How to Revise Ranked Probabilities

Emil Weydert[1]

Abstract. In this paper, we introduce and discuss a new framework for the modeling and revision of probabilistic belief. The epistemic states encode degrees of belief topped by second-order uncertainty using special Spohn-type ranking measures over subjective probability distributions. The revision strategy, which handles incoming information translated into linear probability constraints, is based on variants of Jeffrey-conditionalization and information distance minimization.

1 INTRODUCTION

Belief revision, the modification of the epistemic structure when confronted to new, possibly conflicting information, is a fundamental ingredient of intelligent behaviour. Its formal modeling is therefore a central research theme in artificial intelligence. However, when talking about belief revision, it is important to keep in mind the distinction between reasoning about a changing world (e.g. about the effects of actions or events) and defeasible learning about a static or dynamic world. Both issues are relevant for epistemic kinematics, but their focus differs. The first task is subsumed by the more general problem of nonmonotonic reasoning about action and change. The second task is belief revision in its proper sense and constitutes the subject of this paper.

Traditional work in belief revision has been mainly concerned with qualitative propositional approaches. That is, the epistemic state is assumed to be a propositional belief set structured by some epistemic preference ordering. New information here takes the form of a proposition to be integrated into the belief set without giving up consistency. This is achieved under the control of the preference structure. Originally, not much thought was devoted to the question of how to revise the ordering itself, although this is absolutely necessary for modeling iterated revision in a realistic way. A seminal contribution to this problem was made by Spohn [Spohn 88, 90]. He developed a semi-qualitative procedure for the iterated revision of surprise/disbelief measures (or order-of-magnitude probability valuations) based on Jeffrey-conditionalization. In more recent times, his work has been supplemented by several other proposals for revising prioritized epistemic structures [Boutilier 93, Williams 94, Lehmann 95, Darwiche and Pearl 97]. But these approaches, in addition to their individual problems, share a common drawback - they are hard to operationalize (where do the priorities come from, what do they mean), and they are not fine-grained enough for some practical purposes.

Another line of research, less anchored in AI, has been probabilistic revision or update. Here, an epistemic state is meant to be represented by a subjective probability distribution, or a set of them, and the input usually consists of linear probabilistic constraints. Classical conditionalization on a proposition A, only justifiable if we are fully convinced of its truth, then corresponds to an update with the constraint $Prob(\neg A) = 0$. A more general and cautious approach is Jeffrey-conditionalization, where the degree of belief associated with the new proposition (after some unspecified evaluation or deliberation process) may be any $\alpha \in [0, 1]$, whereas the conditional probabilities $Prob(.|A)$ and $Prob(.|\neg A)$ are preserved. Here the corresponding updating constraint is $Prob(A) = \alpha$.

A very popular approach for dealing with arbitrary linear probabilistic constraints is to pick up the canonical cross-entropy minimizing model. Cross-entropy is a distinguished information-distance concept which measures the gain of information realized when passing from one distribution to another. Shore and Johnson have shown that cross-entropy minimization can be characterized by an intuitively appealing set of postulates [Shore and Johnson 80].

There are several open issues in probabilistic belief revision. For instance, how should we apply cross-entropy minimization, or maybe alternative update rules, to suitable sets of probability measures? How should we proceed if there is a second-order uncertainty valuation on top of these sets? In fact, what is the right level of abstraction for epistemic states? In this paper, we are going to address these questions by developing a general framework for the modeling and iterated revision of probabilistic belief. Our approach exploits cross-entropy minimization in a novel way and implements second-order uncertainty through particular ranking measures. For the sake of clarity, and because we want to concentrate on the conceptual issues, we try to avoid technical details and make some simplifying assumptions. So we restrict ourselves to finitary subjective probability measures.

The plan of the paper is as follows. We start by introducing some basic formal ingredients and sketch a general framework for belief revision together with some examples. Within this context, we discuss several possibilities for defining epistemic states and argue for a particular kind of ranking measures over subjective probability distributions. For these belief states, we define and motivate revision strategies based on variants of Jeffrey-conditionalization and cross-entropy minimization, which we illustrate with a simple example. We conclude the paper with some comparisons and perspectives for future work.

2 FORMAL BACKGROUND

We assume a finite propositional logical background language L. This restriction does not affect the conceptual issues addressed in the paper. L is intended to refer to the external, objective world, as opposed to the internal, epistemic reality. Let $\mathcal{W}$ be the set of propositional valuations over L and $\mathcal{B} = 2^{\mathcal{W}}$ be the corresponding boolean propositional algebra. Propositions are identified with sets of worlds. Let $Prob_{\mathcal{B}}$ be the set of all probability distributions $P : \mathcal{B} \to [0, 1]$.

