ELSEVIER

Artificial Intelligence 98 (1998) 169–208

Artificial Intelligence

Characterizing the principle of minimum cross-entropy within a conditional-logical framework

Gabriele Kern-Isberner [1]

LG Praktische Informatik VIII, Department of Computer Science, FernUniversität Hagen, P.O. Box 940, D-58084 Hagen, Germany

Received October 1996; revised May 1997

Abstract

The principle of minimum cross-entropy (ME-principle) is often used as an elegant and powerful tool to build up complete probability distributions when only partial knowledge is available. The inputs it may be applied to are a prior distribution P and some new information $\mathcal{R}$, and it yields as a result the one distribution P^* that satisfies $\mathcal{R}$ and is closest to P in an information-theoretic sense. More generally, it provides a "best" solution to the problem "How to adjust P to $\mathcal{R}$?"

In this paper, we show how probabilistic conditionals allow a new and constructive approach to this important principle. Though popular and widely used for knowledge representation, conditionals quantified by probabilities are not easily dealt with. We develop four principles that describe their handling in a reasonable and consistent way, taking into consideration the conditional-logical as well as the numerical and probabilistic aspects. Finally, the ME-principle turns out to be the only method for adjusting a prior distribution to new conditional information that obeys all these principles.

Thus a characterization of the ME-principle within a conditional-logical framework is achieved, and its implicit logical mechanisms are revealed clearly. © 1998 Elsevier Science B.V.

Keywords: Probabilistic reasoning; Minimum cross-entropy; Conditionals; Knowledge representation; Nonmonotonic reasoning; Expert systems

[1] Email: gabriele.kern-isberner@fernuni-hagen.de.

PII S0004-3702(97)00068-4

1. Introduction

Within the last decades, knowledge representation and reasoning based upon probability theory have received increasing attention in the area of artificial intelligence. Probability theory provides a solid foundation for nonmonotonic reasoning methods (cf. e.g. [2,3,10,11,28]), and probabilistic expert systems make use of the consistent computability of (quantified) uncertainty (cf. [22,30,31]). Probabilistic methods are applied effectively within domains where statistical databases are available, checking and representing important relationships between the variables investigated, e.g. dependencies between diseases and symptoms in medical diagnosis or between profession and consumption in marketing. Just as well, they may be used to represent commonsense knowledge, as in the most popular example "Birds generally fly, penguins are birds but do not fly", by assigning appropriate probabilities to the assertions. Probabilistic reasoning is nonmonotonic in itself, and so the correct answer to the question "What about Tweety?" where Tweety is a penguin *and* a bird may be derived easily—Tweety does not fly in an adequately specified probabilistic environment.

Usually, probabilistic knowledge is represented by a probability distribution P or by a system of compatible distributions over a set of (discrete or continuous) variables, and inferences are made by calculating conditional probabilities. Thus the notion of conditionals is central to probabilistic inference. On the other hand, conditionals play a major part in knowledge representation and reasoning (cf. e.g. [8,25,29]). The range of their expressiveness includes commonsense knowledge as well as scientific statements, proving them to be quite a natural and fundamental means to formalize important relationships. An appropriate probabilistic representation of (quantified) conditionals would close the circle connecting probability theory as a mathematically established method to handle quantified uncertain knowledge, and conditionals as a popular tool to express knowledge.

Unfortunately, probabilistic representations have to struggle against arbitrariness. Usually, a (consistent) set $\mathcal{R}$ of conditionals, each equipped with a probability, will provide only incomplete probabilistic knowledge, so there will be a lot of distributions which all fulfill the probabilistic conditionals in $\mathcal{R}$. Which of them should be chosen to be the "most adequate" one? More generally, how to proceed if a (prior) probability distribution P is present that has to be adjusted to some new conditional information R, resulting in a posterior distribution P^*?

The aim of this paper is to establish a direct and constructive link between probabilistic conditionals and their suitable representation via distributions, taking prior knowledge into account if necessary. We develop the following four principles which mark the corner-stones for using quantified conditionals consistently for probabilistic knowledge representation and updating:

(P1) the *principle of conditional preservation*: this is to express that prior conditional dependencies shall be preserved "as far as possible" under adaptation;
(P2) the idea of a *functional concept* which underlies the adaptation and which allows us to calculate a posterior distribution from prior and new knowledge;

(P3) the *principle of logical consistency*: posterior distributions shall be used consistently as priors for further inferences; and
(P4) the *principle of representation invariance*: the resulting distribution shall not be dependent on the actual probabilistic representation of the new information.

(P1) links numerical changes to the conditional structure of the new information. (P2) realizes a computable relationship between prior and posterior knowledge by means of appropriate real functions. (P3) forestalls ambivalent results of updating procedures, and (P4) should be self-evident within a probabilistic framework. As we will show, the only method that solves the representation respectively adjustment problem stated above while obeying all of the principles (P1) to (P4) is provided by the *principles of maximum entropy* respectively of *minimum cross-entropy* (ME-principles), both well known from statistics and information theory (cf. e.g. [6,14–16,21]; a short introduction is given in Section 3). The first two axioms (P1) and (P2) will lead to a scheme for adjusting a prior distribution to new conditional information, and the principles of logical consistency and of representation invariance will be applied to this scheme, yielding the desired result. Thus a new characterization of the ME-principles is obtained, completely based on probabilistic conditionals and establishing reasoning at optimum entropy as a most fundamental inference method in the area of quantified uncertain reasoning.

Two earlier papers [26,33] are concerned with characterizing the ME-principles as logically consistent inference methods, too. Shore and Johnson [17,33] succeeded in proving (cross-) entropy to be the only functional the optimization of which satisfies four (respectively five) fundamental axioms of probabilistic inference. A similar result is attained for entropy in [26] by Paris and Vencovská without assuming that inference is performed by optimizing a functional, but heavily relying on solving linear equational systems. Among the properties these authors used for their characterizations are independence and invariance properties in the first place. This justifies ME-inference as an inference procedure of minimal changes, but very few was said about the nature or the extent of changes actually occurring under ME-adjustment.

The present paper points out a more constructive approach to the ME-principles. We show here that ME-inference not only respects (*conditional*) *independencies* but that it is basically determined by *conditional dependencies* (obeying independence properties where no dependency exists), recommending the ME-principles as most adequate methods for reasoning with probabilistic conditionals. Therefore, in contrast to Bayesian networks (cf. e.g. [22]), probabilistic networks based on ME-techniques (cf. [30,31]) do not require lots of probabilities and independence assumptions to process quantified conditional knowledge properly.

Shore and Johnson [33] as well as Paris and Vencovská [26] based their characterizations on more general probabilistic constraints than probabilistic conditionals. Thus the results given in this paper are actually stronger, using only a proper subtype of constraints (cf. [27]). Moreover, the methods used here are quite different from those in [33] and in [26]. In particular, there will be no need to make use of optimization theory, as in [33], or to transfer the problem into the context of linear algebra, as in [26]. Our development explains clearly how the ME-principles may be completely based on probabilistic conditionals. This may improve significantly the explanatory features of computational systems that use these principles for knowledge representation and processing (as, e.g. SPIRIT, cf. [30,31]). For instance, using the representation formula (3) in Section 3 respectively (12) in Section 4.2 revealing the conditional-logical pattern of the ME-distribution, it is possible to indicate which of the conditionals given by the set $\mathcal{R}$ actually make a contribution to a conditional information derived from the posterior distribution (similar to listing active rules in rule based systems).

A certain amount of mathematics and technical details will be necessary to formalize correctly the ideas behind the four principles (P1) to (P4) and to prove the desired results. We will endeavor to give informal reasons for definitions and theorems, enclosing all proofs in an Appendix not to impair the readability of the paper. This paper is organized as follows: The following section provides some preliminaries to probabilistic logic, so as to fix notations and describe fundamental relations. Section 3 is dedicated to a brief presentation of ME-methods, pointing out a first striking parallel between them and conditional logic. Section 4 deals with the *principle of conditional preservation* which will be based on the algebraic representation of *conditional structures*. Section 5 elaborates a functional concept which is to realize the idea of a computable solution. In the following sections, the last two postulates are dealt with, that of *logical consistency* in Section 6 and that of *representation invariance* in Section 7. We prove how they both influence the type of the functions involved in the functional concept so as to determine them uniquely. The adaptation scheme based on this distinguished concept indeed yields a unique posterior distribution, as is proved in Section 8. Finally, we are able to state the main result of this paper: The characterization of the ME-adjustment operator $*_e$ (Theorem 31). Section 9 presents some easy but essential properties of $*_e$, and in the concluding Section 10 we address connections to nonmonotonic reasoning and theory revision. All proofs may be found in the Appendix.

2. Probabilistic conditionals

We consider probability distributions P over a finite set $\mathcal{V} = \{V_1, V_2, V_3, \ldots\}$ of propositional variables V_i which are assumed to be binary. The dotted *literal* $\dot{v}_i \in \{v_i, \bar{v}_i\}$ stands for one of the two possible outcomes of the corresponding variable: v_i symbolizes "V_i is true", and negation is indicated by barring, i.e. $\bar{v}_i = \neg v_i$. P is uniquely determined by the values of its probability function p applied to the elementary events $\omega = \dot{v}_1\dot{v}_2\dot{v}_3 \ldots$, $p(\omega) = p(\dot{v}_1\dot{v}_2\dot{v}_3 \ldots) = P(V_1 = \dot{v}_1, V_2 = \dot{v}_2, V_3 = \dot{v}_3, \ldots)$. The distinction between a probability measure P and its probability function p is not essential but used for the sake of correctness throughout this paper. Let Ω denote the set of all elementary events: $\Omega = \{\omega = \dot{v}_1\dot{v}_2\dot{v}_3 \ldots \mid \dot{v}_i \in \{v_i, \bar{v}_i\}\}$.

A propositional language $\mathcal{L} = \mathcal{L}(\mathcal{V})$ is defined in the usual way, using the letters of the alphabet $\mathcal{V}$ and the classical connectives $\wedge$ (respectively juxtaposition) and $\neg$. Its formulas are denoted by capital Roman letters A, B, C, ...

To each propositional formula $\mathrm{A} \in \mathcal{L}$ a probability may be assigned via $p(\mathrm{A}) = \sum_{\omega:\, \mathrm{A}(\omega)=1} p(\omega)$, where the sum is taken over all elementary events ω, and $\mathrm{A}(\omega) = 1$

means that the complete conjunction corresponding to ω is a disjunct in the canonical disjunctive normal form of A. $p(\mathrm{A})$ reflects the probability that an arbitrary element of the population has properties which are described by A. Thus the correspondence between complete conjunctions and elementary events induces a probabilistic interpretation of $\mathcal{L}$ (on the base of the distribution P).

Now $\mathcal{L}$ is extended to a probabilistic conditional language $\mathcal{L}^*$ by introducing a conditional operator $\rightsquigarrow$ and probabilities in the following way: A *probabilistic conditional* (or a *probabilistic rule*, both terms are used synonymously) is an expression $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$ with *antecedent* $\mathrm{A} \in \mathcal{L}$, *conclusion* $\mathrm{B} \in \mathcal{L}$ and probability $x \in [0,1]$. It is to represent syntactically non-classical conditional assertions $\mathrm{A} \rightsquigarrow \mathrm{B}$ weighted with a degree of certainty x. A *probabilistic fact* has the form $\mathrm{B}[x]$, $\mathrm{B} \in \mathcal{L}$, $x \in [0,1]$ and is considered to be equivalent to the conditional $\top \rightsquigarrow \mathrm{B}[x]$, where $\top$ is tautological.

$\mathcal{L}^* = \{\mathrm{A} \rightsquigarrow \mathrm{B}[x] \mid \mathrm{A},\mathrm{B} \in \mathcal{L},\ x \in [0,1]\}$ is a flat conditional language, the conditional operator $\rightsquigarrow$ must not be nested. Antecedent and conclusion of a conditional are propositional formulas.

A semantic interpretation of probabilistic conditionals is given by conditional probabilities: If P is a distribution, P *fulfills* $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$, $P \models \mathrm{A} \rightsquigarrow \mathrm{B}[x]$, iff $p(\mathrm{A}) > 0$ and $p(\mathrm{B}|\mathrm{A}) = p(\mathrm{AB})/p(\mathrm{A}) = x$. In the sequel, we will tacitly assume that all antecedents of conditionals have positive probabilities.

In general, we have

$$x = p(\mathrm{B}|\mathrm{A}) \quad \text{iff} \quad p(\mathrm{A}) > 0 \text{ and } (1-x)p(\mathrm{AB}) = xp(\mathrm{A\overline{B}}),$$

so the quotient $p(\mathrm{AB})/p(\mathrm{A\overline{B}})$ determines the probability of the conditional $\mathrm{A} \rightsquigarrow \mathrm{B}$. It represents the proportion of individuals or objects with property A which also have property B to those that B is not true of. Thus it is crucial for the acceptability of the conditional, not only within a probabilistic framework (cf. [25]).

For a probability distribution P, let $Th(P) = \{\mathrm{A} \rightsquigarrow \mathrm{B}[x] \in \mathcal{L}^* \mid P \models \mathrm{A} \rightsquigarrow \mathrm{B}[x]\}$ denote the set of all probabilistic conditionals which are valid in P. $Th(P)$ explicitly represents the conditional knowledge embodied in P.

The problem this paper is going to deal with can now be described in a more formal manner:

(*) Given a prior distribution P and some set of probabilistic conditionals $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\} \subset \mathcal{L}^*$, how should P be modified to yield a posterior distribution P^* with $P^* \models \mathcal{R}$?

To maintain compatibility between prior and posterior distributions, P^* has to be P-continuous ($P^* \ll P$), i.e. $p(\omega) = 0$ implies $p^*(\omega) = 0$. Thus to avoid obvious inconsistencies, the set $\mathcal{R}$ is supposed to be *P-consistent*, that means there is some distribution Q with $Q \ll P$ and $Q \models \mathcal{R}$. Throughout this paper, we will assume without further mentioning that the necessity of zero posterior probabilities is stated explicitly in $\mathcal{R}$, i.e. if for any $Q \ll P$, $Q \models \mathcal{R}$ implies $q(\omega) = 0$ then $p(\omega) = 0$, or there is a conditional $\mathrm{A} \rightsquigarrow \mathrm{B}[x] \in \mathcal{R}$ such that either $x = 1$ and $\mathrm{A\overline{B}}(\omega) = 1$ or $x = 0$ and $\mathrm{AB}(\omega) = 1$.

In the next section, we are going to present a special solution to the adjustment problem (*): the distribution at optimum entropy.

3. The principles of entropy

The entropy $H(P) = -\sum_\omega p(\omega)\log p(\omega)$ (where the sum is taken over all elementary events ω, using the convention $0\log 0 = 0$) of a distribution P first appeared as a physical quantity in statistical mechanics and was later interpreted by Shannon as an information-theoretic measure of the uncertainty inherent to P (for a historical review, cf. [16]). It is generalized by the notion of *cross-entropy* (also called *relative entropy*) $R(Q,P) = \sum_\omega q(\omega)\log(q(\omega)/p(\omega))$ (with $0\log(0/0) = 0$ and $q(\omega)\log(q(\omega)/0) = \infty$ for $q(\omega) \neq 0$) between two distributions Q and P. If P_0 denotes the uniform distribution $p_0(\omega) = 1/m$ for all elementary events ω, then

$$R(Q,P_0) = -H(Q) + \log m$$

relates absolute and relative entropy. So maximizing absolute entropy under some given constraints is equivalent to minimizing relative entropy to the uniform distribution under the same constraints. Therefore the *principle of minimum cross-entropy*

$$\begin{aligned}&\min R(Q,P) = \sum_\omega q(\omega)\log\frac{q(\omega)}{p(\omega)}\\ &\text{s.t. } Q \text{ is a probability distribution with } Q \models \mathcal{R}\end{aligned} \tag{1}$$

can be regarded as more general than the *principle of maximum entropy*

$$\begin{aligned}&\max H(Q) = -\sum_\omega q(\omega)\log q(\omega)\\ &\text{s.t. } Q \text{ is a probability distribution with } Q \models \mathcal{R}.\end{aligned} \tag{2}$$

We refer to both principles as the *ME-principle*, where the abbreviation ME stands both for *M*inimum cross-*E*ntropy and for *M*aximum *E*ntropy.

Cross-entropy is a well-known information-theoretic measure of dissimilarity between two distributions and has been studied extensively (for a brief, but informative introduction and further references cf. [32]; cf. [34]). In particular, optimizing entropy is known to yield best expectation values in statistics (cf. [14,16]). Cross-entropy is also called *directed divergence* for it lacks symmetry, i.e. $R(Q,P)$ and $R(P,Q)$ differ in general, so it is not a metric. But cross-entropy is *positive*, that means we have $R(Q,P) \geqslant 0$, and $R(Q,P) = 0$ iff $Q = P$ (cf. [6,32]).

For a distribution P and some P-consistent set $\mathcal{R}$ of probabilistic rules there is a distribution $P_e = P_e(P,\mathcal{R})$ that fulfills $\mathcal{R}$ and has minimal relative entropy to the prior P (cf. [6]), i.e. P_e solves (1).

The condition $Q \models \mathcal{R}$ imposed on a distribution Q can be transformed equivalently into a system of linear equality constraints for the probabilities $q(\omega)$. Using the Lagrangian techniques, we may represent P_e in the form

$$p_e(\omega) = \alpha_0 p(\omega) \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^{1-x_i} \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} \alpha_i^{-x_i}, \tag{3}$$

with the α_i's being exponentials of the Lagrange multipliers, one for each conditional in $\mathcal{R}$, and $\alpha_0 = \exp(\lambda_0 - 1)$, where λ_0 is the Lagrange multiplier of the constraint $\sum_\omega q(\omega) = 1$.

