A possibility theory-based approach to desire change

Didier Dubois and **Emiliano Lorini** and **Henri Prade**[1]

Abstract. Desire is quite different from belief. While the accumulation of beliefs tend to reduce the remaining possible worlds they point at, the accumulation of desires tend to increase the set of states of affairs tentatively considered as satisfactory. Indeed beliefs are expected to be closed under conjunctions, while one can argue that endorsing $\varphi \vee \psi$ as a desire means to desire both φ and ψ. Still desiring φ and $\neg\varphi$ at the same time is not usually regarded as rational, since it does not make much sense to desire one thing and its contrary at the same time. Thus when a new desire is added to the set of desires of an agent, a revision process may be necessary. Just as belief revision relies on an epistemic entrenchment relation, desire relation is based on a hedonic entrenchment relation satisfying other properties, due to the different natures of belief and desire. Epistemic entrenchment relations are known to be qualitative necessity relations. In this paper it is shown that a well-behaved desire revision operation obeying a set of reasonable postulates is underlied by a qualitative guaranteed possibility relation in the sense of possibility theory. Then the general framework of possibilistic logic provides a syntactic setting for encoding desire change.

1 Introduction

Desires constitute the primitive form of motivational attitude that drives an agent to plan her action aimed at satisfying them. Specifically, taking into account her beliefs about the world, the agent chooses what to do in the pursuit of her desires. The result of the agent's choice constitutes her intentions to which she is then committed . Such a simplified schema is for instance advocated in [26] taking inspiration from the philosophical and psychological literature [7]. This is also the building blocks of BDI agents, where B, D, I, respectively stand for Beliefs, Desires, and Intentions [28].

Desires and intentions are sometimes used more or less interchangeably in the literature. However, desires and intentions should be carefully distinguished. For instance, let us reconsider an example adapted from [23]: namely, an agent has a taste for (i.e. in this paper, a desire of) eating sushi. Today, she has the intention to go to restaurant "The Japoyaki" and to eat sushi (after making the choice of the restaurant on the basis of what she heard about). Then learning that the available sushi are made with fish that may be not fresh enough, she is led to revise her plans and to order something else. Here, her intention changes, although she keeps her taste (and, consequently, the desire) for sushi. In case she rather decides to go to another sushi restaurant, she would revise her intention, but not her desire.

In this paper, we do not consider intentions, but only desires. More precisely, we consider positive desires only, namely those that it would be really satisfactory to concretize, as opposed to negative desires corresponding to situations to be avoided because they are unsatisfactory, unbearable for the agent.

We advocate that an agent cannot simply cumulate desires without never making any revision, since it does not make sense to desire everything (at least according to the wisdom of mankind). This means that sometimes an agent has to revise her desires, not on the basis of some believed information about the state of the world which would trigger a change of intention, but just because the acceptance of a new desire altogether with her previous desires would lead her to desire everything and its contrary.

Such a situation is clearly similar to the revision of her beliefs by an agent receiving a new piece of information that she considers to be true, since she has to preserve the consistency of her beliefs. But desires and beliefs behave differently. Indeed, while believing φ and believing ψ amounts to believing $\varphi \wedge \psi$, both desiring φ and desiring ψ amounts to desiring $\varphi \vee \psi$, and conversely.

The difference of behavior between desire and belief has been pointed out by several authors [9, 10], which led them to propose possibility theory as a setting appropriate for modeling desires in terms of guaranteed possibilities, while beliefs can be represented in terms of necessity measures in this setting [17]. More recently, in [12], a modeling of desire change has been briefly outlined, which mirrors to some extent the way belief change can be represented in the framework of possibility theory [13, 14, 4], without proposing any postulates nor representation results. In this paper, we provide postulates for desire revision, contrast them with belief revision postulates [18], and show how desire revision (as well as expansion and contraction) can be implemented in possibility theory in agreement with our postulates, both semantically and syntactically.

The paper is organized as follows. In Section 2, we highlight the main intuitions behind the concept of desire change in contrast with the concept of belief change, from a philosophical and AI perspective. Section 3 introduces the idea of a hedonic entrenchment relation that rank-order desires, and provide axioms for such a relation, whose unique numerical counterpart is a guaranteed possibility distribution, associated with a guaranteed possibility measure in the sense of possibility theory. In Section 4, desires are then represented in this setting. The guaranteed possibility distribution enables us to associate any set of desires with a level of unacceptability, which is the counterpart of the level of inconsistency for a set of beliefs represented in a possibilistic logic manner. Section 5 provides axioms for desire revision and Section 6 presents the revision of sets of prioritized desires axiomatically, semantically, and syntactically using a special type of possibilistic logic. Expansion and contraction of desires are also characterized and discussed.

2 Conceptual framework

An important and general distinction in philosophy of mind is between epistemic attitudes and motivational attitudes. This distinction

[1] IRIT, 118 route de Narbonne, 31062 Toulouse Cedex 09, France, email:{dubois, lorini, prade}@irit.fr.

is in terms of the *direction of fit* of mental attitudes to the world. While epistemic attitudes aim at being true and their being true is their fitting the world, motivational attitudes aim at realization and their realization is the world fitting them [27, 1, 21]. Searle [29] calls "mind-to-world" the first kind of *direction of fit* and "world-to-mind" the second one. Desire is representative of the family of motivational attitudes, while belief is representative of the family of epistemic attitudes. Other kinds of motivational and epistemic attitudes exist with different functions and properties such as preferences, goals and moral values, knowledge and opinions (cf. [26] for a logical theory of the relationship between desires, moral values and preferences).

Beliefs are mental representations aimed at representing how the physical, mental and social worlds are. In contrast, following the Humean conception, a desire can be viewed as an agent's attitude consisting in an anticipatory mental representation of a pleasant state of affairs (representational dimension of desires) that motivates the agent to achieve it (motivational dimension of desires). The motivational dimension of an agent's desire is realized through its representational dimension, in the sense that, a desire motivates an agent to achieve it *because* the agent's representation of the desire's content gives her anticipatory pleasure, following John Locke's intuition. For example when an agent desires to eat sushi, she imagines herself eating sushi and this representation gives her pleasure. This pleasant representation motivates her to go to the "The Japoyaki" restaurant in order to eat sushi.

Desire and belief have also different origins. Belief revision is triggered either via direct sensing from the external environment (e.g., I believe that there is a fire in the house since I can see it) or via communication (e.g., I believe that there is a fire in the house since you told me this and I trust what you say). Desire change is triggered under other conditions. In the case of human agents, these conditions might be physiological or epistemic. For example, the desire of drinking a glass of water could be activated by the feeling of thirst (physiological condition) and the desire of going outside for a walk might be activated by the belief that it is a sunny day (epistemic condition). In the case of artificial agents, conditions of desire activation should be specified by the system's designer. For example, a robotic assistant who has to take care of an old person could be designed in such a way that every day at 4 pm the desire of giving a medicine to the old person is activated in its mind. This highlights that belief change and desire change have different interpretations and meanings.

From the AI perspective, having a formal theory of desire change — and desire revision, as a kind of desire change operation — is important for at least two reasons: (i) desire change is a the heart of the concept of autonomous agent, (ii) a theory of desire change is required to design artificial systems who are expected to interact with humans in the appropriate way. Indeed, one important aspect of the concept of *autonomy* is the fact of being endowed with a mechanism responsible for the generation of internal motivations. From this perspective, an intelligent system (e.g., a robot, a virtual agent) is autonomous insofar it can generate its own desires on the basis of such a mechanism.

Moreover, an artificial agent interacting with a human should be capable of both ascribing desires to the human and understanding how the desires of the human evolve over time.