A traditional way to describe the belief structure of an agent has been to attach probabilities to the propositions he may consider. This

[1] Max Planck Institute for Computer Science, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, emil@mpi-sb.mpg.de

subjective or epistemic probability is assumed to measure his degree of belief in these propositions, in particular to express his commitment to accept certain bets on their truth. However, it is important to make a distinction between statistical, objective probability, i.e. relative frequencies in the real world (e.g. of start-ups making profit after one year), and epistemic probability, i.e. strength of belief in the mind of the agent (e.g. of x.com making profit after one year). Probabilistic belief valuations have several advantages. Their philosophical justification, their high granularity, and the possibility to directly exploit them for decision-making, using for instance the maximum expected utility paradigm.

In some situations, where exact probabilities are unavailable or irrelevant, it is sufficient to estimate the degree of (im-)plausibility or the order-of-magnitude probability of a proposition. In the subjective context, ranking measures constitute a coarse-grained but viable alternative to probability valuations [Weydert 94]. For our purposes, we only need to consider standard ranking measures.

Definition 2.1 (Standard ranking measures)
A standard ranking measure ($\kappa\pi$-measure) r is a real-valued map $r : \mathcal{B} \rightarrow [0, \infty]$ with $r(\mathcal{W}) = 0$, $r(A \cup B) = min\{r(A), r(B)\}$, and $r(\emptyset) = \infty$. The conditional measure is given by $r(A|B) = r(A \cap B) - r(B)$ for $r(B) \neq \infty$, and $r(A|B) = \infty$ otherwise.

Ranking measures may be seen as measures of surprise or disbelief. A very natural interpretation is offered by the order-of-magnitude probability reading. It sets $r(A) = a$ iff $P(A) \sim \varepsilon^a$, where P is a nonstandard probability distribution and ε an arbitrary but fixed infinitesimal. Ranking measures generalize Spohn's natural conditional functions/κ-rankings, which he used for modeling belief and iterated revision [Spohn 88,90]. Finite standard ranking measures, which are characterized by their singleton values $r(\{w\})$ (written $r(w)$), are formally equivalent to real-valued possibility measures with multiplicative conditionalization [Dubois and Prade 88].

3 REVISION FRAMEWORK

To begin with, we want to discuss the revision of belief states on a more general level, abstracting away from any specific epistemic ordering or representational paradigm. There are four major ingredients. First, the basic objects of belief, propositions A about the real world expressed in some object language L. Secondly, the epistemic states e organizing and structuring beliefs and meta-beliefs. Thirdly, the information items i able to trigger a revision process. Fourthly, the revision function $\star$ itself, which for each e and i picks up a revised epistemic state $e \star i$ meant to give a possibly better account of reality and therefore a better base for decision-making.

Epistemic states can be represented in different ways. In a fully comprehensive framework, they would have to encode desires and intentions as well. However, in this paper, we put these considerations aside and focus on the handling of pure belief. What is actually believed or accepted, the old-fashioned belief set, is only a minor part of the overall epistemic structure. Among others, there also must be an epistemic ordering used to measure the strength of belief and to guide decisions and revisions. Examples are epistemic entrenchment orders, ranking measures and probability distributions.

Revision may be triggered by internal (thoughts) or external (observations, communications) information items. It is conceptually useful to split the overall revision process into two parts. First, the interpretation and deliberation of the input in the context of e and i, followed by its translation into an epistemic constraint Φ, i.e. a set of epistemic states. Secondly, the update of the original epistemic structure e by choosing among the admissible candidates – collected in Φ – the most reasonable successor state, i.e. $e \star i$. In the present paper, we concentrate on the second step, which also has been - implicitly or explicitly - the focus of most previous research.

Definition 3.1 (Revision system)
A revision system $(E, E_0, I, \star)$ consists of a set of epistemic states E, a subset $E_0 \subseteq E$ of initial prior states, a class $I \subseteq 2^E$ of input sets, and a revision function $\star : E \times I \rightarrow E$.

We illustrate this notion with two simple revision systems, which also allows us to introduce further important concepts.

