Comparing Strengths of Beliefs Explicitly

SUJATA GHOSH, *Indian Statistical Institute, Chennai.*
Email: sujata@isichennai.res.in

DICK DE JONGH, *Institute for Logic, Language and Computation,*
University of Amsterdam. E-mail: D.H.J.deJongh@uva.nl

Abstract

Inspired by a similar use in provability logic, formulas $p \succ_B q$ and $p \succcurlyeq_B q$ are introduced in the existing logical framework for discussing beliefs to express that the strength of belief in p is greater than (or equal to) that in q. Besides its usefulness in studying the properties of the concept of greater strength of belief itself this explicit mention of the comparison in the logical language aids in defining several other concepts in a uniform way, namely, older and rather clear concepts like the operators for universality (the totality of possibilities considered by an agent), together with newer notions like plausibility (in the sense of 'more plausible than not') and disbelief. Relative expressive powers of the proposed logics are also discussed. A major role is played in our investigations by the relationship between the standard plausibility ordering of the worlds and the strength of belief ordering. If we try to define the strength of belief ordering in terms of the world plausibility ordering we get some undesirable consequences, so we have decided to keep the relation between the two orderings as light as possible to construct a system that allows for widely different interpretations. In fact, we start with considering these orderings to be independent of each other and towards the end we provide a discussion on their possible inter-relationship. Finally, we provide an extension of the framework to the multi-agent setting, and we discuss the possibilities of extending our system to a dynamic one.

Keywords: doxastic logic, belief, disbelief, plausibility

1 Introduction

Being subject to doubts and dilemmas while making decisions is like second nature to the human mind. The difference in the strengths of beliefs of an agent regarding the occurrence of different events may clear doubts of this kind. In betting on games, people make their choices for putting their money on different teams, based on their strengths of beliefs about which team will win. Similarly, when voting, one's preference for the candidates is again based on the strength of beliefs about one candidate's ability to perform compared to the others. Thus, this notion is inherently present in various fields of research like decision theory, game theory and others.

Before proceeding further, let us first consider the following real life situation where comparison of strength of beliefs plays a key role in decision-making for recruitments.

Alice often has applications for jobs in her departmental store. The first time Burt and Cora apply. Alice believes both can do the job, but her belief in Cora being able to do it is stronger than that Burt will be able to do it. She chooses Cora.

The second time Deirdre and Egon apply. She believes that Egon can do the job whereas she is ambiguous about Deirdre: she neither has the belief that Deirdre can do it, nor that she cannot. She chooses Egon.

The third time Fiona and Gregory apply. About both she is ambiguous, but her strength of belief in Gregory being able to do it is stronger than that in Fiona. She chooses Gregory, maybe she has to help him along a little.

The fourth time the applicants are Harold and Irma. She believes neither can do the job. She decides not to take one of those two and hold another round of applications. When she finds out that time is too short for that she thinks again, decides that she believes even less in Irma being able to do it than in Harold, and she takes Harold.

Let us point out one possible misunderstanding. Alice does not judge how well she thinks the applicants will perform, she just judges whether they will be able to do the job or not, a simple yes-no question. Of course, to combine our set up with beliefs in graded abilities would be highly interesting but that is a matter for future work.

All the discussed situations regarding the belief states of Alice can be aptly described, if we talk not only about her beliefs but also compare the strength of her beliefs in the applicants. One sees here how a stronger belief can induce a preference.

One can argue that these situations can be described by the well-studied notion of *preference*, but the essence of describing the mental states of Alice will be lost then. This paper adds a new notion to this line of work, viz. comparing the strengths of beliefs, and very pertinently, doing this in a qualitative manner. The relationship with preference will be developed somewhat further in Section 3.3.

The introduction of explicit notions of ordering for comparing strengths of beliefs in the logical language has various applications. Besides its usefulness in studying the properties of the concept of greater strength of belief itself it aids in defining several other concepts in a uniform way. In models concerning knowledge (epistemic logic) equivalence classes of worlds (and in case of one agent: the set of all worlds) are naturally given by the indistinguishability relation connected with the knowledge operator. In models concerning belief (doxastic logic) a universality operator U is often introduced with the same purpose. In our set up this usually somewhat vague operator can be defined in terms of the order, the idea that the worlds the agent considers are the ones she considers possible in some manner is made explicit. In the semantics, the question - which worlds are going to be a part of the model, gets in our approach a clearer formal and intuitive understanding. It also becomes even more evident that the universality operator must not be identified with the knowledge operator even if they both share the $S5$-properties.

Additionally, newer notions like plausibility (in the sense of 'more plausible than not') and disbelief can also be expressed. Above all it has its advantages in an explicit study of the properties of the orderings themselves, semantically and axiomatically. All these investigations can be carried over to a dynamic setting ([12], see also [3]), but we leave this for future work.

Motivated by the ideas of provability logic [1], formulas $\varphi \succ_B \psi$ and $\varphi \succcurlyeq_B \psi$ are introduced in the existing logical framework for discussing beliefs to express that the strength of belief in φ is greater than (or equal to) that in ψ. We should note here that in the Rosser framework [17, 23] proofs of φ and ψ are compared only if at least one of these proofs really exists, whereas strengths of beliefs are also discussed when neither φ nor ψ are really believed. This makes them less concrete, and therefore we express their comparison as $\varphi \succ_B \psi$, rather than $B\varphi \succ B\psi$. As mentioned earlier, these formulas can be used to express notions like 'disbelief' (the inclination to believe in $\neg\varphi$ is greater than the inclination to believe in φ), and its dual 'more plausible than not', which can be represented by $\neg\varphi \succ_B \varphi$ and $\varphi \succ_B \neg\varphi$, respectively.

1.1 Related work

In [29, 10], orderings of formulas are considered but their interpretations are probabilistic in nature. A binary sentential operator is introduced in the language with the intended interpretation 'at least as probable as'. While [10] takes the explicit ordering operator in a simple language consisting of the truth-functional connectives only, [29] discusses this issue in a modal setting. As the interpretation suggests, the semantics is based on probability measures over worlds.

The notion of epistemic entrenchment [11] gives a syntactic ordering of formulas, which is studied in connection with belief revision. The ordering influences the abandoning and retaining of formulas when a belief contraction or revision takes place. Whereas this ordering of the formulas is on a meta-level, our goal in this work is to propose an object-level ordering of formulas.

Ordering of worlds provide an intuitive way to model various kinds of logical operators, specially the epistemic ones. To mention a few, Lewis [26] proposed a plausibility ordering of worlds to provide a semantics for the counterfactual statements. With the goal of representing qualitative frameworks of belief in terms of the corresponding probabilistic ones, Spohn defines a plausibility ordering of possible worlds in terms of ordinal functions [30]. More recently, such orderings have been discussed in the economics literature [4].

Our basic focus lies on giving a qualitative framework for differing strengths of beliefs that an agent may have on different propositions (possibly, individuals), which influence her decision making process. Semantically, rather than modeling in terms of *world ordering*, we rely on *set ordering* for comparing belief strengths.

We should mention here that the idea of modeling epistemic notions in terms of set orders is not really new. In [18], preferential structures are considered to give semantics to a logic of relative likelihood. A preference ordering over worlds is lifted to an ordering of sets of worlds. Plausibility measures over sets of worlds are considered in [9] to give a semantics of default logic. These measures induce a set ordering which provides an interpretation of the notion of belief (similar to our notion of plausibility in Section 3.1). For a detailed overview, see [19].

While our interpretation of belief is given in terms of world ordering, a set ordering is used to interpret strength of belief. This distinguishes our work from the ones mentioned above. Moreover, this ordering of sets of worlds is only partly determined by the ordering of the worlds. We discuss our reasons for this in later sections, especially in Section 7.

1.2 Overview of the paper

With this background, we now provide a brief summary regarding the structure of this paper. Explicit belief-ordering over formulas is introduced in Section 2, giving a complete axiomatization of this belief logic with explicit ordering ($KD45-O$). Several possible interpretations of the belief-ordered formulas, viz. *plausibility*, *disbelief*, and *preference* are discussed in Section 3. Complete axiomatizations of plausibility logic (P-logic), logic of belief and plausibility (BP-logic) and logic of belief and disbelief (BD-logic) are also provided under their interpretations with respect to belief-ordering. Relative expressive powers of the proposed logics are discusssed in Section 4. In Section 5, the inter-relationship of these ordering formulas and *safe belief* is discussed, with a brief look at the relative expressive powers. Section 6 lifts the whole framework to a multi-agent setting. A discussion on the relationship of the world ordering and the corresponding set ordering is held in Section 7. Some conclusions are drawn in Section 8.

2 Comparing strengths of beliefs explicitly

Modal logic is a useful tool to study knowledge and belief of human agents, which has been a main issue of concern to philosophers as well as computer scientists. Von Wright's work [32] is generally accepted as initiating this line of research, which was further extended by [22]. Subsequently a huge research area has been developed, trying to provide answers to various philosophical issues as well as aiding into the development of several areas of computer science, like distributed systems, security protocols, database theory and others.

Possible-world semantics [25] has been used to model knowledge as well as belief. An extensive discussion together with all pre-requisite definitions can be found in [20]. In this work we are only concerned with *beliefs* of agents, comparison of their strengths as well as some related notions like *universality*, *safe beliefs*, *plausibility*, *disbelief* and others. Various debates and discussions are still going strong among the philosophers regarding the axioms that characterize belief - for this paper we will stick to the *KD45*-model of belief.

In the following, we talk about Kripke structures as well as the plausibility models [3, 2] as and when needed while talking about beliefs. The readers should note that plausibility models are more general in nature in the sense that one can always build up a *KD45* Kripke structure from them, as described in [2].

With this brief overview, we now move on to introduce explicit ordering of beliefs in the logical language, which is the essential new feature of this paper. This explicit mention of such comparison of beliefs provides an informative and uniform way to discuss certain relevant issues like disbelief, plausibility and others.

To introduce this comparison of strengths of beliefs explicitly in the logical language, we add new relation symbols to the existing modal language of belief to form the language of *Belief logic with explicit ordering* (*KD45*$-O$), whose language is defined as follows:

Definition 2.1
Given a countable set of atomic propositions Φ, formulas φ are defined inductively:

$$\varphi := \bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B\varphi \mid \varphi \succcurlyeq_B \varphi \mid$$

where $p \in \Phi$.

The intuitive reading of the formula $B\varphi$ is "φ is believed", and that of $\varphi \succcurlyeq_B \psi$ is "belief in φ is at least as strong as belief in ψ". We introduce the notations $\varphi \succ_B \psi$ for $(\varphi \succcurlyeq_B \psi) \wedge \neg(\psi \succcurlyeq_B \varphi)$ and $\varphi \equiv_B \psi$ for $(\varphi \succcurlyeq_B \psi) \wedge (\psi \succcurlyeq_B \varphi)$. Intuitively, they can be read as “belief in φ is stronger than that in ψ” and “belief in φ and ψ are of same strength”, respectively. We now move on to define a model for this logic.