3 Hedonic states as desirability relations

In this paper, desires are represented by means of a finite set D of sentences, denoted in the following by φ, ψ, χ or ν that belong to a Boolean algebra $\mathcal{B}$. Hence $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$ belong to $\mathcal{B}$ as well. As usual $\top$ and $\bot$ are the top and bottom elements of $\mathcal{B}$ and denote the tautology and the contradiction respectively; $\varphi \rightarrow \psi =_{def} \neg\varphi \vee \psi$; $\varphi \equiv \psi =_{def} (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$. $\vdash$ denotes the entailment defined as usual by $\varphi \vdash \psi$ if and only if $\varphi \rightarrow \psi \equiv \top$.

In this section we first present the notion of hedonic state. We introduce this terminology, since volition is the name of the cognitive process by which an agent decides on and commits to a particular course of action, and since ultimately this process that takes into account the agent's beliefs about the world, relies on the desires of the agent. Then, hedonic states will be described by means of an ordering relation acknowledging the fact that desire is a matter of relative strength. This relation should obey particular axioms, and has guaranteed possibility measures [15] as a unique numerical counterpart, as we shall see. This leads to represent a hedonic state by means of a guaranteed possibility distribution.

3.1 Axioms for desirability relations

If an agent is satisfied by having a cup of coffee or a cup of tea, she should be satisfied by having a cup of coffee. This simple example clearly suggests that desires behave in a reverse way with respect to logical entailment.

Due to this reverse behavior, we should note that desiring φ for an agent entails that she desires ψ as well, as soon as $\psi \vdash \varphi$. If we prefer, desiring φ means desiring any situation where φ is true. For example, since having a cup of coffee (χ) logically entails having a cup of coffee or a cup of tea ($\chi \vee \nu$), having a desire for a cup of coffee or a cup of tea ($\chi \vee \nu$) means that the agent would enjoy a cup of coffee (χ), as well as she would enjoy a cup of tea (ν). We have to keep in mind that desires are considered as tentative in nature, and have not reached the step to be adopted as goals to pursue. Being pleased to have at least tea or coffee served does not mean that one has the goal to drink both in the case where both would be available. Here 'desiring' just means 'finding satisfactory', 'finding enjoyable', 'having a taste for' and so on. Note that the behavior of desires with respect to logical entailment is in full contrast with respect to beliefs, where believing φ for an agent entails that she believes ψ as well as soon as $\varphi \vdash \psi$.

This suggests that desires obey a *reversed entailment*, namely a desire for ψ entails a desire for φ, i.e.,

$$\psi \vdash_{des} \varphi \text{ if and only if } \varphi \vdash \psi.$$

This agrees with the fact that if all the models of ψ are found satisfactory, then any interpretation taken in the subset made of the models of φ is a satisfactory state. As a consequence, in the same way as $\varphi \vdash \top$ trivially holds for any belief φ, we have $\psi \vdash_{des} \bot$ for any desire ψ. In other words, desiring nothing is fully satisfactory, as believing tautologies is compulsory.

Moreover desires are a matter of strength. Some situation may be more strongly desired than another one by an agent. Thus, a hedonic state of an agent will be described by an ordering relation on $\mathcal{B}$, denoted by $\geq_{\Delta}$, called *desirability relation*. Such a relation compares sentences in terms of satisfaction they provide to the agent if made true. $\varphi \geq_{\Delta} \psi$ should be read "φ is at least as desirable as ψ"; it means that concretizing φ should be at least as satisfactory as concretizing ψ, or if we prefer that φ is desired more strongly than ψ in the broad sense. As usual, $\varphi >_{\Delta} \psi$ when $\varphi \geq_{\Delta} \psi$ but not $\psi \geq_{\Delta} \varphi$; and $\varphi \sim_{\Delta} \psi$ means $\varphi \geq_{\Delta} \psi$ and $\psi \geq_{\Delta} \varphi$.

This leads to suppose that $\geq_{\Delta}$ should satisfy the following axioms:

(A0) $\bot >_\Delta \top$
(A1) $\varphi \geq_\Delta \psi$ or $\psi \geq_\Delta \varphi$
(A2) $\varphi \geq_\Delta \psi$ and $\psi \geq_\Delta \chi$ imply $\varphi \geq_\Delta \chi$
(A3) $\bot \geq_\Delta \varphi$
(Pos) $\forall \varphi$, if $\psi \geq_\Delta \chi$ then $\varphi \vee \psi \geq_\Delta \varphi \vee \chi$

Once recognized that desires behave in a reverse way with respect to entailment, (A0) expresses non triviality, while axiom (A3) is a limit condition; it states that having no desire (here represented by $\bot$) cannot be unsatisfactory; in other words, having no desire for an agent should lead her to be ever satisfied. The other axioms are perfectly neutral with respect to a reverse, or a normal, behavior with respect to entailment. Axioms (A1) and (A2) simply say that relation $\geq_\Delta$ is complete and transitive respectively. Axiom (A1) is a working assumption; considering more generally a partial order is left for further investigation. Axiom (Pos) states that if "ψ is at least as desirable as χ", this preference in the broad sense cannot be altered by enlarging the scope of the comparaison on both sides in the same way by φ. Indeed if you find more desirable (in the broad sense) to drink tea than to drink coffee, then you should find at least as desirable to drink tea or orange juice as to drink coffee or orange juice (even if your actual preference is for orange juice). Axiom (Pos) that makes sense for desires can be encountered in other modeling problems such as conditional logics and comparative possibility theory [25, 8]

3.2 Properties of desirability relations

The previous set of axioms entail noticeable properties for desirability relations that agree with intuition. First, we can establish the following result.

Proposition 1 *Under axioms (A0)-(A3), axiom (Pos) is equivalent to* (Δ) *if $\varphi \geq_\Delta \psi$ then $\varphi \vee \psi \sim_\Delta \psi$.*

Proof
$(\Delta) \Rightarrow$ (Pos).
Assume $\psi \geq_\Delta \chi$.
- If $\varphi \geq_\Delta \psi \geq_\Delta \chi$, $\varphi \vee \psi \sim_\Delta \psi \geq_\Delta \chi \sim_\Delta \varphi \vee \chi$;
- If $\psi \geq_\Delta \varphi \geq_\Delta \chi$, $\varphi \vee \psi \sim_\Delta \varphi \geq_\Delta \chi \sim_\Delta \varphi \vee \chi$;
- If $\psi \geq_\Delta \chi \geq_\Delta \varphi$, $\varphi \vee \psi \sim_\Delta \varphi \geq_\Delta \varphi \sim_\Delta \varphi \vee \chi$.
(Pos) $\Rightarrow (\Delta)$.
Let $\chi = \varphi$ in (Pos). Then (Pos) $\Rightarrow \forall \varphi$, if $\psi \geq_\Delta \varphi$ then $\varphi \vee \psi \geq_\Delta \varphi$. But (Pos) applied with $\psi = \bot$ (and $\chi = \psi$) leads to $\forall \varphi, \varphi \geq_\Delta \varphi \vee \psi$. Hence, if $\varphi \leq_\Delta \psi$, then $\varphi \sim_\Delta \varphi \vee \psi$. □

Clearly, (Δ) agrees with the reverse behavior of comparative desirability with respect to entailment, and expresses that desiring $\varphi \vee \psi$ has the same strength as desiring the least desired of φ and ψ. Indeed if the agent desires coffee (φ) more strongly than tea (ψ), it means that concretizing $\varphi \vee \psi$ has the same appeal as concretizing ψ. Indeed, it seems intuitively satisfactory that the strength of desire of $\varphi \vee \psi$ should be *at most* equal to the minimum of the desire strengths of φ and ψ (when dealing with positive desires).

Moreover, under axioms (A0)-(A3) and (Pos), it can be easily shown that the following properties hold.

Proposition 2
[a] If $\varphi \vdash \psi$ then $\varphi \geq_\Delta \psi$.
[b] Either $\varphi \sim_\Delta \top$ or $\neg\varphi \sim_\Delta \top$ or both.
[c] $\forall \varphi, \varphi \geq_\Delta \top$.

