The Strength of Desires: a Logical Approach

Didier Dubois, Emiliano Lorini, Henri Prade

The Strength of Desires: A Logical Approach

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

Abstract The aim of this paper is to propose a formal approach to reasoning about desires, understood as logical propositions which we would be pleased to make true, also acknowledging the fact that desire is a matter of degree. It is first shown that, at the static level, desires should satisfy certain principles that differ from those to which beliefs obey. In this sense, from a static perspective, the logic of desires is different from the logic of beliefs. While the accumulation of beliefs tend to reduce the remaining possible worlds they point at, the accumulation of desires tends to increase the set of states of affairs tentatively considered as satisfactory. Indeed beliefs are expected to be closed under conjunctions, while, in the positive view of desires developed here, one can argue that endorsing $\varphi \vee \psi$ as a desire means to desire φ and to desire ψ. However, 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 revision is based on a hedonic entrenchment relation satisfying other properties, due to the different natures of belief and desire. While epistemic entrenchment relations are known to be quali-tative necessity relations (in the sense of possibility theory), hedonic relations obeying a set of reasonable postulates correspond to another set-function in possi-bility theory, called guaranteed possibility, that drive well-behaved desire revision operations. Then the general framework of possibilistic logic provides a syntactic

✉ Henri Prade
prade@irit.fr

Didier Dubois
dubois@irit.fr

Emiliano Lorini
lorini@irit.fr

[1] IRIT-CNRS, Université Paul Sabatier, 31062 Toulouse Cedex 09, France

setting for encoding desire change. The paper also insists that desires should be carefully distinguished from goals.

Keywords Desire · Revision · Possibility theory

1 Introduction

Desires constitute the primitive form of a motivational attitude that drives an agent to plan her action in order to satisfying them. Specifically, taking into account her beliefs about the world, the agent is able to choose 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 by Lorini (2014) taking inspiration from the philosophical and psychological literature (Castelfranchi and Paglieri 2007). These concepts are also the building blocks of so-called BDI agents, where B, D, I, respectively stand for Beliefs, Desires, and Intentions (Rao and Georgeff 1991).

Desires and intentions are sometimes used more or less interchangeably in the literature. However, they should be carefully distinguished. For instance, let us reconsider an example adapted from Lang and van der Torre (2008): 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 also revise her intention, but not her desire. In this paper, we do not consider intentions, but only desires.

Besides, when modeling desires the paper does not consider the case of unbearable situations for the agent, namely those she primarily wants to avoid or is afraid of. More precisely, we consider positive desires only, namely those that it would be really satisfactory to concretize, as opposed to negative desires (fears) corresponding to situations to be avoided because they are very unsatisfactory, or even unbearable for the agent. In fact, modelling both desires and the fear of unbearable situations would require a bipolar setting already discussed by Benferhat et al. (2006) and Dubois and Prade (2009b) leaving room for both positive and negative desires. Namely, the agent first tries to avoid being in a world that is insufferable for her, then she expresses desires among the remaining non-rejected worlds. Note that if a world is not rejected as unbearable or fearful, it does not mean that it is desired. The agent may just be indifferent about it. Cast in the bipolar setting, desires pertain to worlds the agent is at least indifferent to, and are not concerned with fear attitudes as caused by unbearable situations. This view is akin to prospect theory of Tversky and Kahneman (1992) where gains and losses are separately handled by different utility functions. Here, we only focus on positive desires.

Moreover, desires and beliefs also behave differently. Indeed, as we shall claim in this paper, 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. This is because when an agent discovers new desires, it enlarges the number of desirable situations, while accumulating beliefs reduces the number of possible worlds. This difference of behavior between desire and belief has been pointed out by Casali et al. (2011), 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 (Dubois and Prade 2009a). This point of view was then investigated by Dubois et al. (2013), revisiting the representation of positive preferences proposed earlier on by Benferhat et al. (2006).

We also claim 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 (including the possibility of being in an unbearable world), which would trigger a change of intention, but just because of the activation of a new desire, which, together with her previous desires, would lead her to desire anything and its contrary. New desires may make previously desirable situations less attractive.

Such a situation is apparently 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. In (Dubois et al. 2015), a modeling of desire change has been briefly outlined without proposing any axiomatic foundation nor representation results. In this paper, we provide postulates for desire revision, contrast them with belief revision postulates (Gärdenfors 1988), 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. This proposal mirrors to a large extent the way belief change can be represented in the framework of possibility theory (Dubois and Prade 1991, 1992; Benferhat et al. 2002b). Moreover, note that belief revision in its original formulation does not consider impossible worlds (e.g., violating an integrity constraint). This is similar to the assumption of ruling out unbearable situations in our desire modeling framework.

The paper is organized as follows. In Sect. 2, we highlight the main intuitions behind the concept of desire in contrast with the concept of belief. Section 3 presents an elementary approach for representing desires in a logical setting, from a syntactic and semantical point of view, including a minimal modal logic representation. Section 4 refines this setting by means of a desirability relation accounting for the strength of desires, and provide axioms for such a relation. It is shown that the unique numerical counterpart is a guaranteed possibility measure in the sense of possibility theory. In Sect. 4.4, the representation of graded desires in this setting is presented in detail. 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 possibilistic logic. Section 5 provides axioms for desire revision and then presents the revision of sets of prioritized desires axiomatically, semantically, and syntactically using a special form of possibilistic logic. Expansion and contraction of desires are also discussed.

This paper builds on several preliminary works. Dubois et al. (2013), following a suggestion by Casali et al. (2011), investigated the use of a specific set function in

possibility theory (guaranteed possibility) for modeling the idea of desire, and outline a modal logic for desire and beliefs. A simple version of this modal logic, restricted to desires, is detailed in the present paper. In (Dubois et al. 2015), a revision rule for desires was proposed, based on conditional guaranteed possibility measures. More recently, axioms for comparative desire and desire revision were proposed, with a preliminary discussion, in the workshop paper (Dubois et al. 2016). The present paper is a refined, updated, and completed synthesis, rewritten to a large extent, of these preliminary contributions, including a discussion of the relevant philosophical literature, and logical foundations.

2 Conceptual Foundation

An important and general distinction in the philosophy of mind is between epistemic attitudes and motivational attitudes. This distinction is in terms of the *direction of fit* of mental attitudes to the world. While the aim of epistemic attitudes is truth and their being true refers to their conformity to the world as it stands, motivational attitudes aim at realization and their realization consists in making the world fit them (Platts 1979; Anscombe 1957; Humberstone 1992). Searle (1979) 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. Lorini 2014 for a logical theory of the relationship between desires, moral values and preferences).

Beliefs are mental representations whose aim is to represent the physical, mental and social worlds as they stand. In contrast, following the Humean conception (Hume 1978), a desire can be viewed as an agent's attitude consisting in an anticipatory mental representation of a pleasant state of affairs (hedonic 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 pleasure by anticipation. For example when an agent desires to eat sushi, she is pleased to imagine herself eating sushi. This pleasant representation motivates her to go to the "The Japoyaki" restaurant in order to eat sushi. This view of desires unifies the standard theory of desire (STD)—focused on the motivational dimension—and the hedonic theory of desire (HTD)—focused on the hedonic dimension. A third theory of desire has been advanced in the philosophical literature (see Schroeder 2004), the so-called reward theory of desire (RTD). According to RTD what qualifies a mental attitude as a desire is the exercise of a capacity to represent a certain fact as a reward.[1]

[1] According to Dretske (1988), desire is also a necessary condition for reward. In particular, desire determines what counts as a reward for an agent. For example, a person can be rewarded with water only if she is thirsty and she desires to drink.

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 desires and desire change—and desire revision, as a kind of desire change operation—is important for several reasons: (i) desire is at the heart of the concept of autonomous agent, (ii) a theory of desires and 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.

Before going into the logical representation of desires and desire change, we want to discuss two crucial properties of desires that drive our present analysis. The first property is what we call the *longing aspect of desires*. The idea is that for an agent to desire something, the agent should be in a situation in which she does not have what she desires and she yearns for it. In other words, a state of affairs is desired by an agent only if the agent conceives it as *absent*. The following quotation from Locke (1975, Book II, Chap. XXI) makes this point clear:

> To return then to the inquiry, what is it that determines the will in regard to our actions? And that...is not, as is generally supposed, the greater good in view: but some (and for the most part the most pressing) uneasiness a man is at present under. This that which successively determines the will, and sets us upon those actions, we perform. This uneasiness we may call, as it is, desire; which is uneasiness of the mind for want of some *absent good*...

This quotation seems to be at odds with what we claimed above, namely, that desire is based on anticipatory pleasure. However, the anticipated pleasure associated with a desire is all the stronger as its current lack of fulfillment—the term "uneasiness" in the previous quotation—is felt as more painful, as in the case of longing for a drink when thirsty, for instance. So the contradiction is only apparent. This aspect of uneasiness described by Locke should not be confused with the the concept of aversion which is traditionally opposed to the concept of desire [see (Schroeder

2004, Chap. 5)]. As emphasized above, if an agent desires a certain fact to be true, then she has a mental representation of this fact motivating her to make the fact true and associated with a positive *pleasant* feeling. On the contrary, if an agent is averse to something, then she has a mental representation of certain fact motivating her to prevent the fact from being true and associated with a negative *unpleasant* feeling. The distinction between desire and aversion is not studied in the present paper. Indeed, the joint handling of both aspects requires a bipolar setting that, as emphasized in the introduction, goes beyond the scope of the present work.

The second property, called *unconditional aspect of desires*, reflects the idea that "an agent desires something to be true, no matter the circumstances (assuming they are acceptable)", while other desires have a relative value in the sense that "an agent desires something to be true, conditional to some specific circumstances". For example, an agent may desire to drink red wine only in the circumstance in which she eats meat and, eventually, desire to drink white wine in the circumstances in which she eats fish or a dessert. In this sense, the desire to drink red wine is *relative*. On the contrary, another agent may enjoy drinking red wine, regardless of the proposed dishes. That is, she may desire to drink red wine while eating meat, fish or anything else. This kind of desire to drink red wine is *unconditional*.

The focus of the present paper concerns unconditional desires. It presupposes that the agent expresses desires in a sufficiently specific manner, e.g., expresses a joint desire for red wine and meat rather than for red wine at large, or rules out unpleasant situations where red wine is inappropriate, prior to expressing desires. Clearly, this model of desire has a limited scope. Going beyond these assumptions would require the explicit handling of relative and/or negative desires that is beyond the ambition of this paper. The study of relative desires calls for a *nonmonotonic* version of desires that we have started to investigate in Dubois et al. (2014).

Lastly, desires only express a general propensity, while goals are elements of a planning process aiming at making desired things true. For instance, the agent may have the desires to visit a Picasso exhibition and a Klee exhibition, which means that attending any of them will please her. When she organizes herself, she may adapt only the goal of visiting the Klee exhibition, which is the closest for her, due to a lack of time. In case she would have had more time, she might have taken the goal of visiting both exhibitions, which is clearly equivalent to the conjunction of the goal of visiting the Picasso exhibition and of the goal of visiting the Klee exhibition. Note that the *desire* of visiting both exhibitions has a different meaning: what is satisfactory in this latter case is to see both exhibitions, while it is not said if seeing only one of them would be satisfactory. Such a distinction between desires and goals has been already advocated in Castelfranchi and Paglieri (2007) and Lorini (2014).

Like for beliefs, any theory of desires should have a static aspect and a dynamic aspect. While the idea of coherence is at the root of a theory of rational beliefs (Rott 2001) since a rational agent cannot endorse a contradictory set of beliefs, an agent cannot find desirable all the possible states of the worlds, since the agent should be in one of these states, and desires are only about what we do not have. This is one more clue in favor of the fact that beliefs and desires behave in opposite ways, as we

shall see. We first start with the static aspect of our steps to a logical theory of desires.