The first example is Spohn-type revision of ranking measures. Following Spohn's interpretation, ranking measures are measures of disbelief. Obviously, r may only support belief in A if $r(A) < r(\neg A)$, i.e. $r(\neg A) > 0$. Because plausibility thresholds for belief acceptance offer more flexibility, we use a stronger notion of belief expressed by $r(\neg A) \geq 1$. Let E^{sp} be the set of all ranking measures $r : 2^W \rightarrow [0, \infty]$. The canonical non-informative prior state is the uniform ranking measure r_0 with $r_0(A) = 0$ for $A \neq \emptyset$, which supports only tautological belief. So we set $E_0^{sp} = \{r_0\}$. To accept incoming information represented by $A \subseteq W$, we have to ensure $r(\neg A) \geq 1$. Hence, the minimal class of input sets is $I^{sp} = \{I_A \mid A \subseteq W\}$, with $I_A = \{r \in E^{sp} \mid r(\neg A) \geq 1\}$. For revision, the idea is to obtain $r \star_{sp} I_A$ by uniformly shifting the A-worlds downwards until one of them reaches 0, and uniformly shifting the $\neg A$-worlds upwards until all of them reach 1, if possible. If $r(\neg A) \geq 1$, $r \star_{sp} I_A = r$. If $r(A) = \infty$, revision with A is impossible and we set $r \star_{sp} I_A = r$. Otherwise, $r \star_{sp} I_A(A) = 0$ and $r \star_{sp} I_A(\neg A) = 1$, whereas the conditional ranking values stay invariant. That is, for all $B \subseteq W$, $r \star_{sp} I_A(B|\neg A) = r(B|\neg A)$ and $r \star_{sp} I_A(B|A) = r(B|A)$. Existence and uniqueness are guaranteed.

The second example is probabilistic update based on cross-entropy minimization. Cross-entropy is a binary function H measuring the information gained when passing from a prior distribution Q to a new distribution P.

Definition 3.2 (Cross-entropy)
Let $Q, P \in Prob_{\mathcal{B}}$. If for all $A \in \mathcal{B}$, $Q(A) = 0$ implies $P(A) = 0$, the cross-entropy from Q to P is

$$H(P, Q) = \sum_{\omega \in \mathcal{W}} P(\omega) \log P(\omega) / \log Q(\omega).$$

Otherwise, $H(P, Q) = \infty$.

We have $H(P, Q) \geq 0$ and $H(P, Q) = 0$ iff $P = Q$. But H is not a classical distance function, it violates symmetry and the triangle inequality. Cross-entropy is a well-behaved and well-motivated relative information measure. In particular, it allows us to define a distinguished probabilistic update concept which can be characterized by a small set of intuitively appealing rationality postulates for probabilistic revision functions [Shore and Johnson 80]. More precisely, given some prior distribution Q representing the initial beliefs and some closed convex set of distributions Φ determined by the new evidence, it seems reasonable to stay as uncommitted or unbiased as possible. This is done by picking up a revised successor belief state P in Φ which minimizes the information gain $H(P, Q)$. In fact, if there is a $P \in \Phi$ with $H(P, Q) < \infty$, then there is a unique P minimizing $H(P, Q)$. This gives us the following revision system. $E^{ce} = Prob_{\mathcal{B}}$, $E_0^{ce} = \{P_0\}$, where P_0 is the uniform distribution, I^{ce} is the collection of closed convex subsets of E^{ce}, and $\star_{ce}$ is the function which associates with $Q \in E^{ce}$ and $\Phi \in I^{ce}$ the unique MCE-model P, if it exists, and otherwise Q.

4 EPISTEMIC STATES

According to the Bayesian perspective, an epistemic state should be characterized by a single all-encompassing subjective probability distribution. However, this looks like a crude simplification of the real thing. Even if our goal is not cognitive modeling but designing rational agents in the best possible way, it seems unreasonable to ignore the uncertainty associated with individual probability judgments as well as the impracticality of assigning a single value to every imaginable proposition. Also we should note that in the real world, agents have limited epistemic ressources. In particular, the revision formalism may fail to specify a unique valuation, bringing in additional uncertainty. Therefore, we have to consider valuation sets, just as in a more qualitative context, we have considered world sets. A minimal assumption is that any epistemic structure e should determine an epistemic set $Epi(e)$, i.e. the set of those subjective probability distributions it considers admissible. This doesn't mean that the agent cannot choose a single, coherent probability distribution for decision-taking purposes. It only means that this step is accompanied by a loss of information about the currrent epistemic reality. But when it comes to revise the epistemic state in the light of new evidence, the agent should be able to exploit the full epistemic structure, not only the synoptic decision probability.

Not every valuation set seems to be a reasonable carrier of epistemic uncertainty. Consider for instance an epistemic state e supporting two distributions P and P'. If the agent cannot decide by himself which one is more appropriate - otherwise he would have made his choice - or which probability to attribute to each one - otherwise he would have picked up the mixture -, every weighted combination $\lambda P + (1-\lambda)P'$ for $\lambda \in [0,1]$ appears to be equally reasonable and should therefore be admissible for e. Consequently, each epistemic set $Epi(e)$ should be closed under weighted mixtures, i.e. it should be convex. Convexity has several advantages. In particular, together with topological closedness, it ensures a unique solution for consistent cross-entropy minimization. While the convexity assumption looks reasonable for sets of epistemic distributions, it turns out to be counterintuitive for sets of statistical distributions. For instance, we may very well believe that a coin has been manipulated in one of two ways and that the relative frequency of getting head is therefore either 0.5 or 0.9, but not 0.7 or any value in between. On the other hand, our degree of belief that the next throw will give us head may wobble between 0.5 and 0.9 if we feel unable to judge the relative probability of the different modifications.