Proof

[a] As already observed letting $\psi = \bot$ in (Pos), the following holds $\forall \varphi, \varphi \geq_\Delta \varphi \vee \chi$. If $\varphi \vdash \nu$, ν can be rewritten as $\varphi \vee \chi$.

[b] It is an immediate consequence of (A2) and (Δ), since letting $\psi = \neg\varphi$ in (Δ), we get $\top \sim_\Delta \neg\varphi$ if $\varphi \geq_\Delta \neg\varphi$, and $\top \sim_\Delta \varphi$ if $\neg\varphi \geq_\Delta \varphi$.

[c] Since $\forall \varphi, \forall \chi, \varphi \geq_\Delta \varphi \vee \chi$, letting $\chi = \neg\varphi$ yields the result. □

Property [a] expresses the reverse behavior of $\geq_\Delta$ with respect to logical entailment (i.e., decreasingness with respect to entailment). Note that in particular, $\bot \geq_\Delta \psi, \forall \psi$, as requested by axiom (A3). Property [b] states that one cannot desire φ and $\neg\varphi$ at the same time, at least one the two options should not be desired more than what is the least desired, which is the tautology, whose non-desirability is stated by [c]. This expresses nothing but the fact that one cannot desire everything at the same time. This contrasts with the fact that desiring nothing ($\bot$) is no problem at all. Indeed there is no harm to desire $\bot$, since you are then eversatisfied. Indeed it is not at all compulsory to desire something.

3.3 Desirability relations vs. epistemic entrenchments

It is worth noticing that the set of axioms (A0)-(A3) and (Pos) depart from the ones characterizing *epistemic entrenchment* relations $\geq_{epis}$ that underly any well-behaved belief revision process [18]. It has been established that epistemic entrenchment relations are nothing but comparative necessity relations $\geq_N$ up to a minor difference, namely axiom $\top >_N \bot$ is strengthened into $\top >_{epis} \varphi, \forall \varphi$ for epistemic entrenchment [13]. Comparative necessity relations, and thus epistemic entrenchment relations satisfy (A1) and (A2), but they obey counterparts of the other axioms, namely (A'0): $\top >_N \bot$ (and the above-mentioned strengthening for $\geq_{epis}$) and (A'3) $\varphi \geq_N \bot, \forall \varphi$. They both satisfy the characteristic property of comparative necessity relations:

$$\text{if } \varphi \geq_N \psi \text{ then } \varphi \wedge \chi \geq_N \psi \wedge \chi,$$

which under axioms (A'0)-(A1)-(A2)-(A'3) is equivalent to

$$\text{if } \varphi \geq_N \psi \text{ then } \varphi \wedge \psi \sim_N \psi,$$

where $\varphi \geq_N \psi$ means that φ is at least as certain as ψ.

By duality, comparative necessity relations $\geq_N$ are associated with comparative possibility relations $\geq_\Pi$ through the equivalence $\varphi \geq_\Pi \psi \Leftrightarrow \neg\psi \geq_N \neg\varphi$. Comparative possibility relations [8] satisfy axioms (A'0), (A1)-(A2), together with $\varphi \geq_\Pi \bot$ and $\psi \geq_\Pi \chi$ implies $\varphi \vee \psi \geq_\Pi \varphi \vee \chi$, i.e., axiom (Pos)! It is remarkable that switching from comparative possibility relations to desirability relations comes down to only changing axiom $\varphi \geq_\Pi \bot$ for comparative possibility to axiom (A3) $\bot \geq_\Delta \varphi$ for desirability relations.

3.4 Desirability relations and possibility theory

We have seen that an ordering relation obeying axioms (A0)-(A3) and (Pos) may be appropriate for modeling (positive) desirability in a relative way. A natural question is then to wonder what are the absolute scale-valued functions, if any that agree with a desirability relation.

A numerical function F from $\mathcal{B}$ to $[0, 1]$ is said to agree with a relation $\geq_R$ if $\forall \varphi, \psi, \varphi \geq_R \psi \Leftrightarrow F(\varphi) \geq F(\psi)$. In the following we assume that the set of literals φ, ψ, ... of the considered language is finite. Thus the set W of associated interpretations is finite. We denote by $|w|$ the proposition whose unique model is $w \in W$.

The only numerical functions compatible with the desirability relation ordering $\geq_\Delta$ are guaranteed possibility measures [15] in the sense of possibility theory, as shown now. A guaranteed possibility measure Δ, from $\mathcal{B}$ to $[0,1]$, is characterized by the limit conditions $\Delta(\bot) = 1$ and $\Delta(\top) = 0$, and by the decomposability property:

$$\Delta(\varphi \vee \psi) = \min(\Delta(\varphi), \Delta(\psi)), \forall \varphi, \psi \in \mathcal{B}. \quad (1)$$

We first establish the following proposition, before proving the announced result.

Proposition 3 $\forall \varphi \neq \bot, \exists w \vDash \varphi, |w| \sim_\Delta \varphi.$

Proof If $\varphi = |w|$, this is obvious. If $\varphi \neq |w|$, let $\varphi_1 \neq \bot$ be a strict implicant of φ, such that $\varphi \wedge \neg\varphi_1 \geq_\Delta \varphi_1$. It is always possible to find such a φ_1 thanks to axiom (A1). Using (Δ), since $(\varphi \wedge \neg\varphi_1) \vee \varphi_1 = \varphi$, we conclude $\varphi \sim_\Delta \varphi_1$. If φ_1 has not a unique model, we define $\varphi_2 \neq \bot$ as a strict implicant of φ_1 such that $\varphi_1 \wedge \neg\varphi_2 \geq_\Delta \varphi_2$, and so on. The sequence $\varphi, \varphi_1, \varphi_2, ..., \varphi_i, ...$ is a chain of implicants which is strictly decreasing (in terms of number of models). Since we assume a finite setting, $\exists n, \exists w, \varphi_n = |w|$. Then from axiom (Δ), $\varphi \sim_\Delta \varphi_1 \sim_\Delta \varphi_2 \sim_\Delta \cdots \sim_\Delta \varphi_n$. □

Proposition 4 *Any numerical function F, from $\mathcal{B}$ to $[0,1]$, agreeing with an ordering relation $\geq_\Delta$ obeying axioms (A0)-(A3) and (Pos) is a guaranteed possibility measure. Conversely any guaranteed possibility measure from a Boolean algebra $\mathcal{B}$ to $[0,1]$ satisfying $\Delta(\bot) > \Delta(\top)$ induces a qualitative relation satisfying (A0)-(A3) and (Pos).*

Proof

($\Rightarrow$) From Proposition 3 and its proof, we know that $\forall \varphi \neq \bot, \exists w \vDash \varphi$, such that $\varphi \sim_\Delta \varphi_1 \sim_\Delta \varphi_2 \sim_\Delta \cdots \sim_\Delta \varphi_n = |w|$, where φ is decomposed in a chain of implicants φ_1, φ_2, ... and $\varphi \wedge \neg\varphi_1 \geq_\Delta \varphi_1$, $\varphi_1 \wedge \neg\varphi_2 \geq_\Delta \varphi_2, \cdots, \varphi_{n-1} \wedge \neg\varphi_n \geq_\Delta \varphi_n$. Taking $\varphi = \top$, $\varphi \wedge \neg\varphi_1$, $\varphi_1 \wedge \neg\varphi_2$, ..., $\varphi_{n-1} \wedge \neg\varphi_n$, $|w|$ make a partition of W. Starting from $\varphi_{n-1} \wedge \neg|w| \geq_\Delta |w|$, any model w' of $\varphi_{n-1} \wedge \neg|w|$ is either such that $|w'| \sim_\Delta |w|$ or $|w'| >_\Delta |w|$. Let φ' be the proposition whose set of models is exactly $W \setminus \{w' \text{ s.t. } |w'| \sim_\Delta |w|\}$. Let us apply Proposition 3 to φ' and find w'' such that $\varphi' \sim_\Delta |w''|$. We can iterate this process until we reach k with $\varphi'_{k-1} = \bot$. As a result, we can organize W into a set of layers of strictly increasing desirability (two interpretations in the same layer having the same desirability), and associate the value of a numerical function δ to each layer. Then it is possible to build a function $\Delta(\varphi) = \min_{w \vDash \varphi} \delta(w)$. Due to axiom (Δ), Δ is an agreeing function, and it is clear that Δ is satisfy (1). Moreover, $\Delta(\top) = \min_{w \in W} \delta(w)$ can be taken to be equal to 0.