Definition 2.2
A *KD45−O* model is defined to be a structure $\mathcal{M} = (S, \leq, \geq_B, V)$, where S is a non-empty finite set of states, V is a valuation assigning truth values to atomic propositions in states, $\leq$ is a quasi-linear[1] order relation (a plausibility ordering) over S, and $\geq_B$ is a quasi-linear order relation over $\mathcal{P}(S)$, satisfying the conditions:

1. If $X \subseteq Y$, then $Y \geq_B X$.
2. If $\mathcal{B}$ is the set of all $\leq$-minimal worlds (the set of most-plausible worlds, called the *center*), then $\mathcal{B} \subseteq X$ and $\mathcal{B} \nsubseteq Y$ imply $X >_B Y$, where $X >_B Y$ iff $X \geq_B Y$ and not $(Y \geq_B X)$.
3. If X is non-empty, then $X >_B \emptyset$.

The first condition says that larger sets of worlds are at least as plausible, the second one, which we call the *sufficient belief condition*, that the sets containing the center are more plausible than those not containing it; the third one that non-empty sets are more plausible than the empty set. Truth on the center suffices to make an assertion to be believed. Note that all the models are considered to be finite. This assumption ensures the existence of minimal worlds in terms of the plausibility ordering of the model. The truth definition for formulas φ in a *KD45−O* model $\mathcal{M}$ is as usual with the following clauses for the belief and ordering modalities.

$\mathcal{M}, s \models B\varphi$ iff $\mathcal{M}, t \models \varphi$ for all $\leq$-minimal worlds t.
$\mathcal{M}, s \models \varphi \succcurlyeq_B \psi$ iff $\{t \,|\, \mathcal{M}, t \models \varphi\} \geq_B \{t \,|\, \mathcal{M}, t \models \psi\}$.

We consider $\succcurlyeq_B$ to be a global notion, if $\varphi \succcurlyeq_B \psi$ is true anywhere in the model, it is true everywhere. So, it is either true or false throughout the whole model; $\succcurlyeq_B$ is a global notion like B. Of course, being global in the model is strongly connected with introspection. From the definition of $\succ_B$, it follows that,

$$\mathcal{M}, s \models \varphi \succ_B \psi \text{ iff } \{t \,|\, \mathcal{M}, t \models \varphi\} >_B \{t \,|\, \mathcal{M}, t \models \psi\}.$$

Thus, $\succ_B$ is also a global notion. We will now show that the universal modality U can also be expressed in the language. The modality $E\varphi$ (the abbreviated form of $\neg U \neg \varphi$) can be defined as follows:

$$E\varphi := \varphi \succ_B \bot,$$

Hence $U\varphi$ $(= \neg E \neg \varphi)$ itself can be defined as $\bot \succcurlyeq_B \neg\varphi$: $U\varphi$ expresses that φ is true in all possible worlds in the model, whereas $E\varphi$ stands for existence of a possible world in the model where φ is true. The formula $\varphi \succ_B \bot$, which defines $E\varphi$, expresses the intuition that those worlds should be considered in the model of which the existence is expressed by a positive strength of belief, those possibilities which the agent does not want to exclude. Evidently, we have,

$$\mathcal{M}, s \models U\varphi \text{ iff } \mathcal{M}, t \models \varphi \text{ for all worlds } t.$$

[1] A binary relation $\leq$ on a non-empty set S is said to be quasi-linear if it is reflexive, transitive and linear, i.e. a total pre-order. That we do take the order to be quasi-linear, but not more generally a pre-order is not a matter of principle but rather of convenience.

Alice's belief states (as described in the introduction) can now be formally presented as follows: Suppose each of the applicants' names denotes the proposition that "he (she) can do the job". Cora $\succ_B$ Burt in the first case; B(Egon) $\wedge$ ($\neg$B(Deirdre) $\wedge$ $\neg$B($\neg$Deirdre)) implies that Egon $\succ_B$ Deirdre in the second case, with the third case simply being Gregory $\succ_B$ Fiona again, and the fourth one, B($\neg$Harold)$\wedge$B($\neg$Irma), and later Harold $\succ_B$ Irma. The readers can easily see that in the second case there is some reasoning going on which leads to Egon being given the job, because Alice's belief in the ability of Egon is stronger than her belief in the ability of Deirdre.

2.1 Definability

We have introduced two orderings, a world ordering and a set ordering in the definition of our *KD45−O* model (cf. Definition 2.2). We now show that we indeed need those two orderings separately, that is, one cannot always be defined in terms of the other.

Let us first consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:

$$\begin{array}{ccccc} s_1 & & s_2 & & s_3 \\ \bullet & <^1 & \bullet & <^1 & \bullet \\ & & \mathcal{M}_1 & & \end{array} \qquad \begin{array}{ccccc} s_1 & & s_2 & & s_3 \\ \bullet & <^2 & \bullet & <^2 & \bullet \\ & & \mathcal{M}_2 & & \end{array}$$

In both these models $\{s_1\}$ is the center. The respective set orderings $\geq^1_B$ and $\geq^2_B$ are given as follows:

$\geq^1_B$: $\{s_1, s_2, s_3\} >^1_B \{s_1, s_2\} >^1_B \{s_1, s_3\} >^1_B \{s_1\} >^1_B \{s_2, s_3\} >^1_B \{s_2\} >^1_B \{s_3\} >^1_B \emptyset$

$\geq^2_B$: $\{s_1, s_2, s_3\} >^2_B \{s_1, s_3\} >^2_B \{s_1, s_2\} >^2_B \{s_1\} >^2_B \{s_2, s_3\} >^2_B \{s_2\} >^2_B \{s_3\} >^2_B \emptyset$

We have that the world orderings $\leq^1$ and $\leq^2$ are the same, where as the different set orderings $\geq^1_B$ and $\geq^2_B$ satisfy the conditions of Definition 2.2. Thus we have shown that $\geq_B$ is not definable in terms of $\leq$.

In the same way, consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:

$$\begin{array}{ccccc} s_1 & & s_2 & & s_3 \\ \bullet & <^1 & \bullet & <^1 & \bullet \\ & & \mathcal{M}_1 & & \end{array} \qquad \begin{array}{ccccc} s_1 & & s_3 & & s_2 \\ \bullet & <^2 & \bullet & <^2 & \bullet \\ & & \mathcal{M}_2 & & \end{array}$$

Once again, in both these models $\{s_1\}$ is the center. The respective $\geq^1_B = \geq^2_B$ is given as follows:

$\geq_B$: $\{s_1, s_2, s_3\} >_B \{s_1, s_2\} >_B \{s_1, s_3\} >_B \{s_1\} >_B \{s_2, s_3\} >_B \{s_2\} >_B \{s_3\} >_B \emptyset$

Thus $\geq_B$ satisfies the conditions of Definition 2.2, whereas $\leq^1$ and $\leq^2$ are different from each other. Thus we have shown that $\leq$ is not definable in terms of $\geq_B$.

We should note here that the matter of definability of the plausibility ordering in terms of the set ordering or vice versa is not hereby closed. It is possible, in fact in both directions, to give definitions that by means of one of the orderings we can define an ordering of the other type with the right properties. For example, we will see that defining an ordering of the worlds s and t by $\{s\} \leq_B \{t\}$, under certain conditions gives a plausibility ordering. What we have shown here is that this plausibility ordering is not uniquely determined by the set ordering, no choice of ordering is forced on us in either direction. A more detailed discussion can be found in Section 7.

2.2 Axioms and Completeness

In this subsection we introduce the proof system *KD45−O*, and discuss its relationship to the *KD45−O*-models. The system consists of the following axioms and rules.

Definition 2.3
The system *KD45−O* consists of

a) all *KD45* axioms and rules for B

b) ordering axioms:
$\varphi \succcurlyeq_B \varphi$ (reflexivity)
$(\varphi \succcurlyeq_B \psi) \wedge (\psi \succcurlyeq_B \chi) \rightarrow \varphi \succcurlyeq_B \chi$ (transitivity)
$(\varphi \succcurlyeq_B \psi) \vee (\psi \succ_B \varphi)$ (linearity)
$(B\varphi \wedge \neg B\psi) \rightarrow (\varphi \succ_B \psi)$ (center)
$(\varphi \succcurlyeq_B \psi) \rightarrow B(\varphi \succcurlyeq_B \psi)$ (introspection1)
$(\varphi \succ_B \psi) \rightarrow B(\varphi \succ_B \psi)$ (introspection2)
$\perp \succcurlyeq_B \neg(\varphi \rightarrow \psi) \rightarrow (\psi \succcurlyeq_B \varphi)$ ($U \succcurlyeq_B$-axiom)
$\varphi \rightarrow (\varphi \succ_B \perp)$ (existence)
$(B\varphi \succ_B \perp) \rightarrow B\varphi$ (unique-center)

c) inclusion rule:

$$\frac{\varphi \rightarrow \psi}{\psi \succcurlyeq_B \varphi} \quad \text{(inclusion rule)}$$

Let us discuss these axioms, and, more or less informally, their soundness for the models introduced above. On the way, we will make clear that the $S5$-properties of the universal modality U [16] are all provable in *KD45−O*. We will discuss possible additional axioms in the subsection immediately after this one.

Since, by the set-ordering relation in the *KD45−O* model, $\geq_B$ is a reflexive, transitive and connected relation over $\mathcal{P}(S)$, and $>_B$ is the corresponding strict ordering, the three basic ordering axioms are evidently sound. From these axioms it immediately follows that $>_B$ is a transitive relation and also that $U\psi$ and $\neg E\neg\psi$ are provably equivalent.

The soundness of the inclusion rule follows from the condition (1) that larger sets in the models are at least as plausible as smaller sets. It has two important consequences. The first one is the *equivalence rule*:

$$\frac{\varphi \leftrightarrow \psi}{\varphi \equiv_B \psi}$$

which implies that substituting logically equivalent formulas for each other in ordering formulas leads to logically equivalent formulas. Since the rest is modal logic, this is the only thing needed to show that substituting them for each other anywhere leads to logically equivalent formulas. The second important consequence is the necessitation rule for U (Ugen-rule). It holds because, if $\vdash_{KD45-O} \varphi$, then $\vdash_{KD45-O} \neg\varphi \rightarrow \bot$. The inclusion rule now gives $\vdash_{KD45-O} \bot \succcurlyeq_B \neg\varphi$, i.e., $\vdash_{KD45-O} U\varphi$.

Three axioms, by definition, involve U and/or E. We have the $U \succcurlyeq_B$ -axiom, which can be reformulated as:

$$U(\varphi \rightarrow \psi) \rightarrow (\psi \succcurlyeq_B \varphi),$$

and which is also connected to condition 1 in Definition 2.2. In fact, it expresses that formulas that are equivalent in the model when replacing each other lead to formulas that are equivalent in the model. Moreover, this axiom can be used to prove the K-axiom for U $\big(U(\varphi \rightarrow \psi) \rightarrow (U\varphi \rightarrow U\psi)\big)$. For, assume we have $U\varphi$ and $U(\varphi \rightarrow \psi)$, but not $U\psi$, i.e., $\neg\psi \succ_B \bot$. Then, by the first assumption we have $\bot \succcurlyeq_B \neg\varphi$, and hence $\neg\psi \succ_B \neg\varphi$. By the second assumption, we have also $U(\neg\psi \rightarrow \neg\varphi)$ (equivalence!), and using the $U \succcurlyeq_B$ -axiom, $\neg\varphi \succcurlyeq_B \neg\psi$, a contradiction.