However, modeling epistemic states with convex distribution sets is not the final word. There are at least two reasons for considering strucures on top of the basic epistemic valuations. First of all, we may want to exploit the revision history, which is hardly reflected in the epistemic sets. This information may be irrelevant for current decisions, but it could be useful for guiding future revisions. Secondly, we may want to express the plausibility of different epistemic valuations or corresponding probabilistic assertions.

An obvious generalization would be to consider hyperdistributions, i.e. probability measures over the basic subjective distributions. However, these higher-order entities are even more cumbersome, elusive and hard to grasp than their lower-order counterparts. Furthermore, the same reasons promoting sets of epistemic valuations, would be valid here as well. And why stopping at this level, why not hyperhyperdistributions, or sets of them? If there is a need for higher-order preferences, and we think so, the only sensible way to avoid this type of infinite regress is to assume a lower granularity for higher levels, i.e. to consider coarser-grained alternatives to a probabilistic structure over epistemic distributions. This can be achieved through so-called hyperrankings, a new epistemic valuation concept which marries our representational needs with our demand for simplicity.

Hyperrankings are special ranking measures r over the space of subjective probability distributions $Prob_{\mathcal{B}}$ which take into account some peculiarities of epistemic probability, e.g. the role of convexity. Those epistemic distributions with maximal plausibility, i.e. rank 0, are assumed to form the epistemic set $Epi(r)$. The epistemic distributions with lower plausibility only come into play when revision has to occur. Hyperrankings can be specified from the outside, but we see them primarily as byproducts of the revision process, partly reflecting its history. Similarly to epistemic sets, hyperrankings also have to meet specific requirements. In particular, weaker plausibility thresholds – i.e. $r \in Epi_{\alpha}(r)$ iff $r(P) \leq \alpha$ $(\alpha > 0)$ – should also define convex epistemic sets over hyperrankings. But we do not ask for topological closure.

Definition 4.1 (Convex ranking measures)
A ranking measure $r : 2^{Prob_{\mathcal{B}}} \rightarrow [0, \infty]$ *is called convex iff* $\{P \mid r(P) \leq \alpha\}$ *is convex for each* $\alpha \in [0, \infty]$.

There is another substantial condition. If P is considered epistemically possible, i.e. $R(P) \neq \infty$, we should also allow any P' which is conditionalization accessible from P, i.e. which can be reached from P by Jeffrey-conditionalization. The reason is that we do not want to preclude any consistent future evidence. That is, epistemic possibility should be closed under conditionalization accessibility. This is equivalent to the following.

Definition 4.2 (Conditionalization accessibility)
$P' \in Prob_{\mathcal{B}}$ *is conditionalization accessible from* $P \in Prob_{\mathcal{B}}$ *iff for all* $A \in \mathcal{B}$, $P(A) = 0$ *implies* $P'(A) = 0$.

In addition, for the sake of simplicity and because $Prob_{\mathcal{B}}$ is infinite, we also stipulate a smoothness condition which states that the rank of a set already has to be assumed by one of its singletons.

Definition 4.3 (Hyperranking)
A ranking measure $r : 2^{Prob_{\mathcal{B}}} \rightarrow [0, \infty]$ *is called a hyperranking iff* r *is convex,* $\{P \in Prob_{\mathcal{B}} \mid r(P) < \infty\}$ *is closed under conditionalization accessibility, and* $r(A) = min\{r(P) \mid P \in Prob_{\mathcal{B}}\}$.

Hyperrankings generalize the epistemic set concept by adding an implausibility valuation over all subjective probability distributions. The corresponding epistemic set is $Epi(r) = \{P \mid r(P) = 0\}$. For decision-theoretic purposes, $Epi(r)$ is the only relevant entity. The other distributions and their hyperranks will only become relevant in the context of revision. The need for sets of hyperrankings to express additional uncertainty is less pronounced because ranking measures are already characterized by an inherent vagueness linked to the min-condition. The following result is quite useful.

Theorem 4.4 (Epistemic possibility domain)
If r *is a hyperranking,* $\{P \in Prob_{\mathcal{B}} \mid r(P) < \infty\}$ *has the form* $\{P \in Prob_{\mathcal{B}} \mid P(A) = 0\}$ *for some* $A \in \mathcal{B}$.