($\Leftarrow$) Conversely, a guaranteed possibility measure, in a finite setting, is based on a distribution δ such that $\Delta(\varphi) = \min_{w \vDash \varphi} \delta(w)$, and it is easy to check that it induces an ordering relation that satisfies axioms (A0)-(A3) and (Pos). □

Thus, Property 4 states that the only numerical functions agreeing with a qualitative ordering $\geq_\Delta$ are those obeying decomposability property (1). Note that the range $[0,1]$ may be replaced by any linearly ordered, possibly finite, scale.

4 Modeling desires in possibility theory

Thus, we can interpret $\Delta(\varphi)$, where Δ is a guaranteed possibility measure, as the extent to which the agent desires φ to be true. As suggested in [9], and advocated in [11], a desire φ is properly represented by a constraint of the form $\Delta(\varphi) \geq \alpha$ which stands for "the agent desires φ with strength at least α", while the concept of belief that is properly represented by a constraint of the form $N(\varphi) \geq \alpha$ which stands for "the agent believes φ with strength at least α", where N is a necessity measure. Beliefs, modeled by means of necessity measures, satisfy

$$N(\varphi \wedge \psi) = \min(N(\varphi), N(\psi))$$

i.e., believing φ *and* ψ amounts to believing φ and to believing ψ. As a consequence of property (1) we have

$$\min(\Delta(\varphi), \Delta(\neg\varphi)) = \Delta(\top) = 0$$

which is the numerical counterpart of property $[b]$ in Proposition 2. Moreover $\Delta(\bot) = 1$ by convention, since Δ is monotonically decreasing with respect to entailment. Besides,

$$\Delta(\varphi \wedge \psi) \geq \max(\Delta(\varphi), \Delta(\psi)).$$

This is the consequence that Δ is decreasing with respect to entailment (i.e. property $[a]$ in Proposition 2). This makes perfect sense for motivational attitudes like desires, as suggested by the following example.

Example 1 Suppose Paul has a taste for cheese with strength α (i.e., $\Delta(\text{eat cheese}) = \alpha$) and, at the same time, he likes to drink wine with strength β (i.e., $\Delta(\text{drink wine}) = \beta$). Then, according to the preceding property, Paul likes to eat cheese and drink wine with strength at least $\max(\alpha, \beta)$ (i.e., $\Delta(\text{eat cheese} \wedge \text{drink wine}) \geq \max(\alpha, \beta)$). This is a reasonable conclusion because the situation in which Paul achieves his two desires is (for Paul) at least as pleasant as the situation in which he achieves only one desire.

One might object that if it is generally the case that satisfying simultaneously two desires is at least as good as satisfying one of them, there may exist exceptional situations where it is not the case. Just imagine, in the above example, the case where the wine is corked, and so Paul would not like to drink it with his cheese. This is a situation of nonmonotonic desires that could be also coped with in this setting; see [11] for a preliminary proposal in the possibilistic reasoning setting, but this is out of the scope of the present paper.

4.1 Hedonic states as guaranteed possibility distributions

As in Gärdenfors [18], we assume that the content of a sentence can be described by a subset of possible worlds $w \in W$. Namely let $||\varphi|| \subseteq W$ denote the subset of worlds (corresponding to interpretations) in which the propositional formula φ is true. In other words, φ is put in disjunctive normal form, $\varphi = \bigvee_{i=1,f(\varphi)} \omega^i_\varphi$ and $\forall i, \exists w^i \in W, ||\omega^i_\varphi|| = \{w^i\}$. Due to property (1), a guaranteed possibility measure. Δ in a finite setting can always be written as

$$\Delta(\varphi) = \min_{w^i \in ||\varphi||} \delta(w^i) \quad (2)$$

where $\delta(w^i) = \Delta(\omega^i_\varphi)$ with $||\omega^i_\varphi|| = \{w^i\}$. The function δ is called a guaranteed possibility distribution; its domain W and its range is

$[0, 1]$, or more generally any linearly ordered scale S. Thus, $\delta(w)$ represents the degree of desirability of a given world $w \in W$. We assume that δ satisfies the following normality constraint: there exists $w \in W$ such that $\delta(w) = 0$ (i.e., at least one state of the world is not desired at all). This ensures that $\Delta(\top) = 0$ since $||\top|| = W$. Thus the normality constraint of δ ensuring that not everything is desired, entails that 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 she does not desire at all φ to be false. This is a form of consistency requirement. Clearly, the distribution δ is just the numerical, or more generally the graded counterpart, of the qualitative ordering $\geq_\Delta$.

A desire φ with strength α is expressed by a constraint of the form $\Delta(p) \geq \alpha$. It will be denoted $[p, \alpha]$. A set D of desires $[\varphi_i, \alpha_i]$ (for $i = 1, \ldots, m$) is semantically associated to a guaranteed possibility distribution

$$\delta_D(w) = \max_{i=1,\ldots,m} \min(||\varphi_i||(w), \alpha_i). \quad (3)$$

where $||\varphi_i||(w) = 1$ if w is a model of φ, and $||\varphi_i||(w) = 0$ otherwise. δ_D is the smallest possibility distribution (maximum specificity principle) such that $\Delta(\varphi_i) \geq \alpha_i$ for $i = 1, \ldots, m$. This maximum specificity principle may be understood here as a minimal desire principle: there is no more desire that those expressed in the desire set D. The distribution δ_D rank-orders the interpretations of the language induced by the φ_i's according to their satisfaction level on the basis of the strength of the desires in D. A hedonic state can then be viewed as a fuzzy (or graded) subset of worlds.

Because we should have $\Delta(\top) = 0$, $\min_w \delta_D(w) = 0$ should hold. More generally,

$$una(D) = \min_w \delta_D(w)$$

may be viewed as a level of *unacceptability* of D. The larger $una(D)$, the more unacceptable the set of desires D.

4.2 Desires vs. beliefs in possibility theory

Expression (2) can be contrasted with the expression of a necessity measure in terms of a possibility distribution π

$$N(\varphi) = 1 - \max_{w \in ||\neg\varphi||} \pi(w)$$

which estimates the extent to which the agent believes φ to be true, all the more as $\neg\varphi$ is found impossible in the sense of π. Indeed, the necessity measure of N is the dual of the possibility measure Π, namely $\Pi(\varphi) = 1 - N(\neg\varphi)$ (where $1 - (\cdot)$ denotes the order-reversing map of S).

Formula (3) can be contrasted with the possibilistic representation of a belief set B expressed by a set of possibilistic logic formulas (ψ_j, γ_j) (for $j = 1, \ldots, n$) encoding constraints of the form $N(\psi_j) \geq \gamma_j$. B is semantically associated with a possibility distribution [16]

$$\pi_B(w) = \min_{i=1,\ldots,n} \max(||\psi_j||(w), 1 - \gamma_j).$$

π_B is the largest possibility distribution (minimum specificity principle) such that $N(\psi_j) \geq \gamma_j$ for $j = 1, \ldots, n$. The distribution π_B rank-orders the interpretations of the language induced by the ψ_j's according to their plausibility on the basis of the strength of the beliefs in B. If the set of beliefs $B^* = \{\psi_j, j = 1, \ldots, n\}$ is consistent then the distribution π_B is normalized in the sense that $\exists w, \pi_B(w) = 1$. More generally the level of inconsistency of B is defined by $inc(B) = 1 - \max_w \pi_B(w)$. Thus $una(D)$ should play the same role in desire revision as $inc(B)$ in belief revision [3, 4]. As can be seen in the expression of π_B, a belief set is in underlain by a (weighted) conjunctive view of the pieces of beliefs, while (3) shows that a desire set should be understood through a (weighted) disjunctive view of the desires, in agreement with the intuition.