By construction, P_e satisfies all conditionals in $\mathcal{R}$: $p_e(\mathrm{B}_i|\mathrm{A}_i) = x_i$, which is equivalent to $(1-x)p_e(\mathrm{A}_i\mathrm{B}_i) = xp_e(\mathrm{A}_i\overline{\mathrm{B}_i})$ for all i, $1 \leqslant i \leqslant n$. So $\alpha_1, \ldots, \alpha_n$ are solutions of the nonlinear equations

$$\alpha_i = \frac{x_i}{1-x_i} \frac{\sum_{\omega:\mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1} p(\omega) \prod_{j\neq i,\mathrm{A}_j\mathrm{B}_j(\omega)=1} \alpha_j^{1-x_j} \prod_{j\neq i,\mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1} \alpha_j^{-x_j}}{\sum_{\omega:\mathrm{A}_i\mathrm{B}_i(\omega)=1} p(\omega) \prod_{j\neq i,\mathrm{A}_j\mathrm{B}_j(\omega)=1} \alpha_j^{1-x_j} \prod_{j\neq i,\mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1} \alpha_j^{-x_j}}, \tag{4}$$

$$\text{with} \quad \alpha_i \begin{cases} > 0, & x_i \in (0,1), \\ = \infty, & x_i = 1, \\ = 0, & x_i = 0, \end{cases} \qquad 1 \leqslant i \leqslant n,$$

using the conventions $\infty^0 = 1$, $\infty^{-1} = 0$ and $0^0 = 1$. α_0 arises simply as a normalizing factor. Each α_i symbolizes the impact of the corresponding rule when P is modified. It depends on the prior distribution P, the other rules and probabilities in $\mathcal{R}$ and—in a distinguished way—on the probability of the conditional it corresponds to.

Though the formulas above appear deterrently complex at first sight, (3) shows rather clearly how the ME-adjustment to a rule is carried out: Apart from the normalizing factor α_0, at most the probabilities of those complete conjunctions ω are changed which satisfy the antecedent of this rule. And the new probability depends in addition on whether ω satisfies the conclusion or not. In particular, the probabilities of all conditionals of $Th(P)$ whose antecedents do not fulfill an antecedent of any of the rules in $\mathcal{R}$ remain unchanged. This means that the ME-adaptation respects one of the fundamental principles of conditional logics: Asserting a conditional should only affect the (conditional) knowledge about states which the conditional may be applied to.

This intuitive and reasonable *principle of conditional preservation* will be elaborated more deeply in the next section.

But first, we will illustrate the use of the ME-approach and the benefits of the representation formulas (3) and (4) by two simple but informative examples. The first example shows knowledge processing in the case of conflicting information, whereas the second example deals with transitive inference. All numerical results were obtained by using the probabilistic expert system SPIRIT which realizes knowledge processing at optimum entropy (cf. [31]).

Example 1 (*Conflicting information*). A knowledge base is to be built up representing "Typically, students are adults", "Usually, adults are employed" and "Mostly, students are not employed" with probabilities subjectively attached to of 0.99, 0.8 and 0.9, respectively. Let A, S, E denote the propositional variables A = Being an Adult, S = Being a Student, and E = Being Employed. The quantified conditional information may be written as $\mathcal{R} = \{s \rightsquigarrow a[x_1], a \rightsquigarrow e[x_2], s \rightsquigarrow e[x_3]\}$, $x_1 = 0.99$, $x_2 = 0.8$, $x_3 = 0.1$. No prior information is at hand, so we start from the uniform distribution.

We are interested in the probability of the conditional $as \rightsquigarrow e$ the antecedent of which combines the evidences a and s conflicting with respect to E.

SPIRIT calculates $p_e(e|as) = 0.1009$, which is much more closer to x_3 than to x_2. So the more specific information s dominates a clearly but not completely, as we should expect. If we set $x_1 = 1$, assuming the set of students to be definitely a subset of the set of adults, this preference of the more specific knowledge conveyed by the third rule is a probabilistic necessity, and may also be seen clearly by using (3) and (4):

Let P'_e denote the ME-solution to the same problem as before, except that now $x_1 = 1$ instead of $x_1 = 0.99$. For arbitrary x_2, x_3, we obtain

$$\frac{p'_e(ase)}{p'_e(as\bar{e})} = \frac{\alpha_2^{1-x_2}\alpha_3^{1-x_3}}{\alpha_2^{-x_2}\alpha_3^{-x_3}} = \alpha_2\alpha_3$$

with

$$\alpha_2 = \frac{x_2}{1-x_2}\frac{\alpha_3^{-x_3}+1}{\alpha_3^{1-x_3}+1}, \qquad \alpha_3 = \frac{x_3}{1-x_3}\frac{\alpha_2^{-x_2}}{\alpha_2^{1-x_2}} = \frac{x_3}{1-x_3}\alpha_2^{-1}.$$

This implies at once $\alpha_2\alpha_3 = x_3/(1-x_3)$ (note that the normalizing factor and the constant priors are being canceled, and that $\alpha_1 = \infty$). This shows $p'_e(e|as) = x_3$.

Thus ME-inference solves in an elegant way the problem of conflicting evidences. Specific information dominates more general knowledge by virtue of the inherent mechanisms, without any external preferential or hierarchical structures as in [4,20], and without rankings as in [10,12]. The weight of a rule is encoded by its logical structure and its probability, its dependency on other rules being given implicitly. It is only the application of the ME-principle which combines the probabilistic rules to yield inferences, thus allowing a convenient modularity of knowledge representation.

In the second example, we will make once more use of symbolic calculations to reveal knowledge processing.

Example 2 (*Transitivity*). Let Y, S, C denote the three propositional variables Y = Being *Y*oung, S = Being *S*ingle and C = Having *C*hildren. We know (or assume) that young people are usually singles (with probability 0.9) and that mostly, singles do not have children (with probability 0.85). Here we have $\mathcal{R} = \{y \rightsquigarrow s[x_1], s \rightsquigarrow \bar{c}[x_2]\}$ with $x_1 = 0.9$, $x_2 = 0.85$. Again we take the uniform distribution as prior information. A calculation with SPIRIT shows $y \rightsquigarrow \bar{c}[0.815]$, connecting both rules transitively. By use of the formulas (3) and (4), a more general transitive inference rule can be proved:

For arbitrary x_1, x_2, we obtain for the ME-distribution P_e:

$$\frac{p_e(y\bar{c})}{p_e(yc)} = \frac{\alpha_1^{1-x_1}\alpha_2^{1-x_2}+\alpha_1^{-x_1}}{\alpha_1^{1-x_1}\alpha_2^{-x_2}+\alpha_1^{-x_1}} = \frac{\alpha_1\alpha_2^{1-x_2}+1}{\alpha_1\alpha_2^{-x_2}+1}$$

with

$$\alpha_1 = \frac{x_1}{1-x_1}\alpha_2^{x_2}\frac{2}{\alpha_2+1}, \qquad \alpha_2 = \frac{x_2}{1-x_2}.$$

This yields

$$\alpha_1\alpha_2^{-x_2} = \frac{2x_1(1-x_2)}{1-x_1} \quad \text{and} \quad \frac{p_e(y\bar{c})}{p_e(yc)} = \frac{1+2x_1x_2-x_1}{1-2x_1x_2+x_1},$$

proving

$$p_e(\bar{c}|y) = \frac{1}{2}(1+2x_1x_2-x_1). \tag{5}$$

Of course, the correctness of this formula is independent of the particular meanings of the propositional variables involved. So (5) states a general transitive inference rule for problems with an analogous knowledge structure. More ME-deduction rules may be found in [19].

These examples are but to give an idea of the soundness and the power of the ME-principle. The rest of this paper will be dedicated to its development from a conditional-logical point of view.

4. The principle of conditional preservation

4.1. Conditional structures

Following Calabrese [5] (and, earlier, De Finetti [7]), a conditional $A \rightsquigarrow B$ can be represented as a *generalized indicator function* (B|A) on elementary events, setting

$$(B|A)(\omega) = \begin{cases} 1, & \omega \in AB, \\ 0, & \omega \in A\overline{B}, \\ u, & \omega \notin A, \end{cases}$$

where u stands for *undefined*. This definition captures excellently the non-classical character of conditionals within a probabilistic framework. According to it, a conditional is a function that polarizes AB and $A\overline{B}$, leaving $\overline{A}$ untouched. Due to their non-Boolean nature, conditionals are rather complicated objects. In particular, it is not an easy task to handle the relationships between them so as to preserve conditional dependencies "as far as possible" under adaptation. To make the problem plain and to point out a possible way to solve it, we give an example which is taken from [36] and which illustrates a phenomenon also well known under the name "Simpson's paradox".

Example 3 (*Florida murderers*). This example is based on a real-life investigation. During the six year period 1973–1979, about 5000 murder cases were recorded in the US state of Florida, and the following distribution P mirrors the sentencing policy in those years (for further references, cf. [36, pp.46ff]). The propositional variables involved are V = Victim (of the murder) is black respectively white, $\dot{v} \in \{v_b, v_w\}$, M = Murderer is black respectively white, $\dot{m} \in \{m_b, m_w\}$, and D = Murderer is sentenced to Death, $\dot{d} \in \{d, \bar{d}\}$.

P:	$v_w m_w d$	0.0151	$v_w m_w \bar{d}$	0.4353
	$v_w m_b d$	0.0101	$v_w m_b \bar{d}$	0.0502
	$v_b m_w d$	0	$v_b m_w \bar{d}$	0.0233
	$v_b m_b d$	0.0023	$v_b m_b \bar{d}$	0.4637

Thus P implies

$$m_w \rightsquigarrow d[0.0319], \qquad m_b \rightsquigarrow d[0.0236],$$

so justice seemingly passed sentences without respect of color of skin. Differences, however, become strikingly apparent if the third variable V, revealing the color of skin of the *victim*, is also taken into account:

$$v_w m_w \rightsquigarrow d[0.0335], \qquad v_w m_b \rightsquigarrow d[0.1675],$$
$$v_b m_w \rightsquigarrow d[0], \qquad v_b m_b \rightsquigarrow d[0.0049].$$

If e.g. the probability of the conditional $m_b \rightsquigarrow d[0.0236]$ is to change, the probabilities of the rules $\dot{v}\dot{m} \rightsquigarrow d$ containing important information should be preserved in an adequate manner.

This last example illustrates a strange but typical behavior that marginal distributions and the conditionals involved may have. Let us look upon this problem in an abstract environment.

Suppose P is a distribution over a set of variables containing A, B, and suppose $P \models a \rightsquigarrow b[x]$. In which way may a third variable C affect this conditional, i.e. what can be said about the probability of $a\dot{c} \rightsquigarrow b$ in P?

Roughly, there are two possibilities. In the first case, C does not affect $a \rightsquigarrow b[x]$ at all, that is to say we have $p(b|a\dot{c}) = p(b|a)$. We may classify this as a *monotonic behavior*, showing B and C to be *conditionally independent given a* (cf. [36]). By a straightforward calculation, we see that $p(b|a\dot{c}) = p(b|a)$ iff

$$\frac{p(abc)p(a\bar{b}\bar{c})}{p(a\bar{b}c)p(ab\bar{c})} = 1.$$

In the second, more usual case, we have $p(b|a\dot{c}) \neq p(b|a)$, and consequently

$$\frac{p(abc)p(a\bar{b}\bar{c})}{p(a\bar{b}c)p(ab\bar{c})} \neq 1.$$

Thus departures from conditional independence—and thereby the *extent of nonmonotonicity*, to introduce a logical aspect—may be measured by the cross product ratio or *interaction quotient*

$$\frac{p(abc)p(a\bar{b}\bar{c})}{p(a\bar{b}c)p(ab\bar{c})}.$$

A reasonable demand for a posterior distribution P^* adapted to a changed probability of $a \rightsquigarrow b$ then is that posterior interaction should be the same as prior interaction, i.e.

$$\frac{p^*(abc)p^*(a\bar{b}\bar{c})}{p^*(a\bar{b}c)p^*(ab\bar{c})} = \frac{p(abc)p(a\bar{b}\bar{c})}{p(a\bar{b}c)p(ab\bar{c})}. \tag{6}$$

In statistics, logarithms of such expressions are used to measure the interactions between the variables involved (cf. [13,36]).

In the general case, we consider formulas A, B instead of variables A, B, joint influences of groups of variables (instead of one single variable) on the conditional A $\rightsquigarrow$ B, and, last not least, we have to take a *set* of conditionals into account. Thus the notion of (statistical) interaction quotients has to be generalized, involving more elementary events both in the numerators and in the denominators and being based appropriately on $\mathcal{R}$. The comments above following Example 3 give interaction quotients a logical meaning that fits the intention of this paper better than a statistical interpretation and offers a suitable way to carry out the necessary generalization from a conditional-logical point of view:

In (6), two sets of elementary events are related to each other with respect to P and P^*: $\{abc, a\bar{b}\bar{c}\}$ in the numerator, and $\{a\bar{b}c, ab\bar{c}\}$ in the denominator. In both sets, the conditional $a \rightsquigarrow b$ is once confirmed (by abc respectively $ab\bar{c}$) and once refuted (by $a\bar{b}\bar{c}$ respectively $a\bar{b}c$), so both sets show the same behavior with regard to the new conditional $a \rightsquigarrow b[x]$. This idea of a *behavior* or *structure with respect to* $\mathcal{R}$ may be formalized easily for sets or multi-sets of elementary events. We choose a group theoretical presentation.

To each conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ we associate two symbols a_i, b_i. Let $F_{\mathcal{R}} = \langle a_1, b_1, \ldots, a_n, b_n \rangle$ be the free abelian group with generators $a_1, b_1, \ldots, a_n, b_n$, i.e. $F_{\mathcal{R}}$ consists of all elements of the form $a_1^{z_1} b_1^{w_1} \ldots a_n^{z_n} b_n^{w_n}$ with integers $z_i, w_i \in \mathbb{Z}$ (the ring of integers)), and each element can be identified by its exponents so that $F_{\mathcal{R}}$ is isomorphic to $\mathbb{Z}^{2n}$ (cf. [23]). The commutativity of $F_{\mathcal{R}}$ corresponds to the fact that the conditionals in $\mathcal{R}$ shall be effective all at a time, without assuming any order of application.

For each i, $1 \leqslant i \leqslant n$, we define a function $\sigma_i : \Omega \to F_{\mathcal{R}}$ by setting

$$\sigma_i(\omega) = \begin{cases} a_i & \text{if } (\mathrm{B}_i|\mathrm{A}_i)(\omega) = 1, \\ b_i & \text{if } (\mathrm{B}_i|\mathrm{A}_i)(\omega) = 0, \\ 1 & \text{if } (\mathrm{B}_i|\mathrm{A}_i)(\omega) = u. \end{cases}$$

$\sigma_i(\omega)$ represents the manner in which the conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ applies to the elementary event ω. The neutral element 1 of $F_{\mathcal{R}}$ corresponds to the non-applicability of $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in case that the antecedent A_i is not satisfied. The function

$$\sigma = \sigma_{\mathcal{R}} : \Omega \to F_{\mathcal{R}}, \qquad \sigma(\omega) = \omega^{\sigma} := \prod_{1 \leqslant i \leqslant n} \sigma_i(\omega) = \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} a_i \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} b_i \tag{7}$$

describes the all-over effect of $\mathcal{R}$ on ω. ω^{σ} is called the *conditional structure of* ω *with respect to* $\mathcal{R}$. Having the same conditional structure defines an equivalence relation $\equiv_{\mathcal{R}}$

on Ω. For each elementary event ω, ω^{σ} contains at most one of each a_i or b_i, but never both of them because each conditional applies to ω in a well-defined way.

The notion of conditional structure is generalized to *multi-sets* in a straightforward way. A *multi-set* is a collection of elements as a set, except that each element may occur more than once in a multi-set. Multi-sets are also known as *bags*. In the sequel, we will use the following notation:

Notation 4 (*Multi-set*). A *multi-set* containing a finite number of elements $\gamma_1, \gamma_2, \ldots$, each element γ_j occurring with multiplicity m_j, is denoted as $\{\gamma_1 : m_1, \gamma_2 : m_2, \ldots\}$. The cardinality of such a multi-set is the sum $\sum_j m_j$ of its multiplicities.

Definition 5 (*Conditional structure of multi-sets*). Let $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ denote a multi-set of elementary events. The element $\Omega_1^{\sigma} := \prod_{1 \leqslant i \leqslant m_1} \sigma(\omega_i)^{r_i} \in F_{\mathcal{R}}$ is called the *conditional structure of* Ω_1.

Thus the conditional structure Ω_1^{σ} of a multi-set $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ is represented by a group element which is a product of the generators a_i, b_i of $F_{\mathcal{R}}$, with each a_i occurring with exponent $\sum_{k:\, \sigma_i(\omega_k)=a_i} r_k = \sum_{k:\, (\mathrm{B}_i|\mathrm{A}_i)(\omega_k)=1} r_k$, and each b_i occurring with exponent $\sum_{k:\, \sigma_i(\omega_k)=b_i} r_k = \sum_{k:\, (\mathrm{B}_i|\mathrm{A}_i)(\omega_k)=0} r_k$ (note that each of the sums may be zero in case that the corresponding conditional cannot be applied to any of the elements in Ω_1). So the exponent of a_i in Ω_1^{σ} indicates the number of elementary events in Ω_1 which *confirm* the conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i$, each event being counted with its multiplicity, and in the same way the exponent of b_i indicates the number of elementary events that *refute* $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i$.

Ω_1^{σ} encodes this information in an elegant manner. We could have used a simpler representation, e.g. with tuples of natural numbers representing positive and negative applicabilities of conditionals. But making use of a group structure allows us to form products and thus provides a more convenient representation and handling of conditional structures. Moreover, the group element corresponding to an elementary event (7) obviously parallels the structure of its posterior probability, as we will see later on (see Theorem 11).