Next we have the existence axiom, which can be reformulated as:

$$\varphi \rightarrow E\varphi$$

The existence axiom is basically the same as the ordered formula for $U\varphi \rightarrow \varphi$, one of the $S5$-axioms for U. The last of the three is the unique-center axiom, which can be reformulated as:

$$EB\varphi \rightarrow B\varphi \qquad \text{(unique-center)}$$

It derives from the fact that the $KD45-O$ models have a unique center $\mathcal{B}$. It makes B a global property: the principle $B\varphi \rightarrow UB\varphi$ readily follows by first proving $E\neg B\varphi \rightarrow \neg B\varphi$.

Since the ordering formulas are either globally true or globally false in the models, we have the soundness of the two introspection axioms:

$$(\varphi \succcurlyeq_B \psi) \rightarrow B(\varphi \succcurlyeq_B \psi)$$
$$(\varphi \succ_B \psi) \rightarrow B(\varphi \succ_B \psi)$$

It immediately follows that:

$$\neg(\varphi \succcurlyeq_B \psi) \rightarrow B\neg(\varphi \succcurlyeq_B \psi)$$
$$\neg(\varphi \succ_B \psi) \rightarrow B\neg(\varphi \succ_B \psi)$$

The converses of all these implications above follow from the linearity axiom. This means that all these ordering statements can be considered to be B-statements, i.e. $\varphi \succcurlyeq_B \psi$, $\varphi \succ_B \psi$, $U\varphi$, $E\varphi$ are all B-statements (and remember that equivalent formulas can be replaced by each other modulo provable equivalence). As a result, the inclusion formula concerning the belief and the universal modality, viz. $U\varphi \rightarrow B\varphi$ follows. And the $U\psi \rightarrow UU\psi$ and $\neg U\psi \rightarrow U\neg U\psi$ axioms for U follow as well; because of the very significant property of $U\psi$ being a B-statement the unique-center axiom applies to U-statements as well. We have now covered all the $S5$-axioms for U.

We are now ready to prove the following completeness theorem, which is the most basic and important result of this work.

Theorem 2.4
$KD45-O$ is sound and complete with respect to $KD45-O$ models.

Proof. Soundness has been treated above. Moreover we will freely use U meaning its translation into $KD45-O$, and we can assume that U has the $S5$-properties. We will show completeness using finite sets of sentences.

Assume $\nvdash_{KD45-O} \varphi$. We will have to construct a countermodel to φ as a $KD45-O$-model. We take a finite *adequate* set Φ containing φ. An adequate set can be defined as follows: a set of formulas that is closed under subformulas containing with each formula ψ (a formula equivalent in propositional logic to) $\neg\psi$, containing with $B\psi$ and $B\chi$ (a formula equivalent to) $B(\psi \wedge \chi)$ and (a formula equivalent to) $B(\psi \vee \chi)$. We also need Φ to contain with each formula $B\varphi$ the formula $UB\varphi$. Finally, Φ contains $B\top$ and $B\bot$. It is easy to see that any finite set is contained in a finite adequate set. We use the Henkin method restricted to Φ. Consider the m.c. (maximally consistent) subsets of Φ. In particular consider such an m.c. set Φ_0 containing $\neg\varphi$.

The relations $\mathcal{R}_B$ and $\mathcal{R}_U$ are defined as follows:

$$\begin{array}{lll} P\mathcal{R}_B Q & \text{iff} & (1)\ \text{for all } B\varphi \text{ in } P,\ \varphi \text{ as well as } B\varphi \text{ are in } Q, \\ & & (2)\ \text{for all } \neg B\varphi \text{ in } P,\ \neg B\varphi \text{ in } Q. \\ P\mathcal{R}_U Q & \text{iff} & (1)\ \text{for all } U\varphi \text{ in } P,\ \varphi \text{ as well as } U\varphi \text{ are in } Q, \\ & & (2)\ \text{for all } \neg U\varphi \text{ in } P,\ \neg U\varphi \text{ in } Q \end{array}$$

We have to show that $\mathcal{R}_U$ is an equivalence relation and $\mathcal{R}_B$ an Euclidean subrelation of $\mathcal{R}_U$. Finally, within one U-equivalence class there is one, nonempty, set of B-reflexive elements, which forms a B-equivalence class. Since all these things are standard we skip this part.

We now take the submodel generated by $\mathcal{R}_U$ from Φ_0. The set of worlds W of our model will be the set of worlds in this submodel and the $\mathcal{R}_B$ and $\mathcal{R}_U$ the restrictions of the original $\mathcal{R}_B$ and $\mathcal{R}_U$ to this submodel. $\mathcal{R}_U$ is now the universal relation.

As before, we write $\mathcal{B}$ for the set of $\mathcal{R}_B$-reflexive elements. The axiom $B\varphi \to UB\varphi$ implies that this set of formulas is unique and a B-equivalence class. The world plausibility ordering is given as follows: any world in $\mathcal{B}$ is more plausible than any in $W \setminus \mathcal{B}$, and within these two sets, the worlds are equi-plausible. So, with respect to the modal operators B and U the model behaves properly, and we have a proper world-ordering as well. We will now have to order $\mathcal{P}(W)$ in a proper way.

Let us say that ψ *represents* subset X of W if X is the set of nodes where ψ is true, which we may write as $V(\psi) = X$. We say that X is *representable* if for some $B\psi$ in Φ, ψ represents X. By the conditions on Φ the representable sets are closed under unions and intersections, and contain W itself and the empty set.

The representable subsets of Φ are quasi-linearly ordered by the relation $\geq_1$ defined by $V(\psi) \geq_1 V(\chi)$ iff $\psi \succcurlyeq_B \chi$ is true in the model, $V(\psi) >_1 V(\chi)$ iff $\psi \succ_B \chi$ is true in the model. This follow sfrom the first three ordering axioms.

Moreover, if $V(\psi) \subseteq V(\chi)$ then $V(\psi) \geq_1 V(\chi)$ (subset condition), by the axiom: $U(\chi \to \psi) \to\ \psi \succcurlyeq_B \chi$. Finally if $V(\psi)$ properly contains $\mathcal{B}$ and $V(\chi)$ does not, then $V(\psi) >_1 V(\chi)$ (sufficient belief condition) by the axiom $B\psi \wedge \neg B\chi \to\ \psi \succ_B \chi$.

So, $\geq_1$ behaves properly on the representable elements of $\mathcal{P}(W)$. What remains is to extend $\geq_1$ to an ordering $\geq$ with the right properties over all of $\mathcal{P}(W)$.

Take an arbitrary subset X of W. We define $R(X)$ to be the largest subset of X that is representable. That such a set exists follows from the fact that the representable subsets are closed under finite unions and the finiteness of the model.

We now define $X \geq Y$ iff $R(X) \geq_1 R(Y)$. This immediately makes $\geq$ a quasi-linear order. That $\geq$ satisfies the *subset condition* follows from the fact that, if $X \subseteq Y$, then $R(X) \subseteq R(Y)$.

We will conclude this proof with a lemma showing that $\mathcal{B}$ is representable, i.e. $\mathcal{B} = R(\mathcal{B})$. From that result it follows that, if $\mathcal{B} \subseteq X$, then $\mathcal{B} \subseteq R(X)$. This is clearly sufficient to ensure the *sufficient belief condition* (condition 2 in Definition 2.2): if $\mathcal{B} \not\subseteq Y$ then $\mathcal{B} \not\subseteq R(Y) \subseteq Y$. So, once we finish the proof of the following lemma, we are done.

$\mathcal{B}$ **is representable**: Consider w not in $\mathcal{B}$. Then it is not the case that $w\mathcal{R}_B w$. This means that, for some particular $B(\psi_w)$ in Φ, $B(\psi_w)$ is in w but ψ_w is not. (Other possibilities are excluded because we already know that $B(\psi_w)$ and $\neg B(\psi_w)$ true everywhere or nowhere.) Note that we have ψ_w true all over $\mathcal{B}$. Consider the conjunction ψ of all ψ_w for w in the complement of $\mathcal{B}$. $B(\psi)$ is a member of Φ and ψ is true in all elements of $\mathcal{B}$, but it is falsified at all elements u in the complement of $\mathcal{B}$, since ψ implies ψ_u and ψ_u is falsified in u. We have shown that $\mathcal{B}$ is represented by ψ.

This completes the proof of the theorem. $\blacksquare$

Since the countermodel constructed is finite, we also have that the logic *KD45−O* is decidable.

2.3 Additional principles

Before ending this section we would like to discuss some principles that might be added to the system *KD45−O*. For the basic results we wanted to obtain in this paper we did not need them, but they are definitely worth thinking about as possible additions to *KD45−O*. The four following principles have been arranged in order of strength. The second principle will be useful when we discuss in Section 7 the possibilities of defining the plausibility ordering in terms of the set ordering.

1. $(\varphi \succ_B \bot) \leftrightarrow (\top \succ_B \neg\varphi)$,
2. $(B\neg\psi \wedge \neg B\neg\varphi) \rightarrow (\varphi \succ_B \psi)$.
3. $(\varphi \succ_B \psi) \rightarrow (\neg\psi \succ_B \neg\varphi)$.
4. $(\varphi \succ_B \psi) \leftrightarrow (\varphi \wedge \neg\psi) \succ_B (\psi \wedge \neg\varphi)$.

Principle (1) $(\varphi \succ_B \bot) \leftrightarrow (\top \succ_B \neg\varphi)$, says that $\top$ and $\bot$ play a dual role. With the existence axiom it implies $\varphi \rightarrow (\top \succ_B \neg\varphi)$, if φ is true somewhere, then $\neg\varphi$ is less believable than a tautology. The right to left direction is already provable in *KD45−O*, as the reader can check by the semantics (completeness has been proved!). An equivalent formulation is $(\varphi \succcurlyeq_B \top) \leftrightarrow (\bot \succcurlyeq_B \neg\varphi)$. To make (1) true, the models need an extra clause, saying that,

if $S \neq X$ then $S >_B X$.

This condition is dual to Condition 3 of Definition 2.2. Thus (1) seems a very reasonable addition as it makes the models more symmetric. Also $U\varphi$ can by its use be more simply and intuitively defined as $\varphi \succcurlyeq_B \top$.

Principle (2) $(B\neg\psi \wedge \neg B\neg\varphi) \rightarrow (\varphi \succ_B \psi)$ expresses another form of symmetry, this time connected with Condition 2 on the models. It expresses that if a set lies completely outside the center, then it is less plausible than a set that intersects with the center:

if $X \cap \mathcal{B} \neq \emptyset$ and $Y \cap \mathcal{B} = \emptyset$, then $X >_B Y$.

This is of course equivalent to

if $X \not\subseteq \overline{\mathcal{B}}$ and $Y \subseteq \overline{\mathcal{B}}$, then $X >_B Y$.

In the system as it was presented one can only get $\neg\psi \succ_B \neg\varphi$ from $B\neg\psi \wedge \neg B\neg\varphi$. In the final discussion we will use this axiom to define the world order in terms of the set order.

After the above discussion the more general principle (3) $(\varphi \succ_B \psi) \rightarrow (\neg\psi \succ_B \neg\varphi)$ that implies both (1) and (2) immediately springs to mind. It expresses that

$X >_B Y$ iff $\overline{Y} >_B \overline{X}$.

Principle (3), if implemented, would definitely strengthen the probabilistic flavor of the axiomatization. We did not mention this before but, of course, $KD45{-}O$ can be considered to be a weak axiomatization of qualitative probability.

