Comparing strengths of beliefs explicitly

Dick de Jongh

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Email: D.H.J.deJongh@uva.nl

Sujata Ghosh

Institute of Artificial Intelligence
Rijksuniversiteit Groningen
Email: Sujata.Ghosh@rug.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. This explicit mention of the comparison in the logical language aids in defining several other concepts in a uniform way, viz. older and rather clear concepts like the operators for universality (which possibilities ought to be considered), together with newer notions like plausibility (in the sense of 'more plausible than not') and disbelief. Moreover, it assists in studying the properties of the concept of greater strength of belief itself. A heavy part 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. Finally, after a brief discussion on the multi-agent setting, we move on to talk about the dynamics - the change of ordering under the influence of hard and soft information.**

Index Terms—**doxastic logic, belief, disbelief, plausibility**

I. 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 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 to take neither and hold another round of applications.

All these 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 can argue that these situations can be described by the very well-studied notion of *preference*, but the essence of describing the mental states of Alice will be lost then. This paper addresses the notion of comparison of the strength of beliefs of an agent directly. A great volume of literature and extensive philosophical debates are available on reasoning about knowledge and belief of agents. 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 ordering introduced here operates on formulas.

The introduction of explicit notions of ordering for comparing strengths of beliefs in the logical language has various applications. It aids in defining several other concepts in a uniform way, viz. older and rather clear concepts like the operators for universality, together with newer notions like plausibility and disbelief. Moreover it assists in studying the properties of the concept of greater strength of belief itself. 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 more evident that the universality operator cannot be identified with the knowledge operator even if they both share the $S5$-properties. 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. A pleasant fact is that we can fit the system easily into the framework of dynamic epistemic logic ([1], see also [2]) as explained in section III.

As mentioned in a brief interlude later, the explicit belief ordering also aids in providing an additional feather to the already existing close relationship between beliefs and preferences which is thoroughly discussed in [3].

Before entering into the actual study, let us first discuss the previous work on provability logics which inspired this idea of explicit belief ordering. In Provability Logic, an extensive overview of which can be found in [4], $\Box\varphi$ is interpreted as $\exists x\, Proof_{\mathbf{PA}}(x,\varphi)$, which means that there is a proof x in the axiomatization **PA** (Peano Arithmetic) for the statement φ. With **PA** understood, we also write it in the form - $\exists x\, Proof(x,\varphi)$. In this framework, to handle Rosser's form of incompleteness, one introduces witness comparisons of proofs with formulas like $\Box\varphi \preccurlyeq \Box\psi := \exists x(Proof(x,\varphi) \wedge \forall y < x \,\neg Proof(y,\psi))$, which is interpreted as that φ has a proof that is at least as small as any (possible) proof of ψ. Similarly, $\Box\varphi \prec \Box\psi := \exists x(Proof(x,\varphi) \wedge \forall y \leq x \,\neg Proof(y,\psi))$ means that φ has a proof that is smaller than any (possible) proof of ψ. Logics for these witness comparison formulas were successfully introduced by [5], and the completeness proof for this logic was simplified in [6].

Motivated by the ideas above, 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 proofs of φ and ψ are compared only if one of these proofs really exists, whereas strengths of beliefs are also discussed when neither φ nor ψ are really believed, which 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.

Let us now mention some related works in this area. In [7], [8], 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 [8] takes the explicit ordering operator in a simple language consisting of the truth-functional connectives only, [7] discusses this issue in a modal setting. Related as well are Lewis's proposal for a plausibility ordering of formulas [9] and Spohn's work on giving plausibility ordering of possible worlds in terms of ordinal functions [10]. More recently, we find some similar work in the economics literature [11]. Also, the notion of epistemic entrenchment [12] 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. In our setting the dynamics operates rather differently, as the reader will see.

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. Several possible interpretations of the belief-ordered formulas, viz. *plausibility* and *disbelief* are discussed, together with the inter-relationship of these ordering formulas and *safe belief*, and also, *preference*. Complete axiomatizations of the new belief logic with explicit ordering ($KD45-O$), with safe belief added ($KD45-OS$), plausibility logic (P-logic), logic of belief and plausibility (BP-logic) and logic of belief and disbelief (BD-logic) are provided. The section ends with a short discussion on the multi-agent setting. Section III brings in dynamics to the whole framework and discusses the influence of hard information as well as soft information over these ordering formulas, and provides complete axiomatizations to the dynamic logics under consideration. The conclusions are drawn in section IV.

II. COMPARING STRENGTH 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 [13] is generally accepted as initiating this line of research, which was further extended by [14]. 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 [15] has been used to model knowledge as well as belief. An extensive discussion together with all pre-requisite definitions can be found in [16]. 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 [2], [17] 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 [17].

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 disbeliefs, 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
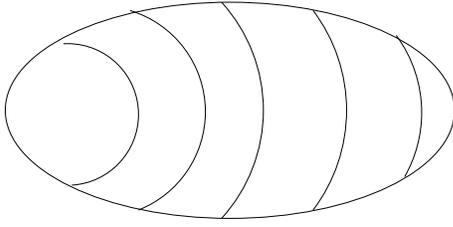
Fig. 1.   Plausibility ordering

of *Belief logic with explicit ordering* ($KD45{-}O$), whose language is defined as follows:

Definition II.1 *Given a countable set of atomic propositions Φ, formulas φ are defined inductively:*

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

where $p \in \Phi$.