5 Desire revision without explicit desire strengths

There are two slightly different views of a belief set. In the dominant one initiated by Gärdenfors [18], the belief set is just a collection of propositional sentences (assumed to be closed by logical entailment), while the revision process is driven by an epistemic entrenchment relation. Then the agent is described from the outside. The observer only sees the agent belief set, not the entrenchment. He sees the agent beliefs evolve due to inputs. The belief revision axioms are a model of the principles guiding the observed changes of the belief sets. The observer concludes that there is an epistemic entrenchment driving the process.

A more practical approach [4] views epistemic states as a collection of prioritized pieces of belief, which are thus associated to priorities that enables us to compute their entrenchment level as the value of a necessity measure.

These two points of view similarly exist when revising a desire set. A desire set D is a collection of propositional sentences, closed under reversed entailment, i.e., $D = \{\psi \mid D \vdash_{des} \psi\} = \{\psi \mid \psi \vdash D\}$, or may be a set of propositions associated with desire strengths, similarly closed.

It is clear that the two views are of interest. They lead to state the axioms governing revision in two different ways. We start by the view without explicit desire strengths.

5.1 Axioms for desire revision

As already said, one cannot desire φ and desire $\neg\varphi$ at the same time, without being led to a meaningless plethora. This parallels the fact that one cannot believe ψ and believe $\neg\psi$ at the same time, without being led to inconsistency. Revising beliefs copes with this constraint. Similarly, revising desires should cope with the previous constraint.

Having in mind the reverse behavior of desires with respect to beliefs, one is naturally led to state axioms that parallel the AGM axioms [18] of belief revision, but cope with the reverse behavior. Here are these axioms:

- [(D*1)] for any sentence and any desire set D, D^*_φ is a desire set.
- [(D*2)] $\varphi \in D^*_\varphi$.
- [(D*3)] $D^+_\varphi \supseteq D^*_\varphi$
- [(D*4)] If $\neg\varphi \notin D$ then $D^*_\varphi \supseteq D^+_\varphi$
- [(D*5)] $D^*_\varphi = \top$ if and only if $\varphi \equiv \top$
- [(D*6)] If $\vdash \varphi \equiv \psi$ then $D^*_\varphi = D^*_\psi$
- [(D*7)] $D^*_{\varphi\vee\psi} \subseteq (D^*_\varphi)^+_\psi$
- [(D*8)] If $\neg\psi \notin D^*_\varphi$ then $D^*_{\varphi\vee\psi} \supseteq (D^*_\varphi)^+_\psi$

where the expansion D^+_φ is just defined by a "reverse logical closure" of D together with φ, in agreement with the intuition underlying the idea of desire:

$$D^+_\varphi = \{\psi \mid \psi \vdash D \cup \{\varphi\}\} \quad (4)$$

(D*1) is a closure property. (D*2) is a success postulate: the new desire should enter in the desire set. (D*3) and (D*4) guarantee that the revision is an expansion that amounts to add the new desire φ to the desire set when $\neg\varphi$ is not already in the closure of the desire set. (D*5) states that the revision cannot result into desiring everything except if the new desire would be to desire everything. (D*6) is the independence with respect to syntax. (D*7) and (D*8) clearly parallel (D*3) and (D*4) when revision is decomposed in two steps.

As for guaranteeing the existence of an epistemic entrenchment in belief revision where the last two AGM axioms are necessary, (D*7) and (D*8) are required for ensuring the existence of a hedonic entrenchment relation in the sense of the postulates of subsection 2.1. This can be established following a route very similar to the one of Grove [20] epistemic entrenchment, taking into account the reverse behavior of hedonic entrenchment, and remembering the very close relationship between sphere systems and possibility distributions.

5.2 Semantic view of desire revision

Since the approach is syntax-free, it is advantageous to write the above axioms on the possible worlds. Below D and the input A are sets of possible worlds. D^+_A is the expanded set, D^*_A the revised set:

- [(D*1)] Trivial: D^*_A is a set of desired possible worlds.
- [(D*2)] $A \subseteq D^*_A$.
- [(D*3)] $D^+_A \supseteq D^*_A$
- [(D*4)] If $\overline{A} \not\subset D$ then $D^*_A \supseteq D^+_A$
- [(D*5)] $D^*_A = W$ if and only if $A = W$
- [(D*6)] Trivial (syntax-free approach)
- [(D*7)] $D^*_{A\cup B} \subseteq (D^*_A)^+_B$
- [(D*8)] If $\overline{B} \notin D^*_A$ then $D^*_{A\cup B} \supseteq (D^*_A)^+_B$

In the set-version,

- one immediately sees that under the axioms but for the two last ones, the revision rule is of the form:

$$D^*_A = \begin{cases} D \cup A \text{ if it is not } W. \\ \text{some } C \neq W, C \supset A \text{ otherwise.} \end{cases}$$

  Besides, $D^+_A = D \cup A$.
- The two last axioms come from the choice function area, and specify that C is selected with respect to an ordering on W (the most desired states outside A) [5]. But it is not clear what it means in practice. Either we consider that this setting uses all-or-nothing desires and it is not clear what the ordering means, or we consider graded desires and it is not clear why the input should be supposed to be fully desired.

One could think of applying here the maximum specificity principle for desires, counterpart of the minimum specificity principle in belief representation. Namely, unless desire is explicit, one assume states are not desirable. Under this assumption, C should be A in the above set revision rule (since there is no desire strength for discriminating the states outside A). This is clearly too drastic, and in the next section we investigate desire revision with explicit desire strengths.

6 Desire revision with explicit desire strengths

We now turn towards the case where the hedonic entrenchment can be computed from the desires given with their explicit strength. We first briefly recall how belief revision works in the possibility theory setting. Indeed it has been recognized early that the epistemic entrenchment relations underlying any well-behaved belief revision process obeying AGM postulates [18] are qualitative necessity relations [13], thus establishing a link between belief revision and possibility theory [15]. In the possibility theory view of belief revision, the epistemic entrenchment is explicit and reflects a confidence-based priority ranking between pieces of information. This ranking is revised when a new piece of information is received.

We first need to recall the possibilistic expression of conditioning underlying belief revision and its counterpart for guaranteed possibility measure, before considering the revision of beliefs, and then the revision of desires.

6.1 Two conditionings in possibility theory

In qualitative possibility theory [15], conditioning is defined by means of equation

$$\Pi(\varphi \wedge \psi) = \min(\Pi(\psi|\varphi), \Pi(\varphi)).$$

The quantitative version would use the product instead of min, but here we prefer a qualitative setting which agrees with the nature of the hedonic entrenchment. Applying the minimum specificity principle which leaves the possibility degrees as high as possible given the constraints (for avoiding arbitrary restrictions of the possible states), we get the possibility distribution $\pi(\cdot|\varphi)$ associated with the possibility measure $\Pi(\cdot|\varphi)$:

$$\pi(w|\varphi) = \left\{ \begin{array}{cc} 1 & \text{if } \pi(w) = \Pi(\varphi) \text{ and } w \vDash \varphi \\ \pi(w) & \text{if } \pi(w) < \Pi(\varphi) \text{ and } w \vDash \varphi \\ 0 & \text{if } w \vDash \neg\varphi \end{array} \right\}.$$

The conditioning of a strong possibility measure Δ contrasts with the previous view, and obeys the equation [2]:

$$\Delta(\varphi \wedge \psi) = \max(\Delta(\psi|\varphi), \Delta(\varphi)). \quad (5)$$

Now applying the *maximum* specificity principle, we get the smallest (i.e., corresponding to the least committed conditional desires) possibility distribution $\delta(w|\varphi)$ obeying (5):

$$\delta(w|\varphi) = \left\{ \begin{array}{cc} 0 & \text{if } \delta(w) = \Delta(\varphi) \text{ and } w \vDash \varphi \\ \delta(w) & \text{if } \delta(w) > \Delta(\varphi) \text{ and } w \vDash \varphi \\ 1 & \text{if } w \vDash \neg\varphi \end{array} \right\}.$$

As can be seen, what is no longer reachable (conditioning by φ means that, for some reason, the possible states are restricted to be those where φ is true) is fully desirable by default ($\Delta(\neg\varphi|\varphi) = 1$), while what we have is no longer desired since $\Delta(\varphi|\varphi) = 0$, but still preserving what is strictly above $\Delta(\varphi)$.

