed possibility measures. Since they are decreasing, it should work in a reversed way w.r.t. Π. Namely, conditioning obeys the following equation

$$\Delta(\varphi \wedge \psi) = \max(\Delta(\psi|\varphi), \Delta(\varphi)).$$

Since $\Delta(\varphi) = \min(\Delta(\varphi \wedge \psi), \Delta(\varphi \wedge \neg\psi))$, it expresses that

- either the minimal desirability level over the models of φ is reached on $\varphi \wedge \psi$, so, $\neg\psi$ is preferred in the context φ and $\Delta(\psi|\varphi)$ should be small enough (e.g. 0, for normalisation purposes),
- or that $\Delta(\varphi) = \Delta(\varphi \wedge \neg\psi) < \Delta(\varphi \wedge \psi)$ and then $\Delta(\psi|\varphi)$ should be equal to $\Delta(\varphi \wedge \psi)$.

Thus, we have

$$\Delta(\psi|\varphi) = \begin{cases} \Delta(\varphi \wedge \psi) & \text{if } \Delta(\varphi) < \Delta(\psi \wedge \varphi) \\ = 0 & \text{if } \Delta(\varphi \wedge \neg\psi) \geq \Delta(\varphi \wedge \psi) = \Delta(\varphi) \end{cases}$$

It means that φ and ψ are simultaneously desired because φ is so, and in context φ, ψ is desired as well.

It can be checked that if $\Delta(\psi|\varphi) = 0$ then $\Delta(\varphi \wedge \psi) = \max(\Delta(\varphi), \Delta(\psi))$, i.e., it is the case when the agent desires φ and ψ simultaneously at least as much as φ or ψ separately.

Moreover, as recalled in the previous section, we have $N(\psi|\varphi) > 0$ iff $\Pi(\varphi \wedge \psi) > \Pi(\varphi \wedge \neg\psi)$ iff $N(\varphi \to \psi) > N(\varphi \to \neg\psi)$,
where $\to$ denotes material implication. It expresses that ψ is somewhat certain in context φ iff $\varphi \wedge \psi$ is strictly more plausible than $\varphi \wedge \neg\psi$. An analogous relation holds for guaranteed possibility:

$$\Delta(\psi|\varphi) > 0 \text{ iff } \Delta(\varphi \wedge \psi) > \Delta(\varphi \wedge \neg\psi) \text{ iff } \nabla(\varphi \to \psi) > \nabla(\varphi \to \neg\psi).$$

This points out the fact that the definition of conditioning enforces the following normalization condition: $\min(\Delta(\psi|\varphi), \Delta(\neg\psi|\varphi)) = 0$. It means that everything (i.e. ψ and $\neg\psi$) cannot be simultaneously desired in any context, and the fact that ψ is desired in context φ means that it is more desired than $\neg\psi$ in this context.

Nonmonotonic constraints: an example We are now in a position to investigate nonmonotonic constraints expressed by means of Δ functions. Consider the example in the abstract, where φ and ψ are supposedly logically independent. Assume we have several apparently conflicting pieces of desire

- φ and ψ are separately desired (rather than $\neg\varphi, \neg\psi$), which translates into $\Delta(\varphi) > \Delta(\neg\varphi)$ and $\Delta(\psi) > \Delta(\neg\psi)$;

- in context ψ, $\neg\varphi$ is desired rather than φ, and likewise for context φ, which translates into

$$\Delta(\neg\varphi \wedge \psi) > \Delta(\varphi \wedge \psi); \Delta(\varphi \wedge \neg\psi) > \Delta(\varphi \wedge \psi).$$

Letting $\Delta(\varphi \wedge \psi) = x$, $\Delta(\varphi \wedge \neg\psi) = y$, $\min(\Delta(\neg\varphi \wedge \psi) = z$, $\Delta(\neg\varphi \wedge \neg\psi) = t$, and since $\Delta(\varphi) = \min(\Delta(\varphi \wedge \psi), \Delta(\varphi \wedge \neg\psi))$, we get the system of constraints

$$\begin{aligned} \min(x, y) &> \min(z, t); \\ \min(x, z) &> \min(y, t); \\ z &> x. \\ y &> x \end{aligned}$$

This is equivalent to $z > x > t$ and $y > x > t$. We can reason assuming that a situation is never desired beyond what it is explicitly claimed (this is also called the maximal specificity principle), we obtain desirability values $z = y > x > t = 0$, i.e.,

$$\Delta(\neg\varphi \wedge \psi) = \Delta(\varphi \wedge \neg\psi) > \Delta(\varphi \wedge \psi) > \Delta(\neg\varphi \wedge \neg\psi) = 0.$$

As can be seen, we solve the apparent conflict resulting from desiring φ and ψ separately but not at the same time, by giving the highest desirability to $\neg\varphi \wedge \psi$ and $\varphi \wedge \neg\psi$, while preserving the ones of φ and ψ; interestingly, it implies that $\neg\varphi \wedge \neg\psi$ is not desired at all, while $\varphi \wedge \psi$ remains slightly desirable (otherwise, it would contradict the claims that φ is desirable on the one hand and ψ too on the other hand).