5 EPISTEMIC DYNAMICS

We now turn to our main task, the design of suitable revision systems for hyperrankings. Let E^{hy} be the set of all hyperrankings over $Prob_{\mathcal{B}}$ and $E_0^{hy} = \{r_0\}$, r_0 being the uniform hyperranking. What about the input space I^{hy}? As for MCE-revision, we assume that the external evidence can be translated into a closed convex $\Phi \subseteq Prob_{\mathcal{B}}$. According to the general definition, Φ then is meant to determine a suitable input set $I_\Phi \in I^{hy}$ collecting the candidate hyperranking updates. If the new evidence is considered epistemically possible, i.e. $r(\Phi) < \infty$, revision should verify the success postulate, that is $Epi(r \star I_\Phi) \subseteq \Phi$. Inconsistent evidence, i.e. Φ with $r(\Phi) = \infty$, may be ignored. So, let $I_\Phi = \{r \in E^{hy} \mid Epi(r) \subseteq \Phi\}$. Then, within our formal framework, we could set $I^{hy} = \{I_\Phi \mid \Phi \subseteq Prob_{\mathcal{B}}, \Phi$ closed, convex$\}$. However, given the close correspondence between Φ and I_Φ, we stipulate $I^{hy} = \{\Phi \subseteq Prob_{\mathcal{B}} \mid \Phi$ closed, convex$\}$. The most straightforward revision function $\star_{hy} : E^{hy} \times I^{hy} \rightarrow E^{hy}$ for hyperrankings is Spohn-type revision extended by a convexification step. This gives us the following definition.

Definition 5.1 (Convex Spohn revision)
Let $r \in E^{hy}$ and $\Phi \in I^{hy}$. Then $r \star_{csp} \Phi$ is defined to be the unique hyperranking r' such that for all $\alpha \in [0, \infty]$, $\{P \mid r'(P) \leq \alpha\}$ is the smallest convex superset of $\{P \mid (r \star_{sp} \Phi)(P) \leq \alpha\}$.

We may illustrate $\star_{csp}$ with a small example. Suppose $\mathcal{W} = \{\omega_x, \omega_y, \omega_z\}$, i.e. $\mathcal{B}$ has only three atomic propositions. Then $Prob_{\mathcal{B}}$ can be represented by $\{(x, y, z) \in R^3 \mid 0 \leq x, y, z,\ x+y+z = 1\}$. We may imagine hyperrankings restricted to singletons as functions with global, but without local minima. Let $\Phi_{x \geq 0.5} = \{P \mid P(\omega_x) \geq 0.5\}$ and $\Phi_{x=0} = \{P \mid P(\omega_x) = 0\}$, which are closed convex. First we revise with $\Phi_{x \geq 0.5}$. Let $r_1 = r_0 \star_{csp} \Phi_{x \geq 0.5}$. Then $r_1(P) = 0$ for $P \in \Phi_{x \geq 0.5}$ and $r_1(P) = 1$ otherwise. This is also the result we would have gotten with $\star_{sp}$.

Next we revise r_1 with $\Phi_{x=0}$. Here $\star_{csp}$ differs from $\star_{sp}$ because convexification flattens the ranks in $\Phi_{x>0}$. Let $r_2' = r_1 \star_{sp} \Phi_{x=0}$. We obtain $r_2'(P) = 0$ over $\Phi_{x=0}$, $r_2'(P) = 1$ over $\Phi_{x \geq 0.5}$, and $r_2'(P) = 2$ elsewhere. Of course, $r_2' = r_1 \star_{sp} \Phi_{x=0}$ is no longer convex. Let $r_2 = r_1 \star_{csp} \Phi_{x=0}$. Because each $P \in \Phi_{0.5 > x > 0}$ can be written as a linear combination of some $P_1 \in \Phi_{x \geq 0.5}$ and $P_2 \in \Phi_{x=0}$, we have $r_2(P) = 0$ over $\Phi_{x=0}$, but $r_2(P) = 1$ over the whole $\Phi_{x>0}$.

Convex Spohn revision is the basic ingredient of our revision strategies for hyperrankings. It is natural, easy to handle, and justifiable by minimal information considerations transferred to the ranking context. However, it is also a very conservative approach which we may want to strengthen by exploiting some probabilistic minimal-distance concept, e.g. cross-entropy. The idea is to give priority to those distributions which are closest to the initial epistemic set. Presumably the most appropriate and well-behaved distance notion for our epistemic purposes is the relative information measure cross-entropy H. To exploit it, we have to extend H to sets of distributions and introduce a suitable binary projection function $\downarrow$. For $P \in Prob_{\mathcal{B}}$ and $\Phi, \Phi' \subseteq Prob_{\mathcal{B}}$, let

- $H(\Phi', \Phi) = inf\{H(P', P) \mid P \in \Phi, P' \in \Phi'\}$,
- $\Phi \downarrow \Phi' = \{P' \in \Phi' \mid H(P', \Phi) = H(\Phi', \Phi) < \infty\}$.