Definition 6 (*$\mathcal{R}$-equivalence of multi-sets*). Two multi-sets $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ and $\Omega_2 = \{\nu_1 : s_1, \ldots, \nu_{m_2} : s_{m_2}\}$ of elementary events with equal cardinalities $\sum_{1 \leqslant k \leqslant m_1} r_k = \sum_{1 \leqslant l \leqslant m_2} s_l$ are *$\mathcal{R}$-equivalent* iff $\Omega_1^{\sigma} = \Omega_2^{\sigma}$, i.e. iff their conditional structures with respect to $\mathcal{R}$ are identical.

The additional prerequisite that both multi-sets must have the same cardinality is due to the fact that actually, the normalizing conditional $\top[1]$ has also to be taken into account, requiring equal numbers of elementary events in both multi-sets.

Recalling the remarks following Definition 5 above, an immediate description of $\mathcal{R}$-equivalent multi-sets can be stated:

Lemma 7. *Two multi-sets* $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ *and* $\Omega_2 = \{\nu_1 : s_1, \ldots, \nu_{m_2} : s_{m_2}\}$ *are $\mathcal{R}$-equivalent iff all of the equations*

$$\sum_{1\leqslant k\leqslant m_1} r_k = \sum_{1\leqslant l\leqslant m_2} s_l,$$

$$\sum_{k:\,(\mathrm{B}_i|\mathrm{A}_i)(\omega_k)=1} r_k = \sum_{l:\,(\mathrm{B}_i|\mathrm{A}_i)(\nu_l)=1} s_l,$$

$$\sum_{k:\,(\mathrm{B}_i|\mathrm{A}_i)(\omega_k)=0} r_k = \sum_{l:\,(\mathrm{B}_i|\mathrm{A}_i)(\nu_l)=0} s_l$$

hold for all $i = 1,\ldots,n$.

4.2. C-adaptations

$\mathcal{R}$-equivalent multi-sets show an equal, indistinguishable behavior with respect to $\mathcal{R}$, so the corresponding posterior probabilities should be, in some sense, quite similar. Conditional structures, however, are abstract objects, independent of probabilities, and we have to relate them to the distributions involved appropriately. This can be achieved by considering the relative change function p^*/p, giving rise to the following idea:

If the changes the prior distribution undergoes when being adapted to $\mathcal{R}$ is to be based on $\mathcal{R}$ in a reasonable and intelligible way there should be no difference in the relative changes of probabilities of $\mathcal{R}$-equivalent multi-sets (as defined below).

Adaptations following this idea will be called *c-adaptations*.

Definition 8 (*C-adaptation*). Let P^* be a distribution which fulfills $\mathcal{R}$. P^* is called a *c-adaptation of* P *to* $\mathcal{R}$ iff it satisfies the following two conditions:

(i) For any $\omega \in \Omega$, $p^*(\omega) = 0$ if and only if $p(\omega) = 0$ or there is a conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in $\mathcal{R}$ such that either $x_i = 1$ and $\mathrm{A}_i\overline{\mathrm{B}_i}(\omega) = 1$, or $x_i = 0$ and $\mathrm{A}_i\mathrm{B}_i(\omega) = 1$.

(ii) The $\mathcal{R}$-equivalence of any two multi-sets $\Omega_1 = \{\omega_1 : r_1,\ldots,\omega_{m_1} : r_{m_1}\}$ and $\Omega_2 = \{\nu_1 : s_1,\ldots,\nu_{m_2} : s_{m_2}\}$ of elementary events ω_k, ν_l with $p(\omega_k), p(\nu_l) > 0$ implies

$$\frac{p^*(\omega_1)^{r_1}\cdots p^*(\omega_{m_1})^{r_{m_1}}}{p(\omega_1)^{r_1}\cdots p(\omega_{m_1})^{r_{m_1}}} = \frac{p^*(\nu_1)^{s_1}\cdots p^*(\nu_{m_2})^{s_{m_2}}}{p(\nu_1)^{s_1}\cdots p(\nu_{m_2})^{s_{m_2}}}, \tag{8}$$

which is equivalent to

$$\frac{p^*(\omega_1)^{r_1}\cdots p^*(\omega_{m_1})^{r_{m_1}}}{p^*(\nu_1)^{s_1}\cdots p^*(\nu_{m_2})^{s_{m_2}}} = \frac{p(\omega_1)^{r_1}\cdots p(\omega_{m_1})^{r_{m_1}}}{p(\nu_1)^{s_1}\cdots p(\nu_{m_2})^{s_{m_2}}}. \tag{9}$$

for $p^*(\omega_k), p^*(\nu_l) \neq 0$, $1 \leqslant k \leqslant m_1$, $1 \leqslant l \leqslant m_2$.

The first condition (i) above states that the posterior distribution is P-continuous and is positive otherwise if not demanded explicitly by $\mathcal{R}$. Note that in particular, two $\mathcal{R}$-equivalent multi-sets have the same cardinality. This precondition ensures that an equal number of factors is involved on both sides of (8) respectively both in numerators and denominators in (9). This allows to interpret (9) in fact as a generalized interaction quotient. So, starting from problems involving interaction quotients, we discovered the concept of conditional structures as an adequate means to describe the behavior of sets of elementary events with respect to $\mathcal{R}$ and to relate their probabilities to one another. By virtue of the equivalence between (8) and (9), we ended up with the intended generalization of interaction quotients.

Another interpretation of conditional structures makes (9) directly intelligible: if two multi-sets have the same conditional structure with respect to $\mathcal{R}$, then they represent equal *conditional weight*, and (9) describes a balanced system.

Notation 9. For a prior distribution P and some P-consistent set $\mathcal{R}$ of probabilistic conditionals, let $C(P,\mathcal{R})$ denote the set of all c-adaptations of P to $\mathcal{R}$:

$$C(P,\mathcal{R}) := \{P_c^* \mid P_c^* \text{ is a c-adaptation of } P \text{ to } \mathcal{R}\}.$$

Theorem 11 respectively Corollary 12 will show that ME-adaptations are c-adaptations, so $C(P,\mathcal{R}) \neq \emptyset$ for P-consistent $\mathcal{R}$ (cf. Corollary 13).

A c-adaptation is completely based both on P and $\mathcal{R}$, using P as a reference point and $\mathcal{R}$ as a guideline for changes. It realizes perfectly a conditional-logical approach to the adaptation problem $(*)$:

Postulate (P1): conditional preservation. The solution P^* of $(*)$ is a c-adaptation: $P^* \in C(P,\mathcal{R})$.

Let us consider the benefits of all these technical definitions when being applied to the example presented in Section 4.1.

Example 3 (*continued*). Assume that in a following year, we observe a slightly changed relationship between m_b and d, say $m_b \rightsquigarrow d[0.03]$, and we want P to be adjusted to this new information. So we have $\mathcal{R} = \{m_b \rightsquigarrow d[0.03]\}$, and let two symbols a_1, b_1 be associated with $\mathcal{R}$. The conditional structures with respect to $\mathcal{R}$ are calculated easily as follows:

$$(v_w m_w d)^\sigma = (v_w m_w \bar{d})^\sigma = (v_b m_w d)^\sigma = (v_b m_w \bar{d})^\sigma = 1,$$
$$(v_w m_b d)^\sigma = (v_b m_b d)^\sigma = a_1,$$
$$(v_w m_b \bar{d})^\sigma = (v_b m_b \bar{d})^\sigma = b_1.$$

Consider the (multi-)sets $\Omega_1 = \{v_w m_b d, v_b m_b \bar{d}\}$ and $\Omega_2 = \{v_b m_b d, v_w m_b \bar{d}\}$ with equal conditional structures $\Omega_1^\sigma = a_1 b_1 = \Omega_2^\sigma$. Therefore for P^* to be a c-adaptation of P to $\mathcal{R}$ it has to satisfy

$$\frac{p^*(v_w m_b d)p^*(v_b m_b \bar{d})}{p^*(v_b m_b d)p^*(v_w m_b \bar{d})} = \frac{p(v_w m_b d)p(v_b m_b \bar{d})}{p(v_b m_b d)p(v_w m_b \bar{d})},$$

which corresponds to (6).

Thus the concept of conditional structures helps to get a technically clear and precise formalization of the intuitive idea of conditional preservation.

The definition of a c-adaptation given above, however, is rather a technical and abstract one. It describes the conditional-logical behavior of such a distribution, but conveys hardly any idea of its actual appearance and form. Nevertheless, a simple but important property of c-adaptations may be seen at once from Definition 8 when regarding the case $|\Omega_1| = |\Omega_2| = 1$:

Lemma 10. *Let $P^* \in C(P, \mathcal{R})$. Then for any two elementary events $\omega_1, \omega_2 \in \Omega$, $p(\omega_1), p(\omega_2) \neq 0$, with equal conditional structures $\omega_1^{\sigma} = \omega_2^{\sigma}$, we have*

$$\frac{p^*(\omega_1)}{p(\omega_1)} = \frac{p^*(\omega_2)}{p(\omega_2)}.$$

The next theorem provides a catchy and easy characterization of c-adaptations:

Theorem 11 (Characterization of c-adaptations). *Suppose P is a distribution and $\mathcal{R} = \{A_1 \rightsquigarrow B_1[x_1], \ldots, A_n \rightsquigarrow B_n[x_n]\}$ is a P-consistent set of probabilistic conditionals. Let P^* denote a distribution.*

P^* is a c-adaptation of P to $\mathcal{R}$ if and only if there are real numbers $\alpha_0, \alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ with $\alpha_0 > 0$ and $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ satisfying the positivity condition

$$\alpha_i^+, \alpha_i^- \geqslant 0, \qquad \alpha_i^+ = 0 \textit{ iff } x_i = 0, \qquad \alpha_i^- = 0 \textit{ iff } x_i = 1 \tag{10}$$

and the adjustment condition

$$\begin{aligned}(1 - x_i)\alpha_i^+ &\sum_{\omega:\, A_iB_i(\omega)=1} p(\omega) \prod_{\substack{j \neq i \\ A_jB_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ A_j\overline{B_j}(\omega)=1}} \alpha_j^- \\ = x_i\alpha_i^- &\sum_{\omega:\, A_i\overline{B_i}(\omega)=1} p(\omega) \prod_{\substack{j \neq i \\ A_jB_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ A_j\overline{B_j}(\omega)=1}} \alpha_j^-,\end{aligned} \tag{11}$$

$1 \leqslant i \leqslant n$, such that

$$p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ A_iB_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n \\ A_i\overline{B_i}(\omega)=1}} \alpha_i^- \tag{12}$$

for all elementary events ω.

Thus probability values of c-adaptations parallel the conditional structures of the corresponding elementary events (cf. (7)). The proof of this theorem can be found in the Appendix.

Comparing (3) and (4) to (12), (11) and (10), we get as an immediate consequence:

Corollary 12. *Any ME-adaptation is a c-adaptation.*

Because ME-adaptations exist for priors P and P-consistent sets $\mathcal{R}$, we have

Corollary 13. *For any prior distribution P and any P-consistent set $\mathcal{R}$ of probabilistic conditionals, $C(P, \mathcal{R}) \neq \emptyset$.*

So c-adaptations generalize the concept of ME-adaptations and embed it into a conditional-logical environment. We will make use of c-adaptations in the form (12). Distributions of this type will play a major part in the rest of this paper.

Notation 14. Let P be a distribution, and let $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ be nonnegative real numbers such that

$$\sum_{\omega} p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} \alpha_i^- > 0.$$

Then $P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-]$ denotes the distribution with probability function

$$p[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-](\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} \alpha_i^-,$$

where

$$\alpha_0 = \Bigl(\sum_{\omega} p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} \alpha_i^-\Bigr)^{-1}.$$

The normalizing factor α_0 is completely determined by P and $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$. Note that $P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-]$ is P-continuous.

According to Theorem 11, for any c-adaptation P^* of P to $\mathcal{R}$, there are nonnegative real *weight factors* $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ satisfying (10) and (11) such that $P^* = P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-]$. Define

$$\begin{aligned}\mathrm{wf}(P^*) := \{(\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-) \in \mathbb{R}^{2n} \mid\ & \\ (\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-) & \text{ satisfies } (10) \text{ and } (11) \\ \text{and } P^* & = P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-]\}\end{aligned}$$

for any $P^* \in C(P, \mathcal{R})$. In general, weight factors of c-adaptations are not uniquely determined, so that $\mathrm{card}(\mathrm{wf}(P^*)) \geqslant 1$.

As the proof of Theorem 11 shows, (10) ensures that all premises A_i occurring in $\mathcal{R}$ have positive probabilities in $P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-]$, and (11) then is equivalent to $P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-] \models \mathcal{R}$.

Corollary 15. *Let P be a distribution, and suppose $\mathcal{R}$ is a P-consistent set of probabilistic rules. If $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ are reals satisfying* (10) *then $P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-] \models \mathcal{R}$ iff $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ fulfill* (11).

Therefore

$$\begin{aligned}\mathrm{WF}(P, \mathcal{R}) &:= \bigcup_{P^* \in C(P,\mathcal{R})} \mathrm{wf}(P^*) \\ &= \{(\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-) \in \mathbb{R}^{2n} \mid (\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-) \\ &\qquad \text{satisfies } (10) \text{ and } (11)\}. \qquad (13)\end{aligned}$$

So, c-adaptations actually realize quite a simple idea of adaptation to new conditional information: When calculating the posterior probability function p^*, one only has to check the conditional structure of each elementary event ω with respect to $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\} \subset \mathcal{L}^*$, set up $p^*(\omega)$ according to (12) with unknown quantities $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ and then determine appropriate unknowns using (10) and (11), i.e. so that $\mathcal{R}$ is satisfied. Finally, α_0 is computed as a normalizing factor to make P^* a probability distribution.

Example 16 briefly illustrates this adaptation scheme.

Example 16. Let P be a positive distribution over two variables A, B, and suppose $\mathcal{R} = \{a \rightsquigarrow b[x]\}$ with $x \in (0,1)$. Applying the formulas above, any c-adaptation P_c^* of P to $\mathcal{R}$ may be written as

$$p_c^*(ab) = \alpha_0 p(ab)\alpha_1^+, \qquad p_c^*(a\bar{b}) = \alpha_0 p(a\bar{b})\alpha_1^-,$$
$$p_c^*(\bar{a}b) = \alpha_0 p(\bar{a}b), \qquad p_c^*(\bar{a}\bar{b}) = \alpha_0 p(\bar{a}\bar{b})$$

with

$$\alpha_1^+, \alpha_1^- > 0, \qquad (1-x)\alpha_1^+ p(ab) = x\alpha_1^- p(a\bar{b}),$$
$$\alpha_0 = \left(\frac{1}{x} p(ab)\alpha_1^+ + p(\bar{a}b) + p(\bar{a}\bar{b})\right)^{-1}.$$

C-adaptations provide a straightforward scheme to calculate solutions to the adjustment problem $(*)$. ME-adaptations are a special instance of this scheme, and it is of interest to investigate which of the characteristics of ME-distributions also hold for c-adaptations in general.

The author proved in [18] that c-adaptations possess the properties of *system independence* and of *subset independence* which both played an outstanding part in Shore and Johnson's [33] characterization of the ME-principle. They also cope in an elegant manner with irrelevant information in that posterior marginals are determined only by conditionals involving the respective variables (cf. [18]; cf. [26]). All this is due to their modular, conditional-logical structure.

There is another principle that ME-adaptations actually seems to fail at first sight and that can now be formulated adequately and proved in terms of c-adaptations: it is the *Atomicity Principle* stating that substituting formulas for variables shall not affect the adjustment process (cf. [27]):

Theorem 17 (Atomicity principle). *Let $\mathcal{V} = \{V_1, V_2, \ldots\}$ and $\mathcal{V}' = \{V_1', V_2', \ldots\}$ be two finite disjoint sets of binary propositional variables with corresponding sets of elementary events Ω respectively Ω', and let V be another binary variable not contained in either of them. Suppose $\Delta \in \mathcal{L}(\mathcal{V}')$ is a propositional formula that is neither a tautology nor a contradiction, using only variables in $\mathcal{V}'$. Let $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ be a set of probabilistic conditionals with antecedents A_i and consequences B_i in $\mathcal{L}(\mathcal{V} \cup \{V\})$. Let A_i^Δ respectively B_i^Δ denote the formulas that arise when each occurrence of V in A_i respectively B_i is replaced by Δ, $1 \leqslant i \leqslant n$, and so $\mathcal{R}^\Delta = \{\mathrm{A}_1^\Delta \rightsquigarrow \mathrm{B}_1^\Delta[x_1], \ldots, \mathrm{A}_n^\Delta \rightsquigarrow \mathrm{B}_n^\Delta[x_n]\} \subset \mathcal{L}^*(\mathcal{V} \cup \mathcal{V}')$.*

Consider the two distributions P' over $\mathcal{V} \cup \mathcal{V}'$ and P over $\mathcal{V} \cup \{V\}$, respectively, that are related via $p(\dot{v}\omega) = \sum_{\omega' \in \Omega',\ \Delta(\omega')=1} p'(\omega'\omega)$, and suppose $\mathcal{R}$ to be P-consistent.

Then $\mathcal{R}^{\Delta}$ is P'-consistent, and $\mathrm{WF}(P, \mathcal{R}) = \mathrm{WF}(P', \mathcal{R}^{\Delta})$.

We omit the straightforward but technical proof. This result emphasizes the importance of the weight factors $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ as logical representatives of an adaptation scheme.

The concept of c-adaptations, however, is not perfect—it fails to satisfy uniqueness: Example 16 shows that, even in the simple case when dealing with two variables and one conditional to be adjusted to, the resulting c-adaptation is not uniquely determined. In general, $\mathrm{WF}(P, \mathcal{R})$ will contain lots of elements, and there will be many different posterior c-adaptations. Demanding uniqueness means to assume a *functional concept* that guides the finding of a "best solution" so that a unique distribution of type (12) arises in dependence of the prior knowledge P and the new conditional information $\mathcal{R}$.