Principle (4) $(\varphi \succ_B \psi) \leftrightarrow ((\varphi \wedge \neg\psi) \succ_B (\psi \wedge \neg\varphi))$ is in the same vein and even stronger. To see that (4) implies (3) just write (4) out for $\neg\psi$ and $\neg\varphi$. This principle is based on the idea that if φ is more believable than ψ, then that can only be based on the non-intersecting parts of the extensions of φ and ψ: the intersection of φ and ψ should be irrelevant in the estimation of their relative believability. In terms of the semantics:

$X >_B Y$ iff $X - Y >_B Y - X$.

The completeness proof for a system including one of these principles would hardly change, it would suffice to note that a principle is satisfied in the model that is constructed.

3 Applying the explicit ordering framework

We now show that the explicit notions of ordering for comparing strengths of beliefs in the logical language aid in expressing several other related concepts in a uniform way, viz. plausibility, disbelief, and preference.

3.1 Plausibility

Comparing the strength of beliefs explicitly has its various advantageous applications. In this subsection we concentrate on plausibility. By plausibility of a proposition we generally mean that we tend to believe in its happening rather than its not happening. That is the interpretation we take here. Hence, in terms of ordered formulas, $P\varphi$ can be expressed as $\varphi \succ_B \neg\varphi$. Of course, there are other possible notions of plausibility, but here we interpret $P\varphi$ as 'more plausible than not'. We now explore this notion of 'plausibility' in terms of belief ordering.

An important principle that will be valid for the *plausibility* operator P under this interpretation is $U(\varphi \to \psi) \to (P\varphi \to P\psi)$. This holds because if $U(\varphi \to \psi)$, not only $\psi \succcurlyeq_B \varphi$, but $U(\varphi \to \psi)$ implies $U(\neg\psi \to \neg\varphi)$, so also $\neg\varphi \succcurlyeq_B \neg\psi$. So, if $P\varphi$, i.e., $\varphi \succ_B \neg\varphi$, then $\psi \succ_B \neg\varphi$, so $\psi \succ_B \neg\psi$, i.e., $P\psi$. This principle leads to consequences like $P(\varphi \wedge \psi) \to P\varphi$.

An important principle that we do not want to be valid is $P\varphi \wedge P\psi \to P(\varphi \wedge \psi)$, This would make the modal logic of P a normal modal logic (of weak belief), because it is equivalent to $P\varphi \wedge P(\varphi \to \psi) \to P\psi$, the K-axiom of modal logic, which makes a logic normal. The principle $P\varphi \wedge P\psi \to P(\varphi \wedge \psi)$ is not intuitively valid. For example, you may judge it more plausible than not that your next client will be male. Similarly, you may consider it to be plausible that your next client will be a foreigner. But, it does not follow that it is more plausible than not that the next client will be a foreign male, most of one's foreign clients may be female. Our proposed semantics clearly does not make it valid.

This discussion makes plausibility logic an important test case for our semantics. As we will see plausibility logic does turn out to be a monotonic logic and for those logics a semantics using sets of worlds instead of single worlds is standard, e.g. neighborhood models (see for a more complete discussion Subsection 3.1.1).

We now move on to exhibiting an independent axiomatization of the plausibility logic P. The language of the P-logic is given by

$$\varphi := p \mid \neg\varphi \mid \varphi \vee \varphi \mid P\varphi$$

We read $P\varphi$ as "φ is plausible". As mentioned above, the intuitive meaning of $P\varphi$ can be captured by the formula $\varphi \succ_B \neg\varphi$, and as such, the truth definition of $P\varphi$ in a *KD45−O* model is given by,

$$\mathcal{M}, s \models P\varphi \text{ iff } \{t \mid \mathcal{M}, t \models \varphi\} >_B \{t \mid \mathcal{M}, t \models \neg\varphi\}.$$

Definition 3.1
The system of P-logic consists of the following axioms and rules:

(a) all propositional tautologies and inference rules

(b) plausibility axioms:

$$P\psi \wedge P\varphi \to P(\psi \wedge P\varphi) \qquad (P\wedge)$$

$$\neg P\varphi \to P\neg P\varphi \qquad (P5)$$

$$P\varphi \to \neg P\neg\varphi \qquad (PD)$$

$$P\top$$

c) monotonicity rule:

$$\frac{\varphi \to \psi}{P\varphi \to P\psi} \qquad \text{(monotonicity rule)}$$

Before giving a completeness proof of this system let us make some remarks on the axioms. The monotonicity rule implies the necessitation rule $\varphi/P\varphi$, by the axiom $P\top$. Moreover, from the monotonicity rule the equivalence rule,

$$\frac{\varphi \leftrightarrow \psi}{P\varphi \leftrightarrow P\psi} \qquad \text{(equivalence rule)}$$

immediately follows. This, in its turn means that provable equivalents can be substituted for each other without impairing provability. As a second plausibility axiom one might have expected $P\psi \wedge \neg P\varphi \rightarrow P(\psi \wedge \neg P\varphi)$, but this follows from $P5$ and PD; just note that by these axioms $\neg P\psi$ and $P\neg P\psi$, are provably equivalent.

On the way to proving completeness we prove a lemma. In its statement and proof we will use the following notation: $\varphi[\chi/\psi]$ stands for a formula arising from φ by replacing some occurrences of ψ by χ. There may be some unclarity here because it is not a unique formula that is defined in this manner, but since our results are valid for any formula obtained in this way it does not matter. One could be more precise by considering formulas with an additional propositional variable p not used elsewhere. Then one can arrange it so that, if φ is $\theta[\psi/p]$, then $\varphi[\chi/\psi]$ is $\theta[\chi/p]$.

Lemma 3.2
Any formula in P-logic is equivalent to a formula with P-depth at most one.

Proof. For the purpose of the proof we first derive the following schemes:

1. $P\psi \rightarrow (\varphi \leftrightarrow \varphi[\top/P\psi])$
2. $\neg P\psi \rightarrow (\varphi \leftrightarrow \varphi[\bot/P\psi])$

We prove these simultaneously by induction on the complexity of formulas φ with possible occurrences of $\top$ and $\bot$. In the base case, that is, for the atomic propositions, propositional constants and $P\psi$, the result follows immediately.

Induction step. This is trivial for the Boolean connectives. So, it suffices to prove it for $P\varphi$ assuming it holds for φ. The induction hypothesis for the first scheme says

$$(P\psi \wedge \varphi) \leftrightarrow (P\psi \wedge \varphi[\top/P\psi]) \qquad \text{(IH)}$$

Now assume $P\psi$ and $P\varphi$. By use of the axiom $P\wedge$, $P(\varphi \wedge P\psi)$ follows. From IH it follows that $P(\varphi[\top/P\psi] \wedge P\psi)$ and hence $P(\varphi[\top/P\psi])$. The proof for the second scheme is very similar.

To see that these schemes imply that each formula in P-logic is equivalent to a formula with P-depth at most one, just note that $\vdash \varphi \leftrightarrow ((P\psi \wedge \varphi) \vee (\neg P\psi \wedge \varphi))$. Now, if we want to get rid of occurrences of $P\psi$ in φ we can replace φ by $((P\psi \wedge \varphi[\top/P\psi]) \vee (\neg P\psi \wedge \varphi[\bot/P\psi]))$. One applies this of course to $P\psi$ with no occurrences of P in ψ. One consecutively removes all occurrences of such $P\psi$ from φ to obtain the desired result. ∎

Theorem 3.3
P-logic is complete with respect to the $KD45{-}O$ models.

Proof. Using the lemma we now show that any consistent set has a model. Assume we have a consistent set in the P-logic. It can be extended to a maximal P-consistent set, say Γ. Since we can restrict attention to formulas which are Boolean combinations of atoms and formulas of the form $P\varphi$ where φ no longer contains P, a maximal consistent set is essentially only a set of atoms, negations of atoms, and such $P\varphi$'s and $\neg P\varphi$'s.

Let us just take a finite number of atoms to keep things finite, and let us take a maximal consistent set Γ of the form described. We now make a $KD45{-}O$-model where $P\varphi$ gets interpreted as $\varphi \succ_B \neg\varphi$. The worlds will simply be defined by a number of atoms being true in it and the rest of the atoms false. Let us now consider the following model, $\mathcal{M} = (S, \leq, \geq_B, V)$, where S is the set of all such worlds. The ordering of the subsets is as follows: There are 5 equivalence classes in the ordering starting with the highest grade of believability. We take membership of those classes to determine the degree of belief in the sets. As representing formulas we just take purely propositional ones.

(1) The whole set, which is of course represented by $\top$ (or other tautologies).
(2) The sets represented by those φ for which $P\varphi$ is in Γ (except for $\top$).
(3) The sets represented by those φ for which $\neg P\varphi$ is in Γ as well as $\neg P\neg\varphi$.
(4) The non-empty sets represented by φ for which $P\neg\varphi$ is in Γ.
(5) The empty set, which is of course represented by $\bot$.

These are all possibilities because of the axiom $P\varphi \to \neg P\neg\varphi$. Finally we take $\mathcal{B}$, the center, to be the whole set (so, there are no beliefs except the trivial one in $\top$).

The two things we have to check are: First, that, if a set is in class (2), then any larger one will be in (2) as well (or in (1)). This follows from the *monotonicity rule*, since by the fact that the worlds are determined by the atoms true in them all inclusions are logical inclusions. Similarly for the other classes. Second, that, if a set X contains all of $\mathcal{B}$, and another set Y doesn't, then $X >_B Y$. That is trivial: X has to be $\mathcal{B}$, the whole set, and Y is not.

Finally, we see that Γ is satisfied by the world in the model that makes exactly its atoms true. So, for each consistent set we can have a model in *KD45−O*. Thus, the axioms and rules given in Theorem 2.6 axiomatize the P-logic of 'more plausible than not'. It is also worth mentioning why $P\varphi \wedge P\psi \to P(\varphi \wedge \psi)$ will fail in general. There may be sets in (2) the intersection of which is not in (2). ∎

Evidently, $P\varphi$ is a global notion - its value does not vary through the model. Again, P is clearly an introspective notion.

Let us finally note that an interpretation of $P\varphi$ as φ having probability more than 0.5 (or any other fixed number between 0.5 and 1) leads to exactly the P-axioms, provided one considers the probability statements themselves to always have probability 1.

3.1.1 Neighborhood models

It is good to mention that a different, more standard but equivalent semantics for the P-logic exists: neighborhood models [7]. A *neighborhood frame* consists of a set of states S and a function ν that maps each state s onto a set of subsets of S such that, if $X \in \nu(s)$ and $X \subseteq Y$, then $Y \in \nu(s)$. A *neighborhood model* is a neighborhood frame with a valuation as usual. A formula $P\varphi$ will be true in s if $V(\varphi) \in \nu(s)$.

It is clear that in our case the set of states S together with the (constant) function

ν that maps each state to $\{X \mid X >_B S - X\}$ is a neighborhood frame. The corresponding neighborhood models will give exactly the same truth conditions as our models. The special properties that the neighborhood frames for the P-logic have beyond the standard ones mentioned above are:

- The function ν is constant on S,
- If $X \in S$, then $S - X \notin S$,
- $\nu(s)$ is non-empty, it contains S.

