CHANGE RULES FOR HIERARCHICAL BELIEFS

B.WALLISER (CERAS-Ecole Nationale des Ponts et Chaussées)
D.ZWIRN (CREA – Ecole Polytechnique)

August 2001

0. INTRODUCTION

An agent may adopt hierarchical beliefs, not only beliefs about other's beliefs, but even second order beliefs about some basic beliefs (Skyrms, 1980; Baron, 1987; Kyburg, 1989). A first interpretation considers that an agent holds some subjective belief about his own first order subjective beliefs, for instance when he feels some ambiguity about the uncertainty judgement he holds on some phenomenon. The most usual interpretation refers to a subjective belief about a truly random phenomenon assumed to be endowed with objective probability, obtained by information about the distribution of its occurrences and treated as some objective property (Suppes, 1981). For instance, the agent may have a subjective belief on the distribution of a population or a subjective belief on some antecedent events conditioning some consequent events.

An agent may change, according to well studied rules, his one-level beliefs which can be either set-theoretic (Alchourron-Gärdenfors-Makinson, 1985; Katsuno-Mendelzon, 1991) or probabilistic (Walliser-Zwirn, 2000). The message received by the agent is generally a set-theoretic one, and can be interpreted in two different contexts of change, illustrated by the composition of an urn with colored balls. In a revising context, the message makes more precise or contradicts the initial belief about a system considered as unchanged; for instance, it says that there are no red balls in the urn. In an updating context, the message adds to the initial belief some information about the way the system is concretely evolving; for instance, it states that there are no more red balls in the urn.

Combining now hierarchical beliefs and belief revision, some change rules were already proposed for different uncertainty models. For a probability distribution on probability distributions, it was suggested that Bayes rule still applies when considered at the relevant level of belief. For a family of probability distributions or a distribution of events (the dual form of Dempster-Shafer belief functions), several change rules were proposed : Dempster, Fagin-Halpern, Gilboa-Schmeidler, Suppes-Zanotti, Jaffray; however, these rules were only justified by some algebraic properties. Moreover, a new context of change appears, the focusing context, where the message adds to the initial belief about a whole population some information about an object extracted from it; for instance, it assesses that a ball picked in the urn is not red.

The paper introduce several principles intended to clarify the domain of application of all these rules, either as concerns the objective or subjective character of beliefs or as concerns the context of change where they are relevant. The first part makes more precise the formal aspects of hierarchical structures (section 1), its main interpretations (section 2), the general form they can take (section 4) and what transformations can be applied to them (section 5). The second part recalls the contexts of change (section 5) and justifies the rules adequate for belief hierarchies as concerns updating (section 6), revising (section 7) and focusing (section 8). Some synthetic examples like the Ellsberg's urn are finally given (section 9).

The main results for 2-level hierarchies are given by the following table:

Hierarchy Context	distribution of distributions	distribution of events	family of distributions
revising	maximal	geometric	Gilboa-Schmeidler
updating	minimal	Dempster	Fagin-Halpern
focusing	synthetic	Jaffray	Fagin-Halpern

1. FORMAL BELIEF HIERARCHY

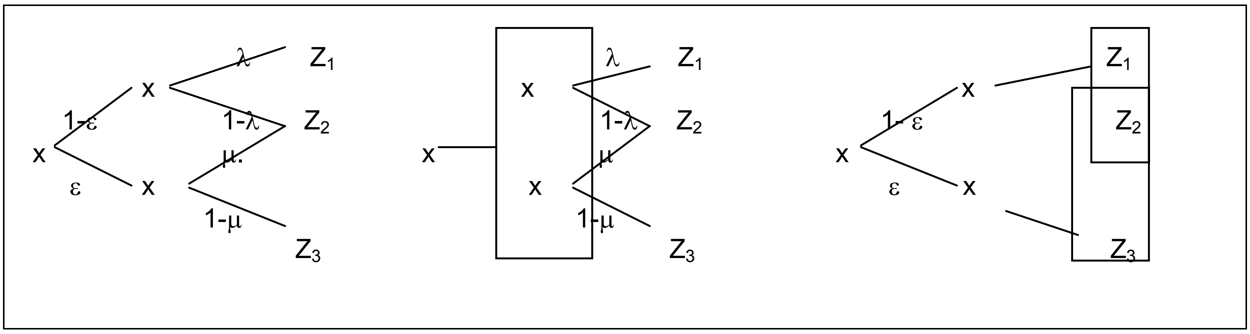
The belief hierarchy is stated exclusively in a semantic framework and more precisely in a possible worlds one, which allows more easily to combine a propositional and a probabilistic view (Billot-Walliser, 1999). A hierarchy is formed of successive layers of worlds (or 0-worlds), meta-worlds (or 1-worlds) and so on (k-worlds for the generic layer), stopping at a final layer of only one world (n-world). The relation R between two layers k and k+1 defines a level of two different types:

- a set-theoretic level, where each (k+1) world is linked to a given set of k-worlds constituting an event.
- a probabilistic level, where each (k+1) world is linked to all k-worlds by a probabilistic distribution on them.

By combination, a two-level hierarchy may be of four different kinds, only three of them being really interesting:

- a distribution of distributions, where a 2-world defines a probability distribution on 1-worlds, each 1-world defining a probability distribution on 0-worlds;
- a family of distributions, where a 2-world defines a set of 1-worlds, each 1-world defining a probability distribution on 0-world.
- a distribution of events, where a 2-world defines a set of 1-worlds, each 1-world defining a set of 0-worlds.

For instance, the following structures can be considered, where the probability distribution is replaced by its support in order to translate a probabilistic layer into a set-theoretic one, and where $\mu > \lambda > 0$:

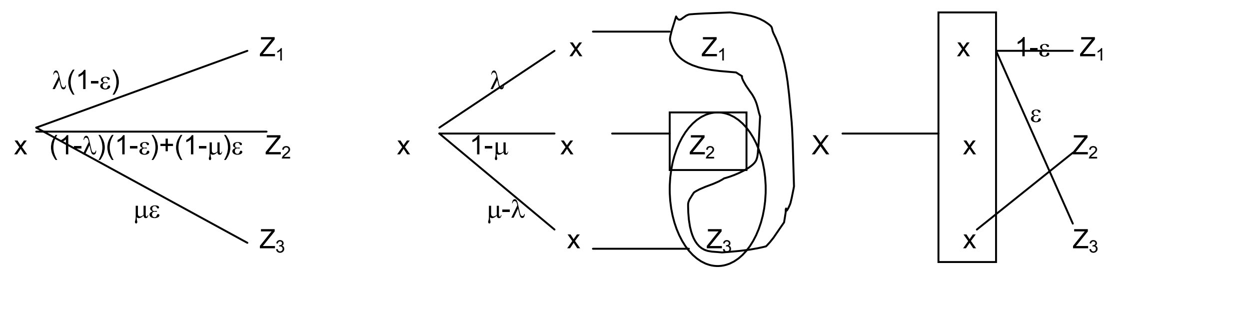


S1	S2	S3
distribution of distributions	family of distributions	distribution of events

From hierarchical structures, it is possible to derive a lower and an upper value associated to any event E on the basic worlds by proceeding recursively from bottom layer 0 to top layer n. In each 0-world, the lower (upper) value of event E is given by its characteristic function (value 1 if the world belongs to it and 0 else). When the relation R_k is set-theoretic, in any k+1 world, the lower (upper) value of E is the maximum (minimum) of the lower (upper) values taken for all the k-worlds in the associated set. When the relation R_k is probabilistic, in any (k+1) world, the lower (upper) value of E is the expected value of the lower (upper) values taken for k-worlds probabilistically associated to it. Moreover, two hierarchical structures are said to be value-equivalent if they give the same lower and upper values to any basic event E, values considered from the point of view of the top-layer n-world.

For instance, given a distribution of events, the lower and upper value are respectively the credibility and plausibility (in Dempster-Shafer sense); the distribution of events is recovered conversely as the Moebius transform of these values. Similarly, given a family of distribution, the lower and upper value are respectively the lower and upper probability (lower and upper envelope); the family of distribution is recovered conversely as their "core". As well known, a credibility function (or belief function) is always value-equivalent to a lower probability function, the converse being true under some specific conditions (Chateauneuf-Jaffray,).

In the example, the three structures S1, S2, S3 are respectively value-equivalent to the following ones, the first being obtained by collapsing two-level probabilities and the last two by using the previous value-equivalence :



S'1
distribution of probabilities

S'2
distribution of events

S'3
family of distributions

2. OBJECTIVE STRUCTURE

Any hierarchy can now be interpreted by distinguishing two types of worlds (all of the same type at each layer) :

- a physical world (represented by a square) corresponds to a material state of some system.
- a psychical world (represented by a circle) corresponds to a mental state of the ongoing agent.

Such a distinction constraints the belief structure since, according to the "**Reflection principle**", psychical worlds are defined on physical worlds and not the converse, excluding the possibility for beliefs to be physically influenced. In order to construct a semantically oriented belief taxonomy, pure probabilistic relations between layers will be favored, the transposition to set-theoretic ones being straightforward.

When considering only physical worlds, a *one-level objective structure* relates one meta-world to a set of basic worlds :

The probabilities involved are objective ones : they are properties of the external world, not of the belief of the agent. They corresponds to what is often labeled as "chance" in the literature (for instance, D.Lewis, 1980), and probability$_2$ by Carnap (1945).

Two major interpretations are usually given to such a structure:

- the **populational interpretation** considers that some external system Y gathers objects of different types Z_i, the probability p_i representing the *proportion* of each type of objects in the whole system. The basic example considers an urn containing balls of three colors, B (blue), Y (yellow), R (red), the probabilities stating in fact the precise type of urn the modeler considers.
- the **causal interpretation** considers that some phenomenon Y induces facts of different types Z_i, the probability p_i representing the *propensity* of each fact to occur. The basic example considers an illness leading to different compound symptoms : -SF (no spot, fever), SF (spots, fever), S-F (spots, no fever), -S-F (no spot, no fever), the probabilities stating in fact some type of illness the modeler considers.

