KYBERNETIKA — VOLUME *50* (2014), NUMBER 2, PAGES 268–283

DEGRADATION IN PROBABILITY LOGIC: WHEN MORE INFORMATION LEADS TO LESS PRECISE CONCLUSIONS

Christian Wallmann and Gernot D. Kleiter

Probability logic studies the properties resulting from the probabilistic interpretation of logical argument forms. Typical examples are probabilistic Modus Ponens and Modus Tollens. Argument forms with two premises usually lead from precise probabilities of the premises to imprecise or interval probabilities of the conclusion. In the contribution, we study generalized inference forms having three or more premises. Recently, Gilio has shown that these generalized forms "degrade" — more premises lead to more imprecise conclusions, i. e., to wider intervals. We distinguish different forms of degradation. We analyse Predictive Inference, Modus Ponens, Bayes' Theorem, and Modus Tollens. Special attention is devoted to the case where the conditioning events have zero probabilities. Finally, we discuss the relation of degradation to monotonicity.

Keywords: probability logic, generalized inference forms, degradation, total evidence, coherence, probabilistic Modus Tollens

Classification: 03B48, 97K50

1. INTRODUCTION

Consider a knowledge base that contains the observations D_1, D_2, D_3. By the knowledge base, we evaluate that $P(H|D_1 \wedge D_2) \in [0.1, 0.12]$. In addition, by the knowledge base, we evaluate that $P(H|D_1 \wedge D_2 \wedge D_3) \in [0.6, 0.9]$. Which one of the two probability intervals should we use to update the probability of H? Three properties may be considered for this choice: (i) The width of the intervals (the interval $[0.1, 0.12]$ is tighter than $[0.6, 0.9]$), (ii) the position of the intervals (the positions of $[0.1, 0.12]$ and $[0.6, 0.9]$ are rather different), (iii) the amount of information ($D_1 \wedge D_2$ is less specific than $D_1 \wedge D_2 \wedge D_3$). The principle of *total evidence* requires to base the updated probability of H on $P(H|D_1 \wedge D_2 \wedge D_3)$. However, this leads to the more imprecise interval. In probability

DOI: 10.14736/kyb-2014-2-0268

logic[1] conflicts between the amount of evidence and the precision of the conclusions are quite typical [4]. Especially in inferences that generalize common inference forms, like generalized Modus Ponens or Modus Tollens, more specific information leads to more imprecise probabilities of the conclusions. The fact that the width of the interval of the conclusion increases as the number of premises increases has been called "degradation in probability logic". In the extreme case, the probability of the conclusion may be noninformative, i. e., may attain any value between zero and one [4, 7, 12, 13].

Probability logic studies the properties resulting from the probabilistic interpretation of logical argument forms. It determines the set of all coherent probability values of the conclusion if a coherent probability assessment on the premises is given. This set is according to de Finetti's Fundamental Theorem [2, 9] an interval or a point value. The probability of a conditional $P(A \Rightarrow B)$ is represented by the conditional probability $P(B|A)$. Consider, for example, Modus Ponens. Its logical form infers from the premises $\{A, A \Rightarrow B\}$ the conclusion B. Accordingly, the probabilistic version of Modus Ponens infers from the premises $\{P(A) = \alpha, P(B|A) = \beta\}$ the conclusion $P(B) \in [\alpha\beta, \alpha\beta + 1 - \alpha]$. Generalized probabilistic Modus Ponens determines the interval $P(H) \in [\delta', \delta'']$ if the premises $\{P(E_1) = \alpha_1, \ldots, P(E_n) = \alpha_n, P(H|\bigwedge_{i=1}^n E_i) = \beta\}$ are given (see Section 2.5 below).

For a generalized inference form we denote by I_n the interval for the conclusion if n premises are given. Let $|I_i|$ be the width of the interval I_i. A generalized inference form *degrades* if and only if for all $i, j \in \mathbb{N}$: If $i < j$, then $|I_i| \leq |I_j|$. A different position of the interval, which is based on more premises, may compensate for a wider interval (compare the intervals $[0.6, 0.9]$ and $[0.1, 0.12]$ in the introductory example). To study degradation in more detail, we therefore distinguish two forms of degradation. A generalized inference form *strongly degrades* if and only if for all $i, j \in \mathbb{N}$: If $i < j$, then $I_i \subseteq I_j$ (i. e., the former interval is included in the latter). A generalized inference form *weakly degrades* if and only if it degrades and if for some $k, l \in \mathbb{N}$ $I_k \not\subseteq I_l$ and $I_l \not\subseteq I_k$ (i. e., the latter interval is wider but does not include the former). In general, however, the preference of intervals with different width and intervals with different positions is a difficult task and both forms of degradation are problematic for the application of probability logic to generalized inference forms.

In this contribution, we analyse generalizations of Modus Ponens, Predictive Inference, Conjunction, Bayes' Theorem, and Modus Tollens for the different kinds of degradation. It is common to all of the inference forms considered in the present paper — with the exception of Modus Tollens — that a certain form of ultimate degradation is observed. If the number of premises is sufficiently high, then the interval of the conclusion is the unit interval $[0, 1]$. The reason is that the lower bound of the conjunction of n events $P(\bigwedge_{i=1}^n E_i)$ quickly becomes zero if the number of conjuncts n increases. The fact that the lower bound of the conjunction is often zero has the consequence that the conditioning event of many conditional events has zero probability. We therefore pay special attention to this case. In particular, we study the generalization of Bayes' Theorem when the prior probability of the hypothesis is zero or when the data has zero

[1]If probabilities of conditionals $P(A \Rightarrow B)$ are represented by conditional probabilities $P(B|A)$, then one usually speaks of conditional probability logic. However, in the remainder of the paper, we use the expression 'probability logic' instead of 'conditional probability logic'.

probability. The case when the conditioning event has zero probability can be treated in the coherence approach of de Finetti [1, 2]. Furthermore, we prove the result for the generalized Modus Tollens stated in [7, 12, 13]. Finally, we discuss the relation of degradation to monotonicity and its significance for uncertain reasoning. In particular, we show that probability logic is weak for decision making and that it should therefore be supplemented by additional principles.