3 A Logical Approach to Static Desires

In this section we move from an informal analysis to a formal analysis of all-or-nothing desires, by showing how the static aspect of desires can be represented either syntactically or semantically and how the two kinds of representations are related.

3.1 A Syntactic Propositional Representation

Let *Atm* be a finite set of atomic propositions. Elements of *Atm* are denoted by $p, q, \ldots$ We denote with $\mathcal{L}_{\mathsf{PL}}$ the propositional language built out of *Atm*. Propositional sentences are denoted in the following by φ, ψ, χ or ν. Hence $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$ belong to $\mathcal{L}_{\mathsf{PL}}$ as well. As usual $\top$ and $\bot$ 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 syntactic entailment defined as usual. Notions of validity and satisfiability are the standard ones from propositional logic.

In this paper, desires are represented by means of a finite set of propositional sentences.

Definition 3.1 *(Desire base)* A desire base $\mathcal{D}$ is a finite subset of the language $\mathcal{L}_{\mathsf{PL}}$.

It contains formulas called *desires* an agent would long for being true. Note that a desire in $\varphi \in \mathcal{D}$ only reflects the taste for φ being true, and does not mean to actually plan to make it true. Equivalently a set of desires can be specified by its characteristic function $f_{\mathcal{D}}$ called a desire function that is, $f_{\mathcal{D}}(\varphi) = 1$ if $\varphi \in \mathcal{D}$, and 0 otherwise. Note that in our framework, $\varphi \notin \mathcal{D}$ means indifference about φ being true or not. It has no negative flavor like disgust or fear.

Definition 3.2 *(Closed desire base)* The desire base $\mathcal{D}$ is said to be *closed* iff it satisfies the following properties:

(R1) if φ is valid then $\varphi \notin \mathcal{D}$;
(R2) if $\psi \in \mathcal{D}$ and $\varphi \rightarrow \psi$ is valid then $\varphi \in \mathcal{D}$;
(R3) if $\varphi \in \mathcal{D}$ and $\psi \in \mathcal{D}$ then $\varphi \vee \psi \in \mathcal{D}$.

The class of closed desire bases is denoted by $\mathbf{D}^{cl}$.

Let us justify these properties, since they generally differ from the usual notion of closed set of formulas.

Property **R1** is justified by the fact that agent cannot desire to achieve something that is necessarily true and cannot be absent. Indeed, this is in contradiction with the *longing aspect of desires* we have discussed in the previous section. There is no point desiring the truth of tautologies, because you can only desire to make true

propositions that you believe to be false in your state of affairs, and tautologies are never believed so by a rational agent.

Property **R2** is justified by the assumption of *unconditional desires* discussed in the previous section. Clearly, if an agent desires a certain proposition ψ to be unconditionally true, then *all* situations in which ψ is true should be desirable for the agent. Suppose $\varphi \rightarrow \psi$ is valid. Thus, all situations in which φ is true are situations in which ψ is true. Consequently, if all situations in which ψ is true are desirable for the agent, then all situations in which φ is true should be desirable for the agent. The latter means that the agent desires φ. For instance, 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. It justifies **R2**. Note that **R2** for desires also echoes the difference of direction of fit of motivational attitudes for desires versus for epistemic attitudes and beliefs.

We have to keep in mind that desires are considered as tentative in nature, and have not reached the step of being adopted as goals to pursue. Here 'desiring' just means 'finding pleasant', 'finding enjoyable', 'having a taste for' and so on. The reverse entailment for desire also relies on the assumption that agents are omniscient, in the sense that, they desire all formulas that are (reversely) entailed by what they desire. This is similar to the omniscience assumption for beliefs. A logical theory of desires for non-omniscient agents goes beyond the scope of the present work.

There is an obvious objection to the reverse entailment for desires. Namely, desiring coffee does not imply desiring coffee while being in a very unpleasant situation, like having a close friend killed. It is clear that you would find this situation unbearable, even with a coffee. However, we have to recall that we do exclude unbearable situations. For instance, if the desire logic does not use propositional variables that refer to unbearable situations, an interpretation in this logic gathers an equivalence class of actual worlds some of which being possibly unbearable. However, the latter worlds cannot be expressed in the language; then they are implicitly assumed to be ruled out. Should some variables describe potentially unbearable situations, then we should use integrity constraints and restrict the rule **R2** to potentially desirable interpretations.

R3 is an acceptable property of any *deductively closed* notion of desire. Informally it means that for example, longing for a cup of coffee (φ) and longing for a cup of tea (ψ) logically entails longing for any of them ($\varphi \vee \psi$). Of course, $\varphi \wedge \psi$ is also deduced, although trivially due to **R2**.

In contrast, for beliefs, deducing $\varphi \wedge \psi$ is informative, and deducing $\varphi \vee \psi$ is not. This has to do with the fact that models of desired formulas do not play the same role as models of believed formulas. Believing φ comes down to considering as impossible all interpretations that make φ false. All models of φ are considered possible by the agent, but only one model of φ is true. In contrast, desiring φ comes down to considering as desirable all interpretations that make φ true. So, believing φ and ψ comes down to believing $\varphi \wedge \psi$ (eliminating all interpretations where either of φ and ψ is false). In contrast, desiring φ and ψ

comes down to desiring $\varphi \vee \psi$ (accumulating all interpretations where either of φ and ψ is true).[2]

Note that the behavior of desires with respect to logical entailment, as expressed by axioms **R1-3** 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 reverse entailment, namely a desire for ψ entails a desire for φ, i.e., $\psi \vdash_{PLD} \varphi$ if and only if $\varphi \vdash \psi$. As a consequence, in the same way as $\varphi \vdash \top$ trivially holds for any belief φ, we have $\psi \vdash_{PLD} \bot$ for any desire ψ. In other words, longing for no state of affairs[3] means to be in a dull state of mind[4], just like believing tautologies is bringing no useful information.

The following proposition highlights that if the desire base is closed then an agent cannot desire φ and its opposite.

Proposition 1 *Let* $\mathcal{D} \in \mathbf{D}^{cl}$. *Then, if* $\varphi \in \mathcal{D}$ *then* $\neg\varphi \notin \mathcal{D}$.

Proof Suppose $\varphi \in \mathcal{D}$ and $\neg\varphi \in \mathcal{D}$. Hence, by Property **R3** in Definition 3.2, $(\varphi \vee \neg\varphi) \in \mathcal{D}$. But, by Property **R1** in Definition 3.2, $(\varphi \vee \neg\varphi) \notin \mathcal{D}$. This leads to a contradiction. □

Due to the blocking of direct entailment in the logic of desire, this property does not rule the case where $\varphi, \psi \in \mathcal{D}$ with $\psi \vdash \neg\varphi$. For instance, I may long for seeing a Picasso exhibition in Paris and a Klee exhibition in New-York, which does not mean that I desire to be in Paris and I desire not to be in Paris.[5]

The following proposition provides an equivalent formulation of the concept of closed desire base.

Proposition 2 *A desire base* $\mathcal{D}$ *is closed iff it satisfies Property* **R1** *in Definition 3.2 plus the following additional property:*

(R2*) *if* $\psi_1, \ldots, \psi_n \in \mathcal{D}$ *and* $\varphi \rightarrow (\psi_1 \vee \ldots \vee \psi_n)$ *is valid then* $\varphi \in \mathcal{D}$.

Proof The fact that **R2** and **R3** together imply **R2*** is clear. The fact that **R2*** implies **R2** is also clear. Let us prove that **R2*** implies **R3**. Suppose $\varphi, \psi \in \mathcal{D}$. Clearly, $(\varphi \vee \psi) \rightarrow (\varphi \vee \psi)$ is valid. Thus, by **R2***, $\varphi \vee \psi \in \mathcal{D}$. □

[2] This is an example where the logical disjunction symbol $\vee$ should be read "and", in natural language, as it translates into a set union, because the interpretations of φ understood as a desire are not mutually exclusive, contrary to when φ represents a belief. In the case where φ is a belief, the models of φ are mutually exclusive as candidates to be the real state of the world. So this set is a *disjunction* of situations, while if φ is a desire, any model of φ is a desirable situation, and the set of models of φ now collects all desirable situations, thus making the set of models a *conjunction* of situations. See Dubois and Prade (2012) and Zadeh (1978) for the importance of the distinction between the conjunctive and disjunctive interpretations of sets in another area.

[3] Remember that the set of models of $\bot$ is $\emptyset$.

[4] Called *anhedonia* as pointed out by a referee.

[5] As noticed by a referee.

3.2 A Semantic Representation

We first present the notion of hedonic state, as opposed to the notion of volitive 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. The semantic representation of desires is based on the following concept of hedonic model.

Definition 3.3 (*Hedonic model*) A hedonic model is a tuple $M = (W, \delta)$ where $W = 2^{Atm}$ is the set of all interpretations[6] and

$$\delta : W \longrightarrow \{0, 1\}$$

is a desirability function such that $\delta(w) = 0$ for some $w \in W$. The set of interpretations $D^{\delta} = \{w : \delta(w) = 1\}$ is called a *hedonic state*.

At the intuitive level, $\delta(w) = 1$ means that w is a desirable state, while $\delta(w') = 0$ indicates mere indifference towards w', not rejection, as we do not model negative desires nor the idea of loathing, fearing and the like.

Satisfaction of propositional desire formulas φ with respect to interpretations $w \in W$ is defined as usual, and denoted by $w \models \varphi$. Let $||\varphi|| \subseteq W$ denote the subset of models of the propositional formula φ. Based on the above discussion and the disjunctive reading of joint desires, the set of models of a desire base $\mathcal{D}$ is understood as the models of the disjunction of formulas in $\mathcal{D} = \{\psi_1, \ldots, \psi_n\}$, that is,

$$||\{\psi_1, \ldots, \psi_n\}|| = ||\psi_1|| \cup \ldots \cup ||\psi_n||$$

with the convention $||\emptyset|| = \emptyset$. The semantic inference for desires, denoted by $\models_D$ is then defined by following the principle according to which if an agent desires a set of possible worlds, she desires any subset of it, that is $\varphi \models_D \psi$ if and only if $||\psi|| \subseteq ||\varphi||$. The semantic entailment from a desire base $\mathcal{D}$ is expressed in a reverse way compared to the usual one (for beliefs), namely,

$$\mathcal{D} \models_D \psi \text{ if and only if } \forall w \in W, \text{ if } w \models \psi \text{ then } w \models \psi_i \text{ for some } \psi_i \in \mathcal{D},$$

that is, for any desired situation according to ψ, there is a desired proposition in the base which accounts for this desired situation. We do get that this is equivalent to $||\psi|| \subseteq ||\psi_1 \cup \ldots \cup \psi_n||$. Noticeably, $\varphi \models_D \psi$ if and only if $\psi \models \varphi$ in the usual sense. Validity is then defined as $\models_D \psi$ if and only if $||\psi|| = \emptyset$. In fact, this semantic inference preserves falsehood, since it can also be expressed as:

$$\mathcal{D} \models_D \psi \text{ if and only if } \forall w \in W, \text{ if } w \models \neg\psi_i \text{ for all } \psi_i \in \mathcal{D} \text{ then } w \models \neg\psi.$$

In other words, we have that

[6] For simplicity, we assume that W does not contain unbearable interpretations.

$$\mathcal{D} \models_D \psi \text{ if and only if } \psi \models \bigvee_{i=1}^{n} \psi_i \text{ if and only if } \mathcal{D}_\neg \models \neg\psi \tag{1}$$

where $\mathcal{D}_\neg = \{\neg\psi_i : \psi_i \in \mathcal{D}\}$. Note the close relationship between property **R2*** and the above equivalence.

