Available online at www.sciencedirect.com
SCIENCE DIRECT®
Engineering Applications of Artificial Intelligence 17 (2004) 345–361
Engineering Applications of
ARTIFICIAL INTELLIGENCE
www.elsevier.com/locate/engappai

Can Bayesian confirmation measures be useful for rough set decision rules?

Salvatore Greco[a,*], Zdzisław Pawlak[b], Roman Słowiński[c]

[a] *Faculty of Economics, University of Catania, Corso Italia, 55, 95129 Catania, Italy*
[b] *Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, and Warsaw School of Information Technology, Newelska 6, 01-447 Warsaw, Poland*
[c] *Institute of Computing Science, Poznań University of Technology, 60-965 Poznań, and Institute for Systems Research, Polish Academy of Sciences, 01-447 Warsaw, Poland*

Abstract

Bayesian confirmation theory considers a variety of non-equivalent confirmation measures which say in what degree a piece of evidence confirms a hypothesis. In this paper, we apply some well-known confirmation measures within the rough set approach to discovering relationships in data in terms of decision rules. Moreover, we discuss some interesting properties of these confirmation measures and we propose a new property of monotonicity that is particularly relevant within rough set approach. The main result of this paper states that only two from among confirmation measures considered in the literature have the desirable properties from the viewpoint of the rough set approach. Moreover, we clarify relationships between logical implications and decision rules, and we compare the confirmation measures to several related measures, like independence (dependence) of logical formulas, interestingness measures in data mining and Bayesian solutions of raven's paradox.

Keywords: Confirmation measures; Bayes' theorem; Rough sets; Decision rules; Decision algorithm; Monotonicity property

1. Introduction

Reasoning from data is the domain of inductive reasoning. Contrary to deductive reasoning, where axioms expressing some universal truths constitute a starting point of reasoning, inductive reasoning uses data about a sample of larger reality to start inference.

Rough set theory (Pawlak, 1991) is a mathematical approach to data analysis. Rough-set-based data analysis starts from a data table, called *information table*. The information table contains data about objects of interest, characterized by a finite set of attributes. It is often interesting to discover some dependency relationships (patterns) in the information table. With this aim, a set of *condition attributes C* and a set of *decision attributes D* are distinguished, in order to analyze how values of attributes from C associate with values of attributes from D. An information table where condition attributes and decision attributes are distinguished is called *decision table*. From a decision table one can induce some patterns in form of "*if* …, *then* …" decision rules. More exactly, the decision rules say that if some condition attributes have given values, then some decision attributes have other given values. For example, in a data table collecting medical information on a sample of patients, we can consider as condition attributes a set of symptoms $S=\{s_1,\dots,s_n\}$, and as decision attributes, a set of diseases $D=\{d_1,\dots,d_m\}$. In the decision table so obtained we can induce decision rules of the form "*if* symptoms $s_{i1},s_{i2},\dots,s_{ih}$ appear, *then* there is disease d_j", with $s_{i1},s_{i2},\dots,s_{ih}\in S$ and $d_j\in D$.

With every decision rule induced from a decision table, three coefficients are traditionally associated: the *strength*, the *certainty factor* and the *coverage factor* of the rule. For example, the decision rule

"*if* symptoms $s_{i1},s_{i2},\dots,s_{ih}$ appear, *then*
there is disease d_j",

*Corresponding author.
E-mail addresses: salgreco@mbox.unict.it (S. Greco), zpw@ii.pw.edu.pl (Z. Pawlak), slowinsk@sol.put.poznan.pl (R. Słowiński).

0952-1976/$ - see front matter
doi:10.1016/j.engappai.2004.04.008

can be characterized as follows (the numbers of % are calculated from a hypothetical data table):

- the patients having symptoms $s_{i1}, s_{i2}, \ldots, s_{ih}$ and disease d_j constitute 15% of all the patients in the sample: in this case, 15% is the *strength* of the rule,
- 91% of the patients having symptoms $s_{i1}, s_{i2}, \ldots, s_{ih}$ have also disease d_j: in this case, 91% is the *certainty factor* of the rule,
- 52% of the patients having disease d_j have also symptoms $s_{i1}, s_{i2}, \ldots, s_{ih}$: in this case, 52% is the *coverage factor* of the rule.

These characteristics are useful to show that discovering patterns in data can be represented in terms of Bayes' theorem (Pawlak, 2002; Greco et al., 2002) in a different way from that offered by standard Bayesian inference techniques, without referring to prior and posterior probabilities, inherently associated with Bayesian inference methodology.