2. DEGRADATION OF INFERENCES IN PROBABILITY LOGIC

2.1. Terminology

Let $\mathcal{F} = \{E_1|H_1, \ldots, E_n|H_n\}$ be a set of conditional events. If H_i is the sure event, i. e., $H_i = \top$, then we write E_i instead of $E_i|H_i$. A possible outcome or a *constituent* is a conjunction of the form $\pm E_1 \wedge \ldots \wedge \pm E_n \wedge \pm H_1 \wedge \ldots \wedge \pm H_n$, where for all events $A \in \{E_1, \ldots, E_n, H_1, \ldots, H_n\}$ $\pm A$ is either A or $\neg A$. If the $2n$ events are logically independent, then there are 2^{2n} constituents $C_1, \ldots, C_{2^{2n}}$. The probability of an event is the sum of the probabilities of the constituents C_r verifying it. Table 1 shows our notation in the case of three events H, E_1, E_2.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	Probability
H	1	1	1	1	0	0	0	0	$P(H) = x_1 + x_2 + x_3 + x_4$
E_1	1	1	0	0	1	1	0	0	$P(E_1) = x_1 + x_2 + x_5 + x_6$
E_2	1	0	1	0	1	0	1	0	$P(E_2) = x_1 + x_3 + x_5 + x_7$

Tab. 1. Constituents $C_1, \ldots, C_8$ and their probabilities x_i for $n = 3$ events.

The interval of the coherent probability values for the conclusion of an inference form can be determined by solving sequences of linear systems. This is a corollary of the following theorem which characterizes coherence [1, p. 81] (original for infinite sets of conditional events).

Theorem 1. (Coletti and Scozzafava, 2002) A probability assessment P on $\mathcal{F} = \{E_1|H_1, \ldots, E_n|H_n\}$ is coherent iff there exists a sequence of compatible systems, with unknowns $x_r^\alpha \geq 0$,

$$\mathcal{S}_\alpha = \begin{cases} \sum\limits_{C_r \subseteq E_i \wedge H_i} x_r^\alpha = P(E_i|H_i) \sum\limits_{C_r \subseteq H_i} x_r^\alpha \\ [\text{if} \sum\limits_{C_r \subseteq H_i} x_r^{\alpha-1} = 0,\ \alpha \geq 1] \quad (i = 1, \ldots, n) \\ \sum\limits_{C_r \subseteq H_0^\alpha} x_r^\alpha = 1 \end{cases}$$

with $\alpha = 0, 1, \ldots, n$, where $H_0^0 = H_1 \vee \ldots \vee H_n$ and H_0^α denotes, for $\alpha \geq 1$, the union of the H_i such that $\sum\limits_{C_r \subseteq H_i} x_r^{\alpha-1} = 0$.

Consider for example probabilistic Modus Tollens. The premises are $P(\neg E_1) = \alpha$, $P(E_1|H) = \beta$. Employing the notation of Table 1, the lower (resp. upper) bound for the conclusion $P(\neg H)$ can be determined by minimizing (resp. maximizing) the sum $x_5 + x_6 + x_7 + x_8$ in the following linear system

$$\begin{aligned} x_3 + x_4 + x_7 + x_8 &= \alpha \\ \beta(x_1 + x_2 + x_3 + x_4) &= x_1 + x_2 \\ \sum_{i=1}^{8} x_i = 1, \qquad x_i &\geq 0 \ . \end{aligned}$$

In the case of Modus Tollens only one linear system is to be considered. However, if we want to determine the probability of a conditional event with a conditioning event that has zero probability, then at least two linear system are to be considered. We demonstrate this in the proof of Bayes' Theorem where the data (corresponding to the conditional event of the conditional) has zero probability (Theorem 6, Theorem 7).

2.2. Conjunction

For the remainder of the paper we suppose that the given probability assessment on the premises is coherent. For the conjunction of n events the following theorem holds (see, for example, [3, 4]).

Theorem 2. (Conjunction of n events) If $P(E_i|H) = \alpha_i$, for $i = 1, \dots, n$, then

$$P(\bigwedge_{i=1}^{n} E_i|H) \in \left[\max\left\{0, \sum_{i=1}^{n} \alpha_i - (n-1)\right\}, \min\{\alpha_i\}\right] \ .$$

The lower bound of $P(\bigwedge_{i=1}^{n+1} E_i|H)$ is less than or equal to that of $P(\bigwedge_{i=1}^{n} E_i|H)$. If the lower bound of $P(\bigwedge_{i=1}^{n} E_i|H)$ is greater than zero, equality holds if and only if $P(E_{n+1}|H) = \alpha_{n+1} = 1$. Moreover, if $n \geq \sum_{i=1}^{n} \alpha_i + 1$, then the lower bound of $P(\bigwedge_{i=1}^{n} E_i|H)$ is 0. We shall soon see that these properties of the conjunction cause the degradation of many other inference forms.

2.3. Disjunction

Gilio [4] investigated the System P rule Or. This rule concerns disjunctions in the conditioning event. In this section, we consider the probability of the disjunction of n events conditional on an event H. The lower and upper bound are easily obtained by the conjunction rule [3].

Theorem 3 (Disjunction of n events)**.** If $P(E_i|H) = \alpha_i$, for $i = 1, \dots, n$, then

$$P(E_1 \vee E_2 \vee \ldots \vee E_n|H) \in \left[\alpha^*, \min\left\{1, \sum_{i=1}^{n} \alpha_i\right\}\right]$$

with $\alpha^* = \max\{\alpha_i\}$.

Proof. We note that $\bigvee_{i=1}^n E_i$ is equivalent to $\neg \bigwedge_{i=1}^n \neg E_i$ and, accordingly, $P(\bigvee_{i=1}^n E_i|H) = 1 - P(\bigwedge_{i=1}^n \neg E_i|H)$. We denote $P(\bigvee_{i=1}^n E_i|H)$ by α. As $P(\neg E_i|H) = 1 - \alpha_i$, we know the probability of the negation of each of the events and we can apply the conjunction theorem to the n negated events to obtain the disjunction. We have

$$1 - \alpha = P(\bigwedge_{i=1}^n \neg E_i|H) \in \left[\max\left\{0, \sum_{i=1}^n (1-\alpha_i) - (n-1)\right\}, \min\{1-\alpha_i\}\right] . \qquad (2.1)$$

Using $P(\bigvee_{i=1}^n E_i|H) = 1 - P(\bigwedge_{i=1}^n \neg E_i|H)$ and (2.1), we obtain

$$\alpha \in \left[\alpha^*, \min\left\{1, \sum_{i=1}^n \alpha_i\right\}\right] . \qquad (2.2)$$

□

An upper bound of the disjunction of n events, that is different from 1, is getting wider, if $\alpha_{n+1} \neq 0$.