The logics corresponding to the neighborhood frames are called *monotonic* logics [21]. The minimal monotonic logic has beyond propositional logic just the axiom $P\top$ and the monotonicity rule.

The fact that this standard semantics for monotonic logics is for this particular case equivalent to our semantics with set ordering strengthens our claim that one needs to involve sets of states in the semantics to discuss strength of belief.

3.1.2 Belief and plausibility

We now consider a system having both belief and the plausibility operator, viz. the BP-system. It is a very natural extension of the P-logic and will provide pointers to discuss logics of belief and disbelief in the next subsection. The language is that of the P-logic, together with the additional modal operator B for belief.

$$\varphi := p \mid \neg\varphi \mid \varphi \vee \varphi \mid P\varphi \mid B\varphi$$

Definition 3.4
BP-logic is axiomatized by the following axioms and rules:

a) all propositional tautologies and inference rules
b) all *KD45* axioms and rules
c) all P axioms and rules
d) special axioms:

$$B\varphi \rightarrow P\varphi$$
$$P\varphi \rightarrow BP\varphi$$

It is easy to see that $\neg P\varphi \rightarrow B\neg P\varphi$ is derivable in BP-logic.

Theorem 3.5
BP-logic is complete with respect to the $KD45 - O$-models.

The proof is very similar to that for the P-logic. It starts with proving that the axioms force all formulas to be equivalent to Boolean combinations of atoms and formulas of the form $P\varphi$ and $B\varphi$, where φ is Boolean. Analogously to the P-logic one first has to prove

1. $P\psi \rightarrow (\varphi \leftrightarrow \varphi[\top/P\psi])$
2. $\neg P\psi \rightarrow (\varphi \leftrightarrow \varphi[\bot/P\psi])$
3. $B\psi \rightarrow (\varphi \leftrightarrow \varphi[\top/B\psi])$

4. $\neg B\psi \to (\varphi \leftrightarrow \varphi[\bot/B\psi])$

One needs the theorem $B\varphi \to PB\varphi$, which follows immediately from $B\varphi \to BB\varphi$ and $B\varphi \to P\varphi$. Instead of five grades of belief we will now have seven, e.g. the second class (in proof of Theorem 3.3) splits into sets represented by φ with $B\varphi$ true and the ones representable by φ with $\neg B\varphi$ and $P\varphi$ true.

It is noteworthy that the principle $B\varphi \wedge P\psi \to P(\varphi \wedge \psi)$ of [5] fails in the BP-logic. Let us consider the model $\mathcal{M}$ as follows:

$$\begin{array}{ccccc} s_1 & & s_2 & & s_3 \\ \bullet & \equiv & \bullet & < & \bullet \\ p, q & & p, \neg q & & \neg p, q \end{array}$$

The center is then $\{s_1, s_2\}$. The set ordering $\geq_B$ is given as follows:

$\geq_B$: $\{s_1, s_2, s_3\} >_B \{s_1, s_2\} >_B \{s_1, s_3\} >_B \{s_2, s_3\} >_B \{s_1\} >_B \{s_2\} >_B \{s_3\} >_B \emptyset$

In this model Bp and Pq hold, but $P(p \wedge q)$ does not hold. Thus we have our required counterexample. Also in this case we have chosen our model in such a way that the additional principles hold in it.

3.2 Disbelief

Disbelief in a proposition is governed by exactly the opposite situation to the one discussed in the previous subsection, $D\varphi$ can be expressed as $\neg\varphi \succ_B \varphi$, that is $P\neg\varphi$.

With the huge amount of work going on in logics of belief and belief revision, consideration of disbelief as a separate epistemic category came to fore in the latter part of last decade [13, 14]. Consideration of changing or revising disbeliefs as a process analogous to belief revision was taken up by [15]. Belief-disbelief pairs i.e. simultaneous consideration of belief and disbelief sets were also taken up [8, 6] through which various connections of possible inter-connectivity of beliefs and disbeliefs have come into focus. As mentioned earlier our notion of explicit belief ordering provides another path into expressing the concept of disbelief.

The basic idea for disbelieving a proposition is that the inclination to believe in its negation is stronger than that to believe it. Consequently, disbelieving is a much weaker notion than believing the negation of the proposition, but it should imply that one does not believe in the proposition. In other words, $D\varphi$ is implied by $B\neg\varphi$ and implies $\neg B\varphi$ but not the other way around in either case.

In general, if a person faces a decision based on whether a certain state of affairs is the case or an event happens, she may not have enough evidence to believe that the state of affairs is the case or is not the case. Then she may base her decision on whether she thinks the state of affairs plausible or disbelieves in it. Only in the case that her strength of belief in the two possibilities is equal, translated into our framework as $\varphi \equiv_B \neg\varphi$, it is a real tossup for her.

Various principles for the 'disbelief' operator together with the 'belief' one have been discussed in [14] in the autoepistemic logic framework of [28]. As such, the possible world semantics provided there which is based on separate sets of worlds for beliefs and disbeliefs is not very interesting, and suffers from 'disjointedness' as well as ‘mirror-image’ problems. These questions will not arise in the semantics we propose here. The basic reason is the fact that ‘disbelief’ is given a global stance in contrast to ‘belief’ which is apparent from their respective interpretations. This also emphasizes the fact that disbelieving something is different from both ‘not believing’ as well as ‘believing the negation’.

We now focus on getting a more feasible logic of belief and disbelief in similar lines to BP-logic introduced earlier. From our formal understanding $D\varphi$ is same as $P\neg\varphi$ and hence we get the following dual axiomatization of the BD-logic .

Theorem 3.6
BD-logic is complete and its validities are completely axiomatized by the following axioms and rules:

a) all propositional tautologies and inference rules
b) all *KD45* axioms and rules
c) disbelief axioms:
$D\psi \wedge D\varphi \rightarrow D(\psi \vee \neg D\varphi)$
$\neg D\varphi \rightarrow DD\varphi$
$D\varphi \rightarrow \neg D\neg\varphi$
$D\bot$
d) special axioms:
$B\varphi \rightarrow D\neg\varphi$
$D\varphi \rightarrow BD\varphi$
e) anti-monotone rule:
if $\varphi \rightarrow \psi$ then $D\psi \rightarrow D\varphi$.

The proof follows similarly as in the case of BP-logic. Some interesting validities of this logic are,

- $B\neg\varphi \rightarrow D\varphi$
- $D\varphi \rightarrow \neg B\varphi$
- $\neg D\varphi \rightarrow B\neg D\varphi$
- $\neg D\varphi \rightarrow DD\varphi$
- $\neg B\varphi \rightarrow DB\varphi$

As in the cases of P-logic and BP-logic, the corresponding intuitively incorrect principle, $D\varphi \wedge D\psi \rightarrow D(\varphi \vee \psi)$ can also be avoided in the BD-logic. It may be very hard to believe that your friend Craig is the traitor and even that another close friend Denis is the traitor, but circumstantial evidence may make it perfectly plausible that one of them is.

3.3 Preference

There is a very close relationship between an agent’s beliefs and her preferences, which has been extensively discussed in [24, 27]. Based on the ideas from *optimality theory*, intrinsic preference on the basis of *priority sequences* $P_1 >> \ldots >> P_n$ is formulated. Here, the P_i's are predicates with exactly one free variable. Preferences over objects can be defined in terms of these sequences. The basic idea is to define *objective preference* by:

$$Pref(d,e) \Leftrightarrow \exists i(P_i d \wedge \neg P_i e) \wedge \forall j < i\,(P_j d \leftrightarrow P_j e)$$

Let us give an example. Alice again has applicants for a simple position. She still judges them on a yes-no basis, but this time in regard to three aspects: are they strong enough (P_1), can they drive a truck sufficiently well (P_2), do they understand English well enough (P_3). These aspects are strictly ordered in the way described above, i.e., if Jennifer is strong but a poor driver who doesn't speak English, she is graded higher objectively than Karl, an excellent driver with fluent english, but a weakling.

If these aspects are subject to belief one can consider *subjective preferences* over objects. Several options to implement this idea are considered in the papers mentioned, their meanings are more or less obvious.

$$Pref(d,e) \Leftrightarrow \exists i(B(P_i d) \wedge \neg B(P_i e) \wedge \forall j < i(B(P_j d) \leftrightarrow B(P_j e)))$$

$$Pref(d,e) \Leftrightarrow \exists i(\neg B(\neg P_i d) \wedge B(\neg P_i e) \wedge \forall j < i(B(\neg P_j d) \leftrightarrow B(\neg P_j e)))$$

$$Pref(d,e) \Leftrightarrow \exists i\,((B(P_i d) \wedge \neg B(P_i e)) \vee (\neg B(\neg P_i d) \wedge B(\neg P_i e)) \wedge \forall j < i\,((B(P_j d) \leftrightarrow B(P_j e)) \wedge (B(\neg P_j d) \leftrightarrow B(\neg P_j e))))$$

The first option directly subjectivizes the original idea, the criteria are made a matter of belief; truth and falsity have been replaced by believing and not believing. If Alice believes that Jennifer is strong but has a low opinion about her other capabilities, while she does not believe that Karl is strong but does believe he can drive a truck well and that he speaks english she will prefer Jennifer.

In the second option 'believing that not' is more central. Returning to the example, let us change things only in so far that Alice is now not able to make up her mind about the strength of Karl, she does not believe he is strong enough but she also doesn't believe he isn't. Under the first option she will still prefer Jennifer. But, under the second option she would only disqualify Karl immediately if she believes he isn't strong enough – she doesn't, so under that option she rates Karl higher than Jennifer whom she believes not to be a good driver.

It is clear that the above three approaches are different ways of expressing that up to a certain level of the priority sequence the degree of belief in the objects d and e having the mentioned properties is the same and that at the next level the degree of belief in d having the right property is greater than that in e having it. With the availability of explicit ordering in the language we can express this in a general way as below, giving one uniform definition.

$$Pref(d,e) \Leftrightarrow \exists i\,(P_i d \succ_B P_i e\ \wedge \forall j < i\,(P_j d \equiv_B P_j e)).$$

In specific models one may then apply this definition of $Pref$ to obtain the effect of one of the three approaches above, or use any other fitting procedure. As in the introduction, we point out that for many decisions involving preference it may be unavoidable to grade the priorities in some way or other. The system described is a basic approach just tailored to decisions involving yes-no questions.

4 Relative expressiveness

Till now we have proposed various logics to describe different but related notions in belief and plausibility. In this section we study the relative expressive powers of the languages of these logics. To aid our discussion, let us first list the different languages as follows:

- $\mathcal{L}_1$: $\bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B\varphi \mid \varphi \succcurlyeq_B \varphi$
- $\mathcal{L}_2$: $\bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B\varphi$
- $\mathcal{L}_3$: $\bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid \varphi \succcurlyeq_B \varphi$
- $\mathcal{L}_4$: $\bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid P\varphi$

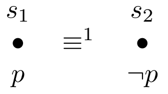
Let $\mathcal{L}_i > \mathcal{L}_j$ denote that the language $\mathcal{L}_i$ is more expressive than $\mathcal{L}_j$. We now do a comparative expressiveness study for these languages.

Theorem 4.1
The only relations $\mathcal{L}_i > \mathcal{L}_j$ that exist between the languages $\mathcal{L}_1, \ldots, \mathcal{L}_4$ are $\mathcal{L}_1 > \mathcal{L}_2$ and $\mathcal{L}_1 > \mathcal{L}_3 > \mathcal{L}_4$.