3.3 Axiomatization

On this basis we are in a position to propose a counterpart to propositional logic for desires, via the following inference system:

Axioms:

($\neg$**PL**) :

1. $\neg(\varphi \rightarrow (\psi \rightarrow \varphi))$;
2. $\neg((\varphi \rightarrow (\psi \rightarrow \mu)) \rightarrow ((\varphi \rightarrow \psi) \rightarrow (\varphi \rightarrow \mu)))$;
3. $\neg((\neg\varphi \rightarrow \neg\psi) \rightarrow (\psi \rightarrow \varphi))$.

Inference rule

$$(\textbf{CMP}) : \quad \frac{\varphi, \neg\varphi \wedge \psi}{\psi}$$

The axioms of this propositional logic of desires (PLD) are those of propositional logic, negated to account for **R1**. The syntactic rule that preserves falsehood is precisely (**CMP**) (while the standard modus ponens preserves truth) since (**CMP**) can be read : if φ is false and $\neg\varphi \wedge \psi$ is false, then ψ is false.

The soundness and completeness of this propositional logic of desires can be obtained based on equivalences (1).

Theorem 1 *The propositional logic of desires PLD defined by axioms* $\neg$*PL and inference rule* **CMP** *is sound and complete with respect to the semantic entailment* $\models_D$.

Proof Using (1), $\mathcal{D} \models_D \psi$ if and only if $\mathcal{D}_\neg \models \neg\psi$, and due to completeness of the propositional calculus, this is equivalent to $\mathcal{D}_\neg \vdash \neg\psi$ in PL. Finally, one can define a one-to-one translation of every proof (in the usual sense) in classical logic of $\mathcal{D}_\neg \vdash \neg\varphi$ into a proof in PLD of $\mathcal{D} \vdash_{PLD} \varphi$. Indeed, if $\phi_1, \ldots, \phi_n$, where $\phi_n = \neg\varphi$ is a proof of $\neg\varphi$ from $\mathcal{D}_\neg$ in PL, then $\neg\phi_1, \ldots, \neg\phi_n = \varphi$ is a proof of φ from $\mathcal{D}$ in PLD, replacing each application of modus ponens by an application of **CMP** to the negated premisses. And vice-versa. □

There is a noticeable inference rule that is admissible in PLD, namely a form of weakening rule for desires:

$$(\textbf{DW}) : \varphi \vdash_{PLD} \varphi \wedge \psi.$$

Indeed, since $\neg\varphi \models \neg\varphi \vee \neg\psi$, we do have that $\varphi \models_D \varphi \wedge \psi$ using (1), hence (**DW**) holds using the completeness theorem; this is the desire counterpart of the weakening rule $\phi \vdash \phi \vee \psi$ in classical logic.

Clearly, property **R2** is exactly encoded by inference rule (**DW**). As for **R3**, we can prove it in PLD if we restrict $\mathcal{D}$ to the negation of propositional axioms, using the fact that φ is a theorem of PLD if and only if $\neg\varphi$ is a theorem of LP in the usual sense. Then since $\vdash \neg\varphi$ and $\vdash \neg\psi$ imply $\vdash \neg\varphi \wedge \neg\psi$, we do get that $\vdash_{PLD} \varphi$ and $\vdash_{PLD} \psi$ imply $\vdash_{PLD} \varphi \vee \psi$.

More generally, the inference from a set of desires $\mathcal{D}$ can be based on the following rule of resolution, which is the desire counterpart of the resolution rule for beliefs (namely, $\{\varphi \vee \psi, \neg\varphi \vee \nu\} \vdash \psi \vee \nu$), that clearly extend **CMP**:

$$(\mathbf{DR}) : \frac{\varphi \wedge \psi, \neg\varphi \wedge \nu}{\psi \wedge \nu}$$

Proposition 3 *The resolution rule* (**DR**) *is admissible in PLD.*

Proof It is obvious, applying (1) and the completeness theorem to $\mathcal{D} = \{\varphi \wedge \psi, \neg\varphi \wedge \nu\}$ and the usual resolution rule to $\mathcal{D}_\neg = \{\neg\varphi \vee \neg\psi, \varphi \vee \neg\nu\}$, which enables $\neg\psi \vee \neg\nu$ to be derived. □

We thus get a propositional logic of desires that is a mirror image of the usual propositional logic (of beliefs).

Besides the application of the resolution rule **DR** in refutation style for desire bases, denoted by $\vdash_{DR}$, is sound and complete with respect to semantic entailment $\mathcal{D} \models_D \psi$:

Theorem 2 $\mathcal{D} \models_D \psi$ *if and only if* $\mathcal{D} \cup \{\neg\psi\} \vdash_{DR} \top$.

Proof Given a desire base $\mathcal{D}$, build the propositional base $\mathcal{D}_\neg = \{\neg\psi_i : \psi_i \in \mathcal{D}\}$. Then $\mathcal{D} \models_D \psi$ if and only if $\mathcal{D}_\neg \models \neg\psi$ if and only if $\mathcal{D}_\neg \cup \{\psi\} \vdash \bot$ using the PL resolution rule, if and only if $\mathcal{D} \cup \{\neg\psi\} \vdash_{DR} \top$ using the resolution rule for desires **DR**. □

3.4 A Minimal Modal Logic of Desires

The idea that an agent desires some state of affairs φ to be true if and only if all situations in which φ is true are desirable can be expressed by means of the following modal operator.

Definition 3.4 *(Desire operator)* Let $M = (W, \delta)$ be a hedonic model, and let $\varphi \in \mathcal{L}_{\mathsf{PL}}$. Then:

$$M \models \mathsf{D}\varphi \Longleftrightarrow \forall w \in W : \text{ if } w \models \varphi \text{ then } \delta(w) = 1$$

$M \models \mathsf{D}\varphi$ means that the hedonic state D^δ contains the set of models of φ. Note that the desire operator works in a reverse way compared to a standard necessity operator in modal logic. This type of modal operator was introduced in Dubois et al. (2000) in the setting of possibility theory.

In this section, we provide a sound and complete axiomatization for the minimal modal logic of desires (MLD) whose language ${\mathcal{L}_{\mathsf{PL}}}^D$ contains atoms of the form $\mathsf{D}\varphi$

for all propositional formulas $\varphi \in \mathcal{L}_{\mathsf{PL}}$ and their propositional combinations via $\neg$ and $\wedge$.

The crucial thing to observe is that the satisfaction relation for the desire operator D for the hedonic model $M = (W, \delta)$ can be reformulated in a equivalent way as follows:

$$M \models \mathsf{D}\varphi \Longleftrightarrow \forall w \in W : \text{ if } \delta(w) = 0 \text{ then } w \models \neg\varphi$$

This means that the formula $\mathsf{D}\varphi$ can be seen as an abbreviation of $\Box\neg\varphi$, where $\Box$ is a normal modal operator of type KD, and $(W, \delta) \models \mathsf{D}\varphi$ is equivalent to $(W, 1 - \delta) \models \Box\neg\varphi$, where $1 - \delta$ is the characteristic function of an epistemic state (a non-empty subset E of W). Consequently, to provide a sound and complete axiomatization of the logic MLD, it is sufficient to borrow the principles of the normal modal system KD, where each instance of a formula $\Box\varphi$ is replaced by $\mathsf{D}\neg\varphi$. In fact, only the simple fragment of KD encoding the all-or-nothing version of possibility theory in (Banerjee and Dubois 2014) is needed. This is what the next theorem highlights.

Theorem 3 *The following four principles plus Modus Ponens provide a sound and complete axiomatization of the logic MLD:*

PL *Axioms of Propositional logic for expressions in* ${\mathcal{L}_{\mathsf{PL}}}^{D}$;
K^c $\mathsf{D}(\neg\varphi \wedge \psi) \rightarrow (\mathsf{D}\varphi \rightarrow \mathsf{D}\psi)$;
N^c *If* $\vdash \neg\varphi$, *then* $\mathsf{D}\varphi$
D^c $\mathsf{D}\varphi \rightarrow \neg\mathsf{D}\neg\varphi$.

Proof The proof is a straightforward adaption of the completeness proof for minimal epistemic logic (MEL) given in Section 3 of Banerjee and Dubois (2014), since MLD is identical to MEL once we change $\mathsf{D}\varphi$ into $\Box\neg\varphi$. □

The first modal axiom is the desire-based translation of Axiom K, while the second corresponds to Axiom N, and the third one to axiom D. Axiom K^c means that if you desire worlds where $\neg\varphi \wedge \psi$ is true and moreover you desire worlds where φ is true, then you also desire worlds where ψ is true (indeed, $\psi \rightarrow (\neg\varphi \wedge \psi) \vee \varphi$ is valid). Interestingly, this logic exactly corresponds to the data-driven modal logic previously proposed by two of the authors with Hajek (Dubois et al. 2000), as a counterpart of KD45. This is because guaranteed possibility measures have an epistemic understanding in terms of observations made, that parallels the representation of desires. However, MLD uses only a fragment of the language in (Dubois et al. 2000) and the semantics of MLD is simpler as it does not use accessibility relations.

The three following principles are valid in MLD (see (Banerjee and Dubois 2014) for the proof of their counterparts):

1. $(\mathsf{D}\varphi \wedge \mathsf{D}\psi) \equiv \mathsf{D}(\varphi \vee \psi)$;
2. $\neg\mathsf{D}\top$;
3. $\mathsf{D}\psi \rightarrow \mathsf{D}\varphi$ whenever $\vdash \varphi \rightarrow \psi$ in propositional logic.

The first principle is the translation of the adjunction rule and the third principle corresponds to the so-called rule of monotony (RM) in the logic KD. They respectively correspond to the properties **R3**, **R1** and **R2** in Definition 3.2.

Finally if we restrict the language ${\mathcal{L}_{\mathsf{PL}}}^{D}$ to conjunctions of atoms $\mathsf{D}\varphi$, we get a copy of the propositional logic of desires of the previous Sect. 3.2:

Corollary 1 $\bigwedge_{i=1}^{n} \mathsf{D}\varphi_i \vdash_{MLD} \mathsf{D}\varphi$ *if and only if* $\{\varphi_1, \ldots, \varphi_n\} \vdash_{PLD} \varphi$.

Proof This is the desire counterpart of Theorem 2 in Banerjee and Dubois (2014). Note that $\bigwedge_{i=1}^{n} \mathsf{D}\varphi_i$ is equivalent to $\mathsf{D} \vee_{i=1}^{n} \varphi_i$, and $\bigwedge_{i=1}^{n} \mathsf{D}\varphi_i \vdash_{MLD} \mathsf{D}\varphi$ is the same as $\varphi \models \vee_{i=1}^{n} \varphi_i$ in classical logic. □

This result shows the complete agreement of the minimal modal logic of desires MLD with our view of a closed propositional desire base in Definition 3.2.

4 Graded Hedonic States: from Desirability Relations to Possibility Theory

In this section, 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 (Dubois and Prade 1998) as a unique numerical counterpart, as we shall see. This leads to representing a gradual hedonic state by means of a guaranteed possibility distribution, thus extending the framework of the previous section. In this section, we represent propositions by subsets of possible worlds $A, B, C \ldots \subseteq W$ which form a (finite) Boolean algebra $\wp(W)$.

4.1 Axioms for Desirability Relations

Desires are a matter of strength. Some situation may be more strongly desired than another one by an agent. Thus, the hedonic state of an agent will be described by an ordering relation on $\wp(W)$, denoted by $\geq_\Delta$, called *desirability relation*. Such a relation compares subsets of W representing propositions in terms of satisfaction the agent expects from them if made true. $A \geq_\Delta B$ should be read "A is at least as desirable as B"; it means that concretizing A should be at least as satisfactory as concretizing B, or if we prefer that A is at least as desired as B, which reflects a gradual view of desire. As usual, $A >_\Delta B$ when $A \geq_\Delta B$ but not $B \geq_\Delta A$; and $A \sim_\Delta B$ means $A \geq_\Delta B$ and $B \geq_\Delta A$.