2.4. Predictive Inference

Predictive Inference is one of the key inference rules in Bayesian statistics. It determines the predictive probability $P(H|E_1 \wedge \ldots \wedge E_r \wedge \neg E_{r+1} \wedge \ldots \wedge \neg E_n)$ of H after having observed r successes and $n - r$ failures in the set $\{E_i\}_{i=1}^n$. It is of main importance if H is regarded exchangeable with the other events. If at least one of the events $\{E_i\}_{i=1}^n$ did not occur, i. e., $r < n$, then the interval obtained for the predictive probability is the unit interval [12]. As observed in [12], the case where all previous trials were successes, i. e., $r = n$, is a special case of the System P rule Cautious Monotonicity. Consequently, the following theorem is a corollary of the result for Cautious Monotonicity stated in [4].

Theorem 4. (Predictive probability) If $P(H) = \beta$ and $P(E_i) = \alpha_i$, for $i = 1, \ldots, n$, then $P(H|E_1 \wedge \ldots \wedge E_n) \in [\gamma', \gamma'']$, with

$$\gamma' = \begin{cases} \max\left\{0, \frac{\beta + \sum_{i=1}^n \alpha_i - n}{\sum_{i=1}^n \alpha_i - (n-1)}\right\} & \text{if } \sum_{i=1}^n \alpha_i - (n-1) > 0 \\ 0 & \text{if } \sum_{i=1}^n \alpha_i - (n-1) \leq 0 \end{cases}$$

$$\gamma'' = \begin{cases} \min\left\{1, \frac{\beta}{\sum_{i=1}^n \alpha_i - (n-1)}\right\} & \text{if } \sum_{i=1}^n \alpha_i - (n-1) > 0 \\ 1 & \text{if } \sum_{i=1}^n \alpha_i - (n-1) \leq 0. \end{cases}$$

We compare this result with the result for the case where the premise $P(E_{n+1}) = \alpha_{n+1}$ is added and the conclusion is $H|E_1 \wedge \ldots \wedge E_n \wedge E_{n+1}$. Observe that in both cases the same event H is predicted. Theorem 4 shows that the upper bound of the conclusion increases and that its lower bound decreases. Thus, the interval gets wider if a new event is added and the old interval is a subset of the new interval. Consequently, in the case of predictive inference we have a strong degradation. Furthermore, if $n \geq \sum_{i=1}^n \alpha_i + 1$, then $P(H|E_1 \wedge \ldots \wedge E_n) \in [0, 1]$. The property that the lower bound of the conjunction decreases is the reason for both, the strong degradation of Predictive Inference and for obtaining the unit interval if n is large.

2.5. Modus Ponens

Modus Ponens is a special case of the System P rule Cut. The following theorem is a corollary of Gilio's result for the generalization of the Cut rule [4].

Theorem 5. (Modus Ponens) If $P(E_i) = \alpha_i$, for $i = 1, \ldots, n$, and $P(H|\bigwedge_{i=1}^{n} E_i) = \beta$, then

$$P(H) \in [\beta\sigma_n \ , \ \beta\sigma_n + 1 - \sigma_n],$$
$$\text{with } \sigma_n = \max\left\{0, \sum_{i=1}^{n} \alpha_i - (n-1)\right\}.$$

We compare this result with the result where $P(E_{n+1}) = \alpha_{n+1}$ is added to the premises and $P(H|\bigwedge_{i=1}^{n} E_i) = \beta$ is replaced by $P(H|\bigwedge_{i=1}^{n+1} E_i) = \gamma$. If $\gamma = \beta$, then Modus Ponens strongly degrades. However, even if $\beta \neq \gamma$, the width of the interval for $P(H)$ normally increases. This follows from the facts that its width is $1 - \sigma_n$ and that σ_n normally decreases. Consequently, that the lower bound of the conjunction decreases causes the degradation of Modus Ponens. However, since β is replaced by $\gamma \neq \beta$, the position of the interval for $P(H)$ may change. Therefore, in the case of Modus Ponens a weak degradation is observed.

Example 1. Consider the premise sets T and T' such that
$T = \{P(E_1) = 0.9,\ P(E_2) = 0.8,\ P(E_3) = 0.95,\ P(H|E_1 \wedge E_2 \wedge E_3) = 0.8\}$ and
$T' = \{P(E_1) = 0.9,\ P(E_2) = 0.8,\ P(E_3) = 0.95,\ P(E_4) = 0.8,\ P(H|E_1 \wedge E_2 \wedge E_3 \wedge E_4) = 0.1\}$.
From T it follows that $P(H) \in [0.52, 0.87]$, whereas from T' it follows that $P(H) \in [0.045, 0.595]$.

The width of the interval $1 - \sigma_n$ depends on the lower bound of the conjunction σ_n. Since this lower bound is zero if $n \geq \sum_{i=1}^{n} \alpha_i + 1$, the interval for $P(H)$ is the unit interval if the number of premises is sufficiently high.

2.6. Bayes' theorem

Suppose that the prior probability of a certain hypothesis $P(H) = \delta$, the likelihoods of the data given both, the hypothesis H, $P(D|H) = \beta$, and the alternative hypothesis $\neg H$, $P(D|\neg H) = \gamma$, are given. The posterior probability of the hypothesis H given the data D is obtained by Bayes' Theorem $P(H|D) = \frac{\beta\delta}{\beta\delta + \gamma(1-\delta)}$. The premises of generalized Bayes' Theorem are $P(H) = \delta$, $P(E_1|H) = \beta_1, \ldots, P(E_n|H) = \beta_n$, $P(E_1|\neg H) = \gamma_1, \ldots, P(E_n|\neg H) = \gamma_n$. In inferential statistics it is often assumed that the E_i's are independent and identically distributed. To be as general as possible, we do neither require conditional independence of the E_i's given H nor do we require that $P(E_i|H) = P(E_j|H)$ for $i \neq j$. The conclusion of generalized Bayes' Theorem is

$P(H|E_1 \wedge \ldots \wedge E_n)$. Observe that if $P(E_1 \wedge \ldots \wedge E_n) > 0$, then

$$P(H|E_1 \wedge \ldots \wedge E_n) = \frac{P(H \wedge E_1 \wedge \ldots \wedge E_n)}{P(E_1 \wedge \ldots \wedge E_n)} = \frac{P(H)P(E_1 \wedge \ldots \wedge E_n|H)}{P(H)P(E_1 \wedge \ldots \wedge E_n|H) + P(\neg H)P(E_1 \wedge \ldots \wedge E_n|\neg H)} . \quad (2.3)$$

To prove the result for the generalization of Bayes' Theorem (Theorem 6 and Theorem 7) we consequently treat two cases for the probability of the data $P(E_1 \wedge \ldots \wedge E_n)$: (i) $P(E_1 \wedge \ldots \wedge E_n) > 0$ and (ii) $P(E_1 \wedge \ldots \wedge E_n) = 0$. In case (ii) it is relevant whether the prior probability $P(H)$ is zero, one, or different from both values. To handle case (ii) properly we make use of Theorem 1. The special case $n = 1$ has been investigated in detail by Coletti and Scozzafava [1, Chapter 16]. The proofs of the next two results are obtained by analogous considerations; hence, we omit the first one.