$H(\Phi', \Phi)$ indicates the greatest lower bound to possible distances between elements of Φ, Φ'. $\Phi \downarrow \Phi'$ picks up those elements of Φ' which are closest to Φ, assuming existence. If Φ' is closed and $H(\Phi', \Phi) < \infty$, then $\Phi \downarrow \Phi' \neq \emptyset$. If Φ' is also convex, $\{P\} \downarrow \Phi'$ is a singleton. Unfortunately, even for closed convex Φ, not very much is known about $\Phi \downarrow \Phi'$. In particular, it doesn't need to be convex.

Based on these definitions, as a prelude to hyperranking revision, we want to find a reasonable probabilistic update function $\circ$ for closed convex $\Gamma, \Gamma' \subseteq Prob_{\mathcal{B}}$. We may distinguish between cautious, pointwise strategies, which are looking at individual distributions, and more adventurous, global strategies, which are looking at distribution sets. The most cautious procedure is pointwise projection $\circ_p$, which is close to Jaeger's approach for direct inference from a statistical knowledge base [Jaeger 94]. The idea is to assume equal importance for all the distributions in Γ, independently from Γ', and to take the convex topological closure of all the projections of $P \in \Gamma$.

$$\Gamma \circ_p \Gamma' = conv(\bigcup \{\{P\} \downarrow \Gamma' \mid P \in \Gamma\}) \quad \text{if } H(\Gamma', \Gamma) < \infty,$$

otherwise, $\Gamma \circ_p \Gamma' = \Gamma$. In the context of cross-entropy projection, this is the largest possible choice for $\Gamma \circ \Gamma'$, i.e. the weakest update. Unfortunately, there is a serious problem with pointwise projection. In fact, $\circ_p$ fails to satisfy the conservation principle, i.e. $\Gamma \cap \Gamma' \neq \emptyset$ does not enforce $\Gamma \circ_p \Gamma' = \Gamma \cap \Gamma'$, even for closed convex Γ, Γ'. That is, we cannot simply conjoin consistent epistemic probability constraints, which is hardly acceptable. To see what may go wrong, consider two closed halfspaces $\Phi_1, \Phi_2 \subseteq Prob_{\mathcal{B}}$ with a non-empty, sufficiently sharp-angled intersection. Then we may find a boundary point P of Φ_1 whose cross-entropy projection $\{P\} \downarrow \Phi_2$ does not belong to $\Phi_1 \cap \Phi_2$.

Consequently, we prefer a stronger approach, global projection $\circ_g$. It picks up the overall Γ-closest elements of Γ' and takes the convex topological closure. That is,

$$\Gamma \circ_g \Gamma' = conv(\Gamma \downarrow \Gamma') \quad \text{if } H(\Gamma', \Gamma) < \infty,$$

otherwise, $\Gamma \circ_g \Gamma' = \Gamma$. Obviously, $\circ_g$ verifies conservation if Γ, Γ' are closed convex. $\circ_p$ and $\circ_g$ are equivalent if Γ is a singleton.

Now, we have all the ingredients – in particular $\star_{csp}$ and $\circ_g$ – for a powerful hyperranking revision concept $\star_{hy}$. But these components can be mixed in different ways. Given a hyperranking $r \in E^{hy}$ and a closed convex input set $\Phi \in I^{hy}$, the basic idea is, first, to use r, $\star_{csp}$ and $\circ_g$ to determine a preferred closed convex $\Psi \subseteq \Phi$, secondly, to realize convex Spohn revision on r and Ψ, which then should give us $r \star_{hy} \Phi$. That is, the interesting part is to get from Φ to Ψ.

Depending on whether we give more importance to the plausibility ordering fixed by r or to the informational closeness to $Epi(r)$, we may choose one of two major strategies. If the ranking is considered more relevant, we start with $\star_{csp}$ and globally project $Epi(r)$ onto the convex closure of the most plausible part of Φ, namely onto $conv(Epi(r \star_{csp} \Phi))$. If cross-entropy minimization is considered more relevant, we directly project $Epi(r)$ onto Φ, exploiting r only afterwards, through the main application of $\star_{csp}$. Interstingly, these approaches, called the conditioning-first resp. projection-first strategy, may provide contradictory results, i.e. the resulting epistemic sets may be disjoint. So, we have to choose.

Because the structure of the prior hyperranking r seems to be more specific and relevant than the information distance to the epistemic set $Epi(r)$ – otherwise r should have reflected this right from the beginning –, we prefer conditioning-first. That is, we proceed in three steps. First, we use convex Spohn revision to determine with r the most plausible closed convex subset Γ of the input set Φ. Secondly, we globally project $Epi(r)$ onto Γ, shrinking it to Γ'. Then we apply convex Spohn revision with input set Γ', giving us $r \star_{hy} \Phi$.