Within inductive reasoning, classical Bayesian theory considers a variety of non-equivalent confirmation measures (see Fitelson (2001) and Kyburg (1983) for surveys) which quantify the degree to which a piece of evidence E provides, "evidence for or against" or "support for or against" a hypothesis H. In this paper, we take into account some of the most relevant of these confirmation measures and apply them within rough set approach to data analysis. Moreover, we discuss some interesting properties of these confirmation measures, which are particularly relevant within rough set approach.

Let us stress how important is the discussion of such a philosophical question related to epistemology within rough set approach and, more generally, within data mining, machine learning and knowledge discovery. Traditionally, the development of a new theory requested that, first, a hypothesis was formulated, and after, confirmation or disconfirmation of the hypothesis was looked for in the data. Nowadays, due to adoption of powerful computer-aided automated processing of a huge amount of data, the order of this process can be reversed: first, some data are collected, and then, practically all possible hypotheses are considered trying to screen the most interesting ones. For example, association rules (Agrawall et al., 1996) in data mining and "*if* …, *then* …" rules within rough set approach (Pawlak, 1982, 1991) can be interpreted as theories induced from data.

In this context, establishing a reliable index, able to discriminate the most interesting hypotheses discovered by induction from an automated data processing, is of fundamental importance for knowledge discovery, data mining and machine learning (see, for example, Fayyad et al., 1996; Michalski et al., 1998; Hajek and Havranek, 1978). Thus, even if the origin of the problem is quite theoretical, the research results have in this context a strong impact on real world operational applications.

Indeed, our research is strongly related to the rich discussion about interestingness measures for decision rules in data mining (see, for example, Hilderman and Hamilton (2001) and Yao and Zhong (1999) for exhaustive reviews of the subject). We shall see, for example, that one of the confirmation measures that we consider has been already proposed in literature as interestingness measure. Until now, however, up to our knowledge, there has not been any discussion about the possibility of using the confirmation measures as interestingness measures. Such a discussion is important from two points of view:

- it permits a systematic construction and analysis of a large class of interestingness measures, i.e. all the interestingness measures which can be expressed as confirmation measures;
- it permits to propose a quantitative confirmation theory for data analysis which brings to rough set approach, knowledge discovery, data mining and machine learning, the results obtained in an important sector of the epistemological debate by researchers as prominent as Carnap (1962), Hempel (1945) and Popper (1959); let us remark that while there were some proposals of a qualitative confirmation theory for knowledge discovery (see, for example, Flach, 1995), no similar proposal has been made with respect to a quantitative confirmation theory.

For reasons raised in the latter point, our research may also bring some interesting results into philosophical debate about confirmation. In fact, quantitative confirmation theory is strongly based on probability functions. However, there is a great and well-known controversy relative to interpretation, origin and status of probability (see, for example, Kyburg, 1970). Consequently, in this paper, we are considering quantitative confirmation based on observed data only, without any consideration of probability functions.

Some confirmation measures considered in this paper may remember statistical independence tests of a contingency table. Indeed, some interestingness measures of decision rules, which are based on these statistical tests, have been proposed in the specialized literature (see, for example, Flach and Lachiche, 2001; Tsumoto, 2002; Zembowicz and Zytkow, 1996). It is worth stressing that our confirmation measures take a different perspective than the statistical approach. First, observe that the independence (dependence) measures are symmetric while decision rules, for which these measures are conceived, are not symmetric. The following example can explain the point. Let us consider variable X and Y. Let us suppose that each of them can take one of two values: 0 or 1. The decision rule "*if* a card is the seven of spades, *then* the card is black" is, of course, quite different from the rule "*if* a card is black, *then* the card is the seven of spades", because the fact that one rule is true does not necessarily mean that the other is true (Eells and Fitelson, 2002). However, from the viewpoint of statistical independence we will obtain the same measure of independence, regardless of the fact which rule is true. Even if some authors tried to generalize classical statistical analysis of a contingency table in order to handle typical asymmetries of rule induction (Flach and Lachiche, 2001), our approach is different in nature because we are interested in some desirable properties of confirmation measures rather than in their statistical properties.