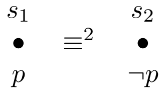
Proof. We prove the inequalities as follows:

- $\mathcal{L}_1 > \mathcal{L}_2$

Let us consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:



$\mathcal{M}_1$



$\mathcal{M}_2$

The center is then $\{s_1, s_2\}$. The respective $\geq_B^1$ and $\geq_B^2$ are given as follows:

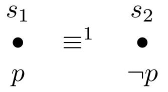
$\geq_B^1$: $\{s_1, s_2\} >_B^1 \{s_1\} >_B^1 \{s_2\} >_B^1 \emptyset$

$\geq_B^2$: $\{s_1, s_2\} >_B^1 \{s_2\} >_B^1 \{s_1\} >_B^1 \emptyset$

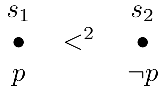
We have that the language $\mathcal{L}_2$ cannot distinguish between the models, whereas $\mathcal{M}_1 \models p \succ_B \neg p$, and $\mathcal{M}_2 \models \neg p \succ_B p$. Thus $\mathcal{L}_1 > \mathcal{L}_2$. It also follows that $\mathcal{L}_2 \not> \mathcal{L}_4$ (because $p \succ_B \neg p$ is Pp) and $\mathcal{L}_2 \not> \mathcal{L}_3$.

- $\mathcal{L}_1 > \mathcal{L}_3$

Let us consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:



$\mathcal{M}_1$



$\mathcal{M}_2$

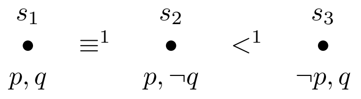
The center of $\mathcal{M}_1$ is then $\{s_1, s_2\}$. The center of $\mathcal{M}_2$ is $\{s_1\}$. The respective $\geq_B^1 = \geq_B^2 = \geq_B$ is given as follows:

$\geq_B$: $\{s_1, s_2\} >_B \{s_1\} >_B \{s_2\} >_B \emptyset$

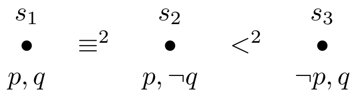
We have that the language $\mathcal{L}_3$ cannot distinguish between the models, whereas $\mathcal{M}_1 \models \neg Bp$, and $\mathcal{M}_2 \models Bp$. Thus $\mathcal{L}_1 > \mathcal{L}_3$. It also follows that $\mathcal{L}_3 \not> \mathcal{L}_2$, and hence $\mathcal{L}_4 \not> \mathcal{L}_2$.

- $\mathcal{L}_3 > \mathcal{L}_4$

Let us consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:



$\mathcal{M}_1$



$\mathcal{M}_2$

The center is then in both cases $\{s_1, s_2\}$. The respective $\geq_B^1$ and $\geq_B^2$ are given as follows:

$\geq_B^1$: $\{s_1, s_2, s_3\} >_B^1 \{s_1, s_2\} >_B^1 \{s_1, s_3\} >_B^1 \{s_2, s_3\} >_B^1 \{s_1\} >_B^1 \{s_2\} >_B^1 \{s_3\} >_B^1 \emptyset$

$\geq_B^2$: $\{s_1, s_2, s_3\} >_B^2 \{s_1, s_2\} >_B^2 \{s_2, s_3\} >_B^2 \{s_1, s_3\} >_B^2 \{s_2\} >_B^2 \{s_1\} >_B^2 \{s_3\} >_B^2 \emptyset$

We have that $\mathcal{M}_1 \models (p \wedge q) \succ_B (p \wedge \neg q)$, and $\mathcal{M}_2 \models (p \wedge \neg q) \succ_B (p \wedge q)$, whereas the language $\mathcal{L}_4$ cannot distinguish between the models, since in both cases all 2 -element sets are more plausible than all 1-element sets and therefore the order between a set and its complement is unchanged. Thus $\mathcal{L}_3 > \mathcal{L}_4$.

We have now made all relevant comparisons of the strength of the four languages. This completes the proof. ■

We do remark that adding one of the additional axioms of Section 2.3 does not change the situation. We did not stress this in the proof but even the additional axiom (4) is satisfied in all the models in the proof above.

5 Safe belief

The notion of 'safe belief' has been introduced in [2]. The authors gave this name to single out those *beliefs* "that are *safe* to hold, in the sense that no future learning of truthful information will force us to revise them." It closely related to "Stalnaker knowledge" [31] where evidence is considered as true information. The safe belief modality is generally denoted by $\Box$. Evidently, 'safe beliefs' are *truthful* ($\Box\varphi \models \varphi$) and *positively introspective* ($\Box\varphi \models \Box\Box\varphi$), but not necessarily *negatively introspective* (in general, $\neg\Box\varphi \not\models \Box\neg\Box\varphi$).

Adding safe belief to our ordering framework is interesting both from the technical as well as intuitive point of view. This is because in the interpretation of [2] there is a very close relationship between the notion of safe belief and the plausibility ordering.

In the *plausibility models*, the truth definition of $\Box\varphi$ is given by the following clause:

$\mathcal{M}, s \models \Box\varphi$ iff $\mathcal{M}, t \models \varphi$ for all worlds $t \leq s$.

which says that φ can be safely believed at some world s if it holds at all the worlds which are at least as plausible as s. In the following we will introduce the safe belief modality in the setting of $KD45-O$, and give a complete axiomatization of this logic. The language of the logic $KD45-OS$ is defined as follows:

Definition 5.1
Given a countable set of atomic propositions Φ, formulas φ are defined inductively:

$$\varphi := \bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B\varphi \mid \Box\varphi \mid \varphi \succcurlyeq_B \psi$$

where $p \in \Phi$.

We now present the axioms of the logic $KD45-OS$ in which the operators U and E are defined as before. Together with the axioms and rules of the $KD45$-logic of beliefs, and the relevant ordering axioms, viz. reflexivity, transitivity, linearity, center, existence, $U \succcurlyeq_B$-axiom and the $S4$-axioms and rules for the safe belief $\Box$ operator, we will have the following extra axioms,

$(\Box\varphi \wedge \neg\Box\psi) \rightarrow \ (\varphi \succ_B \psi)$ ($\Box$-order)
$(\varphi \succcurlyeq_B \psi) \rightarrow \Box(\varphi \succcurlyeq_B \psi)$ ($\Box$-introspection1)
$(\varphi \succ_B \psi) \rightarrow \Box(\varphi \succ_B \psi)$ ($\Box$-introspection2)

The $\Box$-ordering axiom generalizes the center axiom. It expresses that, if a set X contains all worlds with a certain grade of plausibility or higher, and Y does not, then $X >_B Y$. In addition to all of these, the following axiom relates the operator $\Box$ with B.

$\Box\varphi \rightarrow B\varphi$ ($\Box B$-axiom)

The introspection axioms (1-2) and the unique-center axiom of $KD45-O$ are derivable from $KD45-OS$. We can also derive:

$U\varphi \rightarrow \Box\varphi$.

Theorem 5.2
The logic $KD45-OS$ is sound and its validities can be completely axiomatized by the following axioms and rules.

a) all $KD45-O$ axioms and rules

b) $S4$-axioms and rules for the modal operator $\Box$

c) ordering axioms:
$\varphi \succcurlyeq_B \varphi$ (reflexivity)
$(\varphi \succcurlyeq_B \psi) \wedge (\psi \succcurlyeq_B \chi) \rightarrow \varphi \succcurlyeq_B \chi$ (transitivity)
$(\varphi \succcurlyeq_B \psi) \vee (\psi \succ_B \varphi)$ (linearity)
$U(\Box\varphi \rightarrow \Box\psi) \vee U(\Box\psi \rightarrow \Box\varphi)$ ($\Box$-linearity)
$(B\varphi \wedge \neg B\psi) \rightarrow \ (\varphi \succ_B \psi)$ (center)
$(\Box\varphi \wedge \neg\Box\psi) \rightarrow \ (\varphi \succ_B \psi)$ ($\Box$-order)
$(\varphi \succcurlyeq_B \psi) \rightarrow \Box(\varphi \succcurlyeq_B \psi)$ ($\Box$-introspection1)

$(\varphi \succ_B \psi) \rightarrow \Box(\varphi \succ_B \psi)$ ($\Box$-introspection2)

$U(\varphi \rightarrow \psi) \rightarrow (\psi \succcurlyeq_B \varphi)$ ($U \succcurlyeq_B$-axiom)

$\varphi \rightarrow E\varphi$ (existence)

d) $\Box\varphi \rightarrow B\varphi$ ($\Box B$-axiom)

e) inclusion rule:

$$\frac{\varphi \rightarrow \psi}{\psi \succcurlyeq_B \varphi}$$

Proof. Assume $\nvdash_{KD45-OS} \varphi$. We will have to construct a countermodel to φ which is a $KD45-OS$-model. We take a finite adequate set Φ containing φ. Consider the m.c. (maximally consistent) subsets of Φ. In particular consider such an m.c. set Φ_0 containing $\neg\varphi$.

Define the plausibility ordering among m.c. sets as follows: $P \leq Q$ iff for all $\Box\psi$ in the adequate set, if $\Box\psi$ is in P, then $\Box\psi$ and ψ are in Q. Then immediately we have that $\leq$ is reflexive and transitive.

The relations $\mathcal{R}_B$ and $\mathcal{R}_U$ are defined as follows:

$$\begin{array}{lll} P\mathcal{R}_BQ & \text{iff} & (1) \text{ for all } B\varphi \text{ in } P, \varphi \text{ as well as } B\varphi \text{ are in } Q, \\ & & (2) \text{ for all } \neg B\varphi \text{ in } P, \neg B\varphi \text{ in } Q. \\ P\mathcal{R}_UQ & \text{iff} & (1) \text{ for all } U\varphi \text{ in } P, \varphi \text{ as well as } U\varphi \text{ are in } Q, \\ & & (2) \text{ for all } \neg U\varphi \text{ in } P, \neg U\varphi \text{ in } Q \end{array}$$

As in the proof of Theorem 2.4, we can show that $\mathcal{R}_U$ will be an equivalence relation and $\mathcal{R}_B$ an Euclidean subrelation of $\mathcal{R}_U$.

It follows from the $\Box$-introspection axioms that $U\varphi \rightarrow \Box\varphi$ is derivable, and so $\leq$ is a subrelation of $\mathcal{R}_U$. From the axioms relating $\Box$ and B, it follows that $\mathcal{R}_B$ is a subrelation of $\leq$.