The previous discussions in the Boolean case leads to propose that $\geq_\Delta$ should satisfy the following axioms:

(A0) $\emptyset >_\Delta W$
(A1) $A \geq_\Delta B$ or $B \geq_\Delta A$
(A2) $A \geq_\Delta B$ and $B \geq_\Delta C$ imply $A \geq_\Delta C$
(AM) if $A \subseteq B$, then $A \geq_\Delta B$
(Pos) $\forall A$, if $B \geq_\Delta C$ then $A \cup B \geq_\Delta A \cup C$

Admitting that desires behave in a reverse way with respect to entailment, the anti-monotonicity axiom **(AM)** is natural, and **(A0)** expresses the corresponding non triviality condition. Axiom **(AM)** implies the limit condition

$$(\mathbf{A3}) : \emptyset \geq_\Delta A.$$

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 **(Pos)** states that if "B is at least as desirable as C", this preference in the broad sense cannot be altered by enlarging the scope of the comparison on both sides in the same way by A. 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 (Lewis 1973; Dubois 1986)

The following result is worth noticing:

Proposition 4 *Under* **(Pos)**, **(A3)** *implies* **(AM)**.

Proof By (**A3**), $\emptyset \geq_\Delta B \setminus A$ and applying **(Pos)**, we get $A \geq_\Delta B$. □

So desirability relations are characterized as well by **(A0)**–**(A3)** and **(Pos)**.

Desirability Relations Versus 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 (Gärdenfors 1988; Rott 2001). It has been established that epistemic entrenchment relations are nothing but comparative necessity relations $\geq_N$ up to a minor difference, namely axiom $W >_N \emptyset$ is strengthened into $W >_{epis} A, \forall A \neq W$, for epistemic entrenchments (Dubois and Prade 1991), which means that no state of affairs is impossible. Comparative necessity relations (and thus epistemic entrenchment relations) satisfy **(A1)** and **(A2)**, but they obey counterparts of the other axioms, namely

(A′0): $W >_N \emptyset$
(A′3) $A \geq_N \emptyset, \forall A$.
(Nec) If $A \geq_N B$ then $A \cap C \geq_N B \cap C$.

The latter is the characteristic property of comparative necessity relations. Under axioms **(A′0)**–**(A1)**–**(A2)**–**(A′3)**, **(Nec)** is equivalent to

$$\text{if } A \geq_N B \text{ then } A \cap B \sim_N B,$$

where $A \geq_N B$ means that A is at least as certain as B (Dubois 1986).

By duality, comparative necessity relations $\geq_N$ are associated with comparative possibility relations $\geq_\Pi$ through the equivalence $A \geq_\Pi B \Leftrightarrow \overline{B} \geq_N \overline{A}$, where $\overline{A}$ is the complement of A. Comparative possibility relations (Dubois 1986) satisfy axioms **(A′0)**, **(A1)**–**(A2)**, together with axiom **(Pos)**. It is remarkable that switching from comparative possibility relations to desirability relations comes down to only

changing axioms "$W > \Pi \emptyset$" and "$A \geq_\Pi \emptyset$" for comparative possibility into **A0** ($\emptyset > \Delta W$) and **(A3)** ($\emptyset \geq_\Delta A$) for desirability relations.

4.2 Properties of Desirability Relations

The set of axioms **(A0)–(A3)** and (**Pos**) entail noticeable properties for desirability relations that agree with intuition. First, we can establish the following result, which parallels a counterpart for comparative possibility relations in Dubois (1986).

Proposition 5 *Under axioms* (**A0**)–(**A3**), *axiom* (**Pos**) *is equivalent to*

$$(\Delta) \text{ if } A \geq_\Delta B \text{ then } A \cup B \sim_\Delta B.$$

Proof $(\Delta) \Rightarrow$ **(Pos)**.

Assume $B \geq_\Delta C$.

- If $A \geq_\Delta B \geq_\Delta C$, $A \cup B \sim_\Delta B \geq_\Delta C \sim_\Delta A \cup C$;
- If $B \geq_\Delta A \geq_\Delta C$, $A \cup B \sim_\Delta A \geq_\Delta C \sim_\Delta A \cup C$;
- If $B \geq_\Delta C \geq_\Delta A$, $A \cup B \sim_\Delta A \geq_\Delta A \sim_\Delta A \cup C$.

(Δ) follows by transitivity. Note that we do not use (**A3**).

(Pos) $\Rightarrow$ (Δ).

Suppose $A \geq_\Delta B$. Then, by **(Pos)**, $A \cup C \geq_\Delta B \cup C$, in particular $A \cup B \geq_\Delta B$. Now by (**A3**), $\emptyset \geq_\Delta A$ and applying **(Pos)**, we get $B \geq_\Delta A \cup B$. □

Clearly, (Δ) expresses that desiring $A \cup B$ has the same strength as desiring the least desired of A and B. The strength of desire for $A \cup B$ is *conservatively evaluated* by the least desired situation realizing it because at the moment of its realization there may be constraints restricting the choice of the possible world that makes $A \cup B$ fulfilled. For instance, if the agent desires coffee (A) more strongly than tea (B), the strength of desiring coffee and tea is the least of the two desires, because she does not know in advance which one of coffee or tea will be available for drinking at the time where this desire is to be made true. So the axioms of comparative desirability carry a flavor of safeness. Then, it seems intuitively satisfactory that the strength of desire of $A \cup B$ should be *at most* equal to the minimum of the desire strengths of A and B (when dealing with positive desires), since fulfilling $A \cup B$ may let me not better off than fulfilling any of them.

Proposition 6 *Under axioms* (**A0**)-(**A3**) *and* (**Pos**)

a Either $A \sim_\Delta W$ or $\overline{A} \sim_\Delta W$ or both.
b $\forall A, A \geq_\Delta W$.

Proof

a It is an immediate consequence of **(A2)** and (Δ), since letting $B = \overline{A}$ in (Δ), we get $W \sim_\Delta \overline{A}$ if $A \geq_\Delta \overline{A}$, and $W \sim_\Delta A$ if $\overline{A} \geq_\Delta A$.
b Since $\forall A, \forall C, A \geq_\Delta A \cup C$, letting $C = \overline{A}$ yields the result.

□

Proposition 6 states that one cannot desire A and $\overline{A}$ 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 desirability is minimal as stated by Property **b**.

4.3 From Desirability Relations to 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 measuring the strength of desire, if any, that can represent desirability relations.

A numerical function F from $\wp(W)$ to [0, 1] is said to represent a relation $\geq_R$ if $\forall A, B, A \geq_R B \Leftrightarrow F(A) \geq F(B)$. In the following, we assume that the set W of associated interpretations of the considered language, with subsets A, B, etc., is finite.

The only numerical functions compatible with the desirability relation ordering $\geq_\Delta$ are so-called *guaranteed possibility measures*[7] (Dubois et al. 2000; Dubois and Prade 2009b) in the sense of possibility theory, as shown now. A guaranteed possibility measure Δ, from $\wp(W)$ to [0, 1], is characterized by the limit conditions $\Delta(\emptyset) = 1$ and $\Delta(W) = 0$, and by the decomposability property:[8]

$$\Delta(A \cup B) = \min(\Delta(A), \Delta(B)), \forall A, B \in \wp(W). \tag{2}$$

Proposition 7 *Any numerical function F, from $\wp(W)$ to* [0, 1], *representing 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 $\wp(W)$ to* [0, 1] *satisfying $\Delta(\emptyset) > \Delta(W)$ induces a qualitative relation satisfying* **(A0)–(A3)** *and* **(Pos)**.

Proof ($\Rightarrow$) From Proposition 5, any set-function Δ representing $\geq_\Delta$ must be such that $\Delta(A \cup B) = \min(\Delta(A), \Delta(B))$ since $A \cup B \sim_\Delta B$ whenever $A \geq_\Delta B$. As $\emptyset \geq_\Delta A \geq_\Delta W$ and $\geq_\Delta$ is antimonotonic (item a of Proposition 6), and moreover $\emptyset >_\Delta W$ by (A0), one can fix $\Delta(W) = 0$ and $\Delta(\emptyset) = 1$, with $\Delta(A) \in [0, 1]$.

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

[7] The term *guaranteed* comes from the fact that in possibility theory, $\Delta(A) = 1$ means that *all* instances of A can actually be observed. It contrasts with potential possibility, which takes value 1 as soon as one instance of A is observed.

[8] This equality is the same as the one that Spohn ranking functions κ (Spohn 2012) satisfy, but identifying Δ and κ would be misleading: Δ and κ are antimonotonic with inclusion, but the ranges of Δ and κ, respectively [0, 1] and $\mathbb{N}$ (non-negative integers) are directed in opposite ways: $\kappa(A) = 0$ expresses full possibility, and the higher $\kappa(A)$ the more A is impossible, while $\Delta(A) = 1$ expresses full possibility. When mapping $\mathbb{N}$ to [0, 1] via an order-reversing function (e.g., $f(A) = 2^{-\kappa(A)}$), we obtain an increasing maxitive function: $f(A \cup B) = \max(f(A), f(B))$, which departs from the characteristic property of Δ.

This proposition states that the only numerical functions agreeing with a qualitative ordering $\geq_\Delta$ are those obeying decomposability property (2). Note that the range [0, 1] may be replaced by any linearly ordered, possibly finite, scale.

Finally, we can define a function $\delta : W \to [0, 1]$, expressing the strength of desire of possible worlds, that characterizes the set-function Δ:

Corollary 2 $\forall A \neq \emptyset, \exists w \in A, \{w\} \sim_\Delta A$.

Proof Suppose $A = \{w_1, \ldots, w_k\}$ and $w_1 \geq_\Delta \ldots, \geq_\Delta w_k$. It is clear from Proposition 5 that since $\{w_k\}$ is the least element in A,

$$w_k \sim_\Delta \{w_k, w_{k-1}\} \sim_\Delta \{w_k, w_{k-1}, w_{k-2}\} \sim_\Delta A.$$

□

Namely, the gradual desirability function δ can be defined as $\delta(w) = \Delta(\{w\})$. It satisfies a normality constraint $\delta(w) = 0$ for some $w \in W$ since $\Delta(W) = 0$. It is such that

$$\Delta(A) = \min_{w \in A} \delta(w). \tag{3}$$

The function δ is usually called a guaranteed possibility distribution (Dubois et al. 2000; Dubois and Prade 2009b); its range [0, 1], can more generally be any linearly ordered scale S. Here, $\delta(w)$ represents the degree of desirability of a given world $w \in W$. Obviously, the binary function δ_λ such that $\delta_\lambda(w) = 1$ if $\delta(w) \geq \lambda$ and 0 otherwise is a desirability function in the sense of Sect. 3.

Our results thus show that we can interpret guaranteed possibility measures Δ in terms of (guaranteed) desirability. Namely, $\Delta(A)$ is the extent to which the agent can safely desire proposition A to be true (safely in the sense that, no matter how A is fullfilled, the expected level of satisfaction is at least $\Delta(A)$). As a consequence of Proposition 5, we have

$$\min(\Delta(A), \Delta(\overline{A})) = \Delta(W) = 0$$

which is the numerical counterpart of property **b** in Proposition 6, which is $\Delta(A) > 0$ then $\Delta(\overline{A}) = 0$. This means that if an agent safely desires A to be true—i.e., with some strength $\alpha > 0$—then she does not safely desire at all A to be false. This is a form of consistency requirement also expressed by normality constraint of δ ensuring that not everything can be desired.

Besides,

$$\Delta(A \cap B) \geq \max(\Delta(A), \Delta(B)).$$

This is the consequence that Δ is decreasing with respect to entailment (i.e., Axiom **AM**). This makes perfect sense for motivational attitudes like desires, as suggested by the following example.