5. The functional concept

It is not only the abstract property of uniqueness that makes a functional concept desirable. In a fundamental sense, there should be a clear and understandable dependence between prior distribution, new (conditional) information and resulting posterior distribution, i.e. a—somehow well-behaved—function $\mathrm{F} : (P, \mathcal{R}) \mapsto P^*$ that works for all distributions P and all P-consistent sets $\mathcal{R}$. These arguments P and $\mathcal{R}$, however, are quite monstrous. The knowledge represented by them is usually huge and hard to grasp, let alone introducing such concepts as continuity or even differentiability to describe a functional well-behavedness.

Moreover, P^* should depend *significantly* only on the *relevant* parts of the prior P, i.e. relevant with respect to the new information $\mathcal{R}$. Treating this problem requires making clear what relevant information is, and how irrelevant information should be handled.

Let $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$, and suppose P_1, P_2 are two distributions with $p_1(\omega|\mathrm{A}_i) = p_2(\omega|\mathrm{A}_i)$ for all $\omega \in \Omega$ and for all $i = 1, \ldots, n$. Then P_1, P_2 match on all parts which are *relevant* with respect to $\mathcal{R}$, so the difference in their posterior relative changes should be *insignificant*, namely a constant (due to possible differences in irrelevant parts):

Definition 18 (*Relevance condition*). Suppose $\mathrm{F} : (P, \mathcal{R}) \mapsto P^*$ is a function that assigns a P-continuous distribution P^* satisfying $\mathcal{R}$ to any pair $(P, \mathcal{R})$ with P being a distribution and $\mathcal{R}$ representing a P-consistent set of probabilistic rules.

Then F fulfills the *relevance condition* iff the following holds:

Let $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$, and suppose P_1, P_2 are two distributions with $p_1(\omega|\mathrm{A}_i) = p_2(\omega|\mathrm{A}_i)$ for all $\omega \in \Omega$ and for all $i = 1, \ldots, n$, $p_1(\omega) = 0$ iff $p_2(\omega) = 0$ and such that $\mathcal{R}$ is P_1- and P_2-consistent. Let $P_k^* := \mathrm{F}(P_k, \mathcal{R})$, $k = 1, 2$. Then $p_1^*(\omega) = 0$ iff $p_2^*(\omega) = 0$ and

$$\frac{p_1^*(\omega)}{p_1(\omega)} : \frac{p_2^*(\omega)}{p_2(\omega)} = const.$$

for all $\omega \in \Omega$ with $p_k^*(\omega) \neq 0$, $k = 1, 2$.

Let $\mathcal{AP}$ denote the set of all pairs $(P, \mathcal{R})$ representing a solvable adjustment problem $(*)$:

$$\mathcal{AP} := \{(P, \mathcal{R}) \mid P \text{ distribution}, \mathcal{R} \subset \mathcal{L}^*, \mathcal{R}\ P\text{-consistent}\}.$$

According to postulate (P1), the solution to the adaptation problem $(*)$ should be a c-adaptation. So we will now focus on further assumptions that appear useful and reasonable within the special context of c-adaptations.

At first, the following proposition shows that the prerequisites formulated in Definition 18 in fact are able to capture the idea of relevant information for c-adaptations:

Proposition 19. *Let $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$, and suppose P_1, P_2 are two distributions with $p_1(\omega|\mathrm{A}_i) = p_2(\omega|\mathrm{A}_i)$ for all $\omega \in \Omega$ and for all $i = 1, \ldots, n$, $p_1(\omega) = 0$ iff $p_2(\omega) = 0$ and such that $\mathcal{R}$ is P_1- and P_2-consistent. Then* $\mathrm{WF}(P_1, \mathcal{R}) = \mathrm{WF}(P_2, \mathcal{R})$.

Therefore distributions incorporating the same relevant conditional knowledge have the same sets of weight factors occurring in the corresponding c-adaptations.

Let us henceforth assume that there is a function

$$\mathrm{F_c} : \mathcal{AP} \ni (P, \mathcal{R}) \mapsto P_\mathrm{c}^* \in C(P, \mathcal{R}) \tag{14}$$

that assigns to each pair $(P, \mathcal{R}) \in \mathcal{AP}$ a particular c-adaptation. We will describe $\mathrm{F_c}$ by specific properties of the weight factors involved.

Proposition 20. *Assume $(P, \mathcal{R}) \in \mathcal{AP}$, $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ and let $P_\mathrm{c}^* = P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-] \in C(P, \mathcal{R})$ be a c-adaptation of P to $\mathcal{R}$ with weight factors $(\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-) \in \mathrm{wf}(P_\mathrm{c}^*)$. Suppose $I \cup J$ is a partition of $\{1, \ldots, n\}$, and set $P_I := P[\alpha_i^+, \alpha_i^-]_{i \in I}$, $\mathcal{R}_J := \{\mathrm{A}_j \rightsquigarrow \mathrm{B}_j[x_j]\}_{j \in J}$.*

Then $(\alpha_j^+, \alpha_j^-)_{j \in J} \in \mathrm{WF}(P_I, \mathcal{R}_J)$ and $P_I[\alpha_j^+, \alpha_j^-]_{j \in J} = P[\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-]$.

So, once weight factors $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ are chosen to yield a "best" solution $P_\mathrm{c}^* \in C(P, \mathcal{R})$, they should yield "best" solutions in $C(P_I, \mathcal{R}_J)$ (with all notations as stated in the text of Proposition 20). We name this property *continuity* (*of solutions*):

Definition 21 (*Continuity condition*). Let $\mathrm{F_c}$ be as described in (14). $\mathrm{F_c}$ satisfies the *continuity condition* if the following holds:

Suppose $(P, \mathcal{R}) \in \mathcal{AP}$ with $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$. Assume $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^- \in \mathrm{wf}(\mathrm{F_c}(P, \mathcal{R}))$. Let $I \cup J$ be a partition of $\{1, \ldots, n\}$, and set $P_I := P[\alpha_i^+, \alpha_i^-]_{i \in I}$, $\mathcal{R}_J := \{\mathrm{A}_j \rightsquigarrow \mathrm{B}_j[x_j]\}_{j \in J}$. Then $(\alpha_j^+, \alpha_j^-)_{j \in J} \in \mathrm{wf}(\mathrm{F_c}(P_I, \mathcal{R}_J))$, i.e. $\mathrm{F_c}(P_I, \mathcal{R}_J) = P_I[\alpha_j^+, \alpha_j^-]_{j \in J} = \mathrm{F_c}(P, \mathcal{R})$.

Finally, $\mathrm{F_c}$ should obey the principle of atomicity (cf. Theorem 17):

Definition 22 (*Atomicity condition*). Let $\mathrm{F_c}$ be as described in (14). $\mathrm{F_c}$ satisfies the *atomicity condition* if for any $(P, \mathcal{R}), (P', \mathcal{R}^{\Delta}) \in \mathcal{AP}$ as in Theorem 17, $\mathrm{wf}(\mathrm{F_c}(P, \mathcal{R})) = \mathrm{wf}(\mathrm{F_c}(P', \mathcal{R}^{\Delta}))$.

The following proposition derives necessary conditions for a function $\mathrm{F_c}$ to fulfill the conditions of relevance, continuity and atomicity in special but important cases:

Proposition 23. *Assume* $\mathrm{F_c}$ *as described in* (14).

(i) *Suppose* P_1, P_2 *are positive distributions over two variables* A, B *with* $p_1(b|a) = p_2(b|a)$, *and let* $\mathcal{R} = \{a \leadsto b[x]\}$, $x \in (0,1)$. *If* $\mathrm{F_c}$ *satisfies the relevance condition, then the weight factors* α^+, α^- *respectively* β^+, β^- *of* $\mathrm{F_c}(P_1, \mathcal{R})$ *respectively* $\mathrm{F_c}(P_2, \mathcal{R})$ *are equal in pairs, i.e.* $\alpha^+ = \beta^+$ *and* $\alpha^- = \beta^-$.

(ii) *Suppose* $\mathrm{F_c}$ *satisfies the conditions of relevance, continuity and atomicity, and let* $(P, \mathcal{R}) \in \mathcal{AP}$ *with positive prior* P, *and such that no variable occurs both in antecedent and conclusion of a conditional in* $\mathcal{R}$ *and all assigned probabilities in* $\mathcal{R}$ *are different from* 0 *and* 1. *Then the weight factors* α^+, α^- *associated in* $\mathrm{F_c}(P, \mathcal{R})$ *with a conditional in* $\mathcal{R}$ *only depend upon the probability* x *of this conditional and upon their quotient* α^+/α^-, *i.e. for any* $(P, \mathcal{R}), (P', \mathcal{R}') \in \mathcal{AP}$, P, P' *positive*, $\mathcal{R} = \{\mathrm{A} \leadsto \mathrm{B}[x], \mathrm{A}_1 \leadsto \mathrm{B}_1[x_1], \ldots\}$, $\mathcal{R}' = \{\mathrm{A}' \leadsto \mathrm{B}'[x], \mathrm{A}'_1 \leadsto \mathrm{B}'_1[x'_1], \ldots\}$, *both sets finite, all of* $x, x_i, x'_i \in (0,1)$, *no variable occurring both in antecedent and conclusion of any conditional in* $\mathcal{R}$ *and* $\mathcal{R}'$, *and for any weight factors* α^+, α^- *respectively* α'^+, α'^- *associated in* $\mathrm{F_c}(P, \mathcal{R})$ *respectively* $\mathrm{F_c}(P', \mathcal{R}')$ *with the conditional* $\mathrm{A} \leadsto \mathrm{B}[x]$ *respectively* $\mathrm{A}' \leadsto \mathrm{B}'[x]$, $\alpha^+/\alpha^- = \alpha'^+/\alpha'^-$ *implies* $\alpha^+ = \alpha'^+$ *and* $\alpha^- = \alpha'^-$.

Example 16 shows that, in the cases dealt with by Proposition 23(i), all pairs α^+, α^- of weight factors have to fulfill

$$\frac{\alpha^+}{\alpha^-} = \frac{x}{1-x} \frac{p(a\bar{b})}{p(ab)}.$$

The cross ratio on the right-hand side, depending only on prior and new conditional probabilities, represents exactly *relevant knowledge*. The left-hand side is just the quotient of α^+ and α^-. This gives an intuitive reason for this quotient playing a key role, as it is stated in Proposition 23(ii).

Thus in the context of c-adaptations, we identified clearly the parameters weight factors should be dependent on to give rise to a reasonable functional concept: α^+/α^- and (the probability) x incorporate all relevant knowledge for the weight factors. Thus a reasonable functional concept for c-adaptations may be realized by setting

$$\alpha^+ = F^+(x, \alpha), \qquad \alpha^- = F^-(x, \alpha), \tag{15}$$

with two real positive functions F^+ and F^-, defined on $(0,1) \times \mathbb{R}^+$ and related by $F^+(x, \alpha)/F^-(x, \alpha) = \alpha$, i.e.

$$F^+(x, \alpha) = \alpha F^-(x, \alpha). \tag{16}$$

As our global function $\mathrm{F} : \mathcal{AP} \ni (P, \mathcal{R}) \mapsto P^*$ is to work for arbitrary P and $\mathcal{R}$, the functions F^+ and F^- are assumed to be independent of the prior and new information actually present, thus representing a fundamental inference pattern. Moreover, to yield "smooth" inferences we assume them to be continuous on $(0,1) \times \mathbb{R}^+$. The functional concept, designed so far, should also be applied to the extreme probabilities $x \in \{0, 1\}$, incorporating classical logic as a limit case by assuming

$$F^+(0,0) := \lim_{\substack{x \to 0 \\ \alpha \to 0}} F^+(x, \alpha) = 0, \qquad F^+(1, \infty) := \lim_{\substack{x \to 1 \\ \alpha \to \infty}} F^+(x, \alpha) \in \mathbb{R}^+, \tag{17}$$

and

$$F^-(0,0) := \lim_{\substack{x \to 0 \\ \alpha \to 0}} F^-(x, \alpha) \in \mathbb{R}^+, \qquad F^-(1, \infty) := \lim_{\substack{x \to 1 \\ \alpha \to \infty}} F^-(x, \alpha) = 0, \tag{18}$$

in accordance with (10). The resulting posterior distribution P_F^* has the form

$$p_F^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} F^+(x_i, \alpha_i) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} F^-(x_i, \alpha_i) \tag{19}$$

with nonnegative extended real numbers $\alpha_1, \ldots, \alpha_n \in \mathbb{R}^+ \cup \{0, \infty\}$ solving the n equations

$$\alpha_i = \begin{cases} \dfrac{x_i}{1 - x_i} \dfrac{\sum_{\omega: \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1} p(\omega) \prod_{j \neq i, \mathrm{A}_j\mathrm{B}_j(\omega)=1} F^+(x_j, \alpha_j) \prod_{j \neq i, \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1} F^-(x_j, \alpha_j)}{\sum_{\omega: \mathrm{A}_i\mathrm{B}_i(\omega)=1} p(\omega) \prod_{j \neq i, \mathrm{A}_j\mathrm{B}_j(\omega)=1} F^+(x_j, \alpha_j) \prod_{j \neq i, \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1} F^-(x_j, \alpha_j)}, \\ \qquad x_i \neq 0, 1, \\ 0, \quad x_i = 0, \\ \infty, \quad x_i = 1. \end{cases} \tag{20}$$

Note that $\alpha_i \in \mathbb{R}^+$ for $x_i \in (0,1)$, because of the positivity of both functions F^+ and F^- and due to the P-consistency of $\mathcal{R}$. So the positivity condition (10) is satisfied, and (20) corresponds to the adjustment condition (11) here. Thus, for any n nonnegative extended real numbers $\alpha_1, \ldots, \alpha_n \in \mathbb{R}^+ \cup \{0, \infty\}$, $\alpha_1, \ldots, \alpha_n$ is a solution of (20) iff $F^+(x_1, \alpha_1), F^-(x_1, \alpha_1), \ldots, F^+(x_n, \alpha_n), F^-(x_n, \alpha_n)$ is a solution of (11) satisfying (10).

We summarize these remarks for the axiomatization of the second postulate (P2):

Postulate (P2): functional concept for c-adaptations. There is a function $\mathrm{F}^* : \mathcal{AP} \ni (P, \mathcal{R}) \mapsto P_c^* \in C(P, \mathcal{R})$ that assigns to each adjustment problem $(P, \mathcal{R}) \in \mathcal{AP}$ a particular c-adaptation P_F^* fulfilling $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$, and there are two real positive and continuous functions F^+ and F^- defined on $(0,1) \times \mathbb{R}^+$, fulfilling the conditions (17) and (18) and related by (16), such that $P_F^* = \mathrm{F}^*(P, \mathcal{R}) =: P *_F \mathcal{R}$ has the form (19) with $\alpha_1, \ldots, \alpha_n \in \mathbb{R}^+ \cup \{0, \infty\}$ solving (20).

Define for (fixed) F^*, F^+, F^- as in (P2) and for $(P, \mathcal{R}) \in \mathcal{AP}$, $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$,

$$\mathrm{WQ}_F(P, \mathcal{R}) := \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^+ \cup \{0, \infty\} \mid (\alpha_1, \ldots, \alpha_n) \text{ solves } (20)\} \tag{21}$$

to be set of all *weight quotients* that belong to c-adaptations

$$p[\alpha_1, \ldots, \alpha_n]_F(\omega) := \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} F^+(x_i, \alpha_i) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} F^-(x_i, \alpha_i).$$

So $(\alpha_1, \ldots, \alpha_n) \in \mathrm{WQ}_F(P, \mathcal{R})$ iff

$$(F^+(x_1, \alpha_1), F^-(x_1, \alpha_1), \ldots, F^+(x_n, \alpha_n), F^-(x_n, \alpha_n)) \in \mathrm{WF}(P, \mathcal{R}).$$

For any $(P, \mathcal{R}) \in \mathcal{AP}$, $\mathrm{F}^*(P, \mathcal{R}) = P[\alpha_1, \ldots, \alpha_n]_F$ is described by a particular element $(\alpha_1, \ldots, \alpha_n) \in \mathrm{WQ}_F(P, \mathcal{R})$. This disagreeable dependence on a special yet unknown solution of (20) may be overcome by assuming that the functions F^+ and F^- fulfill the *condition of uniqueness*:

Definition 24 (*Uniqueness condition*). Let F^+ and F^- be functions as described in (P2). F^+ and F^- satisfy the *uniqueness condition* iff whenever $(P, \mathcal{R}) \in \mathcal{AP}$ and $(\alpha_1, \ldots, \alpha_n)$, $(\beta_1, \ldots, \beta_n) \in \mathrm{WQ}_F(P, \mathcal{R})$ it holds that

$$P[\alpha_1, \ldots, \alpha_n]_F = P[\beta_1, \ldots, \beta_n]_F.$$

So, if F^+ and F^- satisfy the uniqueness condition then F^* is determined by (20), i.e. $\mathrm{F}^*(P, \mathcal{R}) = P[\alpha_1, \ldots, \alpha_n]_F$ for *any* $(\alpha_1, \ldots, \alpha_n) \in \mathrm{WQ}_F(P, \mathcal{R})$.

The functional concept (P2), describing weight factors of c-adaptations, was initiated by the conditions of relevance, continuity and atomicity. The uniqueness condition ensures recovering these properties from (P2):

Proposition 25. *Let* $\mathrm{F}^* : \mathcal{AP} \ni (P, \mathcal{R}) \mapsto P_F^* \in C(P, \mathcal{R})$ *be as in* (P2) *with associated functions* F^+ *and* F^- *satisfying the uniqueness condition. Then* F^* *fulfills the conditions of relevance, continuity and atomicity.*

We will see in Section 8 that the condition of uniqueness will in fact be satisfied for the special functions F^+ and F^- that will be determined by (P2) together with (P3) and (P4) (cf. Proposition 30).