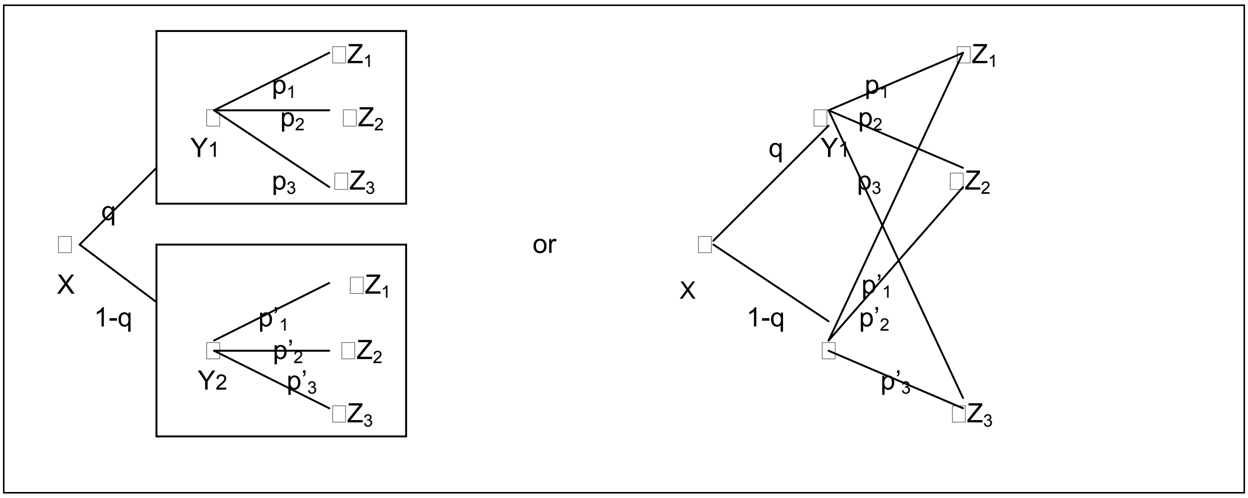
The two interpretations can be linked by a third one, the **frequentist interpretation:**

-for a population, the frequentist interpretation states that an object extracted from the system has a long term frequency for one colour equal to its probability.

-for a causal phenomenon, the frequentist interpretation states that the frequency of each fact when the phenomenon is repeated is equal to its propension.

In fact, in the first case, the probability relates to a generic object (a generic ball) and in the second case to a generic phenomenon (a generic patient). Hence, the urn becomes the fundamental prototype, a generic ball endowed with an objective probability being a ball randomly chosen in the urn (for instance the next ball to be drawn) . Conversely, a specific ball (already drawn) cannot be endowed with an objective probability since a given ball is of only one precise color (even if one doesn't know which) and the objective probability to be of another color is always zero.

When considering two levels of physical worlds, a *2-level objective structure* is naturally symbolized by a nesting of worlds, but can be represented by a lattice since there is no ambiguity:



Two interpretations can again be given:

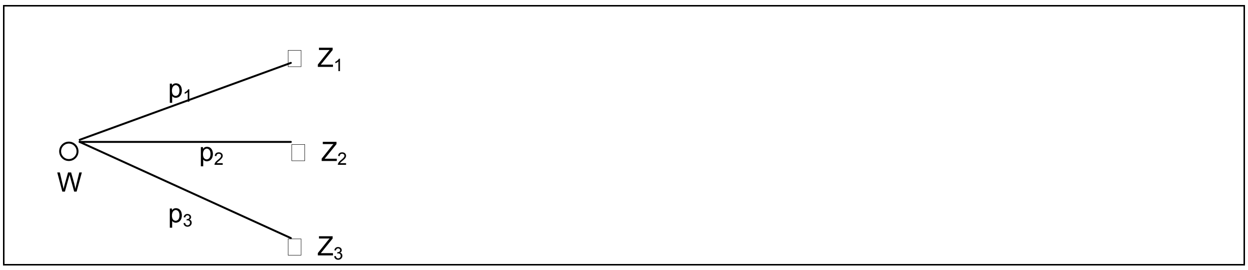
-in the populational interpretation, a meta-system is constituted of several systems. For instance, a meta-urn X is constituted of a population of different urns ($Y_1=U_1$,$Y_2=U_2$), each urn being identified by the proportion of balls of different colors Zi.

-in the causal interpretation, a meta-phenomenon induces a distribution of antecedents of the basic phenomenon. For instance, each illness (defined by its symptoms) is characterized by its occurrence in the whole system.

The two interpretations are again equivalent and can be reduced to the populational one. For instance, it is possible to imagine that the ball of the first urn are spheric and of the second cubic, the meta-urn giving the proportion of both types of balls.

3.SUBJECTIVE STRUCTURE

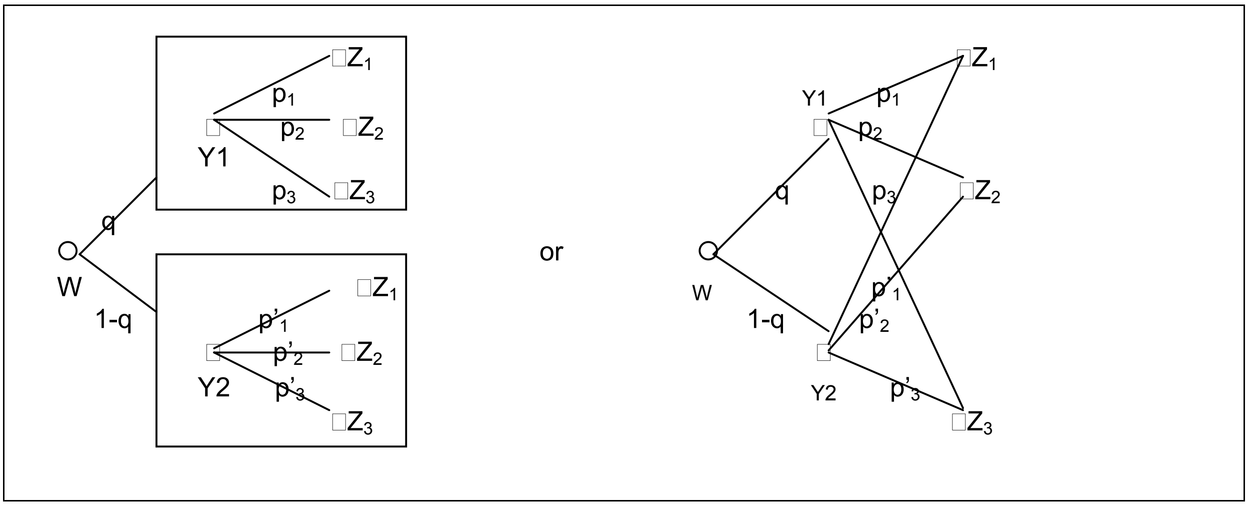
When considering now both physical and psychical worlds, a *one-level subjective structure* relates one psychical meta-world to a set of physical worlds :



The probabilities involved are subjective ones : they characterize the degrees of belief of an agent. They correspond to Lewis's (1980) credence or to Carnap's (1945) $probability_1$.

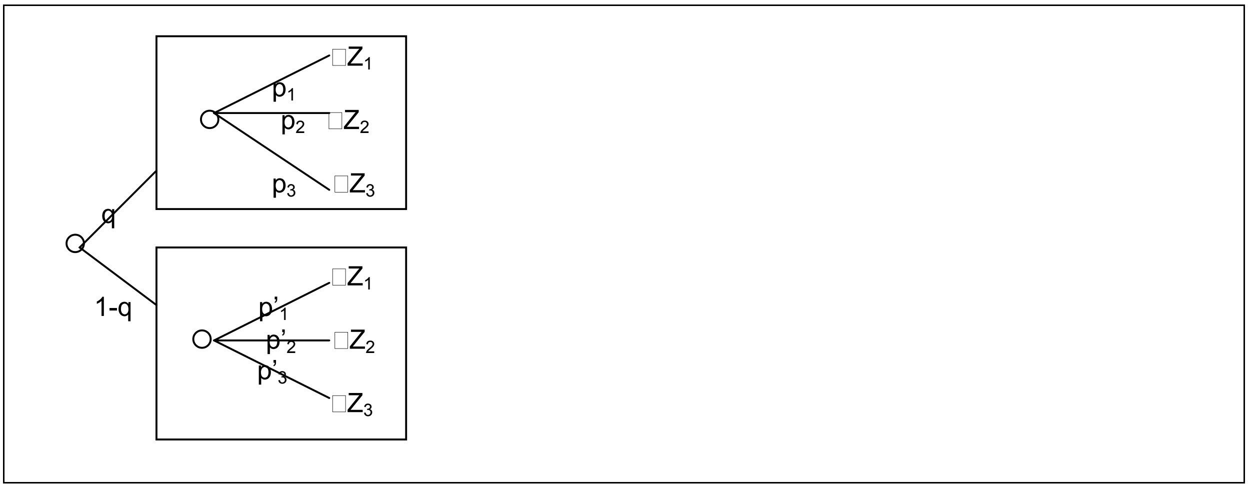
There is only one interpretation saying that the agent considers some basic entity, either an object (a specific ball) or a fact (a specific symptomatology), which can be of types (colors, presence of two different symptoms) Z_i with probabilities p_i. The ball is again the fundamental prototype. The belief of the agent on this entity is uncertain, though its properties are supposed to be well defined (and known by the modeler) : for instance the specific ball is exactly red, but the agent does not know this fact. On the contrary, a generic ball (or an urn) never has a given color.

When the subjective uncertainty bears on a one-level physical structure, a *subjective two-level structure* can again be reduced graphically to a lattice :



For instance, the agent believes that a given urn can be of different types, each type being characterized by a combination of balls of different colors Z_i. He may similarly believe that an illness is of different types, each defined by a set of symptoms.