Example 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. One way out is to restrict desires to good wines explicitly. One referee replaced cheese by beer in the above example, which then becomes a counterexample as one usually does not like to drink beer and (even good) wine together.[9] In our framework one way to avoid this situation is to explicitly say wine alone (i.e., without beer) and beer alone (this is what people usually mean when they say they like beer or wine). It means that in $\mathcal{D}$ we find both $wine \wedge \neg beer$ and $beer \wedge \neg wine$. Alternatively if wine together with beer is a rejected option, it pertains to negative desires that must be expressed prior to describing positive desires.

Such exceptions could be also coped with by means of non-monotonic desires; see Dubois et al. (2014) for a preliminary proposal in the possibilistic reasoning setting, but this is out of the scope of the present paper. In that approach, one could say that in the beer context no wine is desired, in the wine context no beer is desired, and that "wine or beer" is desired, i.e., $A \cap \overline{B} >_\Delta A \cap B$, $B \cap \overline{A} >_\Delta A \cap B$, and $A \cup B >_\Delta \overline{A} \cap \overline{B}$, which are compatible constraints.

4.4 Modeling Desires in Possibilistic Logic

As suggested in Casali et al. (2011), and taken up in Dubois et al. (2014), a desire A with strength α is properly represented by a constraint of the form $\Delta(A) \geq \alpha$ which stands for "the agent desires A with strength at least α". If A is the set of models of a formula φ, it can be syntactically represented by a pair $[\psi, \alpha]$. The corresponding desirability function $\delta_{[\psi,\alpha]}$ on W is defined as $\delta_{[\psi,\alpha]}(w) = \alpha$ if $w \models \psi$ and 0 otherwise. It is the smallest possibility distribution such that $\Delta(||\psi||) \geq \alpha$, according to a maximum specificity principle. This principle says that it is cautious not to assume that models of φ are desired to a strength greater than α, while no countermodels of ψ are desired by lack of information that some of them would be.

Generalizing the logical setting of Sect. 3.1, a desire function will be defined by a function $f_\mathcal{D} : \mathcal{L}_{\mathsf{PL}} \longrightarrow [0, 1]$ (more generally, a bounded, totally ordered chain of levels). A desire function will then induce a prioritized desire base, first introduced in Benferhat et al. (1999) (see also Benferhat and Kaci 2003 for an extensive presentation). Specifically, the prioritized desire base $\mathcal{D}$ induced by the desire function $f_\mathcal{D}$ is a collection of weighted formulas $[\psi_i, \alpha_i]$ where $f_\mathcal{D}(\psi_i) = \alpha_i$, equivalently a collection of nested desire sets $(\mathcal{D}_\lambda)_{\lambda \in [0,1]}$ such that for all $\lambda \in [0, 1]$:

[9] Note that wine and cheese are complementary, while beer and wine can be viewed as redundant, and thus less jointly desirable.

$$\mathcal{D}_\lambda = \{\varphi \in \mathcal{L}_{\mathsf{PL}} : f_{\mathcal{D}}(\varphi) \geq \lambda\}$$

The possibilistic logic of desires does not obey the same inference rules as the possibilistic logic of beliefs. Indeed now the resolution rule **DR** for desires writes in its weighted version:

(**DRw**)**: From** $[\varphi \wedge \psi, \alpha]$ **and** $[\neg\varphi \wedge \nu, \beta]$ **deduce** $[\psi \wedge \nu, \min(\alpha, \beta)]$.

A set $\mathcal{D}$ of weighted desires $[\psi_i, \alpha_i]$ (for $i = 1, \ldots, m$) is semantically associated to a guaranteed possibility distribution

$$\delta_{\mathcal{D}}(w) = \max_{i=1,\ldots,m} \min(||\psi_i||(w), \alpha_i). \tag{4}$$

where $||\psi_i||(w) = 1$ if w is a model of ψ_i, and $||\psi_i||(w) = 0$ otherwise. $\delta_{\mathcal{D}}$ is again the smallest possibility distribution such that $\Delta(||\psi_i||) \geq \alpha_i$ for $i = 1, \ldots, m$, using a minimal commitment principle: there is no other desire than those expressed in the desire set $\mathcal{D}$. The use of max is justified by taking the set union of the sets of models of weighted formulas in the graded desire base $\mathcal{D}$, in agreement with the semantics of the propositional logic of desires outlined in Sect. 3.2. The distribution $\delta_{\mathcal{D}}$ rank-orders the interpretations of the language induced by the ψ_i's according to their satisfaction level on the basis of the strength of the desires in $\mathcal{D}$. A hedonic state can then be viewed as a fuzzy (or graded) subset of worlds.

The quantity

$$una(\mathcal{D}) = \min_w \delta_D(w)$$

may be viewed as a level of inconsistency of $\mathcal{D}$. The larger $una(\mathcal{D})$, the more inconsistent (that is to say, unacceptable) the set of desires $\mathcal{D}$. Ideally, $\min_w \delta_{\mathcal{D}}(w) = 0$ should hold, in which case, the desire base $\mathcal{D}$ is consistent.

Desires Versus Beliefs in Possibility Theory The above modeling of graded desire bases contrasts with the representation of graded beliefs that is properly represented by a constraint of the form $N(A) \geq \alpha$ which stands for "the agent believes A with strength at least α". Function N is a necessity measure (Dubois and Prade 1998) defined from a possibility distribution $\pi : W \to [0, 1]$ by

$$N(A) = 1 - \max_{w \notin A} \pi(w).$$

This expression is the counterpart for beliefs of expression (3) for desires. A possibility distribution π represents a graded epistemic state, such that $\pi(w) = 0$ means that world w is ruled out by our information. The degree $N(A)$ estimates the extent to which the agent believes A to be true, all the more as its complement is found impossible in the sense of π. Indeed, the necessity measure of N is the dual of the possibility measure Π, namely $\Pi(A) = 1 - N(\overline{A})$ (where $\overline{A}$ is a complement of A). Beliefs, modeled by means of necessity measures, satisfy

$$N(A \cap B) = \min(N(A), N(B))$$

i.e., believing A *and* B amounts to believing A and to believing B. It is easy to see that

$$N(A) = \min_{w \notin A} 1 - \pi(w) = \Delta(\overline{A}) = \min_{w \notin A} \delta(w)$$

if $\pi = 1 - \delta$. That is, *at a purely mathematical level*, a necessity measure is closely related to a guaranteed possibility measure.

Desire bases can be contrasted with the possibilistic representation of a belief base $\mathcal{B}$ expressed in possibilistic logic (Dubois and Prade 2004). Then, a set of weighted formulas (φ_j, γ_j) (for $j = 1, \ldots, n$) encodes constraints of the form $N(||\varphi_j||) \geq \gamma_j$. $\mathcal{B}$ is semantically associated with a possibility distribution

$$\pi_{\mathcal{B}}(w) = \min_{j=1,\ldots,n} \max(||\varphi_j||(w), 1 - \gamma_j).$$

This expression is to be compared with (4) for desire bases. Here, $\pi_{\mathcal{B}}$ is the largest (least committed) possibility distribution (minimum specificity principle) such that $N(\varphi_j) \geq \gamma_j$ for $j = 1, \ldots, n$. The distribution $\pi_{\mathcal{B}}$ rank-orders the interpretations of the language induced by the φ_j's according to their plausibility on the basis of the strength of the beliefs in $\mathcal{B}$. As clear in the expression of $\pi_{\mathcal{B}}$, a belief set is underlain by a (weighted) conjunction of the pieces of beliefs, while in (4) a desire set should be understood as a disjunction of formulas, in agreement with the intuition.

If the set of beliefs $\mathcal{B}^* = \{\varphi_j, j = 1, \ldots, n\}$ is consistent then the distribution $\pi_{\mathcal{B}}$ is normalized in the sense that $\exists w, \pi_{\mathcal{B}}(w) = 1$. More generally the level of inconsistency of $\mathcal{B}$ is defined by $inc(\mathcal{B}) = 1 - \max_w \pi_{\mathcal{B}}(w)$. Thus the degree of inconsistency $una(\mathcal{D})$ should play the same role in desire revision as $inc(\mathcal{B})$ in belief revision (Benferhat et al. 2001; Benferhat et al. 2002b).

The weighted resolution rule for desires **(DRw)** contrasts with the better known resolution rule for prioritized beliefs: $(\varphi \vee \psi, \alpha)$ and $(\neg\varphi \vee \nu, \beta)$ entails $(\psi \vee \nu, \min(\alpha, \beta))$ (Dubois and Prade 2004) and follows from it by transforming weighted beliefs (φ, α) into weighted desires of the form $[\neg\varphi, \alpha]$.

4.5 Related Work

An area of research which looks related to our work is preference logic. According to contemporary theories of human motivation both in philosophy and in economics (e.g., Searle 2001; Harsanyi 1982, 1955), preferences of a rational agent may originate either (1) from somatically-marked motivations such as desires (e.g., the preference for a glass of lemonade over a cappuccino originates from the desire of drinking something fresh), or (2) from moral considerations and values (e.g., the preference for helping a poor person, over ignoring her, originates from the moral value of taking care of needy people). More generally, there exist desire-dependent

preferences and desire-independent ones originated from moral values.[10] According to this view, desires and moral attitudes of an agent are two different parameters affecting the agent's preferences. Although, from a conceptual point of view, desire—our object of analysis—is more primitive than preference, existing preference logics share similarities with our approach.

In his seminal work on preference logic (Von Wright 1963, 1972), Von Wright studied dyadic preferences of the form "I prefer φ over ψ" by giving a *ceteris paribus* interpretation of this notion. Specifically, according to Von Wright, φ is preferred to ψ if and only if, *all other things being equal*, any situation in which φ is true is preferred to any situation in which ψ is true. A modal logic account of Von Wright's notion of preference has been recently proposed by van Benthem et al. (2009) who generalize previous work by van Benthem and Liu (2007). They use a total preorder on a set of possible worlds for modelling preferences. A main difference with our work is that the preference relation on possible worlds does not convey any positive nor negative flavor. No world is considered good or bad, only some worlds are preferred to other ones, without absolute judgements about them. In particular, in the paper, there are several ways of lifting the preference relation over to formulas that are equally good and useful. On the contrary, as we consider that desire has a positive flavor, we advocate, via axioms, one way of lifting the desirability relation between worlds to formulas. In our approach, relative desires $\psi \leq_\Delta \varphi$ express the idea that the least desired model of ψ is not worse than the least desired model of ψ; this is less demanding, and we do not make the preferential independence assumption implicit in the *ceteris paribus* method. Doyle et al. (1991) also consider relative desires $\psi \leq \varphi$ that hold if any ψ-world is preferred to any φ-world expressed with propositional variables appearing in φ and ψ, completed by a *ceteris paribus* condition for other propositional variables. Again their notion of relative desire does not seem to differ much from relative preference.

Since in the present work we do not model negative desires (whereby a certain fact φ would be undesirable with a certain strength k), our semantics in terms of possibility distributions over [0, 1] and Van Benthem's qualitative semantics are not equivalent with respect to our notion of desire. Negative desires are not expressible together with positive desires by using a single total preorder over interpretations. We would need a second preference relation expressing fears or unsatisfactory worlds

Lang et al. (2002) study conditional desires of the form $D(\varphi|\psi)$ that put a penalty on interpretations where ψ is true and φ is false and a reward on interpretations where ψ and φ are true. Penalties and rewards are supposed to add and lead to the construction of a utility function on possible worlds (Lang and van

[10] This distinction leads to the identification of two different kinds of moral dilemmas. The first kind of moral dilemma is the one which is determined by the logical conflict between two moral values. The paradigmatic example is the situation of a soldier during a war. As a member of the army, the soldier feels obliged to kills his enemies, if this is the only way to defend his country. But, as a catholic, he thinks that human life should be respected. Therefore, he feels morally obliged not to kill other people. The other kind of moral dilemma is the one which is determined by the logical conflict between desires and moral values. The paradigmatic example is that of Adam and Eve in the garden of Eden. They are tempted by the desire to eat the forbidden fruit and, at the same time, they have a moral obligation not to do it.

der Torre 1996). In contrast, our approach is qualitative (satisfaction degrees do not add), unconditional and focuses on positive desires (there is a positive effect of satisfying φ, no negative effect in not satisfying it) because we consider negative desires (fears) should be handled using a different formalism, even if articulated with desires.