Note that the conciseness of (P2) is essentially due to making use of c-adaptations. So the efforts we invested in developing this conditional-logical concept begin to pay, providing now an elegant functional concept.

After having put the functional dependencies in concrete terms we are now going to study which properties the functions F^+ and F^- should have to guarantee sound probabilistic inferences. To simplify notation, we will usually prefer the operational $P *_F \mathcal{R}$ to the functional $\mathrm{F}^*(P, \mathcal{R})$, where $*_F$ is described by (P2).

6. Logical consistency

Surely, the adaptation scheme (19) will be considered sound only if the resulting posterior distribution can be used as a prior distribution for further adaptations. This is a very fundamental meaning of *logical consistency*.

In particular, if we first adjust P only to a subset $\mathcal{R}_1 \subseteq \mathcal{R}$, and then use this posterior distribution to perform another adaptation to the full conditional information $\mathcal{R}$, we should obtain the same distribution as if we adjusted P to $\mathcal{R}$ in only one step.

We state this demand for logical consistency as Postulate (P3):

Postulate (P3): logical consistency. For any distribution P and any P-consistent sets $\mathcal{R}_1, \mathcal{R}_2 \subset \mathcal{L}^*$, the (final) posterior distribution which arises from a two-step process of adjusting P first to $\mathcal{R}_1$ and then adjusting this intermediate posterior to $\mathcal{R}_1 \cup \mathcal{R}_2$ is identical to the distribution resulting from directly adapting P to $\mathcal{R}_1 \cup \mathcal{R}_2$.

More formally, the operator $*_F$ satisfies (P3) iff the following equation holds:

$$P *_F (\mathcal{R}_1 \cup \mathcal{R}_2) = (P *_F \mathcal{R}_1) *_F (\mathcal{R}_1 \cup \mathcal{R}_2). \tag{22}$$

Theorem 26. *If the adjustment operator $*_F$ satisfies the postulate* (P3) *of logical consistency then*

$$F^-(0,0) = F^+(1,\infty) = 1, \tag{23}$$

and F^- necessarily fulfills the functional equation

$$F^-(x,\alpha\beta) = F^-(x,\alpha)F^-(x,\beta) \tag{24}$$

for all $x \in (0,1)$, $\alpha,\beta \in \mathbb{R}^+$.

Because of (16), F^+ satisfies (24) iff F^- does.

Theorem 26 is proved by checking condition (23) and (24) in the very special case that P is a positive distribution over three variables A, B, C, and $\mathcal{R}_1, \mathcal{R}_2$ are given by $\mathcal{R}_1 = \{a \rightsquigarrow c[x]\}$ and $\mathcal{R}_2 = \{b \rightsquigarrow c[y]\}$ (see the Appendix). These conditions are necessary to guarantee a logical consistent behavior of the adaptation process for this example, and because we assumed the functions F^+, F^- to be independent of the actual case we thus proved the general validity of (23) and (24). In fact, there is little arbitrariness in choosing this special example which such a crucial meaning is being assigned to. The way in which two conditionals with common conclusion should interact is one of the main issues in conditional logic and refers to the *antecedent conjunction problem* (cf. [25,35]). The validity of (23) and (24) ensures a sound probabilistic treatment of this problem.

The functional equation (24) restricts the type of the function F^- (and that of F^+, too,) essentially:

Proposition 27. *Let $*_F$ be an adjustment operator following* (P2) *such that F^+ and F^- satisfy* (23) *and* (24). *Then there is a (continuous) real function $c(x)$, $x \in (0,1)$,* *with*

$$\lim_{x\to 0} c(x) = 0, \qquad \lim_{x\to 1} c(x) = -1 \tag{25}$$

such that

$$F^-(x,\alpha) = \alpha^{c(x)}, \qquad F^+(x,\alpha) = \alpha^{c(x)+1} \tag{26}$$

for any positive real α and for any $x \in (0,1)$. Especially for $\alpha = 1$, this implies

$$F^-(x,1) = F^+(x,1) = 1. \tag{27}$$

In Theorem 26 we showed how the consistency property (22) determines the part the quotients α_i have to play in the adjustment process. We are now left with the investigation of the isolated impact of the numbers x_i which represent posterior conditional probabilities.

7. Representation invariance

By and large, we neglected how (conditional) knowledge is represented in $\mathcal{R}$. Indeed, the principle of atomicity deals with *logical equivalence* of propositional formulas, but what about *probabilistic equivalences*, i.e. equivalences that are due to elementary probability calculus? For instance, the sets of rules $\{\mathrm{A} \rightsquigarrow \mathrm{B}[x], \mathrm{A}[y]\}$ and $\{\mathrm{AB}[xy], \mathrm{A}[y]\}$ are equivalent in this respect because each rule in one set is derivable from rules in the other. We surely expect the result of our adjustment process to be independent of the syntactic representation of probabilistic knowledge in $\mathcal{R}$:

Postulate (P4): representation invariance. If two P-consistent sets of probabilistic conditionals $\mathcal{R}$ and $\mathcal{R}'$ are probabilistically equivalent then the posterior distributions $P * \mathcal{R}$ and $P * \mathcal{R}'$ resulting from adapting the prior P to $\mathcal{R}$ respectively to $\mathcal{R}'$ are identical.

The notion of probabilistic equivalence used here completely corresponds to that introduced in [26]. Using the operational notation, we are able to express (P4) more formally:

The adjustment operator $*_F$ satisfies (P4) iff

$$P *_F \mathcal{R} = P *_F \mathcal{R}' \tag{28}$$

for any two P-consistent and probabilistically equivalent sets $\mathcal{R}, \mathcal{R}' \subset \mathcal{L}^*$.

The demand for independence of syntactic representation of probabilistic knowledge (P4) gives rise to two functional equations for $c(x)$ (cf. Proposition 27):

Proposition 28. *Let the functions F^- and F^+ describing the adaptation operator $*_F$ in* (P2) *be given by* (26) *with a continuous function $c(x)$ fulfilling* (25). *If $*_F$ satisfies postulate* (P4) *respectively* (28) *then for all real $x, x_1, x_2 \in (0,1)$ the following equations hold:*

$$c(x)+c(1-x)=-1, \tag{29}$$

$$c(xx_1+(1-x)x_2)=-c(x)c(x_1)-c(1-x)c(x_2). \tag{30}$$

The most obvious probabilistic equivalence is that of each two rules $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$ and $\mathrm{A} \rightsquigarrow \overline{\mathrm{B}}[1-x]$. This implies (29). Eq. (30) is again proved by investigating a special but crucial adaptation problem. The relation $p(b|a)=p(b|ac)p(c|a)+p(b|a\bar{c})p(\bar{c}|a)$ for arbitrarily chosen propositional variables A, B, C is fundamental to probabilistic conditionals, yielding the probabilistic equivalence of the two sets $\mathcal{R}=\{a \rightsquigarrow c[x], ac \rightsquigarrow b[x_1], a\bar{c} \rightsquigarrow b[x_2]\}$ and $\mathcal{R}'=\{a \rightsquigarrow b[y], ac \rightsquigarrow b[x_1], a\bar{c} \rightsquigarrow b[x_2]\}$ with $y = xx_1+(1-x)x_2$ for real $x, x_1, x_2 \in (0,1)$. The validity of (28) in this case necessarily implies (30) (see Appendix).

As a consequence of (29) and (30), we finally obtain:

Theorem 29. *If the operator* $*_F$ *in* (P2) *is to meet the demands for logical consistency* (P3) *and for representation invariance* (P4), *then* F^+ *and* F^- *necessarily have the forms*

$$F^+(x,\alpha)=\alpha^{1-x} \quad \text{and} \quad F^-(x,\alpha)=\alpha^{-x}, \tag{31}$$

respectively.

The continuity of the functions F^+ and F^-, in particular the continuous integrating of the extreme probabilities 0 and 1, i.e. the seamless encompassing of classical logic, is essential to establish this theorem (see the Appendix).

8. Uniqueness and the main theorem

So far we have proved that the demands for logical consistency and for representation invariance determine the functions which we assumed to underly the adjusting of P to $\mathcal{R}$, as described by the functional concept (P2). Applying (31) to (19) and (20) we recognize that the posterior distribution necessarily is of the same type (3) as the ME-distribution if it is to yield sound and consistent inferences.

Therefore we have nearly reached our goal. But one step is still missing: Is this enough to *characterize* ME-inference within a conditional-logical framework? Are there possibly several different solutions of type (3), only one of which is the ME-distribution? And moreover, if we assume the functions F^+ and F^- to fulfill (31), is this *sufficient* to guarantee that the resulting operator $*_F$ satisfies logical consistency and representation invariance?

The question of uniqueness of the posterior distribution is at the center of all these problems. If it can be answered positively, we will be finished: The unique posterior distribution of type (3) must be the ME-distribution, $*_F$ then corresponds to ME-inference, and ME-inference is known to fulfill (22) and (28) as well as many other reasonable properties, cf. [26,33,34]. Moreover, together with (P2) uniqueness implies the conditions of relevance, atomicity and continuity (cf. Proposition 25).

The uniqueness of the solution $(\alpha_1, \dots, \alpha_n)$ of the fixpoint equation (4) is questionable. Imagine the case that the set $\mathcal{R}$ representing new conditional knowledge contains twice the same rule in different notations. All that can be expected at best is a uniqueness statement for the product $\alpha_i \alpha_j$ of the corresponding factors. Even if we exclude such pathological cases, (4) is not easy to deal with at all.

But remember that we are primarily interested in the uniqueness of the posterior distribution, not in that of the solutions to (4). And indeed, this uniqueness is affirmed by the next theorem. In its proof, we will make use of cross-entropy as an excellently fitting measure of distance for distributions of type (3) (see the Appendix).

Proposition 30. *There is at most one solution of the adaptation problem* $(*)$ *of type* (3), *i.e. the functional concept defined by* (31) *satisfies the uniqueness condition.*

The following theorem summarizes our results in characterizing ME-adjustment within a conditional-logical framework:

Theorem 31 (Main Theorem). *Let* $*_{\mathrm{e}}$ *denote the ME-adjustment operator, i.e.* $*_{\mathrm{e}}$ *assigns to a prior distribution* P *and some* P*-consistent set* $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \dots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ *of probabilistic conditionals the one distribution* $P_{\mathrm{e}} = P *_{\mathrm{e}} \mathcal{R}$ *which has minimal cross-entropy with respect to* P *among all distributions that satisfy* $\mathcal{R}$.

Then $*_{\mathrm{e}}$ *yields the only adaptation of* P *to* $\mathcal{R}$ *that obeys the principle of conditional preservation* (P1), *realizes a functional concept* (P2) *and satisfies the postulates for logical consistency* (P3) *and for representation invariance* (P4). $*_{\mathrm{e}}$ *is completely described by* (3) *and* (4).

In the next section, we are going to prove some properties of $*_{\mathrm{e}}$ which are already known about ME-inference but which can be proved easily within the framework presented here.

9. Some properties of the adjustment operator $*_{\mathrm{e}}$

The uniqueness of the solution of type (3) yields an easy but important corollary which meets fundamental demands for "good solutions":

Corollary 32. $P *_{\mathrm{e}} \mathcal{R} = P$ *if and only if* $P \models \mathcal{R}$.

This corollary makes obvious an idea of minimality of change that is implicitly inherent to all of the postulates (P1)–(P4). We will now make this minimality more explicit by proving two properties known from the area of *nonmonotonic reasoning*. Within that framework, the notions of *idempotence* and *cumulativity* are fundamental to characterize reasonable nonmonotonic inference operations (cf. [20,24]):

Proposition 33. *The adjustment operator* $*_{\mathrm{e}}$ *has the following properties*:

(i) *Idempotence*: $(P *_{\mathrm{e}} \mathcal{R}) *_{\mathrm{e}} \mathcal{R} = P *_{\mathrm{e}} \mathcal{R}$.

(ii) *Cumulativity*: *If* $\mathcal{R}_1 \subseteq \mathcal{R}_2$ *and* $P *_{\mathrm{e}} \mathcal{R}_1 \models \mathcal{R}_2$ *then* $P *_{\mathrm{e}} \mathcal{R}_1 = P *_{\mathrm{e}} \mathcal{R}_2$.

As can be seen at once, cumulativity (ii) is equivalent to

(ii)′ If $P *_{\mathrm{e}} \mathcal{R}_1 \models \mathcal{R}_2$ then $P *_{\mathrm{e}} (\mathcal{R}_1 \cup \mathcal{R}_2) = P *_{\mathrm{e}} \mathcal{R}_1$,

which is stated as Principle 5 in [26].

10. Conclusion and future work

Starting from a conditional-logical point of view we found another characterization of the ME-solution to the problem

> *Given a distribution P and a P-consistent set of probabilistic conditionals $\mathcal{R}$, which way is the best for adjusting P to $\mathcal{R}$?*

The thorough embedding of the problem within the conditional-logical framework presented here conveys a clear understanding what actually makes the ME-distribution to be the best choice—ME-inference and probabilistic conditionals fit perfectly well.

The principle of conditional preservation was used to determine the structure of the posterior distribution. Then we assumed that a (continuous) functional concept extending classical logic should underly the adjustment process, and we isolated the crucial parameters which this concept should depend on. It was represented by means of two functions F^+ and F^- accomplishing the discrimination between these elementary events satisfying the antecedent of a rule that also satisfy its conclusion and those events that do not. So they constitute the decisive components for the extent of distortion the prior distribution is to be exposed to under adjustment.

Only two further preconditions were necessary to arrive at the desired characterization: logical consistency and independence of syntactical representation. Both are usually considered to be fundamental to any reasonable inference procedure. So no exceptional demands had to be made, and the ME-solution arose in a rather natural way. Moreover, the proof of Proposition 30, which states the uniqueness of the solution, illustrates how perfectly well the approach presented here realizes ME-inference in an understandable manner, without imposing any external and abstract minimality demand. Actually, the proper idea of minimality is being made explicit by the four postulates.

In Section 9, we showed that the adjustment operator $*_{\mathrm{e}}$ which we proved to be the best within that framework chosen satisfies two essential axioms of nonmonotonic reasoning. So the nonmonotonic inference operation $C_P(\mathcal{R}) = Th(P *_{\mathrm{e}} \mathcal{R})$ seems to be a good candidate to provide sound *quantitative* inferences.

Introducing a functional concept parallels obviously (and intentionally) the ideas of Gärdenfors [9] concerning theory revision. The problem we were dealing with in this paper may indeed be considered as a theory revision problem, with $Th(P)$ playing the role of the theory to be revised by the new information $\mathcal{R}$.

These connections to nonmonotonic reasoning and theory change are topics of our ongoing research.

Acknowledgements

I thank Christoph Beierle and Wilhelm Rödder for many hours of discussions, and two unknown referees for their helpful comments concerning the presentation and the correctness of this paper. Thanks also to Jeff Paris for sending me preliminary versions of his papers.

The main part of this work was finished while I was supported by a Lise-Meitner-scholarship, Department of Science and Research, North-Rhine-Westfalia, Germany.

Appendix A. Proofs

Proof of Theorem 11. Let P be a distribution and $\mathcal{R} = \{A_1 \leadsto B_1[x_1], \ldots, A_n \leadsto B_n[x_n]\}$ be P-consistent. Let Ω^* denote all elementary events with positive prior probability: $\Omega^* = \{\omega \in \Omega \mid p(\omega) > 0\}$.

Suppose first that P^* is a c-adaptation of P to $\mathcal{R}$. The equivalence relation $\equiv_{\mathcal{R}}$ induces a partition $K_1, \ldots, K_q$ of Ω^* in disjoint classes. According to Lemma 10, $p^*(\omega)/p(\omega)$ is constant on each equivalence class, so assume $p^*(\omega)/p(\omega) = \kappa_j$ for $\omega \in K_j$. Let $\omega_1, \ldots, \omega_q \in \Omega^*$ be a representative system of $K_1, \ldots, K_q$.

For the sake of simplicity of notation, we suppose that $\kappa_1, \ldots, \kappa_{q'} > 0$, $\kappa_{q'+1} = \cdots = \kappa_q = 0$ with $q' \leqslant q$, and furthermore, $x_1, \ldots, x_{n'} \in (0,1)$, $x_{n'+1}, \ldots, x_n \in \{0,1\}$, $n' \leqslant n$.

If $\kappa_j = 0$, $q' < j \leqslant q$, we have $p^*(\omega_j) = 0$ and $p(\omega_j) > 0$. From Definition 8(i) there must be a conditional $A_i \leadsto B_i$ in $\mathcal{R}$, $n' < i \leqslant n$, such that either $x_i = 1$ and $A_i\overline{B}_i(\omega_j) = 1$, or $x_i = 0$ and $A_iB_i(\omega_j) = 1$. In the first case, set $\alpha_i^+ = 1$ and $\alpha_i^- = 0$, in the second case exchange the values of these two factors. In this way, any $\kappa_j = 0$ (and thus any *certain conditional* in $\mathcal{R}$, i.e. any conditional with probability 0 or 1) is dealt with.