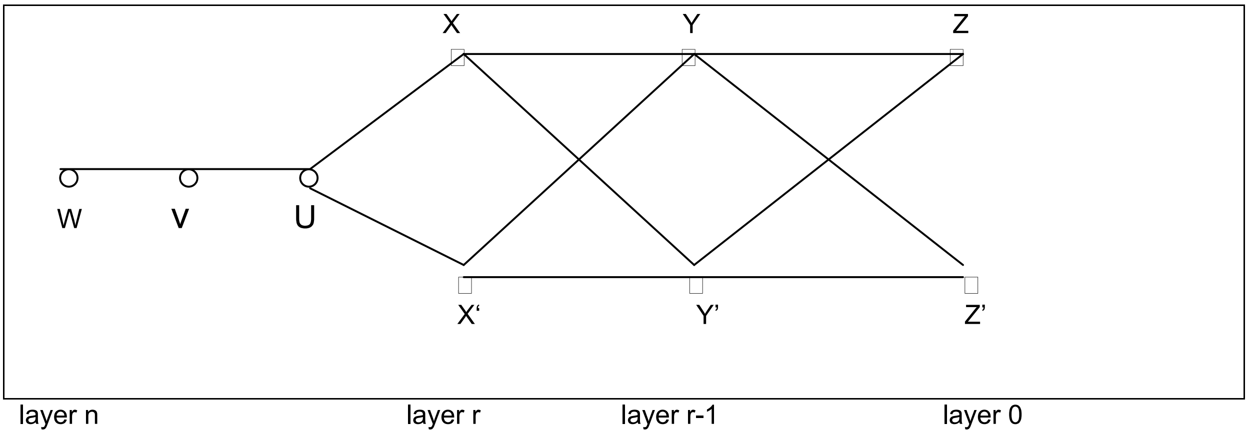
When considering still two types of worlds, a third possibility would be to build a nesting of psychical worlds defined on psychical ones:



In any interpretation, an agent has some belief about alternative beliefs he may have about some object (ball) or fact (symptomatology) possibly of different types (colors, symptoms).

However, the "**No-schizophrenia principle**" states that an agent is unable to believe different beliefs about the same object, if endowed with strong enough cognitive rationality. This principle is proved in annex 1 within epistemic logics (applied to set-theoretic or probabilistic beliefs) for an agent who satisfies logical omniscience, positive and negative introspection.

When the Reflection principle and the No-schizophrenia principle are simultaneously accepted, a multi-level hierarchy is of the following form where three specific layers can be sorted out :

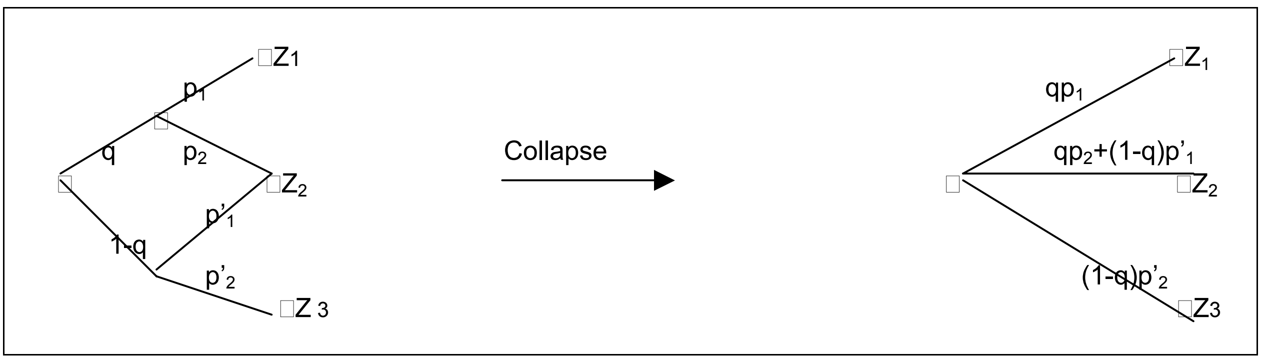


The "reference layer" r characterizes the system the agent is considering; the layers r-1 to 0 are the layers of objects which constitute the compound system (there are several such layers when the knowledge about the properties of basic objects are given on a conditional form); the "fundamental layer" n is the layer of highest hierarchical beliefs, which can be taken to be infinite by positive introspection. For instance, the prototype of a hierarchical structure is a two-level structure where an agent holds a belief about a system of objects (n =2, r=1), for which the three formal types examined in the section 1 can be reinterpreted.

4. TRANSFORMATION OF STRUCTURES

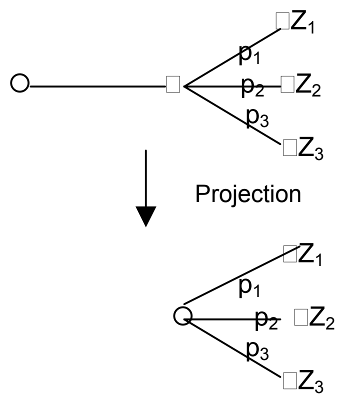
The “**Collapse principle**” states that it is possible to collapse two objective levels in a hierarchical structure, be it objective or subjective, since the initial and final structures are value-equivalent. For instance, one can build a one-level “*synthetic urn*”, which gives the same value to its generic ball than a given meta-urn. When the two levels of the meta-urn are heterogeneous (one set-theoretic and one probabilistic), the synthetic urn can only be described by a set of partial beliefs as represented in the left part of the above graph. But when the two levels of the meta-urn are homogeneous (both probabilistic), the synthetic urn can be described by a single probability distribution combining the values of the two levels.

For instance, starting from a simplified two-level hierarchy, the composition of the associated synthetic urn is represented by the following scheme:



Although conceptually different, it is possible to set up epistemic relations between the knowledge of objective probabilities and subjective probabilities. If an agent is aware of the composition of a given system, he may endorse the objective probability of an occurrence and adopt it as his subjective probability for that occurrence. This "subjectivation principle" is exactly what is expressed by the Miller principle (Miller, 1966), the Principal Principle of Lewis (Lewis, 1980) or the constraint C_2 stated by Skyrms (Skyrms, 1980a). By the same operation, in a populational interpretation, the agent transforms a belief about some k-world (the given system or equivalently a "generic" object of this system) into a belief about a (k-1) world (a "specific" object of the system). This "specification principle" can be formally considered as lowering the reference level of the hierarchical belief structure.

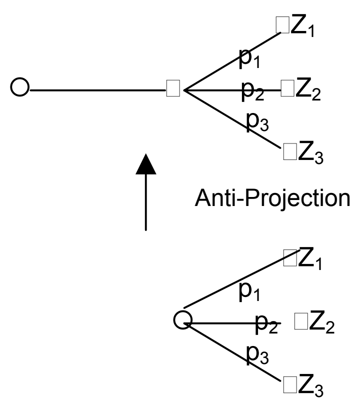
The common idea to the subjectivation principle and the specification principle will be called the "**Projection principle**", which formally turns a physical r-world into a psychical (r-1) one:



The Projection principle may be applied to the transformation of the knowledge of the composition of an urn into the belief on a specific ball contained in the urn . It may be applied to the transformation of the knowledge of objective conditional probabilities relative to a generic ball into a subjective conditional belief relative to a specific one . In the last case, the agent believes with a probability p_1 that if the ball is spherical (S), it is of type Z_1. This conditional belief is not a belief whose degree is conditional to a fact (if the ball is spherical then the agent believes with degree p_1 that it is of type Z_1), but a probabilistic belief on a conditional proposition (if the ball is spherical, then it is of type Z_1). The “if then” conditional of that proposition can be interpreted neither by a material implication nor by a subjunctive conditional. In fact, it has to be interpreted as a "restrictive conditional", i.e. a conditional equivalent to the material implication when the antecedent is true and not defined when it is false.

The subjectivation principle may be reversed in an "objectivation principle" which states that the subjective probability of an object may be considered as stemming from an objective (virtual) probability. It is just a reversal of the Miller Principle, where the agent constructs an objective phenomenon in his mind from which his subjective probability may be derived, even if this phenomenon is a fictitious one. In a populational interpretation, it acts as an "Anti-specification principle", stating that the object is extracted form a virtual system, changing a specific belief into a generic belief. The acceptability of such a principle depends on the origin of the subjective probability: if subjective probability comes from the consideration of situations similar to the situation at hand, some population of cases is implicit; but subjective probability may be grounded on pure intuition.

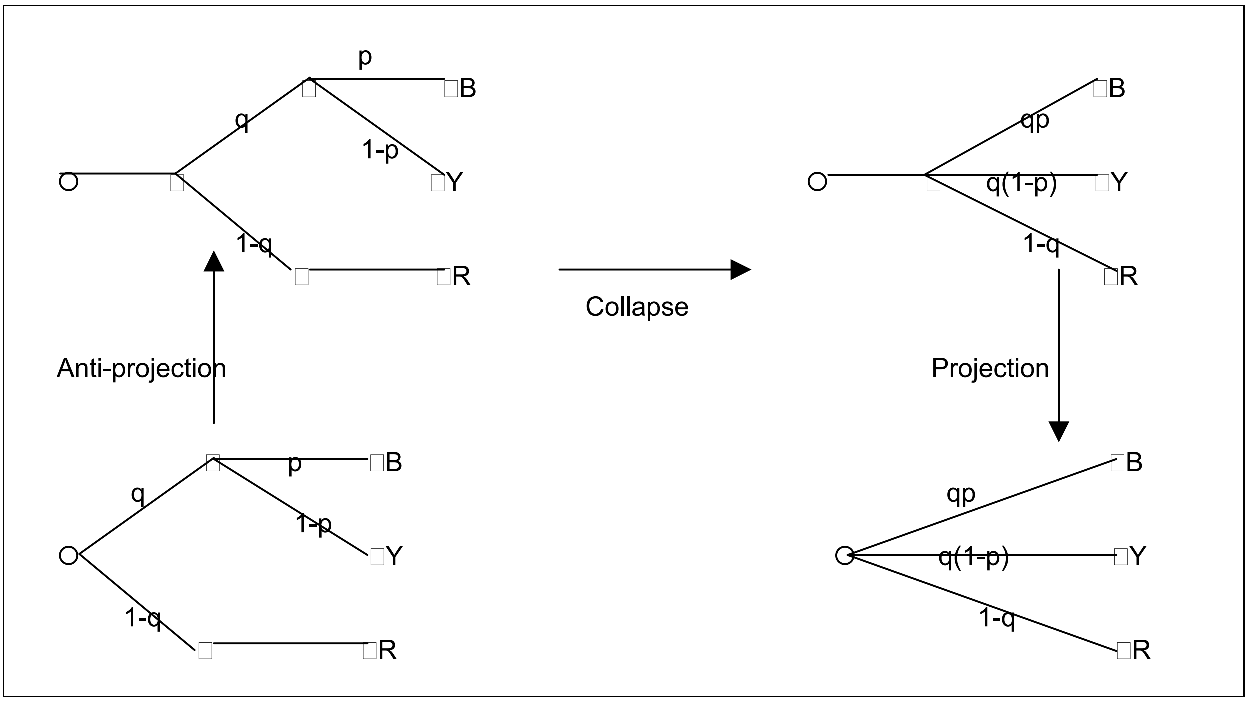
Formally, the common “**Anti-projection principle**” changes psychical (r-1) worlds into physical r-ones, hence translates the layer under consideration for one position to the upper one:



For instance, the Anti-projection principle may be applied to the transformation from the belief on a ball to the knowledge of a fictitious urn. It may also be applied in all the more complex contexts where the Projection principle could be applied, but in the reverse way.