Definition 5.2 (Conditioning-first hyperrevision)
Let r be a hyperranking and $\Phi \subseteq Prob_{\mathcal{B}}$ be closed convex. Then

$$r \star_{hy} \Phi = r \star_{csp} (Epi(r) \circ_g conv(Epi(r \star_{csp} \Phi))).$$

The corresponding revision system is $(E^{hy}, E_0^{hy}, I^{hy}, \star_{hy})$.

We may use the previous example to illustrate our hyperrevision concept. So, we want to compute $(r_0 \star_{hy} \Phi_{x \geq 0.5}) \star_{hy} \Phi_{x=0}$. We start with $r_0 \star_{hy} \Phi_{x \geq 0.5}$ and evaluate

$$r_0 \star_{csp} (Epi(r_0) \circ_g conv(Epi(r_0 \star_{csp} \Phi_{x \geq 0.5}))).$$

First, we have $Epi(r_0) = Prob_{\mathcal{B}}$ and $conv(Epi(r_0 \star_{csp} \Phi_{x\geq 0.5})) = \Phi_{x\geq 0.5}$. Because only the elements of $\Phi_{x\geq 0.5}$ have minimal, i.e. 0 cross-entropy distance to $\Phi_{x\geq 0.5}$, $\Psi_1 = Prob_{\mathcal{B}} \circ_g \Phi_{x\geq 0.5} = \Phi_{x\geq 0.5}$. It follows that $r_0 \star_{hy} \Phi_{x\geq 0.5} = r_0 \star_{csp} \Phi_{x\geq 0.5} = r_1$ with $r_1(P) = 0$ for $P \in \Phi_{x\geq 0.5}$ and $r_1(P) = 1$ otherwise.

Next, we revise r_1 with $\Phi_{x=0}$. We have $Epi(r_1) = \Phi_{x\geq 0.5}$ and $conv(Epi(r_1 \star_{csp} \Phi_{x=0})) = \Phi_{x=0}$. For global projection, we first observe that the smallest informational distance occurs between $P_1 \in \Phi_{x\geq 0.5}$ represented by $(0.5, 0.25, 0.25)$ and $P_2 \in \Phi_{x=0}$ determined by $(0, 0.5, 0.5)$. Therefore $\Psi_2 = Epi(r_1) \circ_g conv(Epi(r_1 \star_{csp} \Phi_{x=0})) = \Phi_{x\geq 0.5} \circ_g \Phi_{x=0} = \{P_2\}$. Accordingly, $r_1 \star_{hy} \Phi_{x=0} = r_1 \star_{csp} \Psi_2 = r_2$ with $r_2(P_2) = 0$, $r_2(P) = 1$ for $P \in conv(\Phi_{x\geq 0.5} \cup \{P_2\}) - \{P_2\}$, and $r_2(P) = 2$ otherwise.

Note that if we start with $\Phi_{x=0}$, we obtain $r_0 \star_{hy} \Phi_{x=0} = r_0 \star_{csp} \Phi_{x=0}$. But then, hyperrevision with $\Phi_{x\geq 0.5}$ is no longer possible because $H(\Phi_{x\geq 0.5}, \Phi_{x=0}) = \infty$. With our definitions, we cannot retract epistemic impossibility, i.e. $P(A) = 0$. To avoid this, we could drop the $< \infty$-restriction in the definition of $\circ_g$.

6 COMPARISONS

Because the revision of epistemic states exploiting second-order valuations has been hardly addressed in the literature, $\star_{hy}$ is a bit more difficult to position. The only competitors are those mentioned in the text, namely convex Spohn revision $\star_{csp}$ and the projection-first variant of hyperrevision $\star_{hy}^{pr}$ discussed above. Convex Spohn revision is the most robust approach. It only exploits informational closeness on the ranking level, not in the context of individual distributions. But it is too cautious if the epistemic set of the prior hyperranking is just a singleton. In particular, convex Spohn revision does not constitute an extension of the well-motivated MCE-revision procedure. Projection-first hyperrevision doesn't share this drawback and is also conceptually simpler than standard, i.e. conditioning-first hyperrevision. However, in addition to its lower "specificity", as explained in the text, it may also conflict with convex Spohn revision in the sense that $Epi(r \star_{hy}^{pr} \Phi) \cap Epi(r \star_{csp} \Phi) = \emptyset$. Of course, given the current absence of rationality postulates for hyperrevision procedures, these considerations about $\star_{hy}$ and its competitors are still preliminary.