The intuitive reading of the formula $B\varphi$ is "φ is believed", and that of $\varphi \succeq_B \psi$ is "belief in φ is at least as strong as belief in ψ". We introduce the notations $\varphi \succ_B \psi$ for $(\varphi \succeq_B \psi) \wedge \neg(\psi \succeq_B \varphi)$ and $\varphi \equiv_B \psi$ for $(\varphi \succeq_B \psi) \wedge (\psi \succeq_B \varphi)$. 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 II.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 $\mathcal{B} \not\subseteq Y$ 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 more plausible, the second one that the sets containing the center are more plausible than those not containing it, and 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.

[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.

$\mathcal{M}, s \models \varphi \succeq_B \psi$ iff $\{t \mid \mathcal{M}, t \models \varphi\} \geq_B \{t \mid \mathcal{M}, t \models \psi\}$.

We consider $\succeq_B$ to be a global notion, if $\varphi \succeq_B \psi$ is true anywhere in the model, it is true everywhere. So, it is either true or false throughout the whole model; $\succeq_B$ 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,

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

Thus, $\succ_B$ is also a global notion. We will now show that the universal modality U can also be expressed in $KD45{-}O$. The modality $E\varphi$ (the abbreviated form of $\neg U \neg \varphi$) can be defined as $\varphi \succ_B \bot$, and hence $U\varphi$ itself as $\bot \succeq_B \neg\varphi$. To clarify matters we should mention here - $U\varphi$ 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 clearly which worlds should be considered in the model: those worlds 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,

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

Thus U is definable in the language above, but to get the properties of the universal modality, we will need to have the $S5$-axioms that hold for U [18] plus the axiom $B\varphi \to UB\varphi$, which expresses that B is a global notion in a model where U expresses universality. In such a model there is only one center.

There are various possible ways of interpreting the formula $\varphi \succeq_B \psi$ in *plausibility models* expressing the belief modality. The foremost question is whether to try to define semantically $\varphi \succ_B \psi$ in terms of the plausibility ordering of the worlds. If one wants to base the strength of belief ordering on the plausibility ordering of the worlds, then immediately the following option comes to mind: the interpretation of $\varphi \succ_B \psi$ can be that there exist φ-worlds which are more plausible than any ψ-world (similar to the proposal in [9]). For $\varphi \succeq_B \psi$ it can be given as follows: for each ψ-world there exist φ-worlds which are at least as plausible. If one does this however, $B\varphi$ becomes equivalent to $\varphi \succeq_B \neg\bot$. In fact, no distinction in strength of belief can be made between propositions which are believed. This is not at all our aim. A more pleasurable consequence of this definition is that conditional belief $B^\psi\varphi$ can equivalently be expressed as - $(\psi \wedge \varphi) \succ_B (\psi \wedge \neg\varphi)$. Ultimately though we think that the price is too high: this interpretation of the ordered formulas gives rise to more unintended validities in the logic $KD45{-}O$ (cf. section II-C). There are more sophisticated reductions of strengths of beliefs to the plausibility ordering of the worlds but we think that they will all have undesirable consequences. This is only partly due to the fact that belief is identified with truth in the most plausible worlds, which

seems necessary for a dynamic interpretation. To do away with this issue we decided to define another *set-plausibility* ordering $\geq_B$ between the sets of worlds in the *plausibility models*. We have put very minimal requirements on this ordering. In fact, as we will see in section II-B, *belief* can be interpreted in terms of *belief ordering* and *safe beliefs*. How this will relate the world ordering and the set ordering is an interesting study, but we leave it for the future.

We should mention here that the idea of modeling epistemic notions in terms of set orders is not really new. In [19], [20], preferential structures are considered where a preference ordering over worlds is lifted to an ordering of sets of worlds. Plausibility measures are considered in [21] to give a semantics of default logic. These measures can be identified with a partial ordering on sets of worlds and they also provide an interpretation of the notion of beliefs. We are not going into a detailed comparison of these works with ours due to lack of space, but just note that an essential difference is that in our case the ordering of the sets of worlds is only partly determined by the ordering of the worlds.

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) $\wedge$ ¬B(Deirdre) $\wedge$ ¬B(¬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(¬Harold)$\wedge$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. Even in the fourth example where Alice believes that both Harold and Irma are unable to perform, her belief in the ability of one might be higher than that of the other. Then, if forced to choose, she could do without another round of applications.

A. Axioms and Completeness

Let us first look into some interesting validities of this logic.

- $(\varphi \succ_B \psi) \rightarrow E(\varphi \wedge \neg\psi)$
- $\varphi \succcurlyeq_B \varphi \wedge \psi$
- $\varphi \vee \psi \succcurlyeq_B \varphi$
- $B(\varphi \succcurlyeq_B \psi) \rightarrow (\varphi \succcurlyeq_B \psi)$

Before providing a complete axiomatization of $KD45-O$, we discuss the motivations behind some of these axioms. 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 following axioms need no introduction:

$\varphi \succcurlyeq_B \varphi$ (refl-axiom)
$(\varphi \succcurlyeq_B \psi) \wedge (\psi \succcurlyeq_B \chi) \rightarrow \varphi \succcurlyeq_B \chi$ (trans-axiom)
$(\varphi \succcurlyeq_B \psi) \vee (\psi \succ_B \varphi)$ (lin-axiom)

From these axioms, it follows that,

$(\varphi \succ_B \psi) \wedge (\psi \succ_B \chi) \rightarrow (\varphi \succ_B \chi)$

and so we get the transitivity of $>_B$. We have already seen that $E\varphi$ can be defined as $\varphi \succ_B \bot$, because of the 3rd condition that $>_B$ satisfies in the model. Using lin-axiom it is easy to show that $U\varphi$ is equivalent to $\bot \succcurlyeq_B \neg\varphi$. The following axiom takes care of the 2nd condition.