Theorem 6. (Bayes' Theorem, lower bound) Suppose that $P(H) = \delta$ and that for all $i = 1, \ldots, n$, $P(E_i|H) = \beta_i$ and $P(E_i|\neg H) = \gamma_i$. Then:

- If $\delta(\sum_{i=1}^n \beta_i - (n-1)) > 0$, then

$$P(H|E_1 \wedge \ldots \wedge E_n) \geq \frac{\delta(\sum_{i=1}^n \beta_i - (n-1))}{\delta(\sum_{i=1}^n \beta_i - (n-1)) + (1-\delta)\min\{\gamma_i\}} .$$

- If $\delta(\sum_{i=1}^n \beta_i - (n-1)) \leq 0$, then $P(H|E_1 \wedge \ldots \wedge E_n) \geq 0$.

Theorem 7. (Bayes' Theorem, upper bound) Suppose that $P(H) = \delta$ and that for all $i = 1, \ldots, n$, $P(E_i|H) = \beta_i$ and $P(E_i|\neg H) = \gamma_i$. Then:

- If $(1-\delta)(\sum_{i=1}^n \gamma_i - (n-1)) > 0$, then

$$P(H|E_1 \wedge \ldots \wedge E_n) \leq \frac{\delta \min\{\beta_i\}}{\delta \min\{\beta_i\} + (1-\delta)(\sum_{i=1}^n \gamma_i - (n-1))} .$$

- If $(1-\delta)(\sum_{i=1}^n \gamma_i - (n-1)) \leq 0$, then $P(H|E_1 \wedge \ldots \wedge E_n) \leq 1$.

Proof. We distinguish two cases.

(I) If $(1-\delta)(\sum_{i=1}^n \gamma_i - (n-1)) > 0$, then $P(E_1 \wedge \ldots \wedge E_n) > 0$. The result is obtained by application of the Conjunction Theorem (Theorem 2) to (2.3).

(II) If $(1-\delta)(\sum_{i=1}^n \gamma_i - (n-1)) \leq 0$, we distinguish two cases (i) $\delta \min\{\beta_i\} > 0$ and (ii) $\delta \min\{\beta_i\} = 0$.

In case (i) the upper bound 1 is obtained by setting $P(H \wedge E_1 \wedge \ldots \wedge E_n)$ to $\delta \min\{\beta_i\} > 0$ and $P(\neg H \wedge E_1 \wedge \ldots \wedge E_n)$ to its minimum 0.

In case (ii) we obtain the upper bound by setting the probability of the data $P(E_1 \wedge \ldots \wedge E_n)$ to 0. We treat the case $n = 2$. The proof generalizes to the case $n > 2$

straightforwardly. We build the sequence of linear systems $\mathcal{S}_\alpha$ (Theorem 1). To improve readability we write x_i instead of x_i^0, y_i instead of x_i^1, and z_i instead of x_i^2.

Using the notation of Table 1, the first linear system $\mathcal{S}_0$ is given by

$$\begin{gathered} x_1 + x_5 = 0 \\ P(H|E_1 \wedge E_2)(x_1 + x_5) = x_1 \\ x_1 + x_2 + x_3 + x_4 = \delta \\ x_1 + x_2 = \beta_1(x_1 + x_2 + x_3 + x_4), \qquad x_1 + x_3 = \beta_2(x_1 + x_2 + x_3 + x_4) \\ x_5 + x_6 = \gamma_1(x_5 + x_6 + x_7 + x_8), \qquad x_5 + x_7 = \gamma_2(x_5 + x_6 + x_7 + x_8) \\ x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 1, \quad x_i \geq 0 \ . \end{gathered}$$

As unique solution of $\mathcal{S}_0$ we obtain $x_1 = x_5 = 0$, $x_2 = \beta_1\delta$, $x_3 = \beta_2\delta$, $x_4 = \delta - (\beta_1 + \beta_2)\delta$, $x_6 = \gamma_1(1-\delta)$, $x_7 = \gamma_2(1-\delta)$, $x_8 = (1-\delta) - (\gamma_1 + \gamma_2)(1-\delta)$. Since $\delta \min\{\beta_1, \beta_2\} = 0$, it holds that $x_4 = \delta(1 - \beta_1 - \beta_2) = \delta(1 - \max\{\beta_1, \beta_2\}) \geq 0$, and since by assumption $\gamma_1 + \gamma_2 \leq 1$, it is $x_8 \geq 0$, so that the solution is admissible.
If $0 < P(H) = \delta < 1$, then $H_0^1 = E_1 \wedge E_2$. The system $\mathcal{S}_1$ is consequently given by

$$\begin{gathered} P(H|E_1 \wedge E_2)(y_1 + y_5) = y_1 \\ y_1 + y_5 = 1, \qquad y_i \geq 0 \ . \end{gathered}$$

So that $P(H|E_1 \wedge E_2) = \frac{y_1}{y_1 + y_5}$ can attain any value in $[0, 1]$.