Let us now consider the constants $\kappa_j \neq 0$. Finding positive factors $\alpha_0, \alpha_1^+, \alpha_1^-, \ldots, \alpha_{n'}^+, \alpha_{n'}^-$ with

$$0 \neq p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n' \\ A_iB_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n' \\ A_i\overline{B}_i(\omega)=1}} \alpha_i^-$$

amounts to solving the following system of q' equations,

$$\alpha_0 \prod_{\substack{1 \leqslant i \leqslant n' \\ A_iB_i(\omega_j)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n' \\ A_i\overline{B}_i(\omega_j)=1}} \alpha_i^- = \kappa_j, \quad j = 1, \ldots, q', \tag{A.1}$$

which can be transformed into a linear equational system

$$\boldsymbol{\Theta\beta} = \boldsymbol{\lambda} \tag{A.2}$$

with $\boldsymbol{\beta} = (\log \alpha_1^+, \log \alpha_1^-, \ldots, \log \alpha_{n'}^+, \log \alpha_{n'}^-, \log \alpha_0)^{\mathrm{T}} \in \mathbb{R}^{2n'+1}$, $\boldsymbol{\lambda} = (\log \kappa_1, \ldots, \log \kappa_{q'})^{\mathrm{T}} \in \mathbb{R}^{q'}$ (where $\mathbb{R}$ denotes the field of real numbers) and a $q' \times (2n'+1)$-matrix Θ with elements in $\{0,1\}$, $\theta_{j,2i} = 1$ iff $\sigma_i(\omega_j) = a_i$, $\theta_{j,2i+1} = 1$ iff $\sigma_i(\omega_j) = b_i$,

$\theta_{j,2n'+1} = 1$ for all $1 \leqslant j \leqslant q'$, $1 \leqslant i \leqslant n'$. Let $\boldsymbol{\theta}_j$, $1 \leqslant j \leqslant q'$, denote the row vectors of Θ. The equational system (A.2) is solvable over $\mathbb{R}$ iff any linear dependencies (over the field of rationals, because each entry of Θ is either 0 or 1) between these row vectors correspond to relations between the $\lambda_j = \log \kappa_j$, i.e. $\sum_k r_{m_k} \boldsymbol{\theta}_{m_k} = \sum_l s_{n_l} \boldsymbol{\theta}_{n_l}$ must imply $\sum_k r_{m_k} \lambda_{m_k} = \sum_l s_{n_l} \lambda_{n_l}$ with rationals r_{m_k}, s_{n_l}.

Arranging and multiplying both sums appropriately, we may assume $\sum_k r_{m_k} \boldsymbol{\theta}_{m_k} = \sum_l s_{n_l} \boldsymbol{\theta}_{n_l}$ with natural numbers r_{m_k}, s_{n_l}. By comparing the vector components, we obtain $\sum_k r_{m_k} \theta_{m_k,2i} = \sum_l s_{n_l} \theta_{n_l,2i}$, $\sum_k r_{m_k} \theta_{m_k,2i+1} = \sum_l s_{n_l} \theta_{n_l,2i+1}$, $1 \leqslant i \leqslant n'$, and $\sum_k r_{m_k} = \sum_l s_{n_l}$, the last equation being valid because of $\theta_{j,2n'+1} = 1$ for all $1 \leqslant j \leqslant q'$. The first two equations imply $\sum_{k:\, \sigma_i(\omega_{m_k})=a_i} r_{m_k} = \sum_{l:\, \sigma_i(\omega_{n_l})=a_i} s_{n_l}$ and $\sum_{k:\, \sigma_i(\omega_{m_k})=b_i} r_{m_k} = \sum_{l:\, \sigma_i(\omega_{n_l})=b_i} s_{n_l}$. Therefore the multi-sets $\{\omega_{m_k} : r_{m_k}\}_k$ and $\{\omega_{n_l} : s_{n_l}\}_l$ are $\mathcal{R}$-equivalent by Lemma 7, and because P^* is assumed to be a c-adaptation we obtain

$$\prod_k \kappa_{m_k}^{r_{m_k}} = \prod_k \frac{p^*(\omega_{m_k})^{r_{m_k}}}{p(\omega_{m_k})^{r_{m_k}}} = \prod_l \frac{p^*(\omega_{n_l})^{s_{n_l}}}{p(\omega_{n_l})^{s_{n_l}}} = \prod_l \kappa_{n_l}^{s_{n_l}}.$$

Applying the logarithm function now yields

$$\sum_k r_{m_k} \lambda_{m_k} = \sum_l s_{n_l} \lambda_{n_l},$$

as desired. Thus the equational system (A.2) is solvable, yielding a solution $\boldsymbol{\beta} = (\beta_1^+, \beta_1^-, \ldots, \beta_{n'}^+, \beta_{n'}^-, \beta_0)^{\mathrm{T}} \in \mathbb{R}^{2n'+1}$. Setting $\alpha_0 = \exp(\beta_0)$ and $\alpha_i^+ = \exp(\beta_i^+)$, $\alpha_i^- = \exp(\beta_i^-)$, $1 \leqslant i \leqslant n'$, we obtain

$$p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n' \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n' \\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \alpha_i^-$$

for $p^*(\omega) \neq 0$. Taking into account the certain conditionals, we thus have

$$p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \alpha_i^-$$

for $\omega \in \Omega^*$ because the non-zero factors belonging to certain conditionals are 1. For $\omega \in \Omega - \Omega^*$, this equation holds trivially because $p(\omega) = p^*(\omega) = 0$ in this case. In particular, these factors $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ satisfy (10). Moreover, $P^* \models \mathcal{R}$ means $p^*(\mathrm{B}_i|\mathrm{A}_i) = x_i$, $1 \leqslant i \leqslant n$ which is equivalent to $(1 - x_i) p^*(\mathrm{A}_i\mathrm{B}_i) = x_i p^*(\mathrm{A}_i\overline{\mathrm{B}}_i)$. This shows (11).

To prove the converse assume

$$p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \alpha_i^-$$

is a distribution with $\alpha_0 > 0$, $\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-$ satisfying (10) and (11). We have to show $P^* \in C(P, \mathcal{R})$.

$$p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1\leqslant i\leqslant n \\ A_iB_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1\leqslant i\leqslant n \\ A_i\overline{B}_i(\omega)=1}} \alpha_i^- = 0$$

implies that $p(\omega) = 0$ or there is a conditional $A_i \rightsquigarrow B_i[x_i]$ in $\mathcal{R}$ such that $\alpha_i^+ = 0$, thus $x_i = 0$ by (10), and $A_iB_i(\omega) = 1$, or $\alpha_i^- = 0$, which implies $x_i = 1$, and $A_i\overline{B}_i(\omega) = 1$. Conversely, if $A_iB_i(\omega) = 1$ for a conditional $A_i \rightsquigarrow B_i[x_i]$ in $\mathcal{R}$ with $x_i = 0$, then, again by (10), $\alpha_i^+ = 0$ and $p^*(\omega) = 0$. The second case $A_i\overline{B}_i(\omega) = 1$ and $x_i = 1$ is similarly dealt with, and $p(\omega) = 0$ implies $p^*(\omega) = 0$ trivially.

Now consider two $\mathcal{R}$-equivalent multi-sets $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ and $\Omega_2 = \{\nu_1 : s_1, \ldots, \nu_{m_2} : s_{m_2}\}$, $\omega_k, \nu_l \in \Omega^*$, $1 \leqslant k \leqslant m_1$, $1 \leqslant l \leqslant m_2$, with identical conditional structures $\Omega_1^\sigma = \prod_{1\leqslant k\leqslant m_1} \sigma(\omega_k)^{r_k} = \prod_{1\leqslant l\leqslant m_2} \sigma(\nu_l)^{s_l} = \Omega_2^\sigma$. Then

$$\sum_{1\leqslant k\leqslant m_1} r_k = \sum_{1\leqslant l\leqslant m_2} s_l,$$

$$\sum_{k:\,(B_i|A_i)(\omega_k)=1} r_k = \sum_{l:\,(B_i|A_i)(\nu_l)=1} s_l,$$

$$\sum_{k:\,(B_i|A_i)(\omega_k)=0} r_k = \sum_{l:\,(B_i|A_i)(\nu_l)=0} s_l$$

hold for all $i = 1, \ldots, n$ according to Lemma 7. Checking condition (8) is now an easy calculation:

$$\begin{aligned}
&\frac{p^*(\omega_1)^{r_1} \cdots p^*(\omega_{m_1})^{r_{m_1}}}{p(\omega_1)^{r_1} \cdots p(\omega_{m_1})^{r_{m_1}}} \\
&= (\alpha_0)^{\sum_{1\leqslant k\leqslant m_1} r_k} \prod_{1\leqslant i\leqslant n} (\alpha_i^+)^{\sum_{k:A_iB_i(\omega_k)=1} r_k} \prod_{1\leqslant i\leqslant n} (\alpha_i^-)^{\sum_{k:A_i\overline{B}_i(\omega_k)=1} r_k} \\
&= (\alpha_0)^{\sum_{1\leqslant l\leqslant m_2} s_l} \prod_{1\leqslant i\leqslant n} (\alpha_i^+)^{\sum_{l:A_iB_i(\nu_l)=1} s_l} \prod_{1\leqslant i\leqslant n} (\alpha_i^-)^{\sum_{l:A_i\overline{B}_i(\nu_l)=1} s_l} \\
&= \frac{p^*(\nu_1)^{s_1} \cdots p^*(\nu_{m_2})^{s_{m_2}}}{p(\nu_1)^{s_1} \cdots p(\nu_{m_2})^{s_{m_2}}}.
\end{aligned}$$

At last we have to show that $P^* \models \mathcal{R}$. (10) implies $p^*(A_i) > 0$ for all $i = 1, \ldots, n$. For assume the contrary, i.e. there is an i such that $p^*(A_i) = 0$. Then for all ω with $A_i(\omega) = 1$,

$$p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1\leqslant i\leqslant n \\ A_iB_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1\leqslant i\leqslant n \\ A_i\overline{B}_i(\omega)=1}} \alpha_i^- = 0,$$

thus for each such ω, we have $p(\omega) = 0$, or there is a j_ω such that $\alpha_{j_\omega}^+ = 0$ and $A_{j_\omega}B_{j_\omega}(\omega) = 1$, or $\alpha_{j_\omega}^- = 0$ and $A_{j_\omega}\overline{B}_{j_\omega}(\omega) = 1$. If $\omega \in \Omega^*$, so either $x_{j_\omega} = 0$ or $x_{j_\omega} = 1$. Suppose Q is a distribution with $Q \models \mathcal{R}$. Then for each $\omega \in \Omega^*$ with $A_i(\omega) = 1$, either $x_{j_\omega} = 0$ and $A_{j_\omega}B_{j_\omega}(\omega) = 1$, or $x_{j_\omega} = 1$ and $A_{j_\omega}\overline{B_{j_\omega}}(\omega) = 1$; in any case, $q(\omega) = 0$. But $Q \models A_i \rightsquigarrow B_i[x_i]$ implies in particular $q(A_i) > 0$, so there must be an $\omega_0 \in \Omega - \Omega^*$

with $A_i(\omega_0) = 1$ and $q(\omega_0) > 0$, i.e. we have $p(\omega_0) = 0$ and $q(\omega_0) > 0$. So Q cannot be P-continuous. This contradicts the P-consistency of $\mathcal{R}$. Therefore $p^*(\mathrm{A}_i) > 0$ for all $i = 1, \ldots, n$.

For $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i] \in \mathcal{R}$ with $x_i = 0$, (10) ensures that $\alpha_i^+ = 0$, therefore $p^*(\mathrm{A}_i\mathrm{B}_i) = 0$, because α_i^+ occurs in each product $p^*(\omega')$ with $\mathrm{A}_i\mathrm{B}_i(\omega') = 1$. Because $p^*(\mathrm{A}_i) > 0$ this implies $p^*(\mathrm{B}_i|\mathrm{A}_i) = 0$. Analogously, $P^* \models \mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ for $x_i = 1$ by virtue of (10). For $x_i \in (0,1)$, $P^* \models \mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ because of (11). □

Proof of Proposition 19. Let $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$, and suppose P_1, P_2 are two distributions with $p_1(\omega|\mathrm{A}_i) = p_2(\omega|\mathrm{A}_i)$ for all $\omega \in \Omega$ and for all $i = 1, \ldots, n$, $p_1(\omega) = 0$ iff $p_2(\omega) = 0$ and such that $\mathcal{R}$ is P_1- and P_2-consistent. According to (13), $(\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-) \in \mathrm{WF}(P_1, \mathcal{R})$ iff $\alpha_i^+, \alpha_i^- \geqslant 0$, $\alpha_i^+ = 0$ iff $x_i = 0$, $\alpha_i^- = 0$ iff $x_i = 1$ and

$$(1-x_i)\alpha_i^+ \sum_{\omega:\, \mathrm{A}_i\mathrm{B}_i(\omega)=1} p_1(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-$$
$$= x_i\alpha_i^- \sum_{\omega:\, \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1} p_1(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-$$

for all $i = 1, \ldots, n$.

For any such i, and because of $p_1(\omega|\mathrm{A}_i) = p_2(\omega|\mathrm{A}_i)$ for all $\omega \in \Omega$, we have for all ω with $\mathrm{A}_i(\omega) = 1$: $p_1(\omega)/p_1(\mathrm{A}_i) = p_2(\omega)/p_2(\mathrm{A}_i)$, so $p_1(\omega) = (p_1(\mathrm{A}_i)/p_2(\mathrm{A}_i))p_2(\omega)$. Consequently, the equations above may be rewritten as

$$(1-x_i)\alpha_i^+ \sum_{\omega:\, \mathrm{A}_i\mathrm{B}_i(\omega)=1} \frac{p_1(\mathrm{A}_i)}{p_2(\mathrm{A}_i)} p_2(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-$$
$$= x_i\alpha_i^- \sum_{\omega:\, \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1} \frac{p_1(\mathrm{A}_i)}{p_2(\mathrm{A}_i)} p_2(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-$$
$$\iff$$
$$(1-x_i)\alpha_i^+ \sum_{\omega:\, \mathrm{A}_i\mathrm{B}_i(\omega)=1} p_2(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-$$
$$= x_i\alpha_i^- \sum_{\omega:\, \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1} p_2(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-$$

because $p_1(\mathrm{A}_i)/p_2(\mathrm{A}_i) > 0$. Together with the positivity condition, this is equivalent to $(\alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^-) \in \mathrm{WF}(P_2, \mathcal{R})$. (Note that all elementary events ω occurring in the sums above satisfy $\mathrm{A}_i(\omega) = 1$.) □

Proof of Proposition 20. Assume that all notations are as stated in the text of the proposition.

$(\alpha_1^+,\alpha_1^-,\dots,\alpha_n^+,\alpha_n^-) \in \mathrm{wf}(P_c^*)$, so in particular, $(\alpha_j^+,\alpha_j^-)_{j\in J}$ satisfy the positivity condition for $\mathcal{R}_J$, and for each $j \in J$, we have

$$(1-x_j)\alpha_j^+ \sum_{\omega:\,\mathrm{A}_j\mathrm{B}_j(\omega)=1} p(\omega) \prod_{\substack{k\neq j\\ \mathrm{A}_k\mathrm{B}_k(\omega)=1}} \alpha_k^+ \prod_{\substack{k\neq j\\ \mathrm{A}_k\overline{\mathrm{B}}_k(\omega)=1}} \alpha_k^-$$

$$= x_j\alpha_j^- \sum_{\omega:\,\mathrm{A}_j\overline{\mathrm{B}}_j(\omega)=1} p(\omega) \prod_{\substack{k\neq j\\ \mathrm{A}_k\mathrm{B}_k(\omega)=1}} \alpha_k^+ \prod_{\substack{k\neq j\\ \mathrm{A}_k\overline{\mathrm{B}}_k(\omega)=1}} \alpha_k^-$$

$$\iff$$

$$(1-x_j)\alpha_j^+ \sum_{\omega:\,\mathrm{A}_j\mathrm{B}_j(\omega)=1} p(\omega) \prod_{\substack{i\in I\\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\mathrm{B}_k(\omega)=1}} \alpha_k^+ \prod_{\substack{i\in I\\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \alpha_i^- \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\overline{\mathrm{B}}_k(\omega)=1}} \alpha_k^-$$

$$= x_j\alpha_j^- \sum_{\omega:\,\mathrm{A}_j\overline{\mathrm{B}}_j(\omega)=1} p(\omega) \prod_{\substack{i\in I\\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\mathrm{B}_k(\omega)=1}} \alpha_k^+ \prod_{\substack{i\in I\\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \alpha_i^- \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\overline{\mathrm{B}}_k(\omega)=1}} \alpha_k^-$$

$$\iff$$

$$(1-x_j)\alpha_j^+ \sum_{\omega:\,\mathrm{A}_j\mathrm{B}_j(\omega)=1} p_I(\omega) \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\mathrm{B}_k(\omega)=1}} \alpha_k^+ \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\overline{\mathrm{B}}_k(\omega)=1}} \alpha_k^-$$

$$= x_j\alpha_j^- \sum_{\omega:\,\mathrm{A}_j\overline{\mathrm{B}}_j(\omega)=1} p_I(\omega) \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\mathrm{B}_k(\omega)=1}} \alpha_k^+ \prod_{\substack{k\in J,k\neq j\\ \mathrm{A}_k\overline{\mathrm{B}}_k(\omega)=1}} \alpha_k^- .$$

Therefore, $(\alpha_j^+,\alpha_j^-)_{j\in J} \in \mathrm{WF}(P_I,\mathcal{R}_J)$, and, by a straightforward calculation, $P_I[\alpha_j^+,\alpha_j^-]_{j\in J} = P[\alpha_1^+,\alpha_1^-,\dots,\alpha_n^+,\alpha_n^-]$. □

Proof of Proposition 23. *Proof of* (i). Let P_1, P_2 be positive distributions over two variables A, B with $p_1(b|a) = p_2(b|a)$, and let $\mathcal{R} = \{a \rightsquigarrow b[x]\}, x \in (0,1)$. Let $P_k^* := \mathrm{F_c}(P_k,\mathcal{R}) \in C(P,\mathcal{R})$ be c-adaptations with weight factors α^+,α^- respectively β^+,β^-, $k = 1,2$. According to (11) and (12), P_1^*, P_2^* have the following forms, respectively:

ω	P_1^*	P_2^*
ab	$\alpha_0\alpha^+p_1(ab)$	$\beta_0\beta^+p_2(ab)$
$a\bar{b}$	$\alpha_0\alpha^-p_1(a\bar{b})$	$\beta_0\beta^-p_2(a\bar{b})$
$\bar{a}b$	$\alpha_0p_1(\bar{a}b)$	$\beta_0p_2(\bar{a}b)$
$\bar{a}\bar{b}$	$\alpha_0p_1(\bar{a}\bar{b})$	$\beta_0p_2(\bar{a}\bar{b})$

with

$$\frac{\alpha^+}{\alpha^-} = \frac{x}{1-x}\frac{p_1(a\bar{b})}{p_1(ab)} = \frac{x}{1-x}\frac{p_2(a\bar{b})}{p_2(ab)} = \frac{\beta^+}{\beta^-}$$

(note that $x \neq 0,1$). Due to the positivity of the prior distributions, the weight factors α^+,α^- and β^+,β^- are uniquely determined by $\mathrm{F_c}$, i.e. $\mathrm{card}(\mathrm{wf}(\mathrm{F_c}(P_k,\mathcal{R}))) = 1$, $k = 1,2$.