$(B\varphi \wedge \neg B\psi) \rightarrow (\varphi \succ_B \psi)$ (center-axiom)

Since the ordering formulas are either globally true or globally false in the model, we have:

$(\varphi \succcurlyeq_B \psi) \rightarrow B(\varphi \succcurlyeq_B \psi)$ (intros-axiom1)
$(\varphi \succ_B \psi) \rightarrow B(\varphi \succ_B \psi)$ (intros-axiom2)

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 inverses of all these implications above follow from the lin-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. As a result, the inclusion formula concerning the belief and the universal modality, viz. $U\varphi \rightarrow B\varphi$ also follows. The following axiom and rule take care of replacement with equivalent formulas in the ordering formulas:

$U(\varphi \rightarrow \psi) \rightarrow (\psi \succcurlyeq_B \varphi)$ ($U \succcurlyeq_B$ -axiom)
$\dfrac{\varphi \rightarrow \psi}{\psi \succcurlyeq_B \varphi}$ (inclusion rule)

We can conclude then

$\dfrac{\varphi}{U\varphi}$ (Ugen-rule)

and the equivalence rule:

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

which means that logically equivalent formulas can be substituted for each other in the ordering formulas as well and hence everywhere. We now have the gen-rule for the universal modality, the K-axiom for U also follows from these principles. We move on to get the other sound ordering axiom, which will help us to get the $S5$-properties of the universal modality U [18]. It is:

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

The existence axiom is basically the equivalent ordered formula for $U\varphi \rightarrow \varphi$. We end with the axiom which forces the $KD45-O$ models to have a unique center $\mathcal{B}$ and hence makes B a global property.

$EB\varphi \rightarrow B\varphi$ (un.center-axiom)

From this the principle $B\varphi \rightarrow UB\varphi$ readily follows. The transitivity and the symmetry axioms for U follow because of the very significant property of $U\varphi$ being a B-statement, the un.center-axiom applies to U-statements as well.

Thus we have the following theorem which is the most basic and important result of this work.

Theorem II.3 $KD45-O$ *is sound and complete with respect to* $KD45-O$ *models and its validities are completely axiomatized by the following axioms and rules:*

a) *all KD45 axioms and rules*
b) *ordering axioms:*

$\varphi \succeq_B \varphi$ *(refl-axiom)*

$(\varphi \succeq_B \psi) \wedge (\psi \succeq_B \chi) \rightarrow \varphi \succeq_B \chi$ *(trans-axiom)*

$(\varphi \succeq_B \psi) \vee (\psi \succ_B \varphi)$ *(lin-axiom)*

$(B\varphi \wedge \neg B\psi) \rightarrow (\varphi \succ_B \psi)$ *(center-axiom)*

$(\varphi \succeq_B \psi) \rightarrow B(\varphi \succeq_B \psi)$ *(intros-axiom1)*

$(\varphi \succ_B \psi) \rightarrow B(\varphi \succ_B \psi)$ *(intros-axiom2)*

$U(\varphi \rightarrow \psi) \rightarrow (\psi \succeq_B \varphi)$ *($U \succeq_B$ -axiom)*

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

$EB\varphi \rightarrow B\varphi$ *(un.center-axiom)*

c) *inclusion rule:*

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

Proof: The readers can easily verify the soundness of these ordering axioms. The $U \succeq_B$ -axiom is covered by property 1 of Definition II.2, existence axiom by property 3, and the center axiom describes property 2. With U and E defined as indicated previously, one can easily show that the $S5$-axioms are derivable for U. It is also not very hard to show completeness using finite sets of sentences.

Assume $\nvdash_{KD45-O} \varphi$. We will have to construct a counter-model to φ as a $KD45-O$-model. We take a finite *adequate* set Φ containing φ. In this case an adequate set will be: a set of formulas that is closed under subformulas containing with each formula ψ (a formula equivalent 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 $B\varphi$ a formula (equivalent to) $UB\varphi$. 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 $\neg\varphi$. When we now refer to $U\psi$ we mean its translation into $KD45-O$. It can be shown that the $S5$-axioms to hold for this translation. The proof is made more perspicuous by referring to U.

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

$P\mathcal{R}_B Q$ iff   (1) for all $B\varphi$ in P, φ as well as $B\varphi$ are in Q,
(2) for all $\neg B\varphi$ in P, $\neg B\varphi$ in Q.

$P\mathcal{R}_U Q$ iff   (1) for all $U\varphi$ in P, φ as well as $U\varphi$ are in Q,
(2) for all $\neg U\varphi$ in P, $\neg U\varphi$ in Q

We have to show that $\mathcal{R}_U$ is an equivalence relation and $\mathcal{R}_B$ a Euclidean sub-relation 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 $B\varphi \rightarrow UB\varphi$ implies that this set 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 $\psi \succeq_B \chi$ is true in the model, $V(\psi) >_1 V(\chi)$ iff $\psi \succ_B \chi$ is true in the model. These follow from the first three ordering axioms.

Moreover, if $V(\psi) \subseteq V(\chi)$ then $V(\psi) \geq_1 V(\chi)$ (subset condition), by the axiom: $U(\chi \rightarrow \psi) \rightarrow \psi \succeq_B \chi$. 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: $B\psi \wedge \neg B\chi \rightarrow \psi \succ_B \chi$.