Let us also remark that the concept of confirmation we are interested in is related to the concept of independence of logical formulas (propositions), as presented by Łukasiewicz (1913). In brief, his definition of independence between two propositions Φ and Ψ amounts to say that the credibility of Ψ given Φ is the same as the credibility of Ψ given $\neg\Phi$. Thus, independence means that the credibility of Φ does not influence the credibility of Ψ. For this definition Łukasiewicz proved the law of multiplication which says that if propositions Φ and Ψ are independent, then the credibility of Ψ given Φ is equal to the product of the individual credibilities of Φ and Ψ. From this law, Pawlak (2003) derived a dependency factor for flows in decision networks and then he applied this formula to decision rules (Pawlak, 2004). From the viewpoint of confirmation, we can say that Φ confirms Ψ if the credibility of Ψ is higher when Φ is true rather than when Φ is false. The difference between the dependency factors derived from the concept of independence proposed by Łukasiewicz and the measures of confirmation we are studying is twofold:

(1) The original concept of independence proposed by Łukasiewicz is qualitative (propositions are independent or not) and do not imply any quantitative measure for the degree of independence (or dependence) between propositions.
(2) The concept of dependence, as presented above, is directional because it is used to check if the credibility of one proposition, Ψ, is affected by the truth or the falsity of another proposition Φ. This is concordant with the concept of confirmation: in general, if evidence Φ confirms hypothesis Ψ, then evidence Ψ does not confirm to the same extent hypothesis Φ. For example, the evidence $\Phi =$"x is a square" confirms conclusively hypothesis $\Psi =$"x is a rectangle" (of course, all squares are rectangles), but the evidence Ψ does not confirm conclusively Φ (in fact, not all rectangles are squares). However, from other parts of the paper by Łukasiewicz (1913), it appears that the concept of dependence he has in mind is substantially non-directional. In fact, he proves that if the credibility of Ψ given Φ is the same as the credibility of Ψ given Φ, then also the credibility of Φ given Ψ is the same as the credibility of Φ given Ψ. The question of the relation between dependence and confirmation measures has recently been raised by Fitelson (2003) who recognized the directionality of the concept of confirmation and the non-directionality of the concept of dependence. For this reason, Fitelson proposed as dependence measure for propositions Φ and Ψ, an average measure between the confirmation of Φ to Ψ and the confirmation of Ψ to Φ.

The article is organized as follows. Section 2 introduces confirmation measures and recalls some desirable properties of symmetry and asymmetry proposed by Eells and Fitelson. Section 3 gives some basic notions concerning decision rules and decision algorithms within rough set approach. Section 4 introduces rough set confirmation measures. Section 5, which deals with some fundamental contributions of Hempel to logics of confirmation, presents the Nicod's criterion, the equivalence condition and the positive instance criterion. In Section 6, we introduce a specific monotonicity property of rough set confirmation measures. This property is extensively discussed and compared to Nicod's criterion, as well as to the positive instance criterion. Section 7 investigates which one among the considered rough set confirmation measures satisfies the monotonicity property. Finally Section 8 presents conclusions and some possible directions of future research.

2. Confirmation measures

According to Fitelson (2001), measures of confirmation quantify the degree to which a piece of evidence E provides, "evidence for or against" or "support for or against" a hypothesis H. Fitelson remarks, moreover, that measures of confirmation are supposed to capture the impact rather than the final result of the "absorption" of a piece of evidence.

Bayesian confirmation assume the existence of a probability Pr. In the following, given a proposition X, $\Pr(X)$ is the probability of X. Given X and Y, $\Pr(X|Y)$ represents the probability of X given Y, i.e.

$$\Pr(X|Y) = \frac{\Pr(X \wedge Y)}{\Pr(Y)}.$$

In this context, a measure of confirmation of a piece of evidence E with respect to a hypothesis H is denoted by $c(E,H)$. $c(E,H)$ is required to satisfy the following minimal property:

$$c(E,H) = \begin{cases} >0 & \text{if } \Pr(H|E) > \Pr(H), \\ =0 & \text{if } \Pr(H|E) = \Pr(H), \\ <0 & \text{if } \Pr(H|E) < \Pr(H). \end{cases}$$