A fundamental question lies then in the possibility to apply the Projection principle to a situation where an agent is no more certain, but uncertain about the composition of a system of objects. Lewis (1980) makes the assumption that it is possible to combine probabilities of the two levels, without giving any clear justification of that assumption (else than to postulate implicitly a “second order Projection principle”). In fact, a direct application of the Projection principle would lead the agent to face a schizophrenic structure (he holds the belief that he has contradictory beliefs). A more complete justification suggests that in order to satisfy the No-schizophrenia principle, the agent apply successively the Anti-Projection principle, the Collapse and the Projection principle to get a belief about a generic object of the system.

For instance, for an urn, the transformation is the following by considering succesively a urn, a meta-urn, a synthetic urn and a ball:



5.CONTEXTS OF CHANGE

A message gives some new insight about the system (or phenomenon) described by the initial belief of the agent, and is generally composed of two complementary parts. The material part relates some physical operation on the system, extracting from it a specific object (sampling) or modifying its structure (evolution). The epistemic part gives some factual information about the resulting system, making the initial belief more precise or contradicting what the agent formerly believed. In a populational interpretation, three contexts of change can be distinguished according to the content of the message in its material as well as epistemic part : updating, revising and focusing (a context occurring only when considering hierarchical beliefs).

The context of "updating" assumes that the system was definitely transformed in its structure and that the message gives some indication on the resulting system (or on the transformation itself). For instance, for an urn, the message indicates that red balls were removed and for a meta-urn, that all urns containing red balls were removed (or all red balls were removed from all urns). In general, we will consider messages that always concern the reference level of the structure. In that case, the problem for the agent is again to combine the initial belief about the system and the new message into a final belief about the transformed system. It is again assumed that the message is true, but this time the initial belief is likely considered as formerly true without contradiction, except if the transformation is not possible with regard to the initial belief.

In a 1-level structure, initial and final beliefs are sets of possible worlds (eventually randomized) while the message is a set of worlds towards which the system has evolved. The updating rules, justified axiomatically (Katsuno-Mendelzon, 1992; Walliser-Zwirn, 2000), are here characterized by a local translation of each accepted world with its weight. In a set-theoretic framework, the final belief gathers the nearest worlds from each world of the initial belief (according to a distance between the two worlds) which are compatible with the message. In a probabilistic framework, the final belief selects the same worlds with associated weights (according to the allocation function from initial to final worlds) in a "General Imaging rule", a family of rules to which belongs the Lewis rule (Lewis, 1976).

The context of "revising" assumes that no transformation was done on the system and that the message just gives more information about the structure of the system. For instance, for an urn, the message indicates that there is no red ball, and for a meta-urn, that there is no urn containing red balls (or containing only red balls or no red balls). Again, we will only consider messages that concern the reference level of the structure. In that case, the problem for the agent is to combine the initial belief about the system and the message into a final belief about the same system. Generally, it is assumed that the message is true in its both parts, but the initial belief may be wrong, revising treating then asymmetrically the two sources of information.

In a 1-level structure, initial belief consists again in a (randomized or not) set of possible worlds, message in a set of worlds and final belief in a (randomized or not) set of possible worlds. The revising rules, justified axiomatically (Alchourron-Gärdenfors-Makinson, 1985; Walliser-Zwirn, 2000) can be characterized by a global translation of the accepted worlds together with their weights. In a set-theoretic framework, the final belief selects the nearest set of worlds from the initial belief (according to a distance between initial belief and worlds) ; when the message does not contradict he initial belief, that nearest set is simply the intersection between the prior set and the message. In a probabilistic framework, the final belief selects the same worlds with associated weights (according to an allocation function from initial belief to final worlds), in a "General Conditioning" rule. It is important to state that Bayes rule is only one among an infinity of possible General Conditioning rules ; the stronger axiomatic principles required to justify the restriction of this set of rules to Bayes rule are very demanding.

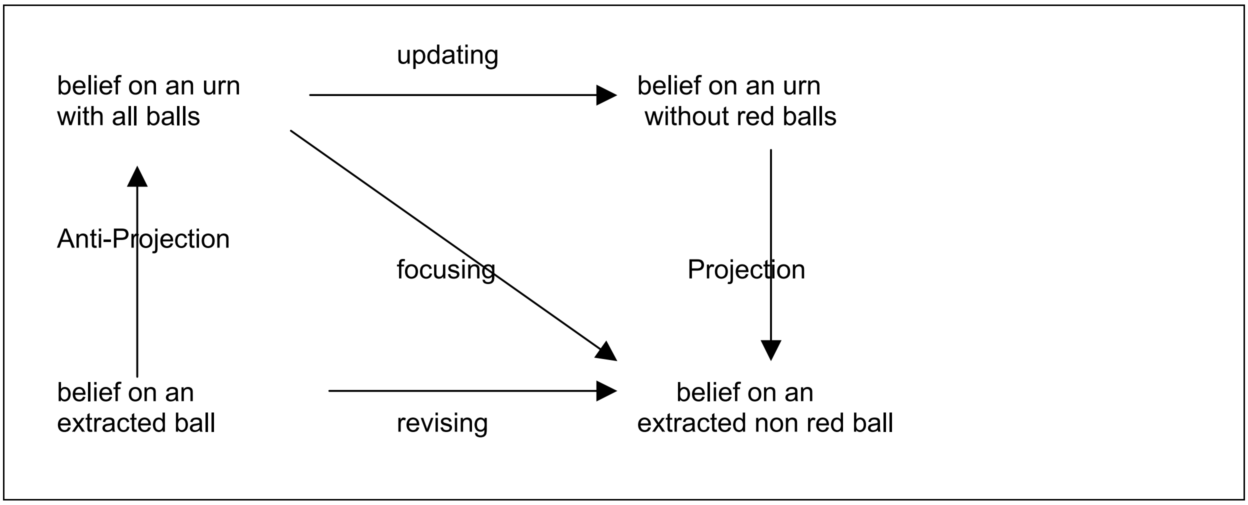
The context of "focusing" assumes that one object is extracted from the system (and replaced afterwards identically) and that the message gives information about the object randomly selected. For instance, for an urn, a ball is extracted and is not red, and for a meta-urn, an urn is extracted and contains no red balls (or a ball is extracted in two steps and is not red). In that case, the problem for the agent is first to form a belief about the extracted object, second to revise that initial belief into a final belief according to the message. It is assumed that the message is true, and the initial belief on the system is likely considered as true without contradiction, except if the message concerns a type of object non available initially.

In a 1-level structure, the message is characterized not only by some information about the object, but by all alternative information an observer could give for all type of objects. For instance, if a blue ball is extracted from some urn, the information may indicate that it is blue, that it is not red or that it is yellow, according to some filter. Focusing has no axiomatic justification since it is considered as separable into a material operation of extraction of an object and an epistemic operation of revising about that object. In fact, the epistemic operation associated to the random extraction is precisely the Projection Principle in so far as it defines what a random extraction is supposed to be : an object is randomly extracted for a population if the distribution of probability on the basic worlds associated with that object is precisely equal to the frequency of the objects of that type in the compound system.

Change rules for belief hierarchies will be studied in these three contexts by extending the rules justified for non hierarchical beliefs. Moreover, it exists some relations between the three contexts for an agent accepting to endorse the Projection and Anti-Projection principles: they consider that updating for a r-level structure corresponds to revising at a (r-1)-level, focusing being a consequence of either. These relations will be systematically exploited beginning : one considers first the updating context where the rules are easy to state independently, then infers the rules for the other contexts.

Hence, the message always gives information about the physical part of the structure and only objective probabilities are changed in belief change.

For a 1-level structure , the relation between revising, updating and focusing is the following :