Calculating all cross-ratios

$$\frac{p_1^*(\dot{a}\dot{b})}{p_1(\dot{a}\dot{b})} : \frac{p_2^*(\dot{a}\dot{b})}{p_2(\dot{a}\dot{b})},$$

we obtain the three values $\alpha_0\alpha^+/\beta_0\beta^+$, $\alpha_0\alpha^-/\beta_0\beta^-$, α_0/β_0. So if $\mathrm{F_c}$ satisfies the relevance condition, then $\alpha_0\alpha^+/\beta_0\beta^+ = \alpha_0\alpha^-/\beta_0\beta^- = \alpha_0/\beta_0$. This implies $\alpha^+ = \beta^+$ and $\alpha^- = \beta^-$.

Proof of (ii). Suppose P, P' are positive prior distributions. Let $\mathcal{R} = \{\mathrm{A} \rightsquigarrow \mathrm{B}[x], \mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots\}$, $\mathcal{R}' = \{\mathrm{A}' \rightsquigarrow \mathrm{B}'[x], \mathrm{A}_1' \rightsquigarrow \mathrm{B}_1'[x_1'], \ldots\}$ be two (finite) P- respectively P'- continuous sets of probabilistic conditionals, all of $x, x_i, x_i' \in (0,1)$, no variable occurs both in antecedent and conclusion of any conditional in $\mathcal{R}$ and $\mathcal{R}'$. Let α^+, α^- respectively α'^+, α'^- be weight factors associated in $\mathrm{F_c}(P, \mathcal{R})$ respectively $\mathrm{F_c}(P', \mathcal{R}')$ with the conditional $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$ respectively $\mathrm{A}' \rightsquigarrow \mathrm{B}'[x]$. Note that the same probability x is assigned to both $\mathrm{A} \rightsquigarrow \mathrm{B}$ in $\mathcal{R}$ and $\mathrm{A}' \rightsquigarrow \mathrm{B}'$ in $\mathcal{R}'$. Let $\mathrm{F_c}$ satisfy the conditions of relevance, continuity and atomicity, and assume $\alpha^+/\alpha^- = \alpha'^+/\alpha'^-$. We have to show: $\alpha^+ = \alpha'^+$ and $\alpha^- = \alpha'^-$.

For $\mathrm{F_c}(P, \mathcal{R})$, we have

$$\frac{\alpha^+}{\alpha^-} = \frac{x}{1-x} \frac{\sum_{\omega:\, \mathrm{A}\overline{\mathrm{B}}(\omega)=1} p(\omega) \prod_{1 \leqslant i \leqslant n,\, \mathrm{A}_i\mathrm{B}_i(\omega)=1} \alpha_i^+ \prod_{1 \leqslant i \leqslant n,\, \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1} \alpha_i^-}{\sum_{\omega:\, \mathrm{AB}(\omega)=1} p(\omega) \prod_{1 \leqslant i \leqslant n,\, \mathrm{A}_i\mathrm{B}_i(\omega)=1} \alpha_i^+ \prod_{1 \leqslant i \leqslant n,\, \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1} \alpha_i^-},$$

where the α_i^+, α_i^- are associated with the remaining conditionals $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in $\mathcal{R}$, $i \in I$. Set $P_1 = P[\alpha_i^+, \alpha_i^-]_{i \in I} = P_I$, in the notation of Definition 21. $\mathrm{F_c}$ is supposed to satisfy the continuity condition, therefore $(\alpha^+, \alpha^-) \in \mathrm{wf}(\mathrm{F_c}(P_1, \{\mathrm{A} \rightsquigarrow \mathrm{B}[x]\}))$. Similarly, $(\alpha'^+, \alpha'^-) \in \mathrm{wf}(\mathrm{F_c}(P_1', \{\mathrm{A}' \rightsquigarrow \mathrm{B}'[x]\}))$ with $P_1' = P_{I'}'$. In particular, we have

$$\frac{\alpha^+}{\alpha^-} = \frac{x}{1-x}\frac{p_1(\mathrm{A}\overline{\mathrm{B}})}{p_1(\mathrm{AB})} \quad \text{and} \quad \frac{\alpha'^+}{\alpha'^-} = \frac{x}{1-x}\frac{p_1'(\mathrm{A}'\overline{\mathrm{B}'})}{p_1'(\mathrm{A}'\mathrm{B}')}.$$

Thus $\alpha^+/\alpha^- = \alpha'^+/\alpha'^-$ implies $p_1(\mathrm{A}\overline{\mathrm{B}})/p_1(\mathrm{AB}) = p_1'(\mathrm{A}'\overline{\mathrm{B}'})/p_1'(\mathrm{A}'\mathrm{B}')$, hence $p_1(\mathrm{B}|\mathrm{A}) = p_1'(\mathrm{B}'|\mathrm{A}')$. By virtue of the atomicity condition, we may replace A, A' and B, B' by new propositional variables $\tilde{A}$ and $\tilde{B}$ (note that we assumed that no variable occurs both in antecedent and conclusion). Suitably marginalizing P_1 and P_1', we thus obtain positive distributions over $\tilde{A}, \tilde{B}$ (which we will denote again by P_1 andP_1') with $p_1(\tilde{b}|\tilde{a}) = p_1'(\tilde{b}|\tilde{a})$, and α^+, α^- respectively α'^+, α'^- being the weight factors of $\mathrm{F_c}(P_1, \{\tilde{a} \rightsquigarrow \tilde{b}[x]\})$ respectively $\mathrm{F_c}(P_1', \{\tilde{a} \rightsquigarrow \tilde{b}[x]\})$. From (i), it follows that $\alpha^+ = \alpha'^+$ and $\alpha^- = \alpha'^-$, as desired. □

Proof of Proposition 25. Let $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$, and suppose P_1, P_2 are two distributions with $p_1(\omega|\mathrm{A}_i) = p_2(\omega|\mathrm{A}_i)$ for all $\omega \in \Omega$ and for all $i = 1, \ldots, n$, $p_1(\omega) = 0$ iff $p_2(\omega) = 0$ and such that $\mathcal{R}$ is P_1- and P_2-consistent. Let $P_k^* = \mathrm{F}^*(P_k, \mathcal{R})$ for $k = 1, 2$,

$$p_1^*(\omega) = p_1[\alpha_1, \ldots, \alpha_n]_F(\omega) = \alpha_0 p_1(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} F^+(x_i, \alpha_i) \prod_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} F^-(x_i, \alpha_i),$$

$$p_2^*(\omega) = p_2[\beta_1,\dots,\beta_n]_F(\omega) = \beta_0 p_2(\omega) \prod_{\substack{1\leqslant i\leqslant n\\ A_iB_i(\omega)=1}} F^+(x_i,\beta_i) \prod_{\substack{1\leqslant i\leqslant n\\ A_i\overline{B_i}(\omega)=1}} F^-(x_i,\beta_i),$$

$(\alpha_1,\dots,\alpha_n) \in \mathrm{WQ}_F(P_1,\mathcal{R}), (\beta_1,\dots,\beta_n) \in \mathrm{WQ}_F(P_2,\mathcal{R})$. Arguing as in the proof of Proposition 19, we see $\mathrm{WQ}_F(P_1,\mathcal{R}) = \mathrm{WQ}_F(P_2,\mathcal{R})$. In particular, $(\alpha_1,\dots,\alpha_n) \in \mathrm{WQ}_F(P_2,\mathcal{R})$. Due to uniqueness, we thus have $P_2^* = P_2[\alpha_1,\dots,\alpha_n]_F$, i.e.

$$p_2^*(\omega) = \alpha_0' p_2(\omega) \prod_{\substack{1\leqslant i\leqslant n\\ A_iB_i(\omega)=1}} F^+(x_i,\alpha_i) \prod_{\substack{1\leqslant i\leqslant n\\ A_i\overline{B_i}(\omega)=1}} F^-(x_i,\alpha_i).$$

Now for any $\omega \in \Omega$, we have $p_1^*(\omega) = 0$ iff $p_2^*(\omega) = 0$ or else

$$\frac{p_1^*(\omega)}{p_1(\omega)} : \frac{p_2^*(\omega)}{p_2(\omega)} = \frac{\alpha_0}{\alpha_0'},$$

a constant. This proofs the relevance property.

With the notations of Definition 21, $(\alpha_j)_{j\in J} \in \mathrm{WQ}_F(P_J,\mathcal{R}_J)$, therefore by the condition of uniqueness, $\mathrm{F}^*(P_J,\mathcal{R}_J) = P_J[\alpha_j, j\in J]_F = P[\alpha_1,\dots,\alpha_n]_F = \mathrm{F}^*(P,\mathcal{R})$. So F^* satisfies the continuity condition.

At last, Theorem 17 implies immediately $\mathrm{WQ}_F(P,\mathcal{R}) = \mathrm{WQ}_F(P',\mathcal{R}^{\Delta})$. Atomicity now follows from uniqueness, using equivalences of classical-logical formulas. We omit the technical details. □

Proof of Theorem 26. If (22) holds in principle for any adaptation carried out by $*_F$, it is surely valid for some special type of P, $\mathcal{R}_1$ and $\mathcal{R}_2$. So let P be any positive distribution over 3 variables A, B and C, let $\mathcal{R}_1 = \{a \rightsquigarrow c\,[x]\}$ and $\mathcal{R}_2 = \{b \rightsquigarrow c\,[y]\}$. Let $p_1,\dots,p_8$ denote the prior probabilities of P, $p_1 = p(abc),\dots,p_8 = p(\bar a\bar b\bar c)$. The following table shows the three adapted distributions $P *_F \mathcal{R}_1$, $(P *_F \mathcal{R}_1) *_F (\mathcal{R}_1 \cup \mathcal{R}_2)$ and $P *_F (\mathcal{R}_1 \cup \mathcal{R}_2)$:

ω	$P *_F (\mathcal{R}_1 \cup \mathcal{R}_2)$	$P *_F \mathcal{R}_1$	$(P *_F \mathcal{R}_1) *_F (\mathcal{R}_1 \cup \mathcal{R}_2)$
abc	$\alpha_0 p_1 F^+(x,\alpha) F^+(y,\beta)$	$\beta_0' p_1 F^+(x,\alpha')$	$\alpha_0'\beta_0' p_1 F^+(x,\alpha') F^+(x,\alpha_1) F^+(y,\beta_1)$
$ab\bar c$	$\alpha_0 p_2 F^-(x,\alpha) F^-(y,\beta)$	$\beta_0' p_2 F^-(x,\alpha')$	$\alpha_0'\beta_0' p_2 F^-(x,\alpha') F^-(x,\alpha_1) F^-(y,\beta_1)$
$a\bar b c$	$\alpha_0 p_3 F^+(x,\alpha)$	$\beta_0' p_3 F^+(x,\alpha')$	$\alpha_0'\beta_0' p_3 F^+(x,\alpha') F^+(x,\alpha_1)$
$a\bar b\bar c$	$\alpha_0 p_4 F^-(x,\alpha)$	$\beta_0' p_4 F^-(x,\alpha')$	$\alpha_0'\beta_0' p_4 F^-(x,\alpha') F^-(x,\alpha_1)$
$\bar a bc$	$\alpha_0 p_5 F^+(y,\beta)$	$\beta_0' p_5$	$\alpha_0'\beta_0' p_5 F^+(y,\beta_1)$
$\bar a b\bar c$	$\alpha_0 p_6 F^-(y,\beta)$	$\beta_0' p_6$	$\alpha_0'\beta_0' p_6 F^-(y,\beta_1)$
$\bar a\bar b c$	$\alpha_0 p_7$	$\beta_0' p_7$	$\alpha_0'\beta_0' p_7$
$\bar a\bar b\bar c$	$\alpha_0 p_8$	$\beta_0' p_8$	$\alpha_0'\beta_0' p_8$.

Postulating $P *_F (\mathcal{R}_1 \cup \mathcal{R}_2) = (P *_F \mathcal{R}_1) *_F (\mathcal{R}_1 \cup \mathcal{R}_2)$ yields $\alpha_0 = \alpha_0'\beta_0'$ and $F^+(y,\beta) = F^+(y,\beta_1), F^-(y,\beta) = F^-(y,\beta_1)$, hence $\beta = \beta_1$ because of (16).

Further for $x = 0$, we see $\alpha = \alpha' = \alpha_1 = 0$ and $F^-(0,0) = F^-(0,0) \cdot F^-(0,0)$. Due to (18), $F^-(0,0) \neq 0$, hence $F^-(0,0) = 1$. Similarly, $F^+(1,\infty) = 1$.

For $x \neq 1$, the weight quotients α, α' and α_1 may be calculated as

$$\alpha = \frac{x}{1-x} \cdot \frac{p_2 F^-(y,\beta) + p_4}{p_1 F^+(y,\beta) + p_3}, \qquad \alpha' = \frac{x}{1-x} \cdot \frac{p_2 + p_4}{p_1 + p_3},$$

$$\alpha_1 = \frac{x}{1-x} \cdot \frac{F^-(x,\alpha')}{F^+(x,\alpha')} \cdot \frac{p_2 F^-(y,\beta_1) + p_4}{p_1 F^+(y,\beta_1) + p_3} = \frac{x}{1-x} \cdot \alpha'^{-1} \cdot \frac{p_2 F^-(y,\beta_1) + p_4}{p_1 F^+(y,\beta_1) + p_3};$$

thus $\alpha = \alpha'\alpha_1$.

Comparing again $P *_F (\mathcal{R}_1 \cup \mathcal{R}_2)$ and $(P *_F \mathcal{R}_1) *_F (\mathcal{R}_1 \cup \mathcal{R}_2)$ we obtain

$$F^-(x,\alpha) = F^-(x,\alpha'\alpha_1) = F^-(x,\alpha')F^-(x,\alpha_1). \tag{A.3}$$

For fixed x, P and y can still be chosen arbitrarily. Choosing $y = 0$ respectively $y = 1$ simplifies the equations for $\alpha, \alpha', \alpha_1$ essentially, making these weight quotients being only dependent on P (and, of course, on x). Straightforward calculations show that indeed any $\alpha', \alpha_1 \in \mathbb{R}^+$ may be represented as weight quotients by setting up P appropriately. Therefore (A.3), i.e. (24), must hold for all positive real α', α_1, and all $x \in (0,1)$. □

Proof of Proposition 27. Let the preconditions of Proposition 27 be satisfied. Assume $x \in (0,1)$ to be held fixed and let for a moment $F_x^-(\alpha) := F^-(x,\alpha)$ be regarded only as a function of α (even if x is held fixed, α still may vary because it generally depends on many parameters other than x, at least on the prior distribution; cf. the proof of Theorem 26).

According to [1, pp.46ff] and by taking account of (24), we see $F_x^-(\alpha) = \alpha^c$ for some real constant c, and again taking into consideration the dependency on x, we obtain $F^-(x,\alpha) = \alpha^{c(x)}$ and, due to (16), $F^+(x,\alpha) = \alpha^{c(x)+1}$, for any positive real α and any $x \in (0,1)$. This proves (26). (27) now is obvious.

According to (23), $1 = F^-(0,0) = \lim_{x\to 0,\,\alpha\to 0} F^-(x,\alpha) = \lim_{x\to 0,\,\alpha\to 0} \alpha^{c(x)}$; this implies $\lim_{x\to 0} c(x) = 0$.

Similarly, by observing (26), (23) and (17), we obtain $\lim_{x\to 1} c(x) = -1$. □

Proof of Proposition 28. All priors P in this proof are assumed to be positive.

Suppose first $\mathcal{R} = \{\mathrm{A} \rightsquigarrow \mathrm{B}[x]\}$ and $\mathcal{R}' = \{\mathrm{A} \rightsquigarrow \overline{\mathrm{B}}[1-x]\}$, $x \in (0,1)$, and let α respectively β be the factor associated with $\mathcal{R}$ respectively $\mathcal{R}'$. Let $P_1^* = P *_F \mathcal{R}$, and $P_2^* = P *_F \mathcal{R}'$. According to the functional concept (P2),

$$p_1^*(\omega) = \alpha_0^{(1)} p(\omega) \begin{cases} F^+(x,\alpha), & \mathrm{AB}(\omega) = 1, \\ F^-(x,\alpha), & \mathrm{A\overline{B}}(\omega) = 1, \\ 1, & \mathrm{A}(\omega) = 0 \end{cases}$$

and

$$p_2^*(\omega) = \alpha_0^{(2)} p(\omega) \begin{cases} F^+(1-x,\beta), & \mathrm{A\overline{B}}(\omega) = 1, \\ F^-(1-x,\beta), & \mathrm{AB}(\omega) = 1, \\ 1, & \mathrm{A}(\omega) = 0, \end{cases}$$

with

$$\alpha = \frac{x}{1-x}\frac{p(\mathrm{A\overline{B}})}{p(\mathrm{AB})} = \beta^{-1}.$$

If (P4) is satisfied then $P_1^* = P_2^*$, thus implying $F^+(x,\alpha) = F^-(1-x,\alpha^{-1})$ and $F^-(x,\alpha) = F^+(1-x,\alpha^{-1})$. Together with (16), this shows $F^-(x,\alpha) = \alpha^{-1}F^-(1-x,\alpha^{-1})$. Using (26) proves (29).