5 Desire Revision

Revision in logic has mainly considered belief sets. In the dominant AGM view described by Gärdenfors (1988), a 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 that is not an explicit part of the epistemic state of the agent.[11] Namely the epistemic state of 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 potential changes of the belief set. The study concludes that belief changes occur as if there were an epistemic entrenchment driving the process.

A more practical approach (Benferhat et al. 2002b) views epistemic states as a prioritized collection of pieces of belief, which are thus associated to priorities that enable us to compute their entrenchment level as the value of a necessity measure.

In this section we study the counterpart of the AGM theory of belief change (Gärdenfors 1988) for desires using closed desired sets, and then weighted desire bases.

5.1 Axioms for Desire Revision

In this subsection, we only consider desire sets $\mathcal{D}$, understood as closed desire bases, namely such that if $\psi \in \mathcal{D}$ and $\varphi \vdash_{PLD} \psi$ then $\varphi \in \mathcal{D}$, if $\varphi \in \mathcal{D}$, $\psi \in \mathcal{D}$ then $\varphi \vee \psi \in \mathcal{D}$. Just recall that inference is reversed with respect to the usual one, i.e., $\varphi \vdash_{PLD} \psi$ means $\psi \vdash \varphi$ with the usual convention.

Suppose a new formula φ is added to a consistent desire base $\mathcal{D}$. As already said, one cannot desire φ and desire $\neg\varphi$ at the same time without being led to desire inconsistency. So, if $\neg\varphi \in \mathcal{D}$ the new desire base becomes inconsistent and should be revised. Let $\mathcal{D}^*_\varphi$ be the result of revising D by φ.

Having in mind the reverse behavior of desires with respect to beliefs, one is naturally led to state for desires axioms that parallel the AGM axioms (Gärdenfors 1988) for belief revision, by in some sense reversing the latter axioms, as follows:

- [(D*1)] for any sentence and any desire set $\mathcal{D}$, $\mathcal{D}^*_\varphi$ is a desire set.
- [(D*2)] $\varphi \in \mathcal{D}^*_\varphi$.
- [(D*3)] $\mathcal{D}^+_\varphi \supseteq \mathcal{D}^*_\varphi$

[11] At least this is so in the book (Gärdenfors 1988) as the epistemic entrenchment is a consequence of the axioms.

- [(D*4)] If $\neg\varphi \notin \mathcal{D}$ then $\mathcal{D}^*_\varphi \supseteq \mathcal{D}^+_\varphi$
- [(D*5)] $\mathcal{D}^*_\varphi = \mathcal{L}_{\mathsf{PL}}$ if and only if $\varphi \equiv \top$
- [(D*6)] If $\vdash \varphi \equiv \psi$ then $\mathcal{D}^*_\varphi = \mathcal{D}^*_\psi$
- [(D*7)] $\mathcal{D}^*_{\varphi\vee\psi} \subseteq (\mathcal{D}^*_\varphi)^+_\psi$
- [(D*8)] If $\neg\psi \notin \mathcal{D}^*_\varphi$ then $\mathcal{D}^*_{\varphi\vee\psi} \supseteq (\mathcal{D}^*_\varphi)^+_\psi$

where the expansion $\mathcal{D}^+_\varphi$ is just defined by a "reverse logical closure" of $\mathcal{D}$ together with φ, in agreement with the intuition underlying the idea of desire:

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

(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 adding 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,[12] (D*7) and (D*8) are required for ensuring the existence of a particular desirability relation called "hedonic entrenchment" in the sense of the postulates of Sect. 4.1. This can be established following a route very similar to the one of Grove (1988) for epistemic entrenchment, taking into account the reverse behavior of hedonic entrenchment, and remembering the very close relationship between sphere systems and possibility distributions.

Alternatively we can use a formal transformation of a desire base $\mathcal{D}$ into another set of formulas on which the AGM belief revision axioms can be equivalently used instead of desire revision axioms. Namely from a closed set of desires $\mathcal{D}$ we can construct the closed belief set $\mathcal{K} = \{\neg\varphi : \varphi \in \mathcal{D}\}$. Then AGM representation theorems can be adapted to the setting of possibilistic desires. For instance, given a desire base $\mathcal{D}$, and an input φ, we can claim that a desire revision operation $*$ satisfies axioms (D*1-D*8) if and only if there exists a desirability relation $\geq_\Delta$ such that $\forall\psi \in \mathcal{D}, \psi >_\Delta \neg\psi$ and the revised desire base is of the form $\mathcal{D}^*_\varphi = \{\psi : \varphi \wedge \psi >_\Delta \varphi \wedge \neg\psi\}$.

Since the approach is syntax-free, it is simpler to write the above axioms in terms of subsets of 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)] $A \subseteq D^*_A$.

[12] In the paper, we use a very restrictive definition of epistemic entrenchment, which is the one in Gärdenfors (1988), that is, a total preorder relation on the language that obeys specific properties, which can only be justified if we take the two last AGM axioms for granted. Clearly one could envisage a less restrictive framework for desire revision, similar to the ones studied by Rott (2001) without (D*7) and (D*8).

- [(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:

$$D_A^* = \begin{cases} D \cup A \text{ if } D \cup A \neq W. \\ \text{some } C \neq W, \text{ such that } C \supset A \text{ otherwise.} \end{cases}$$

- The two last axioms come from the choice function area, and specify that some result $C \supset A$ is selected in agreement with an ordering on W (the most desired states outside A; Bonanno 2009). 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 some desire is explicit, one assume states are not desired (i.e., indifferent for the agent). 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 highlights the interest of possessing a gradual prior hedonic distribution, which may guide the desire revision process, just like a possibility distribution does in the case of beliefs. The next section we investigate desire revision with explicit desire strengths.

5.2 Belief Revision, Expansion and Contraction in Possibility Theory

It has been pointed out early that the epistemic entrenchment relations underlying any well-behaved belief revision process obeying AGM postulates (Gärdenfors 1988) are qualitative necessity relations (Dubois and Prade 1991), thus establishing a link between belief revision and possibility theory (Dubois and Prade 1998). 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, or equivalently the possibility distribution on interpretations underlying it, is revised when a new piece of information is received.

We first need to recall the possibilistic expression of conditioning underlying belief revision. In qualitative possibility theory (Dubois and Prade 1998), a conditional possibility measure $\Pi(B|A)$ is defined, for $A \neq \emptyset$, by means of equation

$$\Pi(A \cap B) = \min(\Pi(B|A), \Pi(A)). \tag{5}$$

In the AGM tradition, no interpretation is impossible, that is, $N(A) = 1$ only if $A = W$, or equivalently $\pi(w) > 0, \forall w \in W$. The quantitative version would use the product instead of min, but here we prefer a qualitative setting in agreement with the modeling assumptions in this paper. Applying the minimum specificity principle which maximizes the possibility degrees given the constraints (for avoiding arbitrary restrictions of the possible states), we get the possibility distribution $\pi(\cdot|A)$ associated with the possibility measure $\Pi(\cdot|A)$:

$$\pi(w|A) = \left\{ \begin{array}{ll} 1 & \text{if } \pi(w) = \Pi(A) \text{ and } w \in A \\ \pi(w) & \text{if } \pi(w) < \Pi(A) \text{ and } w \in A \\ 0 & \text{if } w \in \overline{A}. \end{array} \right\}$$

Conditioning by proposition A acknowledges the fact that according to the input of the new piece of belief A, states where A is false have become impossible.

In the possibilistic setting, the result of *revising* the belief base $\mathcal{B}$ by input formula φ, is expressed at the semantic level by conditioning the possibility distribution $\pi_{\mathcal{B}}$ associated with the belief base $\mathcal{B}$ (recalled in Sect. 4.4; Dubois and Prade 1992):

$$\pi_{\mathcal{B}^*_\varphi}(w) = \pi_{\mathcal{B}}(w|A), \text{with } A = ||\varphi||.$$

The revised base is any belief base $\mathcal{B}^*_\varphi$ whose associated possibility distribution is $\pi_{\mathcal{B}^*_\varphi}$. A natural choice of $\mathcal{B}^*_\varphi$, most respectful of the syntax of $\mathcal{B}$, is of the form (Benferhat et al. 2002b)

$$\mathcal{B}^*_\varphi = \{(\varphi_i, \alpha_i) \in \mathcal{B} \text{ s.t. } \alpha_i > \lambda\} \cup \{(\varphi, 1)\},$$

where $\lambda = inc(\mathcal{B} \cup \{(\varphi, 1)\})$ (inconsistency level).

This expression covers the *expansion* $\mathcal{B}^+_\varphi$ of $\mathcal{B}$ by φ as a particular case, letting $A = ||\varphi||$:

$$\pi_{\mathcal{B}^+_\varphi}(w) = \min(\pi(w), \mathbf{1}_A(w))$$

provided that the consistency condition $core(\pi) \cap A \neq \emptyset$ holds, where $core(\pi) = \{w \mid \pi(w) = 1\}$. Syntactically, $\mathcal{B}^+_\varphi = \mathcal{B} \cup \{(\varphi, 1)\}$.

Besides, the *contraction* $\mathcal{B}^-_\varphi$ of $\mathcal{B}$ by φ is semantically expressed by Dubois and Prade (1992):

$$\pi_{\mathcal{B}^-_\varphi}(w) = \left\{ \begin{array}{ll} 1 & \text{if } \pi(w) = \Pi(\overline{A}) \text{ and } w \notin A \\ \pi(w) & \text{otherwise} \end{array} \right\}.$$

which ensures $\neg\varphi$, with sets of models $\overline{A}$, becomes fully possible. Note that in particular, if $\Pi(A) = \Pi(\overline{A}) = 1$ (which means that we are ignorant about whether φ is true or false), we have $\pi_{\mathcal{B}^-_\varphi}(w) = \pi(w)$. This is the case as soon as $\Pi(\overline{A}) = 1$. Note that we retrieve Harper's identity as $\pi_{\mathcal{B}^-_\varphi} = \max(\pi_{\mathcal{B}}, \pi_{\mathcal{B}^*_{\neg\varphi}})$ (Dubois and Prade 1992).

5.3 Desire Revision, Expansion and Contraction

The conditioning of a guaranteed possibility measure Δ contrasts with the conditioning of possibility distributions. It obeys the equation (Benferhat et al. 2002):

$$\Delta(A \cap B) = \max(\Delta(B|A), \Delta(A)). \tag{6}$$

This equation is easily obtained from (5) using duality, i.e., $\Delta(A) = N(\overline{A}) = 1 - \Pi(A)$ where the possibility distribution π underlying N and Π is taken as $1 - \delta$.