6.2 Belief revision, expansion and contraction

In the possibilistic setting, the *revision* B^*_φ of the belief base B revised by input φ, is expressed at the semantic level as:

$$\pi_{B^*_\varphi}(w) = \pi_B(w|\varphi).$$

where π_B is the possibility distribution associated with the belief base B, as recalled in Section 3.2. Conditioning by φ, acknowledges the fact that according to the input of the new piece of belief φ, states where φ is false have become impossible.

This expression covers the *expansion* B^+_φ of B by φ as a particular case:

$$\pi_{B^+_\varphi}(w) = \min(\pi(w), ||\varphi||(w))$$

provided that the consistency condition $core(\pi) \cap ||\varphi|| \neq \emptyset$ holds, where $core(\pi) = \{w \mid \pi(w) = 1\}$.

Besides, the *contraction* B^{-}_{φ} of B by φ is semantically expressed by [14]:

$$\pi_{B^{-}_{\varphi}}(w) = \left\{ \begin{array}{cc} 1 & \text{if } \pi(w) = \Pi(\neg\varphi) \text{ and } w \vDash \neg\varphi \\ \pi(w) & \text{otherwise} \end{array} \right\}.$$

which ensures $\neg\varphi$ becomes fully possible. Note that in particular, if $\Pi(\varphi) = \Pi(\neg\varphi) = 1$ (which means that we fully ignore if φ is true or false), we have $\pi_{B^{-}_{\varphi}}(w) = \pi(w)$. This is the case as soon as $\Pi(\neg\varphi) = 1$.

6.3 Contraction, expansion and revision of desires

Let D be a set of prioritized desires. Let δ_D be the associated hedonic distribution (as defined by (3) in Section 3.1).

The *contraction* of D by φ amounts to no longer desire φ at all after contraction. It is semantically expressed by:

$$\delta_{D^{-}_{\varphi}}(w) = \left\{ \begin{array}{cc} 0 & \text{if } \delta_D(w) = \Delta(\varphi) \text{ and } w \vDash \varphi \\ \delta(w) & \text{otherwise} \end{array} \right\}.$$

In particular, we have $\delta_{D^{-}_{\varphi}}(w) = \delta_D(w), \forall w$ as soon as $\Delta(\varphi) = 0$.

The *expansion* of a set of desires D by φ amounts to cumulating desire φ with the desires in D, providing that the result is not the desire of everything to some extent (due to the postulate $\Delta(\top) = 0$). Thus, we have

$$\delta_{D^{+}_{\varphi}}(w) = \max(\delta_D(w), ||\varphi||(w))$$

provided that $support(\delta_D) \cup ||\varphi|| \neq W$, where $support(\delta) = \{w | \delta_D(w) > 0\}$.

While the *revision* of a set of beliefs B by φ exactly corresponds to the conditioning of π_B by φ, this is no longer the case with respect to δ_D for the revision of a set of desires D by φ. Indeed, while a belief input $(\varphi, 1)$, i.e., $N(\varphi) = 1$, really means that all the models of $\neg\varphi$ should be impossible, i.e., $\Pi(\neg\varphi) = \max_{w \vDash \neg\varphi} \pi_B(w) = 0$, a desire input $[\varphi, 1]$ means $\Delta(\varphi) = \min_{w \vDash \varphi} \delta_D(w) = 1$, which says that all the models of φ are satisfactory after revision.

Moreover, we have observed in Section 5.1 that $\Delta(\varphi|\varphi) = 0$ and $\Delta(\neg\varphi|\varphi) = 1$. But $\Delta(\varphi|\varphi) = 0$ does not fit with the idea that φ is a new desire, nor $\Delta(\neg\varphi|\varphi) = 1$. Indeed, conditioning by φ does not mean to get a new desire. It means that for some reason, the possible states are restricted to be those where φ is true (which indeed confirms φ is not a new desire). So the agent can only desire such states, which would favor $\Delta_{D^{*}_{\varphi}}(\neg\varphi) = 0$.

Due to this change of focus from $\neg\varphi$ to φ, when moving from beliefs to desires, desire revision is expressed by:

$$\delta_{D^{*}_{\varphi}}(w) = \delta_D(w|\neg\varphi)$$

This leads to

$$\delta_{D^{*}_{\varphi}}(w) = \left\{ \begin{array}{cc} 0 & \text{if } \delta_D(w) = \Delta(\neg\varphi) \text{ and } w \vDash \neg\varphi \\ \delta_D(w) & \text{if } \delta_D(w) > \Delta(\neg\varphi) \text{ and } w \vDash \neg\varphi \\ 1 & \text{if } w \vDash \varphi \end{array} \right\}.$$

As can be seen we have $\Delta_{D^{*}_{\varphi}}(\neg\varphi) = 0$ and $\Delta_{D^{*}_{\varphi}}(\varphi) = 1$.

Having $\Delta_{D^{*}_{\varphi}}(\varphi) = 1$ may be considered as too a strong expression of the *success postulate* when revising the desire set D by the new desire φ. We may think that this interpretation of an input is too strong for revising a gradual desire profile. Introducing a new desire does not necessarily mean that the new desire should be desired with the highest strength. As revision is a merging of two entities of the same nature, we may prefer considering revision by $\Delta(\varphi) \geq \alpha$ (rather than $\Delta(\varphi) = 1$). This leads to

$$\delta_{D^{*}_{(\varphi,\alpha)}}(w) = \left\{ \begin{array}{cc} 0 & \text{if } \delta_D(w) = \Delta(\neg\varphi) \text{ and } w \vDash \neg\varphi \\ \delta_D(w) & \text{if } \delta_D(w) > \Delta(\neg\varphi) \text{ and } w \vDash \neg\varphi \\ \alpha & \text{if } w \vDash \varphi \text{ and } \delta_D(w) < \alpha \\ \delta_D(w) & \text{if } w \vDash \varphi \text{ and } \delta_D(w) \geq \alpha \end{array} \right\}.$$

It can be checked that we now have $\Delta_{D^{*}_{\varphi}}(\varphi) = \alpha$. We may also think of weakening the success postulate into $\Delta_{D^{*}_{\varphi}}(\varphi) > 0$. It can be defined by taking lesson of what is done in belief revision, where this corresponds to the idea of *natural* revision in the sense of Boutilier [6]; see [3]. When using a finite scale, we have just to take α as the smallest non-zero value in the scale.

Let us illustrate the approach by an example.