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*. So, once we finish the proof of the following lemma, we are done.

Lemma. $\mathcal{B}$ is representable.

Proof of Lemma. 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. Note that this implies that ψ_w is 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 while ψ is true in all elements of $\mathcal{B}$, but 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. ∎

Since the counter-model constructed is finite, we also have that the logic $KD45-O$ is decidable. Before ending this section we mention some intuitively true formulas, which we did not need as axioms, but are definitely worth thinking about. One of them is,

$$(\varphi \succ_B \bot) \to (\top \succ_B \neg\varphi),$$

which says that if φ is true somewhere then $\neg\varphi$ is not as much to be believed as a tautology. The other direction of the implication can be derived. An equivalent formulation is,

$$(\varphi \succcurlyeq_B \top) \to (\bot \succcurlyeq_B \neg\varphi).$$

To make this true, the model needs an extra clause 4, saying that,

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

This seems a very reasonable addition as it makes the models more symmetric. A more general version of this possible axiom is,

$$(\varphi \succ_B \psi) \to (\neg\psi \succ_B \neg\varphi).$$

which, if considered definitely increases the already-existing probabilistic flavor of the axiomatization. Another possible principle with a similar flavor is

$$(\varphi \succ_B \psi) \to (\varphi \wedge \neg\psi) \succ_B (\psi \wedge \neg\varphi).$$

This one exemplifies the notion that if φ is more believed than ψ, then that should be based on the idea that the common part of φ and ψ is irrelevant in the estimation of their relative strengths of belief. Readers can note here that if we strengthen this formula above to its bi-implication, then $(\varphi \succ_B \psi) \to (\neg\psi \succ_B \neg\varphi)$ follows.

B. Safe Belief

The notion of 'safe belief' has been introduced in [17] which also corresponds to "Stalnaker knowledge" [22], where evidence is considered as true information. 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." 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. We already have an understanding of the interplay between beliefs and the comparison of strength of beliefs. Our study will be incomplete, if we do not investigate the lively relationship between the very relevant and important issue of safe beliefs together with our notion of belief orderings.

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

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

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 II.4 *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 $KD45-OS$. Together with the axioms and rules of the $KD45$-logic of beliefs, and the relevant ordering axioms, viz. refl, trans, lin, 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,

$$
\begin{array}{ll}
(\Box\varphi \wedge \neg\Box\psi) \to (\varphi \succ_B \psi) & (\Box\text{order-axiom}) \\
(\varphi \succcurlyeq_B \psi) \to \Box(\varphi \succcurlyeq_B \psi) & (\Box\text{intros-axiom1}) \\
(\varphi \succ_B \psi) \to \Box(\varphi \succ_B \psi) & (\Box\text{intros-axiom2})
\end{array}
$$

In addition to all these, the following axiom relates the operator $\Box$ with B.

$$\Box\varphi \to B\varphi \quad (\Box B\text{-axiom})$$

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

$$U\varphi \to \Box\varphi$$

Regarding the $\Box$order-axiom, it should be mentioned that, unlike belief (center axiom), relating safe belief and belief ordering in this manner may be considered questionable. It says that, if φ is safely believed and ψ is more strongly believed than φ, then ψ can also be safely believed. This seems alright at a first glance, but if we consider the subjectivity of the ordering, this axiom may lead to some dispute. Still, technical reasons make it very desirable, the relationship between the world and set orderings becomes much closer, so we decided to keep this axiom. So, we have the following theorem.

Theorem II.5 *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:*

$\varphi \succcurlyeq_B \varphi \quad$ *(refl-axiom)*

$(\varphi \succcurlyeq_B \psi) \wedge (\psi \succcurlyeq_B \chi) \to \varphi \succcurlyeq_B \chi \quad$ *(trans-axiom)*

$(\varphi \succcurlyeq_B \psi) \vee (\psi \succ_B \varphi) \quad$ *(lin-axiom)*

$(\bot \succcurlyeq_B (\neg(\Box\varphi \to \Box\psi))) \vee (\bot \succcurlyeq_B \neg(\Box\psi \to \Box\varphi))$

($\Box$lin-axiom)

$(B\varphi \wedge \neg B\psi) \to (\varphi \succ_B \psi) \quad$ *(center-axiom)*

$$(\Box\varphi \wedge \neg\Box\psi) \rightarrow (\varphi \succ_B \psi) \quad (\Box\text{order-axiom})$$
$$(\varphi \succeq_B \psi) \rightarrow \Box(\varphi \succeq_B \psi) \quad (\Box\text{intros-axiom1})$$
$$(\varphi \succ_B \psi) \rightarrow \Box(\varphi \succ_B \psi) \quad (\Box\text{intros-axiom2})$$
$$U(\varphi \rightarrow \psi) \rightarrow (\psi \succeq_B \varphi) \quad (U \succeq_B \text{-axiom})$$
$$\varphi \rightarrow E\varphi \quad (\text{existence axiom})$$

d) $\Box\varphi \rightarrow B\varphi \quad (\Box B\text{-axiom})$

e) *inclusion rule:*
$$\frac{\varphi \rightarrow \psi}{\psi \succeq_B \varphi}$$

We should mention here that, according to [17], belief and conditional belief can be expressed in terms of knowledge and safe belief as,

$$B^\psi\varphi := \hat{K}\psi \rightarrow \hat{K}(\psi \wedge \Box(\psi \rightarrow \varphi)),$$
$$B\varphi := B^\top\varphi,$$

where $\hat{K}\psi := \neg K\neg\psi$. They gave complete axiomatizations for conditional doxastic logic (logic of conditional belief) as well as the logic of knowledge and safe beliefs. We do not consider knowledge but for this part of the discussion it can be replaced by U. Neither do we talk about conditional belief here, but belief can be defined in terms of the existential modality and safe belief (i.e. in terms of safe belief and belief ordering) as follows:

$$B\varphi := E\Box\varphi$$

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 $\succeq_B$ remains to be seen.