If $P(H) = \delta = 0$ (the case $P(H) = 1$ is treated in the same way), then $x_1 = x_2 = x_3 = x_4 = x_5 = 0$ and consequently $H_0^1 = H \vee (E_1 \wedge E_2)$. In the system $\mathcal{S}_1'$ all constraints that concern conditional events with conditioning event H remain.

$$\begin{aligned} P(H|E_1 \wedge E_2)(y_1 + y_5) &= y_1 \\ y_1 + y_2 = \beta_1(y_1 + y_2 + y_3 + y_4), \qquad y_1 + y_3 &= \beta_2(y_1 + y_2 + y_3 + y_4) \\ y_1 + y_2 + y_3 + y_4 + y_5 &= 1, \qquad y_i \geq 0 \ . \end{aligned}$$

We solve $\mathcal{S}_1'$ in such a way that $P(H) > 0$ and $P(E_1 \wedge E_2) = 0$. The unique solution in this case is $y_2 = \beta_1$, $y_3 = \beta_2$, $y_4 = 1 - (\beta_1 + \beta_2)$. Then $H_0^2 = E_1 \wedge E_2$ and the third system $\mathcal{S}_2'$ is

$$\begin{aligned} P(H|E_1 \wedge E_2)(z_1 + z_5) &= z_1 \\ z_1 + z_5 = 1, \qquad z_i &\geq 0 \ . \end{aligned}$$

So that $P(H|E_1 \wedge E_2) = \frac{z_1}{z_1 + z_5}$ can attain any value in $[0, 1]$.
In both cases, $0 < P(H) < 1$ and $P(H) = 0$, we have constructed a sequence of compatible systems (S_α), with unknowns (x_i^α), $i = 1, \ldots, 8$, $\alpha = 0, 1, 2$, such that $P(H|E_1 \wedge E_2) \in [0, 1]$. According to Theorem 1 $P(H|E_1 \wedge E_2)$ can coherently attain any value in $[0, 1]$. □

In the present paper the non uniquely determined probabilities result from the fact that no knowledge about the association (conditional independence or kind of dependence) of the data given the hypothesis is presumed. For a slightly more general approach to iterated conditioning in Bayes' Theorem and the accordingly also non uniquely determined probabilities see [5].

Bayes' Theorem does not degrade. First of all, Bayes' Theorem does not degrade strongly. The lower bound is not monotonically decreasing, because it depends on the minimum of the set $\{\gamma_i\}$. If for a given n the premises $P(E_{n+1}|\neg H) = \gamma_{n+1} < \min\{\gamma_i\}$ and $P(E_{n+1}|H) = \beta_{n+1}$ are added, the lower bound may increase. Similar considerations show that the upper bound is not monotonically increasing. As a consequence, intervals with rather different positions may result.

Example 2. Suppose that $P(H) = 0.1$, $P(E_1|H) = 0.9$, $P(E_2|H) = 0.8$, $P(E_3|H) = 0.4$, $P(E_1|\neg H) = 0.9999$, $P(E_2|\neg H) = 0.9999$, $P(E_3|\neg H) = 0.001$. Then $P(H|E_1 \wedge E_2) \in [0.072, 0.081]$, but $P(H|E_1 \wedge E_2 \wedge E_3) = [0.917, 0.982]$.[2]

In Bayes' Theorem even no weak degradation is observed. In general, the probability interval of the conclusion need not get wider as the number of premises increases.

Example 3. Suppose that $P(H) = 0.9$, $P(E_1|H) = 0.99$, $P(E_2|H) = 0.99$, $P(E_3|H) = 0.98$, $P(E_1|\neg H) = 0.9999$, $P(E_2|\neg H) = 0.9999$, $P(E_3|\neg H) = 0.001$. Then $P(H|E_1 \wedge E_2) \in [0.898176, 0.89911]$, but $P(H|E_1 \wedge E_2 \wedge E_3) \in [0.999884, 0.999909]$, so that the width of the first interval is 0.000934 and that of the second interval is 0.000025 .[3]

This does by no means imply that additional information makes the situation necessarily better. In many cases the interval does get wider if the number of premises increases. If, for instance, identical probabilities $\beta_i = \beta$ and $\gamma_i = \gamma$ for $i = 1, \ldots, n$, are assumed, then Bayes' Theorem strongly degrades. This case is of main importance because it is implied by the assumption of conditional exchangeability. Moreover, Theorem 6 and Theorem 7 show that even in the case of Bayes' Theorem one ends up with the unit interval. If $n \geq \max\{\sum_{i=1}^{n} \beta_i + 1, \sum_{i=1}^{n} \gamma_i + 1\}$, then $\sum_{i=1}^{n} \beta_i - (n-1) \leq 0$ and $\sum_{i=1}^{n} \gamma_i - (n-1) \leq 0$, so that the interval $[0, 1]$ is obtained.

We therefore claim that Bayesian updating, i. e., the new probability of a hypothesis H after observing $E_1, E_2, \ldots, E_n$ should be measured by $P(H|E_1 \wedge E_2 \wedge \ldots \wedge E_n)$, is accurate only under very restricted assumptions. One such assumption is conditional independence, leading to a point probability of the likelihood $P(E_1 \wedge \ldots \wedge E_n|H)$.

2.7. Modus Tollens

The following holds for the probabilistic Modus Tollens of two events as we show below (Theorem 8). If $P(\neg E_1) = \alpha_1$ and $P(E_1|H) = \beta$, then $P(\neg H) \in [\delta', 1]$, where

$$\delta' = \max\left\{1 - \frac{\alpha_1}{1-\beta}, 1 - \frac{1-\alpha_1}{\beta}\right\} \quad . \tag{2.4}$$

Wagner [11] has shown the result for the lower bound. However, Wagner's upper bound is different from 1. The reason for this is that Wagner defined the conditional probability $P(E_1|H)$ by the fraction $\frac{P(E_1 \wedge H)}{P(H)}$. If $P(\neg H) = 1$, then $P(H) = 0$ and $P(E_1|H)$ would consequently be undefined. As already pointed out, in the coherence approach conditionalizing on events with zero probability is possible, so that the "correct" upper bound $P(\neg H) = 1$ is obtained.

[2] All values are rounded.
[3] All values are rounded.

The result for the generalized Modus Tollens has been presented without proof in [7, 12].

Theorem 8. (Modus Tollens) If $P(\neg E_i) = \alpha_i$, for $i = 1, 2, \dots, n$, and if $P(E_1 \wedge E_2 \wedge \dots \wedge E_n | H) = \beta$, then $P(\neg H) \in [\delta', 1]$, with

$$\delta' = \begin{cases} 1 - \frac{1-\alpha^*}{\beta} & \text{if} \quad \alpha^* + \beta > 1, \\ 1 - \frac{\sum_{i=1}^n \alpha_i}{1-\beta} & \text{if} \quad \sum_{i=1}^n \alpha_i + \beta < 1, \\ 0 & \text{if} \quad \sum_{i=1}^n \alpha_i + \beta \geq 1 \text{ and } \alpha^* + \beta \leq 1, \end{cases}$$

where $\alpha^* = \max\{\alpha_i\}$.