We now take the submodel generated by $\mathcal{R}_U$ from Φ_0. The set of worlds W of our model will be the set of worlds in this submodel and the $\mathcal{R}_B$ and $\mathcal{R}_U$ the restrictions of the original $\mathcal{R}_B$ and $\mathcal{R}_U$ to this submodel. $\mathcal{R}_U$ is now the universal relation. As before, we write $\mathcal{B}$ for the set of $\mathcal{R}_B$-reflexive elements. Because of the $\Box$-linearity axiom $\leq$ becomes linear in this model. So, with respect to the modal operators B and E and $\Box$ the model behaves properly. We will now have to order $\mathcal{P}(W)$ in the proper way, which can be done as in the proof of Theorem 2.4, using the $\Box$-order axiom in addition to the center axiom. ∎

Similar to the work done in [2], belief and conditional belief can be expressed in terms of safe belief and the existential modality as,

$$B^{\psi}\varphi := E\psi \rightarrow E(\psi \wedge \Box(\psi \rightarrow \varphi));$$
$$B\varphi := B^{\top}\varphi.$$

We do not talk about conditional belief here but belief can be defined in terms of the existential modality and safe belief (and therefore, in terms of safe belief and belief ordering) as follows:

$$\begin{aligned} B\varphi &:= B^{\top}\varphi \\ &:= E\top \rightarrow E(\top \wedge \Box(\top \rightarrow \varphi)) \\ &:= E(\top \wedge \Box(\top \rightarrow \varphi)) \\ &:= E\Box(\top \rightarrow \varphi) \\ &:= E\Box\varphi \end{aligned}$$

Once we have in this manner the modal operator B as a defined concept we can easily derive all its well-known properties in $KD45{-}OS$, but if that holds fully for its relations with $\succcurlyeq_B$ remains to be seen.

5.1 Relative expressiveness

We have already seen that belief can be expressed in terms of belief ordering and safe belief, but similar other questions arise. Can safe belief be expressed in terms of belief and belief ordering? Can belief ordering be expressed in terms belief and safe belief? To answer these questions, we first list the different languages as follows:

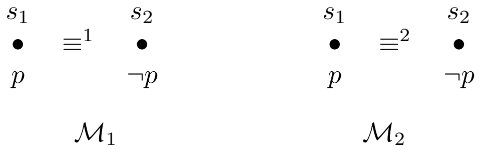
- $\mathcal{L}_5$: $\bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B\varphi \mid \Box\varphi \mid \varphi \succcurlyeq_B \varphi$
- $\mathcal{L}_6$: $\bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B\varphi \mid \Box\varphi$
- $\mathcal{L}_7$: $\bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B\varphi \mid \varphi \succcurlyeq_B \varphi$

Theorem 5.3
The only relations $\mathcal{L}_i > \mathcal{L}_j$ that exist between the languages $\mathcal{L}_5$, $\mathcal{L}_6$ and $\mathcal{L}_7$ are $\mathcal{L}_5 > \mathcal{L}_6$ and $\mathcal{L}_5 > \mathcal{L}_7$.

Proof. We prove the inequalities as follows:

- $\mathcal{L}_5 > \mathcal{L}_6$

Let us consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:



The center is then in both cases $\{s_1, s_2\}$. The respective orderings $\geq^1_B$ and $\geq^2_B$ are given as follows:

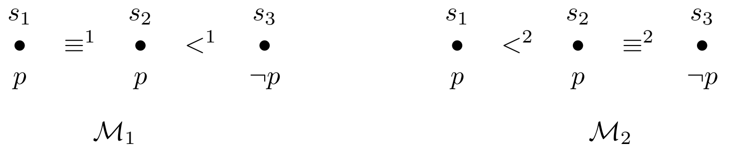
$\geq^1_B$: $\{s_1, s_2\} >^1_B \{s_1\} >^1_B \{s_2\} >^1_B \emptyset$

$\geq^2_B$: $\{s_1, s_2\} >^1_B \{s_2\} >^1_B \{s_1\} >^1_B \emptyset$

We have that the language $\mathcal{L}_6$ cannot distinguish between the models, whereas $\mathcal{M}_1 \models p \succ_B \neg p$, and $\mathcal{M}_2 \models \neg p \succ_B p$. Thus $\mathcal{L}_5 > \mathcal{L}_6$. It also follows that $\mathcal{L}_6 \not> \mathcal{L}_7$.

- $\mathcal{L}_5 > \mathcal{L}_7$

Let us consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows:



The center in $\mathcal{M}_1$ is $\{s_1, s_2\}$, and the center in $\mathcal{M}_2$ is $\{s_1\}$. The respective $\geq^1_B = \geq^2_B = \geq_B$ is given as follows:

$\geq_B$: $\{s_1, s_2, s_3\} >_B \{s_1, s_2\} >_B \{s_1, s_3\} >_B \{s_1\} >_B \{s_2, s_3\} >_B \{s_2\} >_B \{s_3\} >_B \emptyset$

We have that $\mathcal{M}_1, s_2 \models \Box p$, and $\mathcal{M}_2, s_2 \models \neg\Box p$. The language $\mathcal{L}_7$ cannot distinguish between the models. Thus $\mathcal{L}_5 > \mathcal{L}_7$. It also follows that $\mathcal{L}_7 \not> \mathcal{L}_6$.

We have now made all relevant comparisons of the strength of these three languages. This completes the proof. ■

6 Multi-agent system

The main focus of this paper has been on beliefs and strengths of beliefs of a single agent. The whole idea can be generalized to the multi-agent framework which is what we do in the following. The language of the logic of belief ordering in the multi-agent case, *KD45*$-O_M$ can be defined as follows:

Definition 6.1
Given a finite set of agents A, and a countable set of atomic propositions Φ, formulas φ are defined inductively:

$$\varphi := \bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid B_a\varphi \mid \varphi \succcurlyeq_{B_a} \psi$$

where $p \in \Phi$, $a \in A$.

The indices in the belief modality and in the ordering formula denote the agents whose beliefs or strengths of beliefs are considered. The operators $\succ_{B_a}$ and U_a are defined in the usual way. The fact that U is also indexed may surprise the reader for a moment but it is the only coherent way to extend the one agent case. Existence of a location for a proposition to be true meant for us that for the one agent case, belief in the proposition was stronger than belief in a contradiction. With more agents, we may have those who differ in regard to the existence of propositions: more worlds will have to be added to the model, and it will not stop there because there is no reason for E_aE_b to be equivalent to E_a or E_b, etc.

Keeping all these considerations in mind, the models for *KD45*$-O_M$ have to be suitable multi-agent generalizations of those for *KD45*$-O$. The basic idea to consider here is that we can no longer rule out worlds that are *impossible* for an agent a. They might well be possible for another agent b and also have to be considered while talking about agent a's belief about agent b's beliefs and so on. Evidently, the earlier *plausibility ordering* and *set ordering* of worlds will get indexed by agents (one for each agent), and the global concept of belief will give way to more local concepts of beliefs. This fact becomes apparent in the syntax also, with the introduction of formulas like $U_a\varphi$. The notion of *comparative classes* [2] which gives the set of worlds that an agent considers relevant while positioned at her current world comes into play. Formally, a comparative class of some world is just the set of worlds that are related to the current world by the plausibility order. To give meaning to agents' beliefs, strength of beliefs, these relevant worlds are needed to be considered only, unlike the single agent case, where the whole model is taken into account.

Definition 6.2
Given a finite set of agents A, a $KD45\!-\!O_M$ model is defined to be a structure $\mathcal{M} = (S, \{\leq_a : a \in A\}, \{\geq_{B_a} : a \in A\}, V)$, where S is a non-empty finite set of states, V is a valuation assigning truth values to atomic propositions in states, and for each a, $\leq_a$ is a pre-order relation over S, which forms a partition of S given by $\sim_a = \leq_a \cup \geq_a$, an equivalence relation over S. Finally, for each $a \in A$, $\geq_{B_a}$ is a quasi-linear order relation over $\mathcal{P}(T)$ for each equivalence class T of $\sim_a$, satisfying the conditions:

1. If $X \subseteq Y \subseteq T$, then $Y \geq_{B_a} X$
2. If $\mathcal{B}_a \subseteq T$ is the set of a-plausible worlds, truth on which suffices to make an assertion to be believed (that is, the set of all $\leq_a$-minimal worlds in T), then $\mathcal{B}_a \subseteq X \subseteq T \wedge \mathcal{B}_a \not\subseteq Y \subseteq T \Rightarrow X >_{B_a} Y$, where $>_{B_a}$ denotes the corresponding strict ordering.
3. If $X \subseteq T$ is non-empty, then $X >_{B_a} \emptyset$.

For any $s \in S$, let s_a denote the set of all members of S which are $\sim_a$-equivalent to s. The truth definition for formulas φ in a $KD45\!-\!O_M$ model $\mathcal{M}$ is as usual with the following clauses for the belief and ordering modalities.

$\mathcal{M}, s \models B_a\varphi$ iff $\mathcal{M}, t \models \varphi$ for all $\leq_a$-minimal worlds $t \in s_a$.
$\mathcal{M}, s \models \varphi \succcurlyeq_{B_a} \psi$ iff $\{t \in s_a \,|\, \mathcal{M}, t \models \varphi\} \geq_{B_a} \{t \in s_a \,|\, \mathcal{M}, t \models \psi\}$.

We considered $\succcurlyeq_B$ to be a global notion – if $\varphi \succcurlyeq_B \psi$ is true anywhere in the model, it is true everywhere. But in the multi-agent case, $\succcurlyeq_{B_a}$ does become to a certain extent state-dependent, which is intuitive as different agents may perceive the world in different ways. But, of course, the notion does stay a global notion within each $\sim_a$ equivalence class. From the definition of $\succ_{B_a}$ it follows that,

$\mathcal{M}, s \models \varphi \succ_{B_a} \psi$ iff $\{t \in s_a, |\, \mathcal{M}, t \models \varphi\} >_{B_a} \{t \in s_a \,|\, \mathcal{M}, t \models \psi\}$.

Thus, $\succ_{B_a}$ also becomes a more local notion. We will now define the corresponding localized universal modality U_a for each agent $a \in A$. As earlier, the modality $E_a\varphi$ (the abbreviated form of $\neg U_a \neg\varphi$) can be defined as $\varphi \succ_{B_a} \bot$, and hence $U_a\varphi$ as $\bot \succcurlyeq_{B_a} \neg\varphi$. The formula $U_a\varphi$ expresses that φ is true in all a-accessible worlds in the model, whereas $E_a\varphi$ stands for existence of a possible a-accessible world in the model where φ is true. Evidently, we have,

$\mathcal{M}, s \models U_a\varphi$ iff $\mathcal{M}, t \models \varphi$ for all worlds $t \in s_a$.

As earlier, each of these U_a modalities needs to satisfy the $S5$-axioms that hold for U [16] plus the axiom $B_a\varphi \rightarrow U_a B_a\varphi$, which expresses that B_a is a global notion in each of the $\sim_a$-equivalence classes, where U_a expresses this universality.

The logic $KD45\!-\!O_M$ arises from the logic $KD45\!-\!O$ by indexing, for each agent a, in each axiom both the operator B and $\succcurlyeq_B$ by a so that, for each agent the same axioms arise with B_a instead of B and $\succcurlyeq_{B_a}$ instead of $\succcurlyeq_B$.