C. Plausibility

Comparing the strength of beliefs explicitly has its various advantageous applications. 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 is $U(\varphi \rightarrow \psi) \rightarrow (P\varphi \rightarrow P\psi)$. This holds because if $U(\varphi \rightarrow \psi)$, not only will belief in ψ be at least as strong as in φ, but $U(\varphi \rightarrow \psi)$ implies $U(\neg\psi \rightarrow \neg\varphi)$, so belief in $\neg\psi$ is not greater than in $\neg\varphi$. This leads to consequences like $P(\varphi \wedge \psi) \rightarrow P\varphi$.

The reason to take the set semantics for ordering formulas (cf. Definition II.2) becomes clear. If we would adhere to the semantics we may have had for $\succ_B$ in terms of plausibility ordering for worlds (instead of sets of worlds), $P\varphi$ would become equivalent to $B\varphi$, which obviously is undesirable.

One can just subdivide the most plausible worlds (the center) into more and less plausible ones to rectify this, but besides endangering the transition to dynamics this will not yet be really satisfactory in its own right. It will result in interpreting $P\varphi$ into something like 'φ is weakly believed'. This would make the modal logic of P a normal modal logic (of weak belief). In particular $P\varphi \wedge P\psi \rightarrow P(\varphi \wedge \psi)$ would become valid, which is not very intuitive.

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 doesn't 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.

We now move on to showing 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 the $KD45-O$ model is given by,

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

Theorem II.6 *P-logic is complete and its validities are completely axiomatized by the following axioms and rules:*

(a) *all propositional tautologies and inference rules*

(b) *plausibility axioms:*
$$P\psi \wedge P\varphi \rightarrow P(\psi \wedge P\varphi)$$
$$P\psi \wedge \neg P\varphi \rightarrow P(\psi \wedge \neg P\varphi)$$
$$P\varphi \rightarrow \neg P\neg\varphi$$
$$P\top$$

c) *monotonicity rule:*
if $\varphi \rightarrow \psi$ then $P\varphi \rightarrow P\psi$

Proof: First of all, we show that any formula in P-logic is equivalent to a formula with P-depth at most one. For that purpose we first derive the following schemes:

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

Here, $\varphi[\top/P\psi]$ means φ with $\top$ substituted for some occurrences of $P\psi$. We prove 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 and propositional constants, 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. From the induction hypothesis for the first scheme it follows that $(P\psi \wedge \varphi) \leftrightarrow (P\psi \wedge \varphi[\top/P\psi])$ is provable. Now assume $P\psi$ and $P\varphi$. By an axiom $P(\varphi \wedge P\psi)$ follows. From the

fact just proved 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 $((P\psi \wedge \varphi[\top/P\psi]) \vee (\neg P\psi \wedge \varphi[\bot/P\psi]))$. By doing this consecutively for all occurrences of $P\psi$ with no occurrences of P in ψ we obtain the desired result.

Next, we show that any consistent set has a model. Assume we have a consistent set in the P-logic which can be extended to a maximal consistent set Γ, say. 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, such $P\varphi$'s and $\neg P\varphi$'s and their boolean combinations.

We now make a model in our sense where $P\varphi$ gets interpreted as $\varphi >_B \neg\varphi$. The worlds will be simply 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 maximal consistent subsets. 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.
(1) The whole set, which is of course represented by $\top$.
(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 sets represented by those φ for which $P\neg\varphi$ is in Γ (except for $\bot$).
(5) The empty set, which is of course represented by $\bot$. These are all possibilities because of axiom $P\varphi \rightarrow \neg P\neg\varphi$. 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*. Similarly for the other classes. Second, that, if a set X contains all of $\mathcal{B}$, and another set Y doesn't, then $X > Y$. That is trivial: X has to be $\mathcal{B}$, the whole set, and Y isn't.

As earlier, we can induce an ordering over all subsets satisfying the required conditions. All the single worlds have to be taken to be equally plausible, i.e. $s \leq t$, for all $s, t \in S$. So, for each consistent set we can have a model in $KD45-O$. So, the axioms and rules given in Theorem 2.6 axiomatize the P-logic of 'more plausible then not'. It is also worth-mentioning why $(P\varphi \wedge P\psi) \rightarrow P(\varphi \wedge \psi)$ 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. Interestingly, the principles 4 and 5 for the modal operator P are derivable in this P-logic, but the K-axiom is not. That $P\varphi \wedge P\psi \rightarrow P(\varphi \wedge \psi)$ ought not to be a valid principle in the P-logic is clear if we interpret $P\varphi$ as $\varphi \succ_B \neg\varphi$.

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

We now consider a system having both belief and the plausibility operator, viz. the BP-system. This system 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 for belief, B.