Proof. First, we treat the case $n = 1$. Then we use this result and the conjunction rule (Theorem 2) to prove the general case $n > 1$.

One event. If $n = 1$, we employ the following notation

	C_1	C_2	C_3	C_4	Probability
H	1	1	0	0	$P(H) = x_1 + x_2$
E_1	1	0	1	0	$P(E_1) = x_1 + x_3 = 1 - \alpha_1$

and obtain the linear system

$$\beta(x_1 + x_2) = x_1 \quad (2.5)$$

$$x_2 + x_4 = \alpha_1 \quad (2.6)$$

$$x_1 + x_2 + x_3 + x_4 = 1, \qquad x_i \geq 0 \ .$$

To maximize (resp. minimize) $P(\neg H)$, we minimize (resp. maximize)

$$P(H) = x_1 + x_2 \ .$$

Manipulation of (2.5) shows that if $\beta > 0$, then

$$P(H) = \frac{x_1}{\beta} \ . \quad (2.7)$$

If $\beta = 0$, then the solvability of the above linear system requires that $P(H) = x_2 \in [0, \alpha_1]$ and therefore $P(\neg H) \in [1 - \alpha_1, 1]$.
Because of (2.7), to maximize (minimize) $P(H)$ in case of $\beta > 0$, we maximize (minimize) $x_1 = P(H \wedge E_1)$.

Upper bound of $P(\neg H)$. The minimum $x_1 = 0$ is obtained by setting $x_4 = \alpha_1$.

Lower bound of $P(\neg H)$. For the maximum of x_1 observe that $x_1 \leq 1 - \alpha_1$. Furthermore, since $P(H \wedge E_1) \leq P(E_1 | H)$, we have $x_1 \leq \beta$. Therefore, $x_1 \leq \min\{1 - \alpha_1, \beta\}$ and we distinguish two cases: (I) $1 - \alpha_1 \leq \beta$ and (II) $1 - \alpha_1 > \beta$.

In case (I), we set x_1 to its maximum $1-\alpha_1$. By (2.7) we obtain $P(H) = \frac{1-\alpha_1}{\beta}$, so that the minimum of $P(\neg H)$ is $1 - \frac{1-\alpha_1}{\beta}$. It is straightforward to establish that the solution $x_1 = 1-\alpha_1$ is admissible.
In case (II), we cannot set x_1 to its maximum β. Otherwise, we have $x_1 + x_3 + x_4 = \beta + \alpha_1 > 1$. We employ that by (2.5), it is $x_1 = \frac{x_2\beta}{1-\beta}$. Hence, x_1 is maximized if x_2 is maximized. This is the case if $x_2 = \alpha_1$ (and hence $x_4 = 0$), so that $x_1 = \frac{\alpha_1\beta}{1-\beta}$. Consequently

$$x_3 = P(\neg H) = 1 - (x_1 + x_2) = 1 - \frac{\frac{(\alpha_1)\beta}{1-\beta}}{\beta} = 1 - \frac{\alpha_1}{1-\beta} \quad .$$

n events. Employing the case $n = 1$, writing $E := \bigwedge_{i=1}^{n} E_i$, we obtain that if $P(\neg E) = \alpha$ and if $P(E|H) = \beta$, then $P(\neg H) \in [\delta', 1]$, where

$$\delta' = \begin{cases} 1 - \frac{1-\alpha}{\beta} & \text{if} \quad \alpha + \beta > 1, \\ 1 - \frac{\alpha}{1-\beta} & \text{if} \quad \alpha + \beta < 1, \\ 0 & \text{if} \quad \alpha + \beta = 1 \ . \end{cases} \tag{2.8}$$

Applying the disjunction rule (Theorem 3) to $P(\neg E_i) = \alpha_i$, yields for $\alpha = P(\neg E) = P(\bigvee_{i=1}^{n} \neg E_i)$

$$\alpha \in \left[\alpha^*, \min\left\{1, \sum_{i=1}^{n} \alpha_i\right\}\right] \quad . \tag{2.9}$$

Applying (2.9) to (2.8) and distinguishing cases proves the result:

Case 1: If $\alpha^* + \beta > 1$, then $\alpha + \beta > 1$. According to (2.8), since $1 - \frac{1-\alpha}{\beta}$ is monotonically increasing with α, the minimum is $1 - \frac{1-\alpha^*}{\beta}$.
Case 2: If $\alpha^* + \beta \geq 1$, we distinguish two cases.
(2.1): If $\sum_{i=1}^{n} \alpha_i + \beta < 1$, then since $\alpha \leq \sum_{i=1}^{n} \alpha_i$, we have $\alpha + \beta < 1$. Since $1 - \frac{\alpha}{1-\beta}$ is monotonically decreasing in α, $1 - \frac{\sum_{i=1}^{n} \alpha_i}{1-\beta}$ is the minimum.
(2.2): If $\sum_{i=1}^{n} \alpha_i + \beta > 1$, then $1-\beta \in [\alpha^*, \min\{1, \sum_{i=1}^{n} \alpha_i\}]$. Thus, by setting $\alpha = 1-\beta$, from (2.8), we obtain the lower bound zero. □

Modus Tollens has very interesting properties with respect to degradation. Suppose that $P(\neg E_{n+1})$ is added to the premises and $P(E_1 \wedge E_2 \wedge \ldots \wedge E_n|H) = \beta$ is replaced by $P(E_1 \wedge E_2 \wedge \ldots \wedge E_{n+1}|H) = \gamma$. While the upper bound 1 for $P(\neg H)$ is already most "degraded", the lower bound does not decrease as the number of premises n increases. Depending on the values of γ and α^*, we jump back and forth between the cases (a) $\alpha^* + \gamma > 1$ and (b) $\alpha^* + \gamma \leq 1$. In case (b), since $\sum_{i=1}^{n} \alpha_i$ increases as n increases, the lower bound 0 is obtained rapidly. In case (a), the lower bound strongly depends on the values of γ and α^*. As a consequence, it can attain any value $c \in (0, 1]$. If $\alpha^* < 1$, then $P(\neg H) \in [c, 1]$ if $\beta = \frac{1-\alpha^*}{1-c}$. Consequently, Modus Tollens does not degrade. Moreover, contrary to the other inference forms considered in this paper, the unit interval is not necessarily obtained if the number of premises is large.