To complete our picture, we may also have a short look at the postulates $P1-7$ discussed by Grove and Halpern [98] for revision functions Upd updating sets of distributions Γ by propositions A. We can translate their entities into our framework by setting $r_\Gamma = r_0 \star_{sp} \Gamma$ and $\Phi_A = \{P \mid P(A) = 1\}$. That is, we investigate the behaviour of $Upd(\Gamma, A) = r_\Gamma \star_{hy} \Phi_A$. P1 directly follows from our definitions. P2, which states invariance under coarsening of $\mathcal{B}$, is also supported. P3, commutativity, already fails for Spohn-type revision. P4, which says that redundant information doesn't change anything, applies too. P5, which derives the revision of a set from the revision of its elements, is rejected (see global projection). P6, under the version where it requires that a singleton produces a singleton, or an inconsistency, is also valid. P7, a non-triviality postulate, is satisfied as well. Of course, these are only partial results about principles not designed to deal with hyperrankings or similar structures. It is an important task for the future to look for more specific postulates, or maybe characterization results, for hyperrevision mechanisms.

7 CONCLUSIONS

A major goal of this paper has been to give a hint at the actual complexity of more realistic – in the sense of representational power, not in the sense of computational efficiency – epistemic models and their dynamic transformation. To begin with, we have sketched a general revision framework and discussed the nature of epistemic states. In particular, we have introduced hyperrankings, a new type of epistemic structures based on convex ranking measures over subjective probability distributions. We also have proposed a new revision strategy, called hyperrevision, which combines convex Spohn revision with cross-entropy minimization. Our approach tries to find a middle ground between representational expressivity and pragmatism, and may offer a more manageable way to handle higher-order uncertainty and preferences. Of course, this is only a first step.

Conceptually speaking, the present account is mainly a powerful extension to existing, more restricted proposals. There are many issues which have not been addressed. For instance, the handling of independence information, propositional algebras over a first-order language, and the relation between statistical and epistemic probability. We have also dared to ignore the computational issues. Computing the full revised hyperranking, or even the epistemic set, after each revision step appears to be impractical. Therefore, approximation methods and the representation by revision sequences (with partial evaluation on demand) could be useful. On a more general level, what is missing is a better understanding of the first part of the revision process, the step from the real input to some constraint over epistemic distributions. Furthermore, we have to be aware of the fact that belief revision is only a relatively small - although very important - part of the bigger task to model cognitive agents in a realistic way. The handling of information is rarely a purely passive process where incoming inputs are just evaluated and integrated. In practice, agents are actively seeking informations to diminish uncertainty. In addition, beliefs have a certain purpose, namely to help solving problems or reaching specific goals. Decision-taking not only requires belief valuations, but also suitable preferences or utility valuations, which have to be updated as well. It is only in this larger context that the real value of revision strategies will become visible. However, before we can take this road, we need to master the various tools for pure belief revision, in particular hyperrevison procedures.

REFERENCES

[Boutilier 93] C. Boutilier. Revision sequences and nested conditionals. In *Proceedings of IJCAI 93*, 1993.

[Darwiche and Pearl 97] A. Darwiche and J. Pearl. On the logic of iterated belief revision. In *Artificial Intelligence*, 89 : 1-29, 1997.

[Dubois and Prade 88] D. Dubois, H. Prade. *Possibility Theory.* Plenum Press, New York 1988.

[Grove and Halpern 98] A.J. Grove and J.Y. Halpern. Updating sets of probabilities. In *Proceedings of UAI 98*. Morgan Kaufmann, 1998.

[Jaeger 94] M. Jaeger. A logic for default reasoning about probabilities. In *Proceedings of UAI 94*. Morgan Kaufmann, 1994.

[Lehmann 95] D. Lehmann. Belief revision, revised. In *Proceedings of IJCAI 93*. Montreal, Canada, 1995.

[Shore and Johnson 80] J.E. Shore and R.W. Johnson. Axiomatic derivation of the principle of cross-entropy minimization. In *IEEE Transactions on Information Theory*, IT-26(1):26-37, 1980.

[Spohn 88] W. Spohn. Ordinal conditional functions : A dynamic theory of epistemic states. In W.L. Harper and B. Skyrms (eds.), Causation in Decision, Belief Change and Statistics. D. Reidel, Dordrecht, Netherlands, 1988.

[Spohn 90] W. Spohn. A general non-probabilistic theory of inductive reasoning. In R.D.Shachter et al. (eds.), Uncertainty in Artificial Intelligence 4, North-Holland, Amsterdam, 1990.

[Weydert 94] E. Weydert. General belief measures. In *Proceedings of UAI 94*. Morgan Kaufmann, 1994.

[Williams 94] M-A. Williams. Transmutations of knowledge systems. In *Proceedings of KR94*. Morgan Kaufmann, 1994.