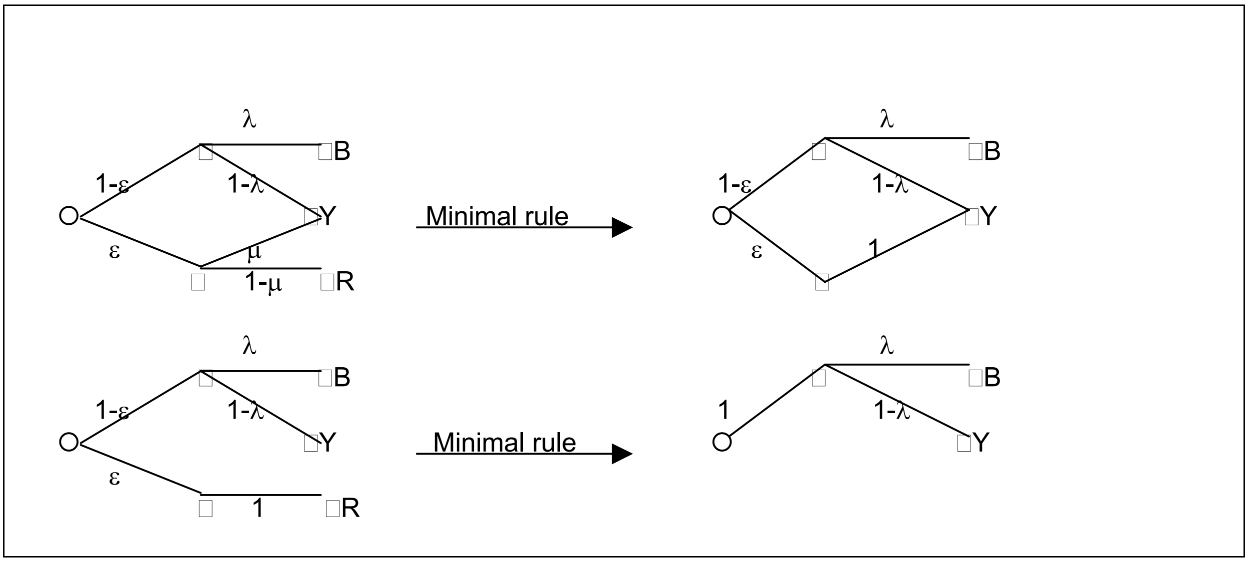


6.UPDATING RULES

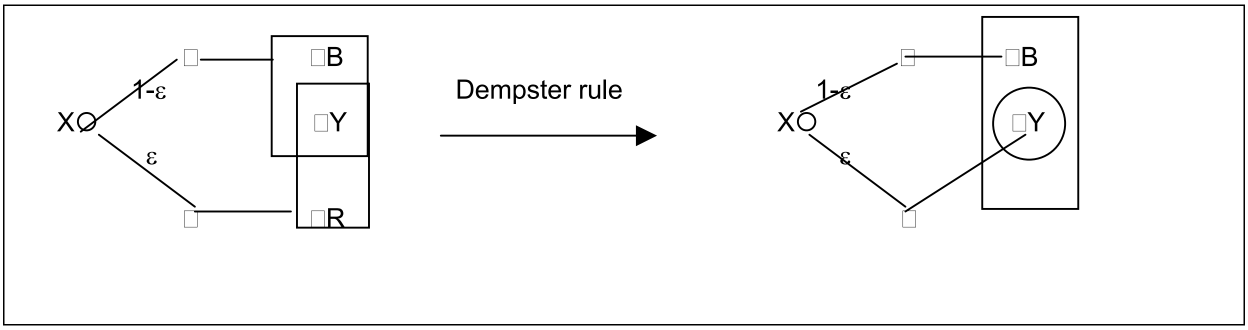
Considering now a belief hierarchy with reference layer r, the same principle than for a one-level hierarchy can be applied to physical r-worlds on which a psychical (r+1)-world defines beliefs. If some r-world of the initial belief is compatible with the message (always expressed in terms of lower layer worlds), this r-world is kept alike in the final belief. If some r-world of the initial belief is incompatible with the message, one considers the "nearest" r-worlds from that world compatible with the message, with an appropriate distance. Moreover, the belief represented by the (r+1)-world is unchanged, be it set-theoretic or probabilistic, but the weight of an initial world have to be allocated between final worlds if many.

For a two-level prototypical hierarchy, the "*minimal method*" applies the Imaging method with a notion of local distance which coincides with Bayes rule when the 1-worlds are probability distributions. Indeed, this is here mathematically justified by the physical change of the population : Bayes rule just expresses the straightforward fact that, in case of such a physical change, the proportion of any two 1-worlds compatible with the message is kept constant. The "minimal method" consists then in revising the basic distributions according to Bayes rule except if the distribution gives a zero probability to the message, in which case the final distribution is undefined. Moreover, the probabilities affected by the 2-worlds to the 1-worlds get unchanged since the physical transformation has no effect on the initial beliefs. However, if a 1-world is incompatible with the message, its weight is reallocated proportionally to the remaining one, a principle conventionally justified by the fact that changes in the external world never result in an empty world : the belief of the agent describes only proportions of types of objects within a population, not its absolute size.

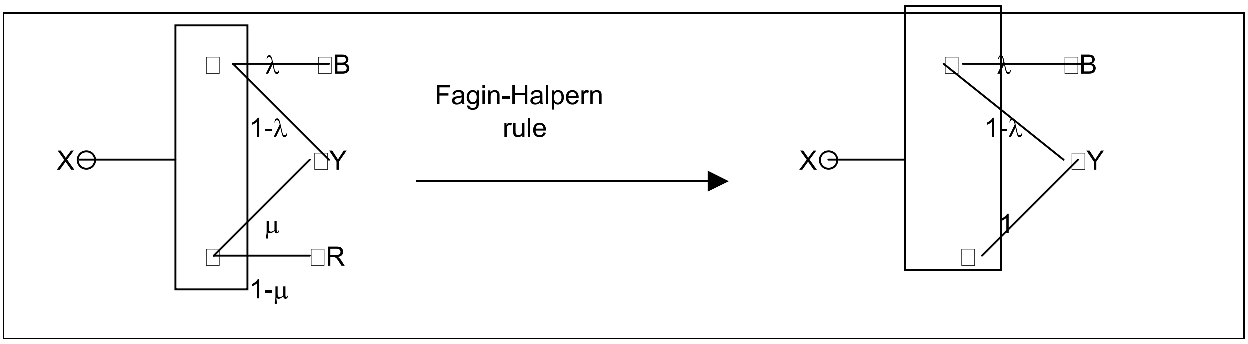
For a *distribution of distribution*s, the “**minimal rule**” applies the minimal method to an urn which is uncertain for the agent and an updating message saying that there is no more red ball in the urn , distinguishing two cases:



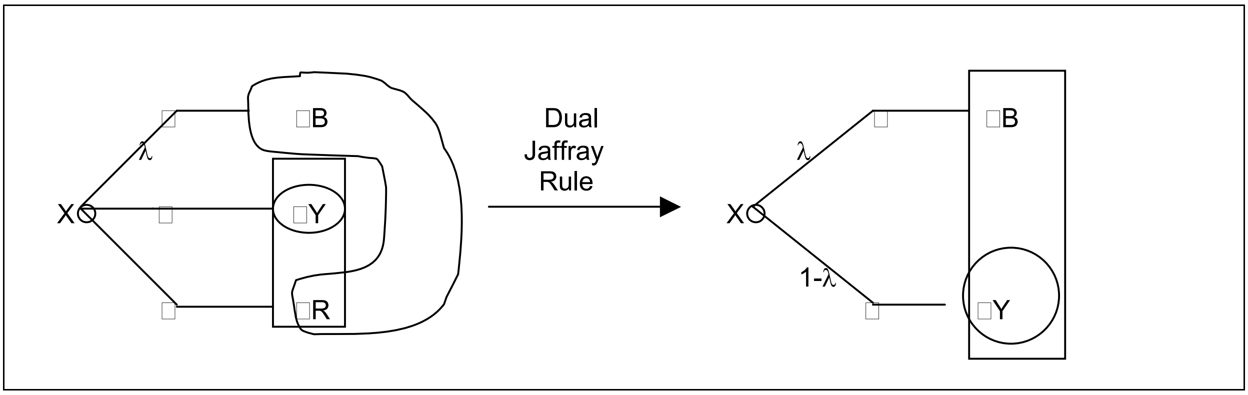
For a *distribution of events*, the "**Dempster rule**" applies the minimal method by just eliminating the excluded worlds:



For a *family of distributions*, the "**Fagin-Halpern rule**" applies again the minimal method to the same type of urn:



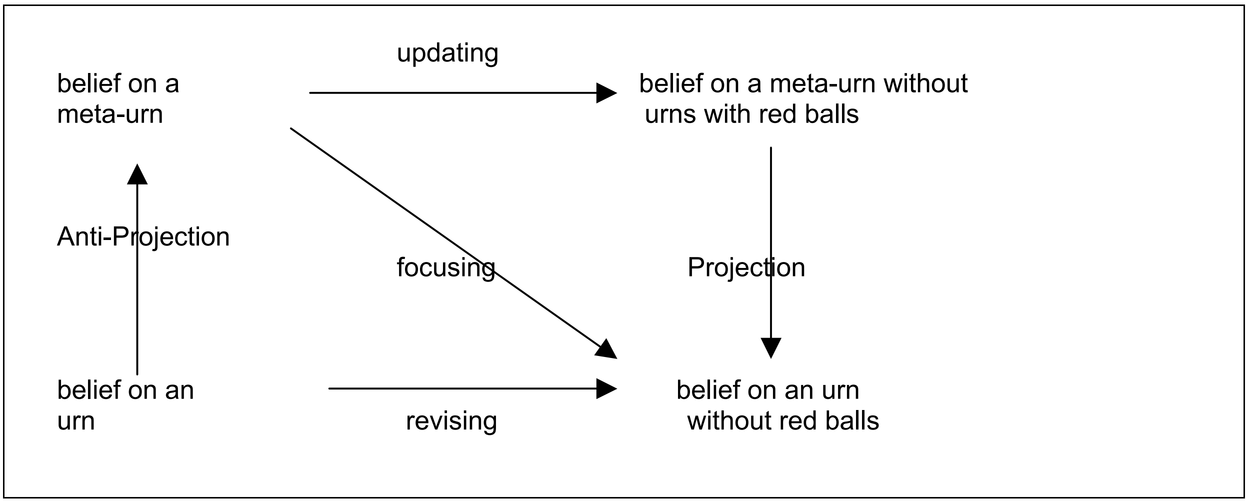
Remark: Fagin-Halpern (1989) and Jaffray (1990) showed that this rule is mathematically equivalent to the Jaffray rule applied to the equivalent distribution of events, given by S'_2 :



7. REVISING RULES

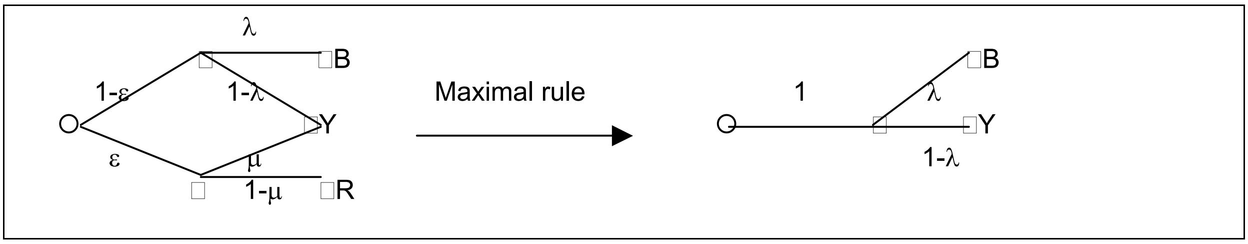
Considering now a belief hierarchy with reference layer r, the same principle can be applied by replacing worlds by r-worlds and looking at the subjective beliefs about them at layer r+1. If there exists some k-worlds which are compatible with the message (expressed in terms of the (r-i)-worlds), these r-worlds are the only ones to be kept in the final belief. If there exists no r-world compatible with the message, one considers the "nearest" set of r-worlds from the initial one compatible with the message, with an appropriate distance. Moreover, the subjective beliefs about the remaining r-world are adjusted according to the fact that level (r+1) is set-theoretic or probabilistic (allocation rule).