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

Some validities of this logic in the $KD45-O$ model are,

- $B\varphi \rightarrow P\varphi$
- $P\varphi \rightarrow BP\varphi$
- $\neg P\varphi \rightarrow B\neg P\varphi$

Theorem II.7 *BP-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) *all P axioms and rules*
  d) *special axioms:*
     $B\varphi \rightarrow (\psi \leftrightarrow \psi[\top/B\varphi])$
     $\neg B\varphi \rightarrow (\psi \leftrightarrow \psi[\bot/B\varphi])$
     $B\varphi \rightarrow P\varphi$

The proof is very similar to that for the P-logic. It uses the fact 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. It is noteworthy that the principle $B\varphi \wedge P\psi \rightarrow P(\varphi \wedge \psi)$ of [23] fails in the BP-logic. It is not difficult to construct a counterexample.

D. 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 beliefs and belief revision, consideration of disbelief as a separate epistemic category came to fore in the latter part of last decade ([24], [25]). Consideration of changing or revising disbeliefs as a process analogous to belief revision was taken up by [26]. Belief-disbelief pairs i.e. simultaneous consideration of belief and disbelief sets were also taken up ([27], [28]) 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.

To exemplify the matter a bit, let us consider the following situation. Due to the unpredictable weather conditions, Pom's belief in that she should not cycle from Amsterdam to Leiden is much stronger than her belief that she should. When options like this are available, it is very natural to have this sort of ordering dilemma playing around people's mind. This can be interpreted as that Pom disbelieves that she should bike, which evidently implies that she does not believe that she should bike. But that 'she believes that she should not bike' is a much stronger statement, which fails to express the finer interplay of doubts that is always prevalent in one's mind.

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 [25] in the autoepistemic logic framework of [29]. 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 from '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 II.8 *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:*
    $D\varphi \to (\psi \leftrightarrow \psi[\top/D\varphi])$
    $\neg D\varphi \to (\psi \leftrightarrow \psi[\bot/D\varphi])$
    $D\varphi \to \neg D\neg\varphi$
    $D\bot$
  d) *special axioms:*
    $B\varphi \to (\psi \leftrightarrow \psi[\top/B\varphi])$
    $\neg B\varphi \to (\psi \leftrightarrow \psi[\bot/B\varphi])$
    $B\varphi \to D\neg\varphi$
  e) *anti-monotone rule:*
    if $\varphi \to \psi$ then $D\psi \to D\varphi$

Some interesting validities of this logic are,

- $B\neg\varphi \to D\varphi$
- $D\varphi \to \neg B\varphi$
- $D\varphi \to BD\varphi$
- $\neg D\varphi \to B\neg D\varphi$
- $\neg D\varphi \to DD\varphi$
- $\neg B\varphi \to DB\varphi$

On the other hand, as in P-logic and BP-logic, the corresponding intuitively incorrect principle, $D\varphi \wedge D\psi \to D(\varphi \vee \psi)$ can also be avoided in the BD-logic.

E. Preference

There is a very close relationship between an agent's beliefs and her preferences which has been extensively discussed in ([30], [3]). 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 first-order formulas with exactly one free variable, which is common to all of them. Preferences over objects can be defined in terms of these sequences. The basic idea is to define objective preference by:

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

For subjective preferences over objects, which in fact are considered to be influenced by beliefs, several options are considered. We mention a few of them for the benefit of the readers, their meanings are more or less obvious.

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

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

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

It is clear that the above three approaches are different attempts to express 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. Here we can express this directly in the language as below, and the way greater strength of belief is to be taken in a particular application is then delegated to the semantics.

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

F. Multi-agent case

We have been focusing on beliefs and strengths of beliefs of a single agent. The whole idea can be generalized to the multi-agent framework. We only give some preliminary ideas here. The technical details need to be worked out, and we leave it for the future. The language of the logic of belief ordering in the multi-agent case, $KD45-O_M$ can be defined as follows:

Definition II.9 *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 $p \in \Phi$.

The indices in the belief and universality 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 belief in the proposition was stronger than belief in a contradiction. With more agents we may have agents 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: there is no reason for $E_a E_b$ to be equivalent to E_a or E_b, etc. It is appropriate to add a real universality operator U that corresponds to the agent's U_a as the common knowledge operator corresponds to the knowledge of the individual agents. With regard to axioms the upshot is for example that the existence axiom $\varphi \rightarrow (\varphi \succ_{B_a} \bot)$ will have to be weakened to $U_a(\varphi \rightarrow (\varphi \succ_{B_a} \bot))$.

Likewise, the models for $KD45-O_M$ have to be 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
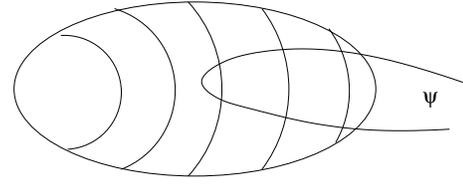

Fig. 2. A public announcement ψ is uttered

also, with the introduction of formulas like $U_a\varphi$. The notion of *comparative classes* [17] 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. As mentioned earlier, we leave the technical details for later.

III. DYNAMICS OF ORDERING FORMULAS

Till now we have been talking about the static language of belief ordering and its corresponding models, representing the information states (possible worlds) of an agent. We move on to discuss the effect of information-update procedures which change the models under consideration.

It should be mentioned here that we will see an extensive use of the notion of *conditional belief* while discussing the dynamics of belief change. Conditional beliefs [2] *pre-encode* beliefs in a static way in situations which arise after new information comes in. Formulas of the form $B^\psi\varphi$ (believing in φ, given ψ) express that we believe in φ, once we assume that ψ is the case. This induces that if ψ is learnt then it is to be believed that φ was the case (before the learning) [17]. This does not literally tell us what happens after ψ is learnt. For example, if φ is unknown but true and one learns that $\neg B\varphi \wedge \varphi$ then afterwards $B\varphi$ will be true, contrary to what was learnt.