3. DEGRADATION IS NOT NON-MONOTONICITY

We have seen that the generalized inference forms which correspond to the most important inference forms in classical logic degrade. "More" premises lead to wider intervals of the conclusion. However, the expression 'more' is metaphoric, since we either *replaced* in all the inference forms considered one of the premises by another one (Modus Ponens, Modus Tollens), or we changed the conclusion (Conjunction-Rule, Predictive Inference, Bayes' Theorem). It might be supposed that if we do not replace premises, but keep and *cumulate* all available premises, the degradation disappears. Indeed, this is a consequence of the fact that the transition from the probability of the premises to the probability of the conclusion is *monotonic*. If a set of premises T already establishes that $P(A) \geq \alpha$ ($P(A) \leq \alpha$), then no additional premises, i. e. restrictions, $T' \supset T$ can yield a smaller (greater) value than α. Although degradation disappears, we are often faced with an even more serious problem then. Paradoxically, if we take into account all premises, most of them are irrelevant. Consider, for instance, generalized Modus Ponens.

Example 4. Let $T_1 = \{P(E_1) = 0.7,\ P(H|E_1) = 0.8\}$ and
$T_2 = \{P(E_1) = 0.7,\ P(E_2) = 0.8,\ P(H|E_1 \wedge E_2) = 0.8\}$.

- From T_1 it follows that $P(H) \in [0.56, 0.86]$.
- From T_2 it follows that $P(H) \in [0.4, 0.9]$.
- From $T_1 \cup T_2 = \{P(E_1) = 0.7, P(E_2) = 0.8, P(H|E_1) = 0.8,\ P(H|E_1 \wedge E_2) = 0.8\}$ it follows that $P(H) \in [0.56, 0.86]$.

The interval obtained by the union $T_1 \cup T_2$ is the same interval as obtained before by T_1. The premises of T_2 are consequently irrelevant in $T_1 \cup T_2$.

Modus Ponens degrades. The interval of $P(H)$ obtained by T_2 $[0.4, 0.9]$ is wider than the interval $[0.56, 0.86]$ obtained by T_1. We add $P(E_2) = 0.8$ to the premises. However, in addition, we replace $P(H|E_1) = 0.8$ in T_1 by $P(H|E_1 \wedge E_2) = 0.8$ in T_2. Thus, it is not the case that $T_1 \subseteq T_2$ and there is no violation of monotonicity.

If we do not replace $P(H|E_1) = 0.8$ by $P(H|E_1 \wedge E_2) = 0.8$, but employ both premises, the degradation disappears. This follows from the fact that the transition from the probability of the premises to the probability of the conclusion is *monotonic*. Since, if T_1 already establishes that the probability $P(H)$ is at least 0.56, then no additional premises, i. e. restrictions, can yield a smaller value for $P(H)$. Equally, $P(H)$ cannot exceed 0.86.

However, if we take into account all premises, two of them are irrelevant. To derive the interval of the conclusion, we can only make use of the premises of the first set T_1. This is equally true for all the inference forms considered in this paper that strongly degrade. If the number of premises is high, not only two but most of them are irrelevant. In the example, the combination of $P(E_2)$ and $P(H|E_1 \wedge E_2)$ with $P(H|E_1)$ does not change the interval $[0.56, 0.86]$. The information given by T_2, i. e., $P(E_2) = 0.8$ and $P(H|E_1 \wedge E_2) = 0.8$ is consequently irrelevant.

Consider, as a second example, the combination of two Modi Ponentes [7]. For the probability interval of the conclusion, we have to take two times the "better" value, i. e., the greater lower bound and the smaller upper bound. The example generalizes to every inference form considered in this contribution that weakly degrades.

Example 5. If from the assessment $P(H|E), P(E)$ it follows $P(H) \in [\alpha_1, \alpha_2]$, and if from the assessment $P(H|F), P(F)$ it follows $P(H) \in [\beta_1, \beta_2]$, and if in addition $[\alpha_1, \alpha_2] \cap [\beta_1, \beta_2] \neq \emptyset$ [4], then it follows from the joint assessment $P(H|E), P(E), P(H|F), P(F)$ that $P(H) \in [\max\{\alpha_1, \beta_1\}, \min\{\alpha_2, \beta_2\}]$.

We would expect, however, that the lower bound (upper bound) is a function of the two lower (upper) bounds which takes into account their positions and their distance. It is, for example, possible to take their mean. Equally, we would expect that in Example 4 the fact that $E_1 \wedge E_2$ is the more specific information than E_1 yields a lower (upper) bound of $P(H)$, derived by $T_1 \cup T_2$, that is closer to, or is even identical with, 0.4 (0.9). Probability logic, however, is often *insensitive* with respect to both, to the specificity of information and to relations between intervals. It is therefore very weak. Simply cumulating evidence, i. e., considering the union of all premise sets, and applying probability logic doesn't leave us with the desired results, since we often have to ignore most of the available information. We are therefore forced to make a *preselection* and replace some of the premises by others, leading to degradation.

4. DISCUSSION

In probability theory and consequently in probability logic, if point probabilities of the conjuncts are given, then only an interval probability for their conjunction can be inferred. If we do not have information about the dependencies between the conjuncts, this interval probability is getting wider as the number of conjuncts increases (compare Section 2.2). As a consequence, many generalized inference forms degrade. We have seen that Predictive Inference strongly degrades, Modus Ponens weakly degrades, and that Bayes' Theorem and Modus Tollens do not degrade. However, even Bayes' Theorem and Modus Tollens tend to degrade. Moreover, in all the inference forms considered — with the exception of Modus Tollens — the unit interval is obtained if the number of premises is sufficiently large. A narrower interval might be considered better than a wider interval and a more complete knowledge might be considered better than a truncated one [8]. In probability logic the number of premises and the precision of the conclusion often *must* conflict.

This conflict between amount of information and precision is persistent and hard-wired. On the one hand, the *principle of total evidence*, i. e., selecting the most "recent" interval obtained by the most specific information, leads to wide intervals. In many cases it even leads to the non-informative interval [0,1]. On the other hand, a *take-the-best-strategy*, i. e., selecting the most precise interval, requires to base the interval of the conclusion on the most unspecific information ($n = 1$). Since all additional premises are discarded, it would be useless to apply probability logic to generalized inference forms. It thereby leads to *overconfidence*, i. e., it suggests precision where no precision would be if we considered all available information. We have seen in Section 3 that the third and maybe most natural way to solve the conflict doesn't work either. Simply cumulating

[4] If $[\alpha_1, \alpha_2] \cap [\beta_1, \beta_2] = \emptyset$, then the joint assessment is incoherent.

evidence, i. e., considering the union of all premise sets, and letting probability logic decide the matter, forces us in many cases to ignore most of the available premises.