For a 2-level hierarchy, the correspondence of revising with updating is the same than for one-level, with a translation of one level:

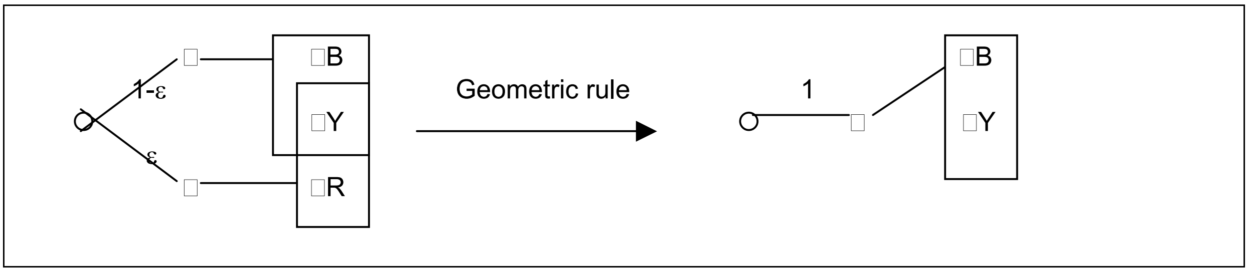


For that prototypical two-level hierarchy, the "*maximal method*" can be derived. It consists in keeping only the 1-worlds compatible with the message and, if the second level is probabilistic, to revise the probabilities at this level according to the Bayes rule ; the value affected to the 0-worlds by the 1-world have never to be changed.

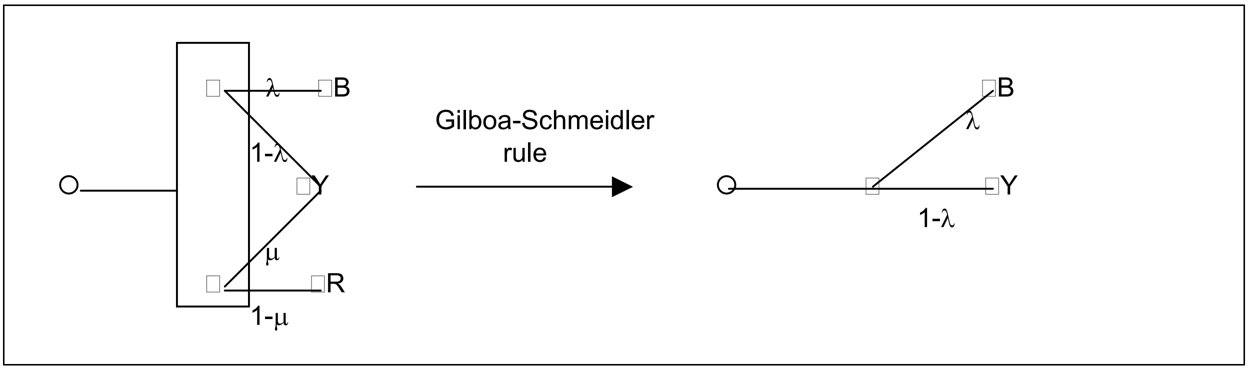
For a *distribution of distributions*, the "**maximal rule**" applies the maximal method to an urn which is uncertain for the agent, and a revising message saying that there is no red ball in the urn:



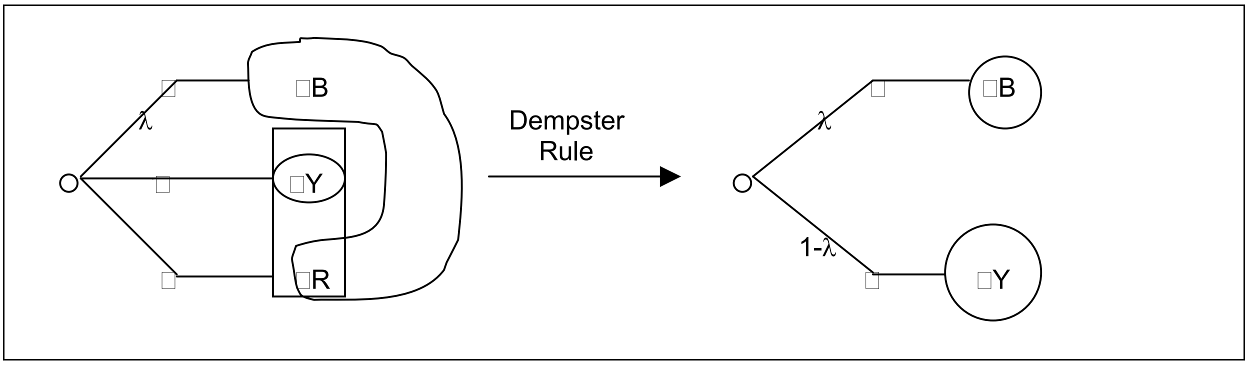
For a *distribution of events*, the "**geometric rule**" (Suppes–Zanotti,1977) only keeps the events compatible with the message (one is assumed to exist) and revises the probabilities about the remaining events according to the Bayes rule :



For a *family of distributions*, the "**Gilboa-Schmeidler rule**" (Gilboa-Schmeidler,1993) keeps the distributions which give maximal probability to the message and revises them with the Bayes rule, hence forming a restricted family:

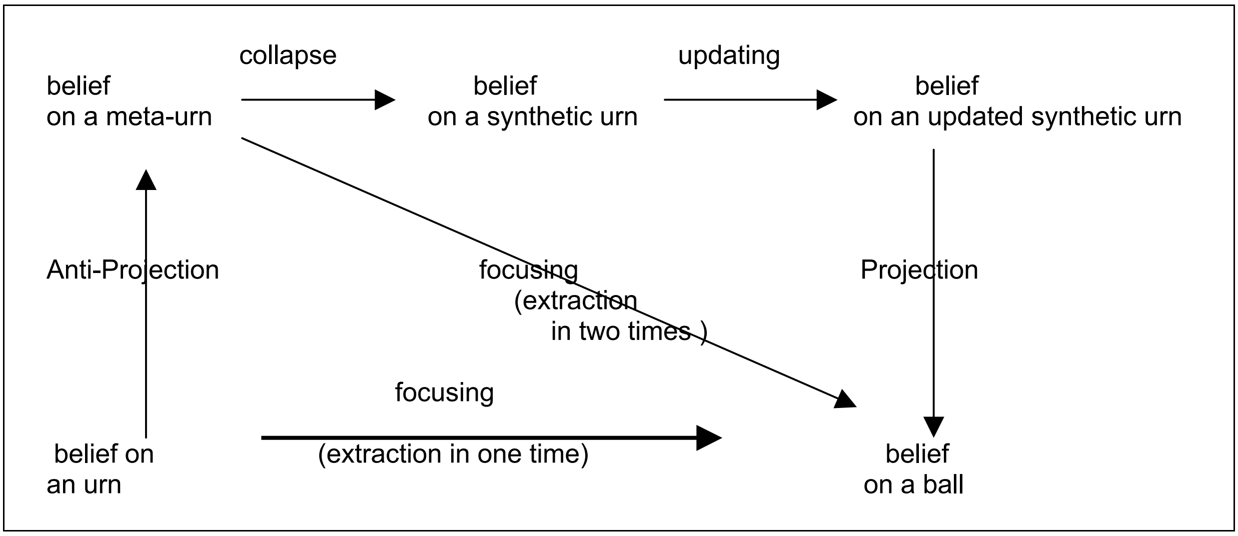


Remark: Gilboa-Schmeidler showed that this rule is equivalent to the Dempster conditioning rule (Shafer, 1976) applied to the equivalent distribution of events, given by S'_2 :



8. FOCUSING RULES

For a prototypical two-level hierarchy, the relation between the contexts of change become more complicated:

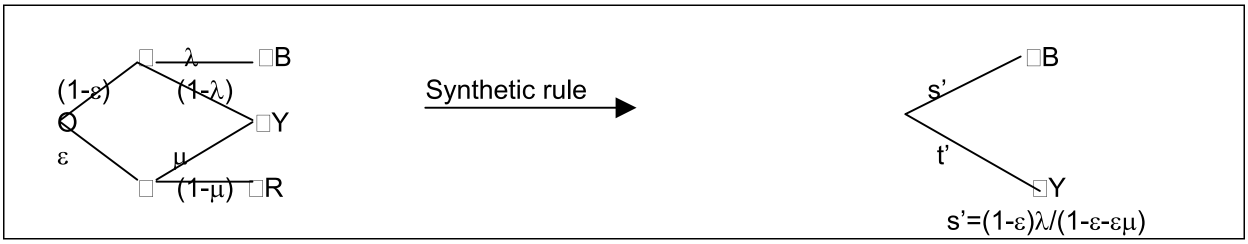


For that 2-level hierarchy, the relevant method is the "*synthetic method*". It can again be obtained from the equivalence between focusing and updating on one hand, between focusing and revising on the other hand, the two ways being proved to be equivalent. In fact, the message about the extracted object may modify or not the belief on the system from which the object was extracted:

-if the agent has a precise belief about the system, that belief is kept unchanged as long as the message does not contradict it (which would be the case if the message says that the ball is green). This property, called "Resiliency" by Skyrms (1977), is also vindicated by Lewis (1980) in the following terms: "To the extent that uncertainty about outcomes is based on certainty about their chances, it is a stable, resilient sort of uncertainty – new evidence won't get rid of it".

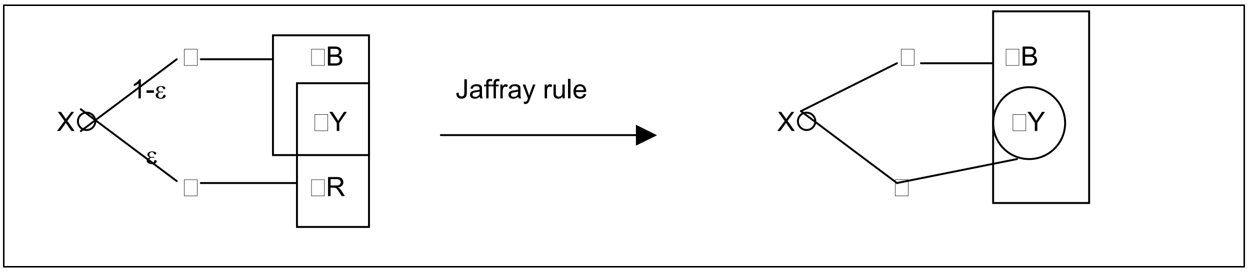
-if the agent has an uncertain belief about the system, it may be modified since the weight accorded to alternative views of the system can be reinforced or weakened by the new message. Such a change is justified by the Anti-projection Principle, since the object appears then as extracted from one of alternative systems and the message indicates from what system it was more likely extracted. In this specific situation, it is possible to "learn from experience", using the Bayes inversion formula, which stems from the Synthetic method.