The most well-known confirmation measures proposed in the literature are the following:

$$d(E,H) = \Pr(H|E) - \Pr(H),$$

$$r(E,H) = \log\left[\frac{\Pr(H|E)}{\Pr(H)}\right],$$

$$l(E,H) = \log\left[\frac{\Pr(E|H)}{\Pr(E|\neg H)}\right],$$

$$f(E,H) = \frac{\Pr(E|H) - \Pr(E|\neg H)}{\Pr(E|H) + \Pr(E|\neg H)},$$

$$s(E,H) = \Pr(H|E) - \Pr(H|\neg E),$$

$$b(E,H) = \Pr(H \wedge E) - \Pr(H)\Pr(E).$$

Measure $d(E,H)$ has been supported by Earman (1992), Eells (1982), Gillies (1986), Jeffrey (1992) and Rosenkrantz (1994). Measure $r(E,H)$ has been defended by Horwich (1982), Keynes (1921), Mackie (1969), Milne (1995, 1996), Schlesinger (1995) and Pollard (1999). Measure $l(E,H)$ and $f(E,H)$ have been supported by Kemeny and Oppenheim (1952), Good (1984), Heckerman (1988), Horvitz and Heckerman (1986), Pearl (1988) and Schum (1994). Fitelson (2001) has advocated for measure $f(E,H)$. Measure $s(E,H)$ has been proposed by Christensen (1999) and Joyce (1999). Measure $b(E,H)$ has been introduced by Carnap (1962).

Many authors have considered, moreover, some more or less desirable properties of confirmation measures. Fitelson (2001) makes a comprehensive survey of these considerations. At the end of his retrospective, Fitelson concludes that the most convincing confirmation measures are $l(E,H)$ and $f(E,H)$. He also proves that $l(E,H)$ and $f(E,H)$ are ordinally equivalent, i.e. for all E, H and E', H',

$$l(E,H) \geqslant l(E',H') \quad \text{if and only if} \quad f(E,H) \geqslant f(E',H').$$

Among the properties of confirmation measures reviewed by Fitelson (2001), there are properties of symmetry introduced by Carnap (1962) and investigated recently by Eells and Fitelson (2000). For all E and H, one can have:

- evidence symmetry (ES): $c(E,H) = -c(\neg E,H)$
- commutativity symmetry (CS): $c(E,H) = c(H,E)$
- hypothesis symmetry (HS): $c(E,H) = -c(E,\neg H)$
- total symmetry (TS): $c(E,H) = c(\neg E,\neg H)$.

Eells and Fitelson (2002) remarked that, given (CS), (ES) and (HS) are equivalent, and that (TS) follows from the conjunction of (ES) and (HS). Moreover, they advocate in favor of (HS) and against (ES), (CS) and (TS). The reason in favor of (HS) is that the significance of E with respect to H should be of the same strength, but of opposite sign, as the significance of E with respect to $\neg H$. The arguments against (ES), (CS) and (TS) can be explained by the following example. A card is randomly drawn from a standard deck. Let E be the evidence that the card is the seven of spades, and let H be the hypothesis that the card is black. Clearly, E is strong and conclusive evidence in favor of H. However, the confirmation of H by E is not corresponding to the negative confirmation of H by $\neg E$. Indeed, the evidence "the card is not the seven of spade" ($\neg E$) is practically useless with respect to the hypothesis "the card is black" (H), which means that (ES) is not valid. Using the same example, we can notice that the evidence "the card is black" does not confirm the hypothesis "the card is the seven of spade", to the same extent as the evidence "the card is the seven of spade" confirms the hypothesis "the card is black". This means that (CS) does not hold. Similarly, one can show that also (TS) does not hold.

Eells and Fitelson (2002) prove that

(1) s and b satisfy (ES), while d,r,l and f do not satisfy (ES),
(2) d,s,b,f and l satisfy (HS), while r does not satisfy (HS),
(3) r and b satisfy (CS), while d,s,f and l do not satisfy (CS),
(4) s and b satisfy (TS), while d,r,f and l do not satisfy (CS).

Thus, assuming that (HS) is a desirable property, while (ES), (CS) and (TS) are not, Eells and Fitelson (2002) conclude that with respect to the property of symmetry, d,f and l are satisfying confirmation measures while s, r and b are not satisfying confirmation measures.

3. Decision rules and decision algorithm

Let $S = (U,A)$ be an information table, where U and A are finite, non-empty sets called the *universe* and the set of *attributes*, respectively. If in the set A two disjoint subsets of attributes, called *condition* and *decision attributes*, are distinguished, then the system is called a *decision table* and is denoted by $S = (U,C,D)$, where C and D are sets of condition and decision attributes, respectively. With every subset of attributes, one can associate a formal language of logical formulas L defined in a standard way and called the *decision language*. Formulas for a subset $B \subseteq A$ are build up from attribute-value pairs (a,v), where $a \in B$ and $v \in V_a$ (set V_a is a domain of a), by means of logical connectives $\wedge$ (*and*), $\vee$ (*or*), $\neg$ (*not*). We assume that the set of all formula sets in L is partitioned into two classes, called *condition* and *decision formulas*, respectively.