Now we are going to prove (30). Because of

$$p(b|a) = p(b|ac)p(c|a) + p(b|a\bar{c})p(\bar{c}|a)$$

for arbitrarily chosen variables A, B and C, the two sets of rules

$$\mathcal{R} = \{a \rightsquigarrow c[x], ac \rightsquigarrow b[x_1], a\bar{c} \rightsquigarrow b[x_2]\},$$
$$\mathcal{R}' = \{a \rightsquigarrow b[y], ac \rightsquigarrow b[x_1], a\bar{c} \rightsquigarrow b[x_2]\} \quad \text{with } y = xx_1 + (1-x)x_2$$

are probabilistically equivalent for $x, x_1, x_2 \in (0,1)$. Because $*_F$ is assumed to satisfy (28), we have $P *_F \mathcal{R} = P *_F \mathcal{R}'$ when applying (19). We list both distributions below. The correspondence between each α_i respectively β_i and the conditional it belongs to should be clear.

ω	$P *_F \mathcal{R}$	$P *_F \mathcal{R}'$
abc	$\alpha_0 p_1 \alpha_1^{c(x)+1}\alpha_2^{c(x_1)+1}$	$\beta_0 p_1 \beta_1^{c(y)+1}\beta_2^{c(x_1)+1}$
$ab\bar{c}$	$\alpha_0 p_2 \alpha_1^{c(x)}\alpha_3^{c(x_2)+1}$	$\beta_0 p_2 \beta_1^{c(y)+1}\beta_3^{c(x_2)+1}$
$a\bar{b}c$	$\alpha_0 p_3 \alpha_1^{c(x)+1}\alpha_2^{c(x_1)}$	$\beta_0 p_3 \beta_1^{c(y)}\beta_2^{c(x_1)}$
$a\bar{b}\bar{c}$	$\alpha_0 p_4 \alpha_1^{c(x)}\alpha_3^{c(x_2)}$	$\beta_0 p_4 \beta_1^{c(y)}\beta_3^{c(x_2)}$
$\bar{a}bc$	$\alpha_0 p_5$	$\beta_0 p_5$
$\bar{a}b\bar{c}$	$\alpha_0 p_6$	$\beta_0 p_6$
$\bar{a}\bar{b}c$	$\alpha_0 p_7$	$\beta_0 p_7$
$\bar{a}\bar{b}\bar{c}$	$\alpha_0 p_8$	$\beta_0 p_8$

with

$$\alpha_1 = \frac{x}{1-x}\cdot\frac{\alpha_3^{c(x_2)}}{\alpha_2^{c(x_1)}}\cdot\frac{p_2\alpha_3 + p_4}{p_1\alpha_2 + p_3}, \tag{A.4}$$

$$\alpha_2 = \frac{x_1}{1-x_1}\cdot\frac{p_3}{p_1}, \qquad \alpha_3 = \frac{x_2}{1-x_2}\cdot\frac{p_4}{p_2}, \tag{A.5}$$

$$\beta_1 = \frac{y}{1-y}\cdot\frac{p_3\beta_2^{c(x_1)} + p_4\beta_3^{c(x_2)}}{p_1\beta_2^{c(x_1)+1} + p_2\beta_3^{c(x_2)+1}}, \tag{A.6}$$

$$\beta_2 = \frac{x_1}{1-x_1}\cdot\frac{p_3}{p_1\beta_1} = \alpha_2\beta_1^{-1}, \qquad \beta_3 = \frac{x_2}{1-x_2}\cdot\frac{p_4}{p_2\beta_1} = \alpha_3\beta_1^{-1}. \tag{A.7}$$

These last equations (A.6) and (A.7) yield

$$\frac{p_3\beta_2^{c(x_1)}}{p_4\beta_3^{c(x_2)}} = \frac{1-x_1}{1-x_2}\cdot\frac{x}{1-x},$$

and putting all these equations together we obtain

$$\alpha_1 = \beta_1^{c(x_2)-c(x_1)}. \tag{A.8}$$

For $P *_F \mathcal{R} = P *_F \mathcal{R}'$ to hold, we necessarily must have

$$\alpha_1^{c(x)} = \beta_1^{c(y)-c(x_2)}, \tag{A.9}$$

and (30) now follows from (29), (A.8) and (A.9). $\square$

Proof of Theorem 29. Properties (P3) and (P4) imply $F^-(x,\alpha) = \alpha^{c(x)}$ with a continuous real function $c(x)$ satisfying $\lim_{x\to 0} c(x) = 0, \lim_{x\to 1} c(x) = -1$ and $c(x) + c(1-x) = -1, c(xx_1+(1-x)x_2) = -c(x)c(x_1) - c(1-x)c(x_2)$ for all real $x, x_1, x_2 \in (0,1)$, due to Propositions 27 and 28. Choosing $x = \frac{1}{2}$ in the first of these equations, we see $c(\frac{1}{2}) = -\frac{1}{2}$. From the second equation, $x_2 \to 0$ yields

$$c(xx_1) = -c(x)c(x_1). \tag{A.10}$$

Using this, we obtain for $x = \frac{1}{2}$: $c(\frac{1}{2}x_1 + \frac{1}{2}x_2) = c(\frac{1}{2}x_1) + c(\frac{1}{2}x_2)$ for $x_1, x_2 \in (0,1)$. Therefore $c(x)$ fulfills a Cauchy functional equation on $(0, \frac{1}{2})$. Similar to the proof in [1, p.44], using (A.10) and $c(\frac{1}{2}) = -\frac{1}{2}$, one sees $c(x) = -x$ for all $x \in (0,1)$. Together with (16), this shows (31). $\square$

Proof of Proposition 30. Suppose $(P,\mathcal{R}) \in \mathcal{AP}$, $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$, and let P_1^*, P_2^* be distributions of type (3),

$$p_1^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^{1-x_i} \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \alpha_i^{-x_i},$$

$$p_2^*(\omega) = \beta_0 p(\omega) \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \beta_i^{1-x_i} \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \beta_i^{-x_i}$$

with nonnegative real numbers $(\alpha_i)_{0\leqslant i\leqslant n}, (\beta_i)_{0\leqslant i\leqslant n}$ fulfilling equations (4). Let $\Omega^* = \{\omega \in \Omega \mid p(\omega) > 0\}$. Without loss of generality, we may assume that all $x_i \neq 0, 1$ (in those cases, $\alpha_i = \beta_i$). So $p_1^*(\omega), p_2^*(\omega) > 0$ for all $\omega \in \Omega^*$.

We calculate the cross-entropy between P_1^* and P_2^*:

$$\begin{aligned} R(P_1^*, P_2^*) &= \sum_\omega p_1^*(\omega) \log\frac{p_1^*(\omega)}{p_2^*(\omega)} = \sum_{\omega\in\Omega^*} p_1^*(\omega)\log\frac{p_1^*(\omega)}{p_2^*(\omega)} \\ &= \sum_{\omega\in\Omega^*} p_1^*(\omega) \log\left(\frac{\alpha_0}{\beta_0} \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \left(\frac{\alpha_i}{\beta_i}\right)^{1-x_i} \prod_{\substack{1\leqslant i\leqslant n\\ \mathrm{A}_i\overline{\mathrm{B}}_i(\omega)=1}} \left(\frac{\alpha_i}{\beta_i}\right)^{-x_i}\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\omega \in \Omega^*} p_1^*(\omega) \Bigg[\log \alpha_0 - \log \beta_0 \\
&\quad + \sum_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i \mathrm{B}_i(\omega)=1}} (1 - x_i) \Big(\log \alpha_i - \log \beta_i \Big) + \sum_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i \overline{\mathrm{B}}_i(\omega)=1}} (-x_i) \Big(\log \alpha_i - \log \beta_i \Big) \Bigg] \\
&= \sum_{\omega \in \Omega^*} p_1^*(\omega) [\log \alpha_0 - \log \beta_0] \\
&\quad + \sum_{\omega \in \Omega^*} p_1^*(\omega) \sum_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i \mathrm{B}_i(\omega)=1}} (1 - x_i)(\log \alpha_i - \log \beta_i) \\
&\quad + \sum_{\omega \in \Omega^*} p_1^*(\omega) \sum_{\substack{1 \leqslant i \leqslant n \\ \mathrm{A}_i \overline{\mathrm{B}}_i(\omega)=1}} (-x_i)(\log \alpha_i - \log \beta_i) \\
&= [\log \alpha_0 - \log \beta_0] \sum_{\omega \in \Omega^*} p_1^*(\omega) \\
&\quad + \sum_{1 \leqslant i \leqslant n} (1 - x_i)(\log \alpha_i - \log \beta_i) \sum_{\omega \in \Omega^* : \mathrm{A}_i \mathrm{B}_i(\omega)=1} p_1^*(\omega) \\
&\quad + \sum_{1 \leqslant i \leqslant n} (-x_i)(\log \alpha_i - \log \beta_i) \sum_{\omega \in \Omega^* : \mathrm{A}_i \overline{\mathrm{B}}_i(\omega)=1} p_1^*(\omega) \\
&= [\log \alpha_0 - \log \beta_0] + \sum_{1 \leqslant i \leqslant n} (1 - x_i)(\log \alpha_i - \log \beta_i) p_1^*(\mathrm{A}_i \mathrm{B}_i) \\
&\quad + \sum_{1 \leqslant i \leqslant n} (-x_i)(\log \alpha_i - \log \beta_i) p_1^*(\mathrm{A}_i \overline{\mathrm{B}}_i) \\
&= [\log \alpha_0 - \log \beta_0] \\
&\quad + \sum_{1 \leqslant i \leqslant n} (\log \alpha_i - \log \beta_i) [(1 - x_i) p_1^*(\mathrm{A}_i \mathrm{B}_i) - x_i p_1^*(\mathrm{A}_i \overline{\mathrm{B}}_i)] \\
&= [\log \alpha_0 - \log \beta_0]
\end{aligned}
$$

because $p_1^*(\mathrm{B}_i | \mathrm{A}_i) = x_i$ for all $1 \leqslant i \leqslant n$.

In the same way,

$$R(P_2^*, P_1^*) = [\log \beta_0 - \log \alpha_0]$$

can be derived. But now both equations together imply

$$R(P_1^*, P_2^*) = 0,$$

for cross-entropy is nonnegative, and by using its positivity (cf. [32]), both distributions must be identical. This proves the proposition. □

Proof of Corollary 32. If $P *_{\mathrm{e}} \mathcal{R} = P$ then $P \models \mathcal{R}$ by definition of any adjustment operator.

Conversely, if $P \models \mathcal{R}$ then $\alpha_i = 1$, $1 \leqslant i \leqslant n$, provide a solution to (4) because of $x_i p(\mathrm{A}_i\overline{\mathrm{B}}_i) = (1 - x_i) p(\mathrm{A}_i\mathrm{B}_i)$, $1 \leqslant i \leqslant n$. These factors leave the prior distribution unchanged, yielding the trivial posterior P of type (3). The uniqueness statement of Proposition 30 now implies $P *_{\mathrm{e}} \mathcal{R} = P$. □

Proof of Proposition 33. Idempotence is clear with Corollary 32.

If $\mathcal{R}_1 \subseteq \mathcal{R}_2$, we have $P *_{\mathrm{e}} \mathcal{R}_2 = P *_{\mathrm{e}} (\mathcal{R}_1 \cup \mathcal{R}_2) = (P *_{\mathrm{e}} \mathcal{R}_1) *_{\mathrm{e}} \mathcal{R}_2$ because of (22). By prerequisite $P *_{\mathrm{e}} \mathcal{R}_1 \models \mathcal{R}_2$, so again Corollary 32 implies $(P *_{\mathrm{e}} \mathcal{R}_1) *_{\mathrm{e}} \mathcal{R}_2 = P *_{\mathrm{e}} \mathcal{R}_1$. □

References

[1] J. Aczél, *Vorlesungen über Funktionalgleichungen und ihre Anwendungen* (Birkhäuser, Basel, 1961).

[2] E.W. Adams, *The Logic of Conditionals* (Reidel, Dordrecht, 1975).

[3] F. Bacchus, *Representing and Reasoning with Probabilistic Knowledge: A Logical Approach to Probabilities* (MIT Press, Cambridge, MA, 1990).

[4] G. Brewka, Preferred subtheories: an extended logical framework for default reasoning, in: *Proceedings IJCAI-89*, Vol. 2 (Morgan Kaufmann, San Mateo, CA, 1989) 1043–1048.

[5] P.G. Calabrese, Deduction and inference using conditional logic and probability, in: I.R. Goodman, M.M. Gupta, H.T. Nguyen and G.S. Rogers, eds., *Conditional Logic in Expert Systems* (North-Holland, Amsterdam, 1991) 71–100.

[6] I. Csiszár, I-divergence geometry of probability distributions and minimization problems, *Ann. Probab.* 3 (1975) 146–158.

[7] B. De Finetti, *Theory of Probability*, Vols. 1 and 2 (John Wiley and Sons, New York, 1974).

[8] D. Dubois and H. Prade, Conditional objects and non-monotonic reasoning, in: *Proceedings 2nd International Conference on Principles of Knowledge Representation and Reasoning* (*KR-91*) (Morgan Kaufmann, Los Altos, CA, 1991) 175–185.

[9] P. Gaerdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (MIT Press, Cambridge, MA, 1988).

[10] H. Geffner, *Default Reasoning: Causal and Conditional Theories* (MIT Press, Cambridge, MA, 1992).

[11] M. Goldszmidt, Research issues in qualitative and abstract probability, *AI Mag.* (Winter 1994) 63–66.

[12] M. Goldszmidt and J. Pearl, Rank-based systems: a simple approach to belief revision, belief update and reasoning about evidence and actions, in: *Proceedings 3rd International Conference on Principles of Knowledge Representation and Reasoning* (*KR-92*), Cambridge, MA (1992) 661–672.

[13] I.J. Good, Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, *Ann. Math. Statist.* 34 (1963) 911–934.

[14] A.J. Grove, J.Y. Halpern and D. Koller, Random worlds and maximum entropy, *J. Artif. Intell. Res.* 2 (1994) 33–88.

[15] P. Hajek, T. Havranek and R. Jirousek, *Uncertain Information Processing in Expert Systems* (CRC Press, Boca Raton, FL, 1992).

[16] E.T. Jaynes, *Papers on Probability, Statistics and Statistical Physics* (Reidel, Dordrecht, 1983).

[17] R.W. Johnson and J.E. Shore, Comments on and correction to "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy", *IEEE Trans. Inform. Theory* 29 (6) (1983) 942–943.

[18] G. Kern-Isberner, Conditional logics and entropy, Informatik Fachbericht 203, FernUniversität Hagen (1996).

[19] G. Kern-Isberner, A logically sound method for uncertain reasoning with quantified conditionals, in: *Proceedings 1st International Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU-FAPR'97* (Springer, Berlin, 1997) 365–379.

[20] S. Kraus, D. Lehmann and M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, *Artificial Intelligence* 44 (1990) 167–207.

[21] S. Kullback, *Information Theory and Statistics* (Dover, New York, 1968).

[22] S.L. Lauritzen and D.J. Spiegelhalter, Local computations with probabilities in graphical structures and their applications to expert systems, *J. Roy. Statist. Soc. Ser. B* 50 (2) (1988) 415–448.
[23] R.C. Lyndon and P.E. Schu, *Combinatorial Group Theory* (Springer, Berlin, 1977).
[24] D. Makinson, General patterns in nonmonotonic reasoning, in: D.M. Gabbay, C.H. Hogger and J.A. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 3 (Oxford University Press, Oxford, 1994) 35–110.
[25] D. Nute, *Topics in Conditional Logic* (Reidel, Dordrecht, 1980).
[26] J.B. Paris and A. Vencovská, A note on the inevitability of maximum entropy, *Internat. J. Approximate Reasoning* 14 (1990) 183–223.
[27] J.B. Paris and A. Vencovská, In defence of the maximum entropy inference process. *Internat. J. Approximate Reasoning*, to appear.
[28] J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann, San Mateo, CA, 1988).
[29] J. Pearl, Bayesian and belief-functions formalisms for evidential reasoning: A conceptual analysis, in: Z.W. Ras and M. Zemankova, eds., *Intelligent Systems—State of the Art and Future Directions* (Ellis Horwood, Chichester, 1990) 73–117.
[30] W. Rödder and G. Kern-Isberner, Representation and extraction of information by probabilistic logic, *Inform. Systems* 21 (8) (1997) 637–652.
[31] W. Rödder and C.-H. Meyer, Coherent knowledge processing at maximum entropy by spirit, in: E. Horvitz and F. Jensen, eds., *Proceedings 12th Conference on Uncertainty in Artificial Intelligence*, Portland, OR (Morgan Kaufmann, San Francisco, CA, 1996) 470–476.
[32] J.E. Shore, Relative entropy, probabilistic inference and AI, in: L.N. Kanal and J.F. Lemmer, eds., *Uncertainty in Artificial Intelligence* (North-Holland, Amsterdam, 1986) 211–215.
[33] J.E. Shore and R.W. Johnson, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inform. Theory* 26 (1980) 26–37.
[34] J.E. Shore and R.W. Johnson, Properties of cross-entropy minimization, *IEEE Trans. Inform. Theory* 27 (1981) 472–482.
[35] M. Spies, Combination of evidence with conditional objects and its application to cognitive modeling, in: I.R. Goodman, M.M. Gupta, H.T. Nguyen and C.S. Rogers, eds., *Conditional Logic in Expert Systems* (North-Holland, Amsterdam, 1991) 181–209.
[36] J. Whittaker, *Graphical Models in Applied Multivariate Statistics* (John Wiley & Sons, New York, 1990).