For a *distribution of distributions*, the "**synthetic rule**" applies the synthetic method to an urn with which is uncertain for the agent and the message that the ball extracted is not red:



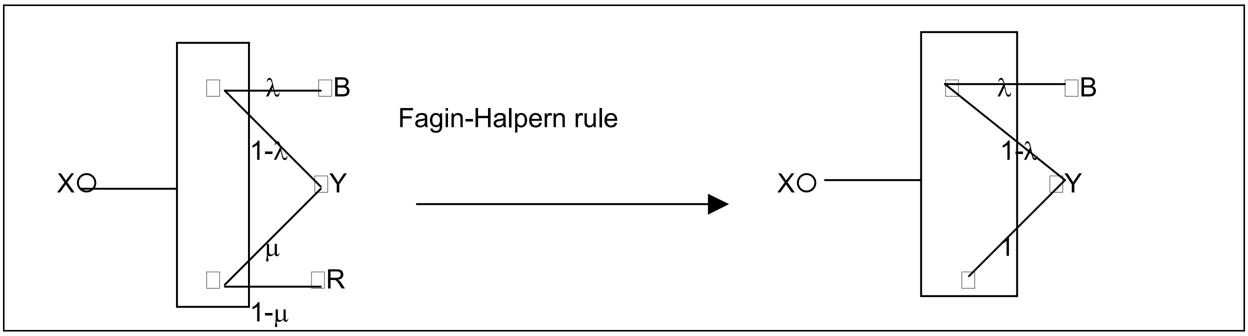
Remark: It is easy to see that this rule corresponds to a simultaneous revision at two levels as shown in annex 2.

For a *distribution of events*, the "**Jaffray rule**" is the relevant one since it can be proved that it give the values of the zero-worlds for a ball extracted in two times from a mental urn resulting along the Anti-Projection principle from an uncertainty about a given urn:



Remark: The relevance of the Jaffray rule can be shown directly by simulating the extraction process in a frequential interpretation of the urn, as shown in annex 3.

For a *family of distributions*, the "**Fagin-Halpern rule**" is again the relevant rule since the message about the ball extracted gives no information about what type of urn it is extracted :



9. SUMMARY EXAMPLES

One considers first the simplified Ellsberg urn T_3 as an example of a family of probability distributions, and the equivalent structure T'_3 as a distribution of events :

The Gilboa-Schmeidler rule for a family of probability distributions being equivalent to the Dempster Rule for the corresponding distribution of events, one obtains :

The rule applies for "revising" and, knowing that the urn contains no red balls, only one urn remains possible.

The Fagin-Halpern rule for a family of probability distributions being equivalent to the Jaffray rule for the corresponding distribution of events, one obtains :

The rule applies for "updating" and, knowing that all red balls have been removed, the possible urns are transformed accordingly.

Cohen et al. (1999) considered such an urn and tested the rule chosen by an agent, but in a focusing context of change, where a ball is extracted form the urn and happens not to be red. In fact, in order to prevent a revising process on the content of the urn itself, the best procedure says that a ball is extracted form the urn having a non red ball. The agent is then invited to bet on the color of the extracted ball, and earns \$1000 if the ball is blue and he bets on blue, and \$1000 if the ball is yellow and they bet on yellow. If he uses the Gilboa-Schmeidler rule, his interest is to bet on yellow (probability 2/3) and if he uses the Fagin-Halpern rule, his interest is to bet on blue (probability between 1/3 and 1). The results of the experiment worked out with two samples show that the agents use significantly more often the Fagin-Halpern rule than the Gilboa-Schmeidler one. These results are in accordance with our analysis showing that the Fagin-Halpern rule is the correct rule in the contexts of "focusing" and of "updating" for families of probability distributions. They accord simultaneously with the conclusion that the Jaffray rule is the correct rule in the context of "focusing" for the equivalent distribution of events, a result contradicting for instance the analysis of Dubois-Prade (....). An intuition associated to these equivalencies in the present case is that it is the same to say that "one extracts a red ball till obtaining a non red one" and "one removes all red balls and then extract a ball".

One considers now the *three prisoners* problem where three prisoners A, B and C are in jail and one at random will be executed, the guard knowing who it is but not being allowed to give information about it.

APPENDIX 1 : THE NON SCHIZOPHRENIA PRINCIPLE

In epistemic logic, the logical omniscience property states that that an agent, endowed with a belief operator B, if he believes some event E, also believes any event F which can be deduced from E :

$\forall E, F, (B(E) \wedge (E \rightarrow F)) \rightarrow B(F)$

Hence, it is possible to define an "extract belief operator" $\mathtt{B}$ such that an agent exactly believes an event E iff that belief cannot be deduced from some other belief :

$\mathtt{B}(E) \equiv_{def} B(E) \wedge [B(F) \rightarrow -(F \rightarrow E) \ \forall\ F \neq E]$

The preceding definition can be stated in an equivalent way, where an agent exactly believes an event E iff any other belief can be deduced from it :

$\mathtt{B}(E) \equiv_{def} B(E) \wedge [B(F) \rightarrow (E \rightarrow F) \ \forall F]$

Proof : if B(E) and B(F), by logical introspection B(E^F). Hence, according to the first definition, G = E ^ F , if different from E, must be such that -(E^F → E), which is impossible. Consequently, E^F = E, which means that E→F.

Consider now the semantic counterpart of the syntactic view, i.e. a Kripke structure defined by a set of possible worlds W and a non empty accessibility domain H(w) associated to each world w. The belief of an event E is then represented by the set of worlds for which all accessible worlds validate the event :

$B(E) = \{w \text{ s.t. } H(w) \subseteq E\}$

The exact belief of an event is represented by the set of worlds for which all accessible worlds - and only them - validate the event :

$\mathtt{B}(E) = \{w \text{ s.t. } H(w) = E\}$

The exact belief operator obeys to the following properties :

- Inclusion (Î) : $\mathtt{B}(E) \rightarrow B(E)$
- Unicity (Û) : $\mathtt{B}(E) \rightarrow - \mathtt{B}(F), \ \forall F \neq E$
- Positive introspection (when B satisfies Positive and Negative introspection) (PÎ) :

$\mathtt{B}(E) \rightarrow \mathtt{B}(\mathtt{B}(E))$

<u>Proof</u> (in semantic) :
$\mathtt{B}(\mathtt{B}(E)) = \{w, \text{ s.t. } H(w) = \{w', \text{ s.t. } H(w') = E\}\}$
$\mathtt{B}(E) = \{w', \text{ s.t. } H(w') = E\}$
Consider $w' \in \mathtt{B}(E)$, then $H(w') = E$ and $w' \in H(w)$, $\forall w \in \mathtt{B}(\mathtt{B}(E))$. By positive and negative introspection, if $w' \in H(w)$, then $H(w') = H(w)$. Hence $w \in \mathtt{B}(\mathtt{B}(E))$.

Finally, one can derive the following property, asserting that an agent cannot exactly believe different exact beliefs :

- No schizophrenia Principle (NS) : $\mathtt{B}(\mathtt{B}(E) \vee \mathtt{B}(F)) = \varnothing \ \forall F \neq E$

<u>Proof :</u>
Consider $E \neq F$, hence by Û, $\mathtt{B}(E) \neq \mathtt{B}(F)$.
Suppose that $\mathtt{B}(\mathtt{B}(E) \vee \mathtt{B}(F))$.
On one side, by (PÎ) : $\mathtt{B}(\mathtt{B}(\mathtt{B}(E) \vee \mathtt{B}(F)))$, and by (Î) : $B(\mathtt{B}(\mathtt{B}(E) \vee \mathtt{B}(F)))$.
On the other side, by (Î) : $B(\mathtt{B}(E) \vee \mathtt{B}(F))$, then by (PÎ) and logical omniscience : $B(\mathtt{B}(\mathtt{B}(E)) \vee \mathtt{B}(\mathtt{B}(F))$.
Then, by (Û) and logical omniscience :
$B(-\mathtt{B}(\mathtt{B}(E) \vee \mathtt{B}(F)) \vee -\mathtt{B}(\mathtt{B}(E) \vee \mathtt{B}(F))) = B(-\mathtt{B}(\mathtt{B}(E) \vee \mathtt{B}(F))$.
Gathering both sides and applying logical omniscience, one gets $\mathtt{B}(\varnothing)$. Hence the result since $\mathtt{B}(\varnothing) = \varnothing$ by the semantic expression of $\mathtt{B}(E)$.

It is trivial to prove directly that the No Schizophrenia Principle holds partially for probabilistic beliefs : an agent cannot have a second order probabilistic belief on probabilistic beliefs with different supports, since it would imply different set-theoretic exact beliefs.

More generally, the extension of this principle to probabilistic beliefs with same supports is very simple :
Let $B_{\alpha}(E) \equiv_{def} P(E) > \alpha$ be the counterpart of belief and $\mathtt{B}_{\alpha}(E) \equiv_{def} P(E) > \alpha$ be the counterpart of exact belief when the agent is endowed with a subjective probabilistic distribution P. Inclusion, Unicity and Positive introspection hold under the following form :

- probabilistic inclusion : $B_\alpha(E) \rightarrow B_\alpha(E)$
- probabilistic unicity : $B_\alpha(E) \rightarrow - B_\beta(E)$, $\forall \alpha \neq \beta$
- probabilistic introspection : $B_\alpha(E) \rightarrow B_1(B_\alpha(E))$

A probabilistic schizophrenic belief would be of the form $B_\beta(B_\alpha(E))$, which would allow to hold simultaneously $B_{\beta'}(B_\alpha(E))$, with $\beta \neq \beta'$. The impossibility of such a belief follows immediately from the two previous properties.