Definition 6.3
The system $KD45-O_M$ consists of, for each agent $a \in A$,

a) all *KD45* axioms and rules for B_a

b) ordering axioms:

$\varphi \succeq_{B_a} \varphi$ (reflexivity)

$(\varphi \succeq_{B_a} \psi) \wedge (\psi \succeq_{B_a} \chi) \rightarrow \varphi \succeq_{B_a} \chi$ (transitivity)

$(\varphi \succeq_{B_a} \psi) \vee (\psi \succ_{B_a} \varphi)$ (linearity)

$(B_a\varphi \wedge \neg B_a\psi) \rightarrow (\varphi \succ_{B_a} \psi)$ (center)

$(\varphi \succeq_{B_a} \psi) \rightarrow B_a(\varphi \succeq_{B_a} \psi)$ (introspection1)

$(\varphi \succ_{B_a} \psi) \rightarrow B_a(\varphi \succ_{B_a} \psi)$ (introspection2)

$\bot \succeq_{B_a} \neg(\varphi \rightarrow \psi) \rightarrow (\psi \succeq_{B_a} \varphi)$ ($U \succeq_{B_a}$-axiom)

$\varphi \rightarrow (\varphi \succ_{B_a} \bot)$ (existence)

$(B_a\varphi \succ_{B_a} \bot) \rightarrow B_a\varphi$ (unique-center)

c) inclusion rule:

$$\frac{\varphi \rightarrow \psi}{\psi \succeq_{B_a} \varphi} \quad \text{(inclusion rule)}$$

As in the single-agent case we have the following result.

Theorem 6.4
$KD45-O_M$ is sound and complete with respect to $KD45-O_M$ models.

The completeness proof is a generalization of the completeness proof for $KD45-O$ by executing within each U_a-equivalence class the same prodeure as in that proof. We refrain from going into the proof details. Evidently, $KD45-O_M$ is also decidable.

7 World ordering versus set ordering: a discussion

We return here to the issue of the definability of the plausibility ordering and set ordering in terms of each other. The discussion will be to a certain extent informal. We have not fully researched it.

There are various possible ways of interpreting $X >_B Y$ in plausibility models. The following option immediately comes to mind: the interpretation of $X >_B Y$ is that there exist X-worlds which are more plausible than any Y-world (similar to the proposal in [26]). This will not do because the sufficient belief condition does not follow, one needs that if X contains all worlds in the center and Y does not, then $X >_B Y$. If one defines $X >_B Y$ as saying that X contains all worlds in the center and Y does not, then $\varphi \succ_B \psi$ becomes equivalent to $B\varphi \wedge \neg B\psi$ erasing all distinctions we would like to make. The disjunction of these two options would lead $X >_B Y$ to be equivalent to $(B\varphi \wedge \neg B\psi) \vee (\neg B\neg\varphi \wedge B\neg\psi)$, not very attractive either.

A more complicated option is the following. Let us call a set in a plausibility model a *layer* if it is an equivalence class w.r.t. $\leq$, i.e. it contains all worlds in the same ordering as one particular one. The center is then the layer with highest plausibility. We now take the disjunction above spread over all layers. Let us call X *better than* Y *in layer* Z if there are either some X-worlds and no Y-worlds in Z, or all worlds in Z are X-worlds and not all are Y-worlds, in symbols $(Z \cap X \neq \emptyset \wedge Z \cap Y = \emptyset) \vee (Z \subseteq X \wedge \neg (Z \subseteq Y))$. And we may call X *equivalent to* Y *in layer* Z if neither X is better than Y nor Y better than X in Z, in symbols $(Z \subseteq X \wedge Z \subseteq Y) \vee (Z \cap X = \emptyset \wedge Z \cap Y = \emptyset) \vee (Z \cap X \neq \emptyset \wedge Z \cap Y \neq \emptyset \wedge Z \cap \overline{X} \neq \emptyset \wedge Z \cap \overline{Y} \neq \emptyset)$. One might then define $X >_B Y$ iff there exists a layer Z such that X is better than Y in Z, whereas X and Y are equivalent in all layers with higher plausibility. This definition will give the right properties to $X >_B Y$ but clearly a flavor of arbitrariness remains (compare the discussion in Section 3.3). We have shown in this paper that an independent set ordering is natural and fruitful in the study of stronger belief, and we do think that attempts as the above to define it in terms of the plausibility ordering are too artificial.

An attempt in the other direction, to define the plausibility ordering $s < t$ as $\{t\} >_B \{s\}$ fails in first instance because we cannot guarantee that if $s \in \mathcal{B}$ and $t \notin \mathcal{B}$, then $\{t\} >_B \{s\}$. This does however follow if we adopt axiom (2) of Section 2.3, as is explained there. This makes us advocate to take the set ordering to be the primary ordering, and to define the plausibility ordering in terms of it, adopting principle (2). There is one catch here, there is no reason that all the worlds in the center get the same maximal degree of plausibility. For our intuitions this is not a great problem, but it does mean a definite obstacle in making our system dynamic, since in the standard plausibility models the center consists of the most plausible worlds, and this fact provides the means to single out the new center after a public announcement or other information has been received. We do have ideas to solve this problem. We can show that any model for $KD45-O$ can be transformed into one that adheres to this standard condition but still satisfies the same formulas. So it may be reasonable to restrict the attention to these models, but that is for a future occasion.

8 Conclusion and further work

An explicit ordering of formulas to compare the strengths of belief is introduced in this paper. A complete axiomatization for this belief logic with explicit ordering is provided with respect to a semantics that includes a set ordering in addition to the standard plausibility ordering. The notion aids in giving intuitive formulations for various related concepts as well as some other epistemic attitudes - much older and thoroughly discussed notions like *universality* and *preference*, together with relatively newer ones like *plausibility* and *disbelief*. Independent axiomatizations for the logics of plausibility, belief and plausibility as well as belief and disbelief are also provided. Interplay of belief ordering with the concept of safe beliefs is discussed. Relative expressive powers of the proposed logics have been discussed as well. Lastly, we lift the proposed framework to a multi-agent setting.

In Section 7 we advocate the usage of set ordering as a more fundamental ordering and provide pointers towards defining plausibility ordering in terms of the set ordering, so that all the intuitive properties of world ordering can still be satisfied. We discuss the possibilities of providing a dynamic version of the present work. This seems definitely promising, but as is indicated there it is connected with how one sees the relationship between the plausibility ordering and the new set ordering.

Acknowledgements: We thank the anonymous referees of FAMAS 2009 as well as the L. J. of the IGPL special issue on FAMAS 2009 for their close reading and extensive comments, which helped us to improve the paper. We also thank Alexandru Baltag and Sonja Smets for presenting our paper at FAMAS 2009, and their subsequent comments which inspired us in writing this version. The second author thanks the Center for Soft Computing Research, Indian Statistical Institute, Kolkata for the congenial atmosphere she had there while doing this work during February, 2008 to February, 2009 and also in August, 2009. She also acknowledges NWO grant # 600.065.120.08N201.

References

[1] S. Artemov and L. Beklemishev, "Provability logic," in *Handbook of Philosophical Logic, 2nd ed.*, D. Gabbay and F. Guenthner, Eds. Kluwer, Dordrecht, 2004, vol. 13.

[2] A. Baltag and S. Smets, "A qualitative theory of dynamic interactive belief revision," in *Logic and the Foundations of Game and Decision Theory, Texts in Logic and Games*, G. Bonanno, W. van der Hoek, and M. Wooldridge, Eds., vol. 3. Amsterdam University Press, 2008, pp. 9–58.

[3] J. van Benthem, "Dynamic logic for belief revision," *Journal of Applied Non-Classical Logic*, vol. 17, no. 2, pp. 129–155, 2007.

[4] O. Board, "Dynamic inetractive epistemology," *Games and Economic Behavior*, vol. 49, pp. 49–80, 2004.

[5] J. Burgess, "Probability logic," *Journal of Symbolic Logic*, vol. 34, no. 2, pp. 264–274, 1969.

[6] M. Chakraborty and S. Ghosh, "Belief-disbelief interface: A bi-logical approach," *Fundamenta Informaticae*, to appear.

[7] B.F. Chellas, *Modal logic: An introduction.* C.U.P., 1980.

[8] S. Chopra, J. Heidema, and T. Meyer, "Logics of belief and disbelief," in *Proceedings of the ninth International Workshop on Non-Monotonic Reasoning*, 2002.

[9] N. Friedman and J. Halpern, "Plausibility measures and default reasoning," *Journal of the ACM*, vol. 48, no. 4, pp. 648–685, 2001.

[10] P. Gärdenfors, "Qualitative probability as an intentional logic," *Journal of Philosophical Logic*, vol. 4, pp. 171–185, 1975.

[11] P. Gärdenfors and D. Makinson, "Revisions of knowledge systems and epistemic entrenchment," in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, M. Vardi, Ed. Los Altos: Morgan Kaufmann, 1988, pp. 83–95.

[12] J. Gerbrandy, "Bisimulation on planet Kripke," Ph.D. dissertation, University of Amsterdam, 1999.

[13] A. Ghose and R. Goebel, "Belief states as default theories: Studies in non-prioritised belief change," in *Proceedings of the 13th European Conference on Artificial Intelligence*, H. Prade, Ed., 1998, pp. 8–12.

[14] A. Gomolinska, "On the logic of acceptance and rejection," *Studia Logica*, vol. 60, pp. 233 – 251, 1998.

[15] A. Gomolinska and D. Pearce, "Disbelief change," *Electronic essays on the occasion of the fiftieth birthday of Peter Gärdenfors*, 1999.

[16] V. Goranko and S. Passy, "Using the universal modality: Gains and questions," *Journal of Logic and Computation*, vol. 2, no. 1, pp. 5–30, 1992.

[17] D. Guaspari and R. Solovay, "Rosser sentences," *Annals of Mathematical Logic*, vol. 16, pp. 81–89, 1979.

[18] J. Halpern, "Defining relative likelihood in partially-ordered preferential structures," *Journal of AI Research*, vol. 7, pp. 1–24, 1997.

[19] ——, *Reasoning About Uncertainty.* MIT Press, 2003.

[20] J. Halpern and Y. Moses, "A guide to completeness and complexity for modal logics of knowledge and belief," *Artificial Intelligence*, vol. 54, pp. 319–379, 1992.

[21] H.H. Hansen, "Monotonic modal logics," *ILLC-prepublication series*, PP-2003-24.

[22] J. Hintikka, *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press, 1962.

[23] D. de Jongh, "A simplification of a completeness proof of Guaspari and Solovay," *Studia Logica*, vol. 46, pp. 187–192, 1987.

[24] D. de Jongh and F. Liu, "Preference, priorities and belief," in *Preference Change*, T. Gruene-Yanoff and S. Ove Hansson, Eds., pp. 85–108, 2009.

[25] S. Kripke, "Semantical considerations on modal logics," *Acta Philosophica Fennica*, vol. 16, pp. 83–94, 1963.

[26] D. Lewis, *Counterfactuals*. Blackwell and Harvard U.P., 1973.

[27] F. Liu, "Changing for the better: Preference dynamics and agent diversity," Ph.D. dissertation, University of Amsterdam, 2008.

[28] R. Moore, "Semantical considerations on nonmonotonic logic," *Artificial Intelligence*, vol. 25, pp. 75–94, 1985.

[29] K. Segerberg, "Qualitative probability in a modal setting," in *Proceedings of the 2nd Scandinavian Logic Symposium*, J. Fenstad, Ed. Amsterdam: North-Holland, 1971.

[30] W. Spohn, "Ordinal conditional functions. a dynamic theory of epistemic states," in *Causation in Decision, Belief Change, and Statistics*, W. Harper and B. Skyrms, Eds. Kluwer, Dordrecht, 1988, vol. II.

[31] R. Stalnaker, "On logics of knowledge and belief," *Philosophical Studies*, vol. 128, no. 1, pp. 169–199, 2006.

[32] G. von Wright, "Deontic logic," *Mind*, vol. 60, pp. 1–15, 1951.

Received 7 January 2011.