Recently, probability logic was used to model human reasoning (see, for example, [10]). The properties of degradation and monotonicity (see Section 3) are of special interest from the perspective of reasoning and judgment under uncertainty. They demonstrate that in some situations less information may be preferred to more information. Degradation shows that adding premises may weaken an inference. Monotonicity shows that additional premises are often irrelevant. Degradation and monotonicity may act as Ockham's razor [7]. In uncertain reasoning and in decision making it may often be rational to keep the number of premises small.

Although additional premises often yield more imprecise intervals, they do not always make inference worse. First, they prevent overconfidence. Second, contrary to strong degradation, in the case of weak degradation obtaining intervals with different positions to a certain degree compensates for obtaining wider intervals (as, for instance, in Example 2). Since the new position is based on more information, it is more "recent" than the old position. The knowledge of the position of the interval is of main importance for decision making, so that it is not reasonable to discard the new information. Solving the conflict between precision and specificity requires to counterbalance (i) the width of an interval, (ii) the amount of information it is based upon, and (iii) the position of the interval.

However, probability logic is insensitive with regard to these respects and therefore too weak for the practical needs of decision making. Whether a take-the-best strategy is rational, which information should be considered, or how to counterbalance (i), (ii), and (iii) are questions of subjective preferences and pragmatic conditions of *decision making*. Although raised by degradation and monotonicity (compare Section 3), they cannot be answered by the formal results of probability logic, because probability logic is too weak. Probability logic is not decision theory. The basic aspect of probability logic is to determine the coherent extensions of a given initial assessment. How this can be done is stated in the Fundamental Theorem of de Finetti: "Given the probabilities $P(E_i)$ $(i = 1, 2, \dots, n)$ of a finite number of events, the probability, $P(E)$ of a further event E, either (a) turns out to be determined (whatever P is) if E is linearly dependent on the E_i, or (b) can be assigned, coherently, any value in a closed interval $p' \leq P(E) \leq p''$ (which can often give a illusory restriction, if $p' = 0$ and $p'' = 1$)." ([2], p. 112). Contrary to decision theory, the only thing that matters in probability logic, is the principle of total evidence. I. J. Good, for instance, claims that the principle of total evidence follows from the principle of rationality [6].

Moreover, the language of probability logic (probabilities defined on formulas of propositional logic) and its tools are too parsimonious. The logical inference forms studied should be supplemented by, for example, assumptions about the probabilistic dependencies (or independencies) of the events. More complex — and often more realistic — problems of probabilistic inference are studied in statistics. Here, problems are embedded in statistical models. Typically, such models specify the structure of a data generating process, they make assumptions about probability distributions, parameters, dependencies, homogeneity of variances, prior distributions etc.

We might also strenghten probability logic by stating the dependencies and correlations between the events considered. Indeed, the assumption of stochastic independence often weakens the degradation. However, independence is a strong and often not justified assumption. We have investigated the more realistic assumption of exchangeability. However, even exchangeability does not prevent degradation [12]. Finally, we remark that degradation is not restricted to inference forms involving conjunctions. The System P rule Or [4] as well as the disjunction of n events (see Section 2.3) degrade.

ACKNOWLEDGEMENT

This work is supported by the Austrian Science Foundation (FWF, I 141-G15) within the LogICCC Programme of the European Science Foundation. We thank two anonymous reviewers for their detailed and valuable comments.

(Received February 28, 2013)

REFERENCES

[1] G. Coletti and R. Scozzafava: Probabilistic Logic in a Coherent Setting. Kluwer, Dordrecht 2002.

[2] B. De Finetti: Theory of Probability. A Critical Introductory Treatment. Volume 1. Wiley, New York 1974.

[3] M. Fréchet: Généralisations du théorème des probabilités totales. Fund. Math. *25*5 (1935), 379–387.

[4] A. Gilio: Generalization of inference rules in coherence-based probabilistic default reasoning. Internat. J. Approx. Reason. *53* (2012), 413–434.

[5] A. Gilio and G. Sanfilippo: Conditional random quantities and iterated conditioning in the setting of coherence. In: Symbolic and Quantitative Approaches to Reasoning with Uncertainty (L. van der Gaag, ed.), Lecture Notes in Comput. Sci. *7958*, Springer (2013), pp. 218–229.

[6] I. J. Good: On the principle of total evidence. British J. Philos. Sci. *17* (1967), 319–321.

[7] G. D. Kleiter: Ockham's razor in probability logic. In: Synergies of Soft Computing and Statistics for Intelligent Data Analysis (R. Kruse, M. R. Berthold, C. Moewes, M. A. Gil, P. Grzegorzewski and O. Hryniewicz, eds.). Adv. in Intelligent Systems and Computation *190*, Springer (2012), pp. 409–417.

[8] H. E. Kyburg and C. M. Teng: Uncertain Inference. Cambridge University Press, Cambridge 2001.

[9] F. Lad: Operational Subjective Statistical Methods. Wiley, New York 1996.

[10] K. I. Manktelow, D. E. Over, and S. Elqayam, eds.: The Science of Reasoning. A Festschrift for Jonathan St B.T. Evans. Psychology Press, New York 2011.

[11] C. G. Wagner: Modus tollens probabilized. British J. Philos. Sci. *55* (2004), 4, 747–753.

[12] C. Wallmann and G. D. Kleiter: Exchangeability in probability logic. In: IPMU 2012, Part IV (S. Greco et al., eds.), CCIS *300* (2012), pp. 157–167.

[13] C. Wallmann and G. D. Kleiter: Probability propagation in generalized inference forms. Studia Logica. In press. Doi= 10.1007/s11225-013-9513-4.

Christian Wallmann, University of Salzburg, Department of Philosophy, Franziskanergasse 1. Austria.

e-mail: christian.wallmann@stud.sbg.ac.at

Gernot D. Kleiter, University of Salzburg, Department of Psychology, Hellbrunnerstr. 34. Austria.

e-mail: gernot.kleiter@sbg.ac.at