A *decision rule* induced from S and expressed in L is presented as $\Phi \to \Psi$, read "*if* Φ, *then* Ψ", where Φ and Ψ are condition and decision formulas in L, called *premise* and *conclusion*, respectively. A decision rule $\Phi \to \Psi$ is also seen as a binary relation between premise and conclusion, called *consequence relation* (see critical discussion about interpretation of decision rules as logical implications at the end of Section 6).

Let $\|\Phi\|$ denote the set of all objects from universe U, having the property Φ in S.

If $\Phi \to \Psi$ is a decision rule, then $\mathrm{supp}_S(\Phi, \Psi) = \mathrm{card}(\|\Phi \wedge \Psi\|)$ will be called the *support* of the decision rule and

$$\sigma_S(\Phi, \Psi) = \frac{\mathrm{supp}_S(\Phi, \Psi)}{\mathrm{card}(U)}$$

will be referred to as the *strength* of the decision rule.

With every decision rule $\Phi \to \Psi$ we associate a *certainty factor*:

$$\mathrm{cer}_S(\Phi, \Psi) = \frac{\mathrm{supp}_S(\Phi, \Psi)}{\mathrm{card}(\|\Phi\|)}$$

and a *coverage factor*:

$$\mathrm{cov}_S = \frac{\mathrm{supp}_S(\Phi, \Psi)}{\mathrm{card}(\|\Psi\|)}.$$

If $\mathrm{cer}_S(\Phi, \Psi) = 1$, then the decision rule $\Phi \to \Psi$ will be called *certain*, otherwise the decision rule will be referred to as *uncertain*.

A set of decision rules supported in total by the universe U creates a *decision algorithm* in S. Pawlak (2002) points out that every decision algorithm associated with S displays well-known probabilistic properties; in particular it satisfies the total probability theorem and Bayes' theorem. As a decision algorithm can also be interpreted in terms of the rough set concept, these properties give a new look on Bayes' theorem from the rough set perspective. In consequence, one can draw conclusions from data without referring to prior and posterior probabilities, inherently associated with Bayesian reasoning. The revealed relationship can be used to invert decision rules, i.e., giving reasons (explanations) for decisions, which is useful in decision analysis.

4. Confirmation measures and decision algorithms

In this section, we translate confirmation measures to the language of decision algorithms.

A preliminary question that arises naturally in this context is the following: why a new measure is required for decision rules in addition to strength, certainty and coverage? In other words, what is the intuition behind the confirmation measure that motivates its use for characterization of decision rules?

To answer this question, it will be useful to recall the following example proposed by Popper (1959). Consider the possible result of rolling a die: $1, 2, 3, 4, 5, 6$. We can built a decision table, presented in Table 1, where the fact that the result is even or odd is the condition attribute, while the result itself is the decision attribute.

Table 1
Decision table

Condition attribute (result odd or even)	Decision attribute (result of rolling the die)
Odd	1
Even	2
Odd	3
Even	4
Odd	5
Even	6

Let us consider the case Ψ = "the result is 6" and the case $\neg\Psi$ = "the result is not 6". Let us also take into account the information Φ = "the result is an even number (i.e. 2 or 4 or 6)". Therefore, we can consider the following two decision rules:

- $\Phi \to \Psi$ ="if the result is even, then the result is 6", with certainty $\mathrm{cer}_S(\Phi,\Psi)=\frac{1}{3}$,
- $\Phi \to \neg\Psi$ ="if the result is even, then the result is not 6", with certainty $\mathrm{cer}_S(\Phi,\neg\Psi)=\frac{2}{3}$.

Remark that the rule $\Phi \to \Psi$ has a smaller certainty than the rule $\Phi \to \neg\Psi$. However, the probability that the result is 6 is $\frac{1}{6}$, while the probability that the result is different from 6 is $\frac{5}{6}$. Thus, the information Φ raises the probability of Ψ from $\frac{1}{6}$ to $\frac{1}{3}$, and decreases the probability of $\neg\Psi$ from $\frac{5}{6}$ to $\frac{2}{3}$. In conclusion, we can say that Φ confirms Ψ and disconfirms $\neg\Psi$, independently of the fact that the certainty of $\Phi \to \Psi$ is smaller than the certainty of $\Phi \to \neg\Psi$. From this simple example, one can see that certainty and confirmation are two completely different concepts, so it advocates for a new index expressing the latter type of information.