ANNEX 2 : THE SYNTHETIC RULE

An agent considers the following belief about an urn :
- he imagines several alternative urns U_i containing balls B_j with objective probabilities $Pr(B_j/U_i)$;
- he adopts subjective probabilities $Pr(U_i)$ about these urns.

On one hand, by projection, agent's synthetic belief about the generic ball is given by :

$$\hat{\Pr}(B_j) = \sum_i \Pr(B_j / U_i) \Pr(U_i) = \sum_i \Pr(B_j \cap U_i)$$

When a message M is given about a ball extracted, agent's belief about that ball becomes :

$$\hat{\Pr}(B_j / M) = \frac{\Pr(B_j \cap M)}{\Pr(M)} = \sum_i \frac{\Pr(B_j \cap U_i \cap M)}{\Pr(M)}$$

On the other hand, when receiving the message M on the extracted ball, the agent revises simultaneously the probability of the urn where it may be extracted and the probability of the balls in each urn :

$$\hat{\hat{\Pr}}(U_i / M) = \frac{\Pr(U_i \cap M)}{\Pr(M)}$$

$$\hat{\hat{\Pr}}(B_j / U_i \cap M) = \frac{\Pr(B_j \cap U_i \cap M)}{\Pr(U_i \cap M)}$$

By collapsing in order to compute the probability of the ball extracted :

$$\hat{\hat{\Pr}}(B_j / M) = \sum_i \Pr(B_j / U_i \cap M) \Pr(U_i / M) = \sum_i \frac{\Pr(B_j \cap U_i \cap M)}{\Pr(M)}$$

The two methods are equivalent, thanks to the properties of the Bayes rule.

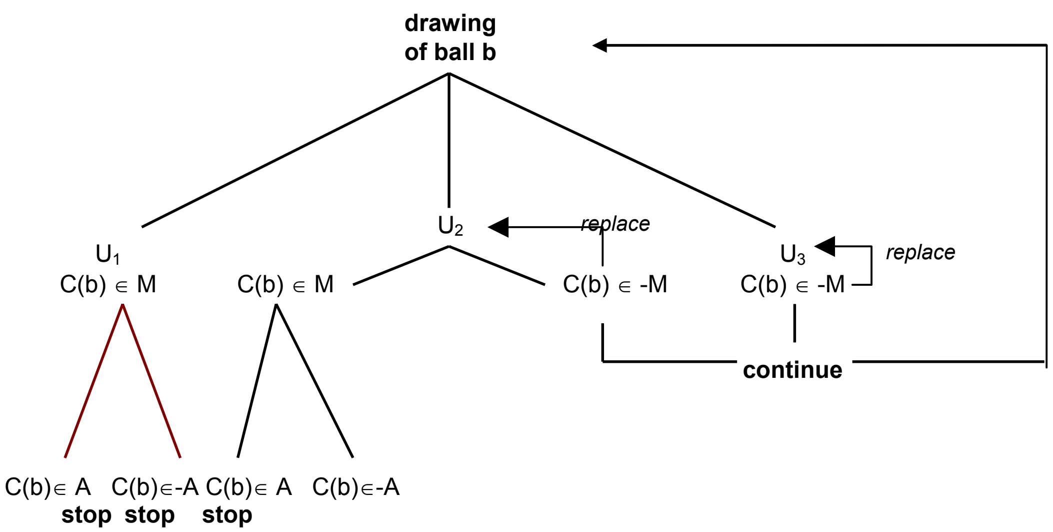
ANNEX 3 : THE JAFFRAY RULE

One considers the situation where an agent has an uncertain belief concerning an urn expressed by a distribution of events, i.e. the dual representation (Moebius inverse) of a belief function (Shafer, 1976, 1981). Accepting the Anti-projection Principle, that belief is equivalent to a certain belief in a meta-urn formed of urns, each represented by an event, i.e. composed of colored balls in unknown proportion. For instance, with the usual urn :

One considers now that a ball is extracted in two steps from the meta-urn, that a message M is given concerning the color of the ball, and that a final belief has to be formed about the ball. That procedure is assumed to be equivalent to the one consisting in extracting a ball in one step from the uncertain urn and in forming a final belief about that ball, i.e. usual focusing. For instance, a ball is extracted from the uncertain urn and the message says that the ball is not red, leading to a modified belief about the ball.

The belief in any event A will be given by a frequentist interpretation, transposed from a probabilistic to a credibilist framework, and grounded on two main assumptions (Shafer, 1981) :
- the first assumption states that the frequency of drawings which validate A is the minimal value of the frequency of drawings which validate A in an infinite sequence of drawings.
- The second assumption states that the frequency of drawings validating A among all drawings validating M in an infinite sequence of drawings is equal to the frequency of drawings validating A in an infinite sequence of drawings with replacing the balls of colors excluded by M.
By combining both assumptions, one has to compute the credibility of event A by the minimal value of the frequency of drawings validating A in an infinite sequence of drawings with replacement if -M.

The procedure can be summarized by considering three types of urns with balls b of color C(b):
- urns U_1 containing only balls with color in M
- urns U_2 containing balls simultaneously in M and in -M
- urns U_3 containing only balls with color in -M



REFERENCES

ALCHOURRON, C. E. - GÄRDENFORS, P. - MAKINSON, D. (1985): On the logic of theory change: partial meet contraction and revision functions, *Journal of Symbolic Logic,* 50, 510-530

BARON, J. (1987) : Second order probabilities and belief functions, *Theory and Decision*, 23,25-26.

BILLOT, A., WALLISER, B. (1999): A mixed knowledge hierarchy, *Journal of Mathematical Economics*, 32, 185-205.

CARNAP, R. (1945) : The two concepts of Probability, *Philosophy and Phenomenological Research*, V, 513-532.

CHATEAUNEUF,A.-JAFFRAY,J.Y. (1989) : Some characterization of lower probabilities and other monotone capacities through the use of Möbius Inversion, *Mathematical Social Sciences*, 17, 263-283.

COHEN,M.-GILBOA,I.-JAFFRAY,J.Y.-SCHMEIDLER,D. (1999) : An experimental study of ambiguous beliefs, 1st International Symposium on Imprecise Probabilities.

DUBOIS,D.-PRADE,H. (1994) :

ELLSBERG, D. (1961) : Risk, ambiguity and the Savage axioms, *quarterly Journal of Economics*, 75, 643-669.

FAGIN,R.- HALPERN,J.Y. (1990) : A new approach to updating beliefs, *Proc.of 6th Conference on Uncertainty in A.I.*

GÄRDENFORS, P. (1988): *Knowledge in Flux*, MIT Press

GILBOA,I.-SCHMEIDLER (1993) : Updating ambiguous beliefs, *Journal of Economic Theory*, 59,39-49.

JAFFRAY, J.Y., (1990) : Bayesian updating and belief functions, *Proc. of the 3d International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU),* Paris, July 2-6, 449-451.

JENSEN, F.V. (1996) : *An introduction to Bayesian Networks*, UCL Press, 1996.

KATSUNO, A. - MENDELZON, A. (1992): On the difference between updating a knowledge base and revising it, in P. Gärdenfors ed.: *Belief Revision*, Cambridge University Press, 183-203

KYBURG, H. (1989) : Higher order degrees of belief, in D.H.Mellor, T.S. Levitt, J.F.Lemmer eds : *Uncertainty in Artificial Intelligence*, Elsevier, 15-22.

LEWIS, D.K. (1976): Probabilities of conditionals and conditional probabilities, *Philosophical Review*, 85, 297-315

LEWIS, D.K. (1980) : A subjectivist guide to objective chance, in R.Jefrey, ed.: *Studies in Inductive Logic and Probability, vol II*, University of California Press; reprinted in *Philosophical Papers*, vol.II, Oxford University Press, 1986.

MILLER, D.(1966) : A paradox of information, *British Journal for the Philosophy of Science*, 17, 59-61.

PEARL, J. (1988) : *Probabilistic Reasoning in Intelligent Systems, Networks of Plausible Inference*, Morgan Kaufman Publishers Inc.

SHAFER, G. (1976) : *A Mathematical Theory of Evidence,* Princeton University Press.

SHAFER, G. (1990) : Perspectives on the Theory and Practice of Belief Functions,

SKYRMS, B. (1977) : Resiliency, Propensities and Causal Necessity, *Journal of Philosophy*, 74, 704-713

SKYRMS, B. (1980a) : Second order probabilities and fallible learning, in B.Skyrms : *Causal Necessity*, Yale University Press, 162-176.

SKYRMS, B. (1980b) : Higher orders degrees of belief, in D.H.Mellor ed.: *Prospects for Pragmatism*, Cambridge University Press, 162-176.

SUPPES, P., ZANOTTI, M., (1977) : « On using random relations to generate upper and lower probability », Synthese, 36,427-440.

SUPPES, P. (1981) : *Logique du probable*, Flammarion.

VAN FRAASSEN, B.(1989) : *Laws and Symmetry*, Oxford University Press, , trad.fr.: *Lois & Symétrie,* Paris, Vrin, 1994.

WALLISER, B.- ZWIRN, D. (2000) : Probabilistic belief change principles - to appear.

ANNEX 4 : THE KULLBACK DISTANCE

Consider a set K of 1-worlds, i.e. a set of probability distributions over three basic 0-worlds :
K = {r s.t. (p_1^r , p_2^r , p_3^r) }
Consider the message the third 0-world is not in the 1-world, and the set of probability distribution r' which are compatible with this message :
K# = {r' s.t. $(p_1^{r'} , p_2^{r} , 0)$ }
The problem is to find the distributions r' which are the nearest to K, the distance between distribution r' $\in$ K# and set K being taken as the minimum distance between distribution r' and all the distributions r $\in$ K.
For each probability distribution r, consider the nearest one compatible with the message, according to

$$\underset{r'}{Min}\, D = \sum_i \overline{p}^{r'} Log \frac{\overline{p_i}^{r'}}{p_i^{r}}$$

the Kullback distance :

As well known, it is precisely the Bayes transform of r :

$p_1^{r'} = p_1^r / (1- p_3^r)$
$p_2^{r'} = p_2^r / (1- p_3^r)$

Moreover, the Kullback distance between the initial distribution r and the final distribution r' is :
D = Log 1/ $(1- p_3^r)$
In order to consider the nearest distributions of the set K, one keeps only the distributions such that $(1- p_3^r)$ is maximal, i.e. the distribution which maximize the probability of the message.