In [2], van Benthem discusses three different update procedures influencing the beliefs of an agent, viz. the influence of *hard* information like *public announcement of ψ* (the not-ψ-worlds get deleted from the model and only the ψ-worlds remain), and of *soft* information with *lexicographic upgrade* (introducing a preference for the ψ-worlds, all the ψ-worlds become more plausible than all the not-ψ-worlds - within these zones, the existing ordering remains), and lastly the impact of soft information with *elite* upgrade (the best ψ-worlds come out on top, beside that the previous ordering remains). We currently just focus on the effect of hard information and soft information with lexicographic upgrade.
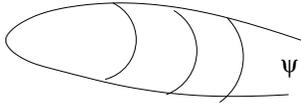
Fig. 3. The influence of hard information

A. Influence of hard information

While discussing the influence of 'hard' information over beliefs, [2] considered the following *reduction axioms* for the logic of public announcement in terms of 'belief' and 'conditional belief', where $[!\psi]\varphi$ is to be read as 'after the public announcement of ψ, φ is true'.

- $[!\psi]q \leftrightarrow (\psi \to q)$
- $[!\psi]\neg\varphi \leftrightarrow (\psi \to \neg[!\psi]\varphi)$
- $[!\psi](\varphi \wedge \chi) \leftrightarrow ([!\psi]\varphi \wedge [!\psi]\chi)$
- $[!\psi]B\varphi \leftrightarrow (\psi \to B^\psi[!\psi]\varphi)$
- $[!\psi]B^\chi\varphi \leftrightarrow (\psi \to B^{\psi\wedge[!\psi]\chi}[!\psi]\varphi)$

In fact, the effect of public announcements of ψ over an *ordered* model should be clear from the figures 2 and 3. Under the announcement of ψ, the earlier model $\mathcal{M}$, say (cf. fig 2), reduces to a model relativized to ψ (cf. fig 3), which is essentially a submodel of $\mathcal{M}$, whose domain set is the set where ψ holds.

Let us now first investigate how the $KD45-O$ model changes under the influence of a public announcement $!\psi$, say. The definition is as follows:

Definition III.1 *A $KD45-O$ model is defined as in definition II.2 as the structure $\mathcal{M} = (S, \leq, \geq_B, V)$. Under the influence of public announcement of ψ the model becomes $\mathcal{M}^{!\psi} = (S^{!\psi}, \leq^{!\psi}, \geq_B^{!\psi}, V^{!\psi})$ where $S^{!\psi} = \{s \in S : \mathcal{M}, s \models \psi\}$, $\leq^{!\psi} = \leq\!\lfloor_{S^{!\psi} \times S^{!\psi}}$, $\geq_B^{!\psi} = \geq_B\!\lfloor_{\mathcal{P}(S^{!\psi}) \times \mathcal{P}(S^{!\psi})}$, and $V^{!\psi} = V\!\lfloor_{S^{!\psi}}$.*

The model satisfies the following conditions:

1) If $X \subseteq Y \subseteq S^{!\psi}$, then $Y \geq_B^{!\psi} X$
2) if $\mathcal{B}^{!\psi}$ is the new set of plausible worlds, truth on which suffices to make an assertion to be believed, then $\mathcal{B}^{!\psi} \subseteq X \subseteq S^{!\psi} \wedge \mathcal{B}^{!\psi} \not\subseteq Y \subseteq S^{!\psi} \Rightarrow X >_B^{!\psi} Y$, where $>_B^{!\psi}$ denotes the corresponding strict ordering.
3) If X is non-empty, then $X >_B^{!\psi} \emptyset$.

The truth definitions of the formula $[!\psi]\varphi$ in a $KD45-O$ model is given by,

$$\mathcal{M}, s \models [!\psi]\varphi \text{ iff if } \mathcal{M}, s \models \psi, \text{ then } \mathcal{M}^{!\psi}, s \models \varphi.$$

Any logic L completing $KD45-O$ to contain conditional beliefs will have to contain

$$B^\psi\varphi \wedge \neg B^\psi\chi \to (\psi \wedge \varphi) \succ_B (\psi \wedge \chi).$$

This axiom ensures that the updated model still has the necessary property 2 of Definition II.2 and thus will be a $KD45-O$ model. We have the following theorem:

Theorem III.2 *If we take L to be a complete axiomatization of $KD45-O$ together with conditional beliefs, then its extension under public announcement is complete and its validities are completely axiomatized by the following axioms and rules in addition to L:*

(a) *PAL reduction axioms for atomic facts, Boolean operations, belief and conditional belief*

(b) *PAL reduction axioms for ordering formulas:*
$[!\psi](\varphi \succeq_B \chi) \leftrightarrow \psi \to ((\psi \wedge [!\psi]\varphi) \succeq_B (\psi \wedge [!\psi]\chi))$
$[!\psi](\varphi \succ_B \chi) \leftrightarrow \psi \to ((\psi \wedge [!\psi]\varphi) \succ_B (\psi \wedge [!\psi]\chi))$

B. Influence of soft information

Following [2], we considered the influence of *hard information* over beliefs. Van Benthem continued by discussing the effect of the arrival of *soft information* on beliefs, which does not influence the existence of the possible worlds, only their plausibility ordering. We now give the definition of the changed models under the influence of soft information with lexicographic upgrade and the corresponding reduction formulas.