Inferring what is desired Given a set of default desires modelled by $\mathcal{D} = \{\Delta_{\mathrm{d}}(\varphi_i \wedge \psi_i) > \Delta_{\mathrm{d}}(\varphi_i \wedge \neg\psi_i) \mid i = 1, m\}$ and a formula χ describing a context of interest, the problem is to infer what is desirable in this context.

Note that the input information is here of two distinct kinds:

i) conditional desires encoded in terms of a guaranteed possibility measure Δ_{d}, and
ii) a given context described by the set of situations where a certain formula χ is true.

The inference procedure is then the following:

Step 1. Compute the maximally specific possibility distribution δ_{d}, solution of the set of constraints $\mathcal{D}$ by introducing the minimal number of distinct levels necessary for satisfying the set of constraints. If the set of constraints is consistent, this qualitative distribution δ_{d} exists and is unique. Compute the corresponding $\Delta_{\mathrm{d}}(\varphi_i \wedge \psi_i)$ associated with δ_{d}, and encode them as a set $\mathcal{T}$ of possibilistic Δ-formulas $[\varphi_i \wedge \psi_i, \Delta_{\mathrm{d}}(\varphi_i \wedge \psi_i)]$ [4].

Step 2. Move the variables pertaining to the description of the context of each rule to the weight part of the possibilistic Δ-formulas, thus defining $\mathcal{T}^*$. Namely, $[\varphi_i \wedge \psi_i, \alpha_i]$ semantically entails $[\psi_i, \min(v(\varphi_i), \alpha_i)]$ where $v(\varphi_i) = 1$ is true, and $v(\varphi_i) = 0$ otherwise [4].

Step 3. Project $\mathcal{T}^*$ on the context χ of interest. Namely if $\chi \vDash p_i$, then $v(\varphi_i)$ is set to 1, otherwise it is set to 0. We thus obtain $\mathcal{T}^*_\chi$.

Step 4. Compute the level of Δ-inconsistency $inc(\mathcal{T}^*_\chi)$ of $\mathcal{T}^*_\chi$, where $inc(\mathcal{T}^*_\chi) = \max\{\alpha | \mathcal{T}^*_\chi \vDash [\top, \alpha]\}$. For doing it, use the cut rule associated with Δ-formulas [14], namely

$$[\varphi \wedge \psi, \alpha], [\neg\varphi \wedge \phi, \beta] \vDash [\psi \wedge \phi, \min(\alpha, \beta)]$$

As a consequence $[\varphi, \alpha], [\neg\varphi, \beta] \vDash [\top, \min(\alpha, \beta)]$.

Step 5. Infer the desires $[\psi_i, \alpha_i]$ such as $\alpha_i > inc(\mathcal{T}^*_\chi)$.

Example:

$\mathcal{D} = \{\Delta_{\mathrm{d}}(\varphi) > \Delta_{\mathrm{d}}(\neg\varphi), \Delta_{\mathrm{d}}(\psi \wedge \neg\varphi) > \Delta_{\mathrm{d}}(\psi \wedge \varphi)\}$

We get $\mathcal{T} = \{[\varphi, \beta], [\psi \wedge \neg\varphi, \alpha]\}$, with $\alpha > \beta$.

Then $\mathcal{T}^* = \{[\varphi, \beta], [\neg\varphi, \min(v(\psi), \alpha)]\}$

Assume $\chi \equiv \top$, one can only infer $[\varphi, \beta]$, since $\mathcal{T}^*_\top = \{[\varphi, \beta]\}$, (indeed $[\neg\varphi, 0]$, equivalent to $\Delta_{\mathrm{d}} \geq 0$, is a trivial formula which can be deleted, since $v(\psi)$ is set to 0).

Now suppose $\chi \equiv \psi$, then $v(\psi) = 1$ and $\mathcal{T}^*_\psi = \{[\varphi, \beta], [\neg\varphi, \alpha]\}$. Since $\mathcal{T}^*_\psi \vDash [\top, \min(\alpha, \beta)] = [\top, \beta]$, only $[\neg\varphi, \alpha]$ is safe from inconsistency, as $\alpha > inc(\mathcal{T}^*_\psi) = \beta$.

We thus get the expected nonmonotonic behavior.

5 Concluding remarks

In this note we have outlined an approach for handling the nonmonotonic behavior of desires, a problem which has apparently not be much considered in AI. Starting with the fact that desires can be appropriately described by means of guaranteed possibility measures in the sense of possibility theory, and keeping the lesson of the encoding of knowledge-oriented nonmonotonic reasoning in the setting of possibility theory, we have proposed a possibilistic approach to the nonmonotonic handling of desires. This approach, based on the notion of conditional guaranteed possibility measures, somewhat parallels the nonmonotonic handling of default knowledge under incomplete information, but with some noticeable differences due to the decreasingness of guaranteed possibility measures wrt entailment, and the need of distinguishing between desires and contexts of interest in which we consider them.