Now applying the *maximum* specificity principle, we get the smallest (i.e., corresponding to the least committed conditional desires) possibility distribution $\delta(w|A)$ obeying (6):

$$\delta(w|A) = \left\{ \begin{array}{ll} 0 & \text{if } \delta(w) = \Delta(A) \text{ and } w \in A \\ \delta(w) & \text{if } \delta(w) > \Delta(A) \text{ and } w \in A \\ 1 & \text{if } w \in \overline{A} \end{array} \right\}.$$

As can be seen, what is no longer reachable (conditioning by A means that, for some reason, the possible states are restricted to be those where A is true) is fully desirable by default ($\Delta(\overline{A}|A) = 1$), while A is no longer desired ($\Delta(A|A) = 0$) because the conditioning set A assumes that the current world lies in A; besides, desired worlds in A with strength strictly above $\Delta(A)$ are preserved. So, conditioning means that for some reason, the possible states are restricted to be those in A thus limiting the set of possible worlds, but clearly it does not mean acquiring a new desire. While the *revision* of a set of beliefs $\mathcal{B}$ by a formula φ exactly corresponds to the conditioning of $\pi_{\mathcal{B}}$ by $A = ||\varphi||$, this is no longer the case with respect to desirability functions, for the revision of a set of desires by a new one.

Let δ be a desirability function. While, in the possibility setting, a belief input A is taken to mean that all the elements in $\overline{A}$ should be impossible, a desire input A means that all states in A should be desirable after revision, in agreement with our logical framework for desires, i.e., $\Delta^*_A(A) = \min_{w \in A} \delta^*_A(w) = 1$. Since conditioning is such that $\Delta(A|A) = 0$ and $\Delta(\overline{A}|A) = 1$, it does not fit with this requirement. As the agent now desires such states in A, it should enforce $\Delta^*_A(\overline{A}) = 0$.

Due to this change of focus from $\overline{A}$ to A, when moving from beliefs to desires, desire revision is better expressed by:

$$\delta^*_A(w) = \delta(w|\overline{A})$$

This leads to

$$\delta^*_A(w) = \left\{ \begin{array}{ll} 0 & \text{if } \delta(w) = \Delta(\overline{A}) \text{ and } w \in \overline{A} \\ \delta(w) & \text{if } \delta(w) > \Delta(\overline{A}) \text{ and } w \in \overline{A} \\ 1 & \text{if } w \in A \end{array} \right\}.$$

As can be seen we have $\Delta^*_A(\overline{A}) = 0$ and $\Delta^*_A(A) = 1$.

The condition $\Delta^*_A(A) = 1$ may be considered as too strong an expression of the *success postulate* when revising the desirability function δ by the new desire A.

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 a mild constraint such as $\Delta_A^*(A) \geq \alpha$ (rather than $\Delta_A^*(A) = 1$). This leads to a variant of the revision operator:

$$\delta^*_{(A,\alpha)}(w) = \left\{ \begin{array}{ll} 0 & \text{if } \delta(w) = \Delta(\overline{A}) \text{ and } w \in \overline{A} \\ \delta(w) & \text{if } \delta(w) > \Delta(\overline{A}) \text{ and } w \in \overline{A} \\ \alpha & \text{if } w \in A \text{ and } \delta(w) < \alpha \\ \delta(w) & \text{if } w \in A \text{ and } \delta(w) \geq \alpha \end{array} \right\}.$$

It can be checked that we now have $\Delta_A^*(A) = \alpha$.

We may also think of weakening the success postulate into $\Delta_A^*(A) > 0$. 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 (1993); see Benferhat et al. (2001). When using a finite scale, we have just to take α as the smallest non-zero value in the scale.

The *expansion* of a set of desires $\mathcal{D}$ by φ amounts to cumulating desire φ with the desires in $\mathcal{D}$, providing that the result is not the desire of everything to some extent (due to the postulate $\Delta(\top) = 0$). Thus, we have at the semantic level, with input $A = ||\varphi||$, with characteristic function $\mathbf{1}_A$:

$$\delta_A^+(w) = \max(\delta_{\mathcal{D}}(w), \mathbf{1}_A(w))$$

provided that $support(\delta) \cup A \neq W$, where $support(\delta) = \{w | \delta_{\mathcal{D}}(w) > 0\}$. Note that $\delta_A^* = \delta_A^+$ in this case.

The *contraction* of $\mathcal{D}$ by φ amounts to no longer desire φ at all after contraction. Thus, we have at the semantic level, with input $A = ||\varphi||$:

$$\delta_A^-(w) = \left\{ \begin{array}{ll} 0 & \text{if } \delta(w) = \Delta(A) \text{ and } w \in A \\ \delta(w) & \text{otherwise} \end{array} \right\}.$$

In particular, we have $\delta_A^-(w) = \delta(w), \forall w$ as soon as $\Delta(A) = 0$.

Let us illustrate the approach by an example.

Example 2 Let $\mathcal{D} = \{[\varphi \wedge \psi, \alpha], [v, \beta]\}$ be a desire base where $\alpha > \beta$, where φ, ψ, v are literals. Applying Eq. (4), we get its semantical counterpart under the form of a desirability function δ:

$$\begin{aligned} \delta(\varphi\psi v) = \delta(\varphi\psi\neg v) &= \alpha; \\ \delta(\varphi\neg\psi v) = \delta(\neg\varphi\psi v) = \delta(\neg\varphi\neg\psi v) &= \beta; \\ \delta(\varphi\neg\psi\neg v) = \delta(\neg\varphi\psi\neg v) = \delta(\neg\varphi\neg\psi\neg v) &= 0. \end{aligned}$$

Clearly, $una(\mathcal{D}) = 0$. Now, assume we want to add desire $[\neg\varphi, 1]$. Let us compute $\delta^*_{\neg\varphi}$. We get:

$$\delta^*_{\neg\varphi}(\neg\varphi\psi v) = \delta^*_{\neg\varphi}(\neg\varphi\neg\psi v) = \delta^*_{\neg\varphi}(\neg\varphi\psi\neg v) = \delta^*_{\neg\varphi}(\neg\varphi\neg\psi\neg v) = 1 (\text{models of } \neg\varphi)$$
$$\delta^*_{\neg\varphi}(\varphi\psi v) = \delta^*_{\neg\varphi}(\varphi\psi\neg v) = \alpha; \delta^*_{\neg\varphi}(\varphi\neg\psi v) = \beta;$$
$$\delta^*_{\neg\varphi}(\varphi\neg\psi\neg v) = 0, \text{ which remain unchanged.}$$

Observe that $una(\mathcal{D} \cup \{[\neg\varphi, 1]\}) = 0$, which means that after addition of the new desire, the set of desires remains acceptable. In fact, we have just performed an expansion here so $\delta^*_{\neg\varphi}$ is the desirability function for $\{[\varphi \wedge \psi, \alpha], [v, \beta], [\neg\varphi, 1]\}$.

Now suppose we only add the desire $[\neg\varphi, \gamma]$. Then the modified part of δ would be now

$$\delta^*_{\neg\varphi}(\neg\varphi\psi v) = \delta^*_{\neg\varphi}(\neg\varphi\neg\psi v) = \max(\beta, \gamma), \delta^*_{\neg\varphi}(\neg\varphi\psi\neg v) = \delta^*_{\neg\varphi}(\neg\varphi\neg\psi\neg v) = \gamma.$$

Suppose now the new desire is $[\neg v, \epsilon]$ and the new desire base $\{[\varphi \wedge \psi, \alpha], [v, \beta], [\neg v, \epsilon]\}$.

The corresponding desirability function for this expansion is $\delta^+_{[\neg v,\epsilon]}$ such that

$$\begin{aligned}
\delta^+_{[\neg v,\epsilon]}(\varphi\psi v) &= \alpha; \\
\delta^+_{[\neg v,\epsilon]}(\varphi\neg\psi v) &= \delta^+_{[\neg v,\epsilon]}(\neg\varphi\psi v) = \delta^+_{[\neg v,\epsilon]}(\neg\varphi\neg\psi v) = \beta; \\
\delta^+_{[\neg v,\epsilon]}(\varphi\neg\psi\neg v) &= \delta^+_{[\neg v,\epsilon]}(\neg\varphi\psi\neg v) = \delta^+_{[\neg v,\epsilon]}(\neg\varphi\neg\psi\neg v) = \epsilon \\
\text{and } \delta^+_{[\neg v,\epsilon]}(\varphi\psi\neg v) &= \max(\alpha, \epsilon).
\end{aligned}$$

Thus $una(\mathcal{D} \cup \{[\neg v, \epsilon]\}) = \min(\alpha, \beta, \epsilon) = \min(\beta, \epsilon) = \beta$ assuming $\epsilon > \beta$ (the new desire is stronger than some of the other desires in $\mathcal{D}$).

We must perform a revision. The result $\delta^*_{[\neg v,\epsilon]}$ of the revision is such that the strength β of those least desired interpretations is decreased to 0:

$$\delta^*_{[\neg v,\epsilon]}(\varphi\neg\psi v) = \delta^*_{[\neg v,\epsilon]}(\neg\varphi\psi v) = \delta^*_{[\neg v,\epsilon]}(\neg\varphi\neg\psi v) = 0$$

while for the other interpretations, we keep $\delta^*_{[\neg v,\epsilon]}(w) = \delta^+_{[\neg v,\epsilon]}(w)$. □

It is easy to write for desire revision the counterpart of the axioms for belief revision presented in the previous section in terms of sets of possible worlds. Namely

- [(Δ*1)] For any input A, δ^*_A represents a hedonic state.
- [(Δ*2)] $\Delta^*_A(A) = 1$. This a (strong) priority to the new desire.
- [(Δ*3)] δ^+_A is not more specific than δ^*_A
- [(Δ*4)] If $\Delta(\overline{A}) = 0$ then $\delta^*_A \geq \delta^+_A$
- [(Δ*5)] $\delta^*_A = 1$ if and only if $A = W$
- [(Δ*7)] $\delta^*_{A\cup B} \leq (\delta^*_A)^+_B$
- [(Δ*8)] If $\Delta^*_A(\overline{B}) = 0$ then $\delta^*_{A\cup B} \geq (\delta^*_A)^+_B$

(Δ*2) may be weakened into $\Delta^*_A(A) > 0$. The above properties ($\Delta^*1 - \Delta$*8) extend ($D^*1 - D^*8$) to the revision of a hedonic distribution δ, but fail to uniquely characterize it. However, it is routine to check that the conditioning-based revision

operation δ_A^* in this section obeys properties ($\Delta^*1 - \Delta^*8$). These axioms are the exact counterpart of the axioms for gradual belief revision (Dubois and Prade 1992). It could be checked that they result from a simple exchange with the latter, under a transformation corresponding to the formal identity $\Delta_\delta(A) = N_{1-\delta}(\overline{A})$, where Δ_δ (resp. $N_{1-\delta}$) is the guaranteed possibility (resp. necessity) measure defined from the distribution δ (resp. $1 - \delta$). However they do not characterize a unique possibilistic revision operation for desires, as it is the case with the possibilistic revision operation for belief (Dubois and Prade 1992).

Analogously to the belief revision case, a syntactic counterpart of desire revision can be performed. It can be checked that only the desires strictly above the level of unacceptability of the expanded desire base with the new input are preserved:

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

the others being lost.

Example 3 Let $\mathcal{D} = \{[\varphi, \alpha], [\nu, \beta]\}$ with $\alpha > \beta$, where $una(\mathcal{D}) = 0$ obviously.

Now, let us add desire $[\neg\varphi, 1]$. We have $una(\mathcal{D} \cup \{[\neg\varphi, 1]\}) = \alpha$ and then $\mathcal{D}_{\neg\varphi}^* = \{[\neg\varphi, 1]\}$.