Example 2
Let $D = \{[\varphi \wedge \psi, \alpha], [\nu, \beta]\}$, be a desire base where $\alpha > \beta$, where φ, ψ, ν are literals. Applying (3), we get its semantical counterpart under the form of the distribution δ_D. Namely we have
$\delta_D(\varphi\psi\nu) = \delta_D(\varphi\psi\neg\nu) = \alpha$;
$\delta_D(\varphi\neg\psi\nu) = \delta_D(\neg\varphi\psi\nu) = \delta_D(\neg\varphi\neg\psi\nu) = \beta$;
$\delta_D(\varphi\neg\psi\neg\nu) = \delta_D(\neg\varphi\psi\neg\nu) = \delta_D(\neg\varphi\neg\psi\neg\nu) = 0$.
Clearly, $una(D) = 0$.
Now, assume we want to add desire $[\neg\varphi, 1]$. Let us compute $\delta_{D^{*}_{\neg\varphi}}$. We get:
$\delta_{D^{*}_{\neg\varphi}}(\varphi\psi\nu) = \delta_{D^{*}_{\neg\varphi}}(\varphi\psi\neg\nu) = \alpha$;
$\delta_{D^{*}_{\neg\varphi}}(\varphi\neg\psi\nu) = \beta$;
$\delta_{D^{*}_{\neg\varphi}}(\varphi\neg\psi\neg\nu) = 0$, which remain unchanged,
while it gives
$\delta_{D^{*}_{\neg\varphi}}(\neg\varphi\psi\nu) = \delta_{D^{*}_{\neg\varphi}}(\neg\varphi\neg\psi\nu) = \delta_{D^{*}_{\neg\varphi}}(\neg\varphi\psi\neg\nu) = \delta_{D^{*}_{\neg\varphi}}(\neg\varphi\neg\psi\neg\nu) = 1$.
Observe that $una(D \cup \{[\neg\varphi, 1]\}) = 0$,
which means that after addition of the new desire, the set of desires remains acceptable. In fact, we have just performed an expansion here.
Now suppose we only add the desire $[\neg\varphi, \gamma]$. Then the modified part of δ_D would be now
$\delta_{D^{*}_{\neg\varphi}}(\neg\varphi\psi\nu) = \delta_{D^{*}_{\neg\varphi}}(\neg\varphi\neg\psi\nu) = \max(\beta, \gamma)$,
$\delta_{D^{*}_{\neg\varphi}}(\neg\varphi\psi\neg\nu) = \delta_{D^{*}_{\neg\varphi}}(\neg\varphi\neg\psi\neg\nu) = \gamma$.
Suppose now D has to be modified by input $[\neg\nu, \epsilon]$. Then we have
$\delta_{D\cup\{[\neg\nu,\epsilon]\}}(\varphi\psi\nu) = \alpha$;
$\delta_{D\cup\{[\neg\nu,\epsilon]\}}(\varphi\neg\psi\nu) = \delta_{D\cup\{[\neg\nu,\epsilon]\}}(\neg\varphi\psi\nu) = \delta_{D\cup\{[\neg\nu,\epsilon]\}}(\neg\varphi\neg\psi\nu) = \beta$;
$\delta_{D\cup\{[\neg\nu,\epsilon]\}}(\varphi\neg\psi\neg\nu) = \delta_{D\cup\{[\neg\nu,\epsilon]\}}(\neg\varphi\psi\neg\nu) = \delta_{D\cup\{[\neg\nu,\epsilon]\}}(\neg\varphi\neg\psi\neg\nu) = \epsilon$ and $\delta_{D\cup\{[\neg\nu,\epsilon]\}}(\varphi\psi\neg\nu) = \max(\alpha, \epsilon)$.
Thus $una(D \cup \{[\neg\nu, \epsilon]\}) = \min(\alpha, \beta, \epsilon) = min(\beta, \epsilon) = \beta$ assuming $\epsilon > \beta$ (the new desire is not less strong than the desires in D). The result of the revision is $\delta_{D^{*}_{[\neg\nu,\epsilon]}}(\varphi\neg\psi\nu) = \delta_{D^{*}_{[\neg\nu,\epsilon]}}(\neg\varphi\psi\nu) = \delta_{D^{*}_{[\neg\nu,\epsilon]}}(\neg\varphi\neg\psi\nu) = 0$, while for the other interpretations, we keep $\delta_{D^{*}_{[\neg\nu,\epsilon]}}(w) = \delta_{D\cup\{[\neg\nu,\epsilon]\}}(w)$. This preserves $una(D^{*}_{[\neg\nu,\epsilon]}) = 0$.

6.4 Axioms for gradual desire revision

It is easy to write the possibilistic counterpart of the axioms for desire revision presented in subsection 5.1. Namely

- $[(\Delta * 1)]$ For any sentence $\varphi \in \mathcal{B}$, $\delta_{D^{*}_{\varphi}}$ represents a hedonic state.

- $[(\Delta{*}2)]$ $\Delta_{D^*_\varphi}(\varphi) = 1$. This a (strong) priority to the new desire.
- $[(\Delta * 3)]$ $\delta_{D^+_\varphi}$ is not more specific than $\delta_{D^*_\varphi}$
- $[(\Delta * 4)]$ If $\Delta_D(\neg\varphi) = 0$ then $\delta_{D^*_\varphi} \geq \delta_{D^+_\varphi}$
- $[(\Delta * 5)]$ $\delta_{D^*_\varphi} = \delta_\top$ if and only if $\varphi \equiv \top$
- $[(\Delta * 6)]$ Equivalent pieces of desires lead to equivalent revisions. We have it for free in the semantic view.
- $[(\Delta * 7)]$ $\delta_{D^*_{\varphi\vee\psi}} \leq \delta_{(D^*_\varphi)^+_\psi}$
- $[(\Delta * 8)]$ If $\Delta_{D^*_\varphi}(\neg\psi) = 0$ then $\delta_{D^*_{\varphi\vee\psi}} \geq \delta_{(D^*_\varphi)^+_\psi}$

$(\Delta * 2)$ may be weakened into $\Delta_{D^*_\varphi}(\varphi) > 0$. It can be easily checked from the definition of $\delta_{D^*_\varphi}$ that they all hold in the possibility theory setting. These axioms are the exact counterpart of the axioms for gradual belief revision [14]. It could be checked that they are exchanged under a transformation corresponding to the formal identity $\Delta_\delta(\phi) = N_{1-\delta}(\neg\phi)$, where Δ_δ (resp. $N_{1-\delta}$) are the guaranteed possibility (resp. necessity) measure defined from the distribution δ (resp. $1-\delta$).

6.5 Syntactic view in possibilistic logic

One interest of the possibility theory setting for belief revision is that possibilistic logic provides a tool for syntactic computation. Indeed the possibilistic base B^*_φ, corresponding to the revision of a belief base B by input φ, can be obtained syntactically as $\{(\varphi_i, \alpha_i) \in B \text{ s.t. } \alpha_i > \lambda\} \cup \{(\varphi, 1)\}$, where $\lambda = inc(B \cup \{(\varphi, 1)\})$ (where inc returns the inconsistency level).

The possibilistic logic of desires does not obey the same rules as the possibilistic logic of beliefs. Indeed now the resolution rule writes $[\varphi \wedge \psi, \alpha]$ and $[\neg\varphi \wedge \nu, \beta]$ entails $[\psi \wedge \nu, \min(\alpha, \beta)]$, which echoes the reverse rule (4) for defining expansion, and contrasts with the more classical resolution rule for prioritized beliefs: $(\varphi \vee \psi, \alpha)$ and $(\neg\varphi \vee \nu, \beta)$ entails $(\psi \vee \nu, \min(\alpha, \beta))$ [16].

Analogously to the belief revision case, it can be checked that only the desires strictly above the level of unacceptability are saved:

$$D^*_\varphi = \{[\varphi_i, \alpha_i] \in D \text{ s.t. } \alpha_i > una(D \cup \{[\varphi, \alpha]\})\} \cup \{[\varphi, \alpha]\}.$$

the others being drown, as it is the case for $[\nu, \beta] \in D'$ in the following example

Example 3
Let $D' = \{[\varphi, \alpha], [\nu, \beta]\}$ with $\alpha > \beta$.
Then $una(D') = 0$ since D' does not entail $\top$ at any non zero degree. Now, let us add desire $[\neg\varphi, 1]$.
We have $una(D' \cup \{[\neg\varphi, 1]\}) = \alpha$ and then $D'^*_{\neg\varphi} = \{[\neg\varphi, 1]\}$. If we rather consider $D'' = \{[\varphi, \beta], [\nu, \alpha]\}$ (always with $\alpha > \beta$),
then we have $una(D'' \cup \{[\neg\varphi, 1]\}) = \beta$, and $D''^*_{\neg\varphi} = \{[\nu, \alpha], [\neg\varphi, 1]\}$.