Although the outlined approach is promising, it remains preliminary in many respects. Wee still need a formal proof that the deduction method in possibilistic Δ-based logic outlined above does compute then inference of $\Delta(\varphi|\chi) > 0$ from the conditional desire base. A postulate-based approach to reasoning about conditional desires in a way that would parallel Kraus, Lehmann and Magidor postulates for reasoning with default knowledge, is a natural objective as well. The nonmonotonic handling of graded desires might also parallel the approach used in [16] for dealing with more or less certain pieces of default knowledge.

References

1. Banerjee, M., Dubois, D.: A simple logic for reasoning about incomplete knowledge. International Journal of Approximate Reasoning 55, 639–653 (2014)
2. Battigalli, P., Siniscalchi, M.: Strong belief and forward induction reasoning. J. of Economic Theory 106(2), 356–391 (2002)
3. Benferhat, S., Dubois, D., Kaci, S., Prade, H.: Bipolar possibilistic representations. In: Darwiche, A., Friedman, N. (eds.) Proc. 18th Conf. in Uncertainty in Artificial Intelligence (UAI '02), Edmonton, Alberta, Aug. 1-4. pp. 45–52. Morgan Kaufmann (2002)
4. Benferhat, S., Dubois, D., Kaci, S., Prade, H.: Modeling positive and negative information in possibility theory. Int. J. Intell. Syst. 23(10), 1094–1118 (2008)
5. Benferhat, S., Dubois, D., Prade, H.: Nonmonotonic reasoning, conditional objects and possibility theory. Artif. Intell. 92(1-2), 259–276 (1997)
6. Benferhat, S., Dubois, D., Prade, H.: Practical handling of exception-tainted rules and independence information in possibilistic logic . Applied Intelligence 9, 101–127 (1998)
7. Brewka, G., Marek, V., Truszczynski, eds., M.: Nonmonotonic Reasoning. Essays Celebrating its 30th Anniversary., Studies in Logic, vol. 31. College Publications (2011)
8. Casali, A., Godo, L., Sierra, C.: A graded BDI agent model to represent and reason about preferences. Artificial Intelligence 175, 1468–1478 (2011)
9. Dubois, D., Hajek, P., Prade, H.: Knowledge-driven versus data-driven logics. Journal of Logic, Language, and Information 9, 65–89 (2000)
10. Dubois, D., Lang, J., Prade, H.: Possibilistic logic. In: Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3, pp. 439–513. Oxford Univ. Press (1994)
11. Dubois, D., Lorini, E., Prade, H.: Bipolar possibility theory as a basis for a logic of desires and beliefs. In: Liu, W., Subrahmanian, V.S., Wijsen, J. (eds.) Proc. 7th Int. Conf. on Scalable Uncertainty Management (SUM'13), Washington, DC, Sept. 16-18. LNCS, vol. 8078, pp. 204–218. Springer (2013)
12. Dubois, D., Prade, H.: Epistemic entrenchment and possibilistic logic. Artificial Intelligence 50, 223–239 (1991)
13. Dubois, D., Prade, H.: Possibility theory: qualitative and quantitative aspects. In: Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, pp. 169–226. Kluwer (1998)

14. Dubois, D., Prade, H.: Possibilistic logic: a retrospective and prospective view. Fuzzy Sets and Systems 144, 3–23 (2004)
15. Dubois, D., Prade, H., Schockaert, S.: Reasoning about uncertainty and explicit ignorance in generalized possibilistic logic. In: Proceedings of the European Conference on Artificial Intelligence (2014)
16. Dupin de Saint-Cyr, F., Prade, H.: Handling uncertainty and defeasibility in a possibilistic logic setting. Int. J. Approx. Reasoning 49(1), 67–82 (2008)
17. Group Léa Sombé: Besnard, P., Cordier, M.O., Dubois, D., Fariñas del Cerro, L., Froidevaux, C., Moinard, Y., Prade, H., Schwind, C., Siegel, P.: Reasoning under Incomplete Information in Artificial Intelligence: A Comparison of Formalisms Using a Single Example. Wiley (1990)
18. Hume, D.: A Treatise of Human Nature. Clarendon Press, Oxford (1978)
19. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence 44, 167–207 (1990)
20. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial Intelligence 55, 1–60 (1992)
21. McDaniel, K., Bradley, B.: Desires. Mind 117(466), 267–302 (2008)
22. Spohn, W.: Ordinal conditional functions: a dynamic theory of epistemic states. In: Causation in Decision, Belief Change and Statistics, vol. 1, pp. 105–134. Kluwer (1988)