Consider now $\mathcal{D}' = \{[\varphi, \beta], [\nu, \alpha]\}$ (always with $\alpha > \beta$). Then, observe that $una(\mathcal{D}' \cup \{[\neg\varphi, 1]\}) = \beta$, and $\mathcal{D}'^*_{\neg\varphi} = \{[\nu, \alpha], [\neg\varphi, 1]\}$. □

Similarly, for the desire base $\mathcal{D}$ used in Example 2, it can be checked that $D_{\neg\varphi}^* = \mathcal{D}_{\neg\varphi}^+$, and the expansion comes down to adding $[\varphi, 1]$ to the base. Moreover, the revision by $[\neg\nu, \epsilon]$ did yield $\mathcal{D}_{[\neg\nu,\epsilon]}^* = \{[\varphi, \alpha], [\neg\nu, \epsilon]\}$ assuming $\epsilon > \beta$. This revision enforces $una(\mathcal{D}_{[\neg\nu,\epsilon]}^*) = 0$.

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. This assumption is similar to the one underlying epistemic entrenchments in the AGM theory. In any case, if such an impossibility exists for some of them, this has to be taken into account by means of integrity constraints, for instance excluding rejected interpretations corresponding to negative desires.

6 Conclusion

The paper has presented a logical approach to the modeling of hedonic states and to the revision of desires. It is worth noticing that this topic has been little considered in the literature before. Although the 'D' of 'BDI agents' refers to desires, there have been very few works until now aiming at modeling desires (independently of goals) in a formal way (see Casali et al. 2011 for a noticeable exception). Desiring a state of affairs is potential, i.e., the fact of longing for a better situation regardless of its actual feasibility, thus distinguishing desires from goals to actually pursue. 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 achieving goal φ and achieving goal ψ should

be the same as achieving both goals, i.e., $\varphi \wedge \psi$). The revision of a set of prioritized goals should be based on a volitive entrenchment, formally similar to an epistemic entrenchment.

The specific nature of desires with respect to beliefs has also been advocated and emphasized. Roughly speaking, desires behave in a reverse way compared to beliefs. This is reflected in the different series of axioms characterizing desirability functions, inference from a set of desires, and also axioms for desire revision that have been proposed in this paper. Several directions remain to investigate, such as studying iterated desire revision. Another issue would be to allow, as outlined in Casali et al. (2011), for a separate handling of desires asserted positively, and desires asserted negatively (as expressing rejections) by means of two distinct functions, one expressing desirability and the other undesirability (rather than using a unique function as in our approach). Indeed, possibility theory is equipped with proper formal tools to handle positive and negative aspects separately, but consistently as outlined in (Dubois and Prade 2009b). It would mean adding, to a desire base, a base of (more or less) compulsory requirements meant to avoid fearful situations, and reasoning from these two sets of information items.

Besides, it is known that belief revision and nonmonotonic reasoning are two sides of the same coin (Gärdenfors 1990; Rott 2001). This remains to be checked for nonmonotonic desires (Dubois et al. 2014) and desires revision. Actually, we plan to extend the static modal logic of belief and desire we proposed in Dubois et al. (2013) 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 Lang et al. (2003).

Acknowledgements The authors are grateful to the referees for a careful reading and insightful remarks, that led us to clarify a number of issues in the paper, in particular, the shaping of the proof of the completeness Theorem 1.

References

Anscombe, G. E. M. (1957). *Intention*. Oxford: Basil Blackwell.

Banerjee, M., & Dubois, D. (2014). A simple logic for reasoning about incomplete knowledge. *International Journal of Approximate Reasoning*, *55*, 639–653.

Benferhat, S., Dubois, D., Kaci, S., & Prade, H. (2002). Bipolar possibilistic representations. In A. Darwiche & N. Friedman (Eds.), *Proceedings of the 18th conference in uncertainty in artificial intelligence (UAI '02)* (pp. 45–52). Edmonton, Alberta: Morgan Kaufmann.

Benferhat, S., Dubois, D., Kaci, S., & Prade, H. (2006). Bipolar possibility theory in preference modeling: Representation, fusion and optimal solutions. *Information Fusion*, *7*, 135–150.

Benferhat, S., Dubois, D., & Prade, H. (1999). Towards a possibilistic logic handling of preferences. In *Proceedings of the 16th conference on artificial intelligence (IJCAI 99)* (pp. 1370–1375). Stockholm: Morgan Kaufmann.

Benferhat, S., Dubois, D., & Prade, H. (2001). A computational model for belief change and fusing ordered belief bases. In M.-A. Williams & H. Rott (Eds.), *Frontiers in belief revision* (pp. 109–134). Dordrecht: Kluwer Academic Publishers.

Benferhat, S., Dubois, D., Prade, H., & Williams, M.-A. (2002). A practical approach to revising prioritized knowledge bases. *Studia Logica*, *70*, 105–130.

Benferhat, S., & Kaci, S. (2003). Logical representation and fusion of prioritized information based on guaranteed possibility measures: Application to the distance-based merging of classical bases. *Artificial Intelligence*, *148*(1–2), 291–333.

Bonanno, G. (2009). Rational choice and AGM belief revision. *Artificial Intelligence*, *173*(12–13), 1194–1203.

Boutilier, C. (1993). Revision sequences and nested conditionals. In *Proceedings of the 13th international joint conference on artificial intelligence (IJCAI'93)* (pp. 519–525). Chambéry: Morgan Kaufmann.

Casali, A., Godo, L., & Sierra, C. (2011). A graded BDI agent model to represent and reason about preferences. *Artificial Intelligence*, *175*, 1468–1478.

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

Doyle, J., Shoham, Y., & Wellman, M. P. (1991). A logic of relative desire (preliminary report). In Z. Ras & M. Zemankova (Eds.), *Methodologies for intelligent systems (ISMIS 1991), lecture notes in computer science* (Vol. 542, pp. 16–31). New York: Springer.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge: MIT Press.

Dubois, D. (1986). Belief structures, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence (Bratislava)*, *5*(6), 403–416.

Dubois, D., Hajek, P., & Prade, H. (2000). Knowledge-driven versus data-driven logics. *Journal of logic, Language and information*, *9*, 65–89.

Dubois, D., Lorini, E., & Prade, H. (2013). Bipolar possibility theory as a basis for a logic of desires and beliefs. In W. Liu, V. S. Subrahmanian, & J. Wijsen (Eds.), *Proceedings of the 7th international conference scalable uncert. Mgmt. (SUM'13), LNCS 8078*. Washington, DC: Springer.

Dubois, D., Lorini, E., & Prade, H. (2014). Nonmonotonic desires–A possibility theory viewpoint. In R. Booth, G. Casini, S. Klarman, G. Richard, & I. J. Varzinczak (Eds.), *Proceedings of the international workshop on defeasible and ampliative reasoning (DARe@ECAI 2014)* (Vol. 1212). Prague: CEUR Workshop Proceedings.

Dubois, D., Lorini, E., & Prade, H. (2015). Revising desires–A possibility theory viewpoint. In T. Andreasen, H. Christiansen, J. Kacprzyk, H. Larsen, G. Pasi, O. Pivert, G. De Tré, M. A. Vila, A. Yazici, & S. Zadrożny (Eds.), *Proceedings of the 11th international conference on flexible query answering systems (FQAS'15)* (Vol. 400, pp. 3–13). Advances in Intelligent Systems and Computing series.

Dubois, D., Lorini, E., & Prade, H. (2016). A possibility theory-based approach to desire change. In R. Booth, G. Casini, S. Klarman, G. Richard, & I. J. Varzinczak (Eds.), *Proceedings of the international workshop on defeasible and ampliative reasoning (DARe@ECAI 2016)* (Vol. 1626). The Hague: CEUR Workshop Proceedings.

Dubois, D., & Prade, H. (1991). Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, *50*, 223–239.

Dubois, D., & Prade, H. (1992). Belief change and possibility theory. In P. Gärdenfors (Ed.), *Belief revision* (pp. 142–182). Cambridge: Cambridge University Press.

Dubois, D., & Prade, H. (1998). Possibility theory: Qualitative and quantitative aspects. In D. Gabbay & P. Smets (Eds.), *Quantified representation of uncertainty and imprecision, handbook of defeasible reasoning and uncertainty management systems* (Vol. 1, pp. 169–226). Dordrecht: Kluwer.

Dubois, D., & Prade, H. (2004). Possibilistic logic: A retrospective and prospective view. *Fuzzy Sets and Systems*, *144*, 3–23.

Dubois, D., & Prade, H. (2009a). Accepted beliefs, revision and bipolarity in the possibilistic framework. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief* (pp. 161–184). New York: Springer.

Dubois, D., & Prade, H. (2009b). An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets and Systems*, *160*(10), 1355–1366.

Dubois, D., & Prade, H. (2012). Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets and Systems*, *192*, 3–24.

Gärdenfors, P. (1988). *Knowledge in flux. Modeling the dynamics of epistemic states*. Cambridge: The MIT Press.

Gärdenfors, P. (1990). Belief revision and nonmonotonic logic: Two sides of the same coin? In *Proceedings of the 9th European conference on artificial intelligence (ECAI'90)* (pp. 768–773). Stockholm.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, *17*, 157–170.

Harsanyi, J. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*, 309–321.

Harsanyi, J. (1982). Morality and the theory of rational behaviour. In A. K. Sen & B. Williams (Eds.), *Utilitarianism and beyond*. Cambridge: Cambridge University Press.

Humberstone, I. L. (1992). Direction of fit. *Mind*, *101*(401), 59–83.

Hume, D. (1978). *A treatise of human nature* (2nd Oxford edn.). L. A. Selby-Bigge & P. H. Nidditch (Eds.), Oxford: Oxford University Press.

Lang, J., & van der Torre, L. (1996). Conditional desires and utilities: An alternative logical approach to qualitative decision theory. In W. Wahlster (Ed.), *Proceedings of the 12th European conference artificial intelligence (ECAI'96)* (pp. 318–322). Budapest: Wiley .

Lang, J., & van der Torre, L. (2008). From belief change to preference change. In M. Ghallab, C. D. Spyropoulos, N. Fakotakis, & N. M. Avouris (Eds.), *Proceedings of the 18th European conference on artificial intelligence (ECAI'08)* (pp. 351–355). Patras: IOS Press.

Lang, J., van der Torre, L., & Weydert, E. (2002). Utilitarian desires. *Journal of Autonomous Agents and Multi-Agent Systems*, *5*, 329–363.

Lang, J., van der Torre, L., & Weydert, E. (2003). Hidden uncertainty in the logical representation of desires. In G. Gottlob & T. Walsh (Eds.), *Proceedings of the 18th international joint conference on artificial intelligence (IJCAI'03)* (pp. 685–690). Acapulco: Morgan Kaufmann.

Lewis, D. (1973). Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, *2*(4), 418–446.

Locke, J. (1975). *An essay concerning human understanding*. Oxford: Oxford University Press. The Clarendon Edition of the Works of John Locke.

Lorini, E. (2014). A logic for reasoning about moral agents. *Logique et Analyse, Centre National de Recherches en Logique (Belgium)*, *58*(230), 177–218 .

Platts, M. (1979). *Ways of meaning*. London: Routledge and Kegan Paul.

Rao, A. S., & Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. In *Proceedings of the 2nd international conference on principles of knowledge representation and reasoning* (pp. 473–484).

Rott, H. (2001). *Change, choice and inference. A study of belief revision and nonmonotonic reasoning*. Oxford: Clarendon Press.

Schroeder, T. (2004). *Three faces of desires*. Oxford: Oxford University Press.

Searle, J. (1979). *Expression and meaning*. Cambridge: Cambridge University Press.

Searle, J. (2001). *Rationality in action*. Cambridge: MIT Press.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

van Benthem, J., Girard, P., & Roy, O. (2009). Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic*, *38*, 83–125.

van Benthem, J., & Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, *17*(2), 157–182.

Von Wright, G. H. (1963). *The logic of preference*. Edinburgh: Edinburgh University Press.

Von Wright, G. H. (1972). The logic of preference reconsidered. *Theory and Decision*, *3*, 140–169.

Zadeh, L. A. (1978). PRUF: A meaning representation language for natural languages. *International Journa of Man-Machine Studies*, *10*, 395–460.