Similarly in Example 2, it can be checked, we have $D^*_{\neg\varphi} = D^+_{\neg\varphi}$, and the syntactic counterpart is $D^*_{\neg\varphi} = \{[\varphi \wedge \psi, \alpha], [\nu, \beta], [\neg\varphi, 1]\}$. Moreover $D^*_{[\neg\nu, \epsilon]} = \{[\varphi, \alpha], [\neg\nu, \epsilon]\}$ assuming $\epsilon > \beta$.

Note that in all the above examples, we have assumed that none of the interpretations induced by the language used for specifying the desire set is impossible in the real world. In any case, if such an impossibility exists for some of them, this has to be taken into account when adopting goals, but not in the revision of desires.

7 Conclusion

The paper has presented a formal approach to the revision of desires. By desire, we mean potential desires and distinguishing them from goals. Goals are desires that have been actualized by the agent and to which she is committed. Their revision is not the same problem as the one of desire revision, and is in fact quite similar to belief revision (since having goal φ and having goal ψ should be the same as having goal $\varphi \wedge \psi$). The goal revision of a set of prioritized goals would be based on a volitive entrenchment, formally similar to an espistemic entrenchment. The particular nature of desires with respect to beliefs has been advocated and emphasized. Roughly speaking, desires behave in a reverse way. This is reflected in the different series of axioms characterizing the hedonic entrenchment and then desire revision that have been proposed. Several directions remain to investigate, such as studying iterated desire revision.

Besides, it is known that belief revision and nonmonotonic reasoning are two sides of the same coin [19]. This remains to be checked for nonmonotonic desires [11] and desires revision. Finally, we plan to extend the static modal logic of belief and desire we proposed in [10] by dynamic operators of belief revision and desire revision. This will provide a unified modal logic framework based on possibility theory dealing with both the static and the dynamic aspects of beliefs and desires, to be compared with the proposal made in [24].

REFERENCES

[1] G. E. M. Anscombe. Intention. Basil Blackwell, 1957.

[2] S. Benferhat, D. Dubois, S. Kaci, H. Prade. Bipolar possibilistic representations. Proc. 18th Conf. in Uncertainty in Artificial Intelligence (UAI '02), (A. Darwiche and N. Friedman, ed.), Edmonton, Alberta, Aug. 1-4, Morgan Kaufmann, 45-52, 2002.

[3] S. Benferhat, D. Dubois, H. Prade. A computational model for belief change and fusing ordered belief bases. Frontiers in Belief Revision (Williams, M.-A. and Rott, H., Eds.), Kluwer Acad. Publ., 109-134, 2001.

[4] S. Benferhat, D. Dubois, H. Prade, M.-A. Williams. A practical approach to revising prioritized knowledge bases. Studia Logica, 70, 105-130, 2002.

[5] G. Bonanno. Rational choice and AGM belief revision. Artificial Intelligence, 173(12-13), 1194-1203, 2009.

[6] C. Boutilier. Revision sequences and nested conditionals. Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93), Chambéry, Aug. 28 - Sept. 3, Morgan Kaufmann, 519-525, 1993.

[7] C. Castelfranchi, F. Paglieri. The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions. Synthese, 155 (2): 237–263, 2007.

[8] D. Dubois. Belief structures, possibility theory and decomposable confidence measures on finite sets. Computers and Artificial Intelligence (Bratislava), 5(5), 403-416, 1986.

[9] A. Casali, L. Godo, C. Sierra. A graded BDI agent model to represent and reason about preferences. Artificial Intelligence,175, 1468–1478, 2011.

[10] D. Dubois, E. Lorini, H. Prade. Bipolar possibility theory as a basis for a logic of desires and beliefs. Proc. 7th Int. Conf. Scalable Uncert. Mgmt. (SUM'13), (W. Liu and V. S. Subrahmanian and J. Wijsen, eds.), Washington, DC, Sept. 16-18, Springer, LNCS 8078, 2013.

[11] D. Dubois, E. Lorini, H. Prade. Nonmonotonic desires - A possibility theory viewpoint. Proc. Int. Workshop on Defeasible and Ampliative Reasoning (DARe@ECAI 2014), (R. Booth, Casini, G., Klarman, S., Richard, G., Varzinczak, I. J., eds.), Prague, Aug. 19, CEUR Workshop Proc., vol. 1212, 2014.

[12] D. Dubois, E. Lorini, H. Prade. Revising desires - A possibility theory viewpoint. Proc. 11th Int. Conf. on Flexible Query Answering Systems(FQAS'15), (T. Andreasen, H. Christiansen, J. Kacprzyk, H. Larsen, G. Pasi, O. Pivert, G. De Tré, M. A. Vila, A. Yazici, S. Zadrożny, eds.) Vol. 400, Advances in Intelligent Systems and Computing series, pp. 3-13, 2015.

[13] D. Dubois, H. Prade. Epistemic entrenchment and possibilistic logic. Artificial Intellig., 50, 223-239,1991.

[14] D. Dubois, H. Prade. Belief change and possibility theory. In: Belief Revision (P. Gärdenfors, ed.), Cambridge University Press, 142-182, 1992.

[15] D. Dubois, H. Prade. Possibility theory: qualitative and quantitative aspects. In: Quantified Representation of Uncertainty and Imprecision, (D. Gabbay, P. Smets, eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, Kluwer, v.1, 169–226, 1998.

[16] D. Dubois, H. Prade. Possibilistic logic: a retrospective and prospective view. Fuzzy Sets and Systems, 144, 3-23, 2004.

[17] Dubois, H. Prade. Accepted beliefs, revision and bipolarity in the possibilistic framework. In : Degrees of Belief, (F. Huber, C. Schmidt-Petri, eds.), Springer, 161-184, 2009.

[18] P. Gärdenfors. Knowledge in Flux. The MIT Press, 1988.

[19] P. Gärdenfors. Belief revision and nonmonotonic logic: Two sides of the same coin? Proc. 9th Europ. Conf. on Artificial Intelligence (ECAI'90), Stockholm, 768-773, 1990.

[20] A. Grove. Two modellings for theory change. J. Philos. Logic, 17, 157-170, 1988.

[21] I. L. Humberstone. Direction of fit. Mind, 101(401):59–83, 1992.

[22] J. Lang. Conditional desires and utilities: An alternative logical approach to qualitative decision theory. Proc. 12th Eur. Conf. Artif. Intellig. (ECAI'96), (W. Wahlster, ed.), Budapest, August 11-16, J. Wiley, 318-322,1996.

[23] J. Lang, L. W. N. van der Torre. From belief change to preference change. Proc. 18th Europ. Conf. on Artificial Intelligence (ECAI'08), (M. Ghallab, C. D. Spyropoulos, N. Fakotakis, N. M. Avouris, eds.), Patras, July 21-25, IOS Press, 351–355, 2008.

[24] J. Lang, L. W. N. van der Torre, E. Weydert. Hidden uncertainty in the logical representation of desires. Proc. 18th Int. Joint Conf. on Artificial Intelligence (IJCAI'03), (G. Gottlob, T. Walsh, eds.), Acapulco, August 9-15, Morgan Kaufmann, 685-690, 2003.

[25] D. Lewis. Counterfactuals and comparative possibility. Journal of Philosophical Logic, 2 (4) (1973) 418-446.

[26] E. Lorini. A logic for reasoning about moral agents. Logique et Analyse, Centre National de Recherches en Logique (Belgium), 58(230), 2014.

[27] M. Platts. Ways of meaning. Routledge and Kegan Paul, 1979.

[28] A. S. Rao, M. P. Georgeff. Modeling rational agents within a BDI-Architecture. Proc. 2nd Int. Conf. on Principles of Knowledge Representation and Reasoning, 473–484, 1991.

[29] J. Searle. Expression and meaning. Cambridge University Press, 1979.