Definition III.3 *A $KD45-O$ model is defined as in definition II.2 as the structure $\mathcal{M} = (S, \leq, \geq_B, V)$. Under the influence of soft information ψ, say the model becomes $\mathcal{M}^{\Uparrow\psi} = (S^{\Uparrow\psi}, \leq^{\Uparrow\psi}, \geq_B^{\Uparrow\psi}, V^{\Uparrow\psi})$ where $S^{\Uparrow\psi} = S$, $\leq^{\Uparrow\psi} = \leq \lfloor_{K\times K} \cup \leq\!\lfloor_{L\times L} \cup \{(u,v) : u \in K \text{ and } v \in L\}$, where $K = \{s \in S : \mathcal{M}, s \models \psi\}$, $L = S \setminus K$, $\geq_B^{\Uparrow\psi} = \geq_B$, and $V^{\Uparrow\psi} = V$.*

Once again, the model $\mathcal{M}^{\Uparrow\psi}$ satisfies the conditions,

1) If $X \subseteq Y$, then $Y \geq_B^{\Uparrow\psi} X$
2) if $\mathcal{B}^{\Uparrow\psi}$ is the new set of plausible worlds, truth on which suffices to make an assertion to be believed, then $\mathcal{B}^{\Uparrow\psi} \subseteq X \wedge \mathcal{B}^{\Uparrow\psi} \not\subseteq Y \Rightarrow X >_B^{\Uparrow\psi} Y$, where $>_B^{\Uparrow\psi}$ denotes the corresponding strict ordering.
3) If X is non-empty, then $X >_B^{\Uparrow\psi} \emptyset$.

The truth definitions of the formula $[!\psi]\varphi$ in a $KD45-O$ model is given by,

$$\mathcal{M}, s \models [\Uparrow \psi]\varphi \text{ iff } \mathcal{M}^{\Uparrow\psi}, s \models \varphi.$$

We have the following theorem:

Theorem III.4 *If we take L to be a complete axiomatization of $KD45-O$ with conditional beliefs, then its extension under announcement of soft information with lexicographic upgrade is complete and its validities are completely axiomatized by the following axioms and rules in addition to L:*

(a) *reduction axioms:*
$[\Uparrow \psi]q \leftrightarrow q$
$[\Uparrow \psi]\neg\varphi \leftrightarrow \neg[\Uparrow \psi]\varphi$
$[\Uparrow \psi](\varphi \wedge \chi) \leftrightarrow ([\Uparrow \psi]\varphi \wedge [\Uparrow \psi]\chi)$
$[\Uparrow \psi]B^\chi\varphi \leftrightarrow (E(\psi \wedge [\Uparrow \psi]\chi) \wedge (B^{\psi\wedge[\Uparrow\psi]\chi}[\Uparrow \psi]\varphi \vee B^{[\Uparrow\psi]\chi}[\Uparrow \psi]\varphi))$

$$[\Uparrow \ \psi](\varphi \succeq_B \chi) \ \leftrightarrow \ E\psi \wedge (((\psi \wedge [\Uparrow \ \psi]\varphi) \succeq_B (\psi \wedge [\Uparrow \ \psi]\chi)) \vee ([\Uparrow \psi]\varphi \succeq_B [\Uparrow \ \psi]\chi))$$

$$[\Uparrow \ \psi](\varphi \succ_B \chi) \ \leftrightarrow \ E\psi \wedge (((\psi \wedge [\Uparrow \ \psi]\varphi) \succ_B (\psi \wedge [\Uparrow \ \psi]\chi)) \vee ([\Uparrow \psi]\varphi \succ_B [\Uparrow \ \psi]\chi))$$

The product-update model [2] or the more general version of the action-priority update model [17] may not be of much use in the modelling of the ordering formulas, because the first order frame conditions of $KD45-O$-models are not all universal Horn sentences (e.g. the linearity condition), which is required for easy passage to this type of update models [2].

IV. Conclusion and Further Work

An explicit ordering of formulas to compare the strengths of beliefs is introduced. A complete axiomatization for this belief logic with explicit ordering is provided. This notion aids in giving intuitive formulations for various related concepts like universality 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. Lastly, we delved into the dynamics of this ordering concept, e.g. the effect of hard as well as soft information over the ordering formulas. A few possible avenues for future work are discussed below.

a) Interpreting the ordered formulas: We have provided different ways of interpreting these ordered belief formulas, and a complete axiomatization is provided with respect to a most general one. It will be interesting to find extensions for other possible applications and their properties, specially to get more interesting inter-connections between the world-ordering and the set-ordering.

b) Dynamic setting: As evident from the discussions in section 3, the whole system fits very well into the dynamic epistemic logic framework. But clearly this is the just the start of serious research in this area. There are various notions in this context that need to be thoroughly investigated, especially the relation to conditional belief. There are also important model-theoretic issues that one can look into. In short, a lot of possibilities have emerged with the introduction of this ordering of beliefs in the already existing dynamic logic framework dealing with epistemic attitudes.

References

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

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

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

[4] 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.

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

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

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

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

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

[10] 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.

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

[12] 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.

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

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

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

[16] 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.

[17] 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.

[18] 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.

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

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

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

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

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

[24] 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.

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

[26] A. Gomolinska and D. Pearce, "Disbelief change," *Electronic essays on the occasion of the fiftieth birthday of Peter Gardenfors*, 2001.

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

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

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

[30] D. d. Jongh and F. Liu, "Optimality, belief and preference," in *Proceedings of the Workshop on Rationality and Knowledge, ESSLLI*, S. Artemov and R. Parikh, Eds., 2006.