Given a decision rule $\Phi \to \Psi$, the confirmation measure we want to introduce should give the credibility of the proposition: *Ψ is satisfied more frequently when Φ is satisfied rather than when Φ is not satisfied.*

Differently from Bayesian confirmation, however, we start from a decision table rather than from a probability measure. In this context, the probability Pr of Φ is substituted by the relative frequency Fr in the considered data table S, i.e.

$$\mathrm{Fr}_S(\Phi) = \frac{\mathrm{card}(\|\Phi\|)}{\mathrm{card}(U)}.$$

Analogously, given Φ and Ψ, $\Pr(\Psi|\Phi)$—the probability of Ψ given Φ—is substituted by the certainty factor $\mathrm{cer}_S(\Phi, \Psi)$ of the decision rule $\Phi \to \Psi$.

Therefore, a measure of confirmation of property Ψ by property Φ, denoted by $c(\Phi, \Psi)$, where Φ is a condition formula in L and Ψ is a decision formula in

L, is required to satisfy the following minimal property

$$c(\Phi,\Psi) = \begin{cases} >0 & \text{if } \mathrm{cer}_S(\Phi,\Psi) > \mathrm{Fr}_S(\Psi), \\ =0 & \text{if } \mathrm{cer}_S(\Phi,\Psi) = \mathrm{Fr}_S(\Psi), \\ <0 & \text{if } \mathrm{cer}_S(\Phi,\Psi) < \mathrm{Fr}_S(\Psi). \end{cases} \qquad \text{(i)}$$

(i) can be interpreted as follows:

- $c(\Phi,\Psi)>0$ means that property Ψ is satisfied more frequently when Φ is satisfied (then, this frequency is $\mathrm{cer}_S(\Phi,\Psi)$), rather than generically in the whole decision table (where this frequency is $\mathrm{Fr}_S(\Psi)$),
- $c(\Phi,\Psi)=0$ means that property Ψ is satisfied with the same frequency when Φ is satisfied and generically in the whole decision table,
- $c(\Phi,\Psi)<0$ means that property Ψ is satisfied less frequently when Φ is satisfied, rather than generically in the whole decision table.

Let us also remark that

$$\mathrm{cer}_S(\Phi,\Psi)\times\mathrm{Fr}_S(\Phi) + \mathrm{cer}_S(\neg\Phi,\Psi)\times\mathrm{Fr}_S(\neg\Phi) = \mathrm{Fr}_S(\Psi). \qquad \text{(ii)}$$

The proof of (ii) is as follows:

$$\mathrm{cer}_S(\Phi,\Psi) = \frac{\mathrm{supp}_S(\Phi,\Psi)}{\mathrm{card}(\|\Phi\|)}, \quad \mathrm{Fr}_S(\Phi) = \frac{\mathrm{card}(\|\Phi\|)}{\mathrm{card}(U)},$$

$$\mathrm{cer}_S(\neg\Phi,\Psi) = \frac{\mathrm{supp}_S(\neg\Phi,\Psi)}{\mathrm{card}(\|\neg\Phi\|)},$$

$$\mathrm{Fr}_S(\neg\Phi) = \frac{\mathrm{card}(\|\neg\Phi\|)}{\mathrm{card}(U)}$$

and, therefore,

$$\begin{aligned}&\mathrm{cer}_S(\Phi,\Psi)\times\mathrm{Fr}_S(\Phi)+\mathrm{cer}_S(\neg\Phi,\Psi)\times\mathrm{Fr}_S(\neg\Phi)\\ &\quad= \frac{\mathrm{supp}_S(\Phi,\Psi)}{\mathrm{card}(\|\Phi\|)}\times\frac{\mathrm{card}(\|\Phi\|)}{\mathrm{card}(U)} + \frac{\mathrm{supp}_S(\neg\Phi,\Psi)}{\mathrm{card}(\|\neg\Phi\|)}\times\frac{\mathrm{card}(\|\Phi\|)}{\mathrm{card}(U)}\\ &\quad= \frac{\mathrm{supp}_S(\Phi,\Psi)}{\mathrm{card}(U)}+\frac{\mathrm{supp}_S(\neg\Phi,\Psi)}{\mathrm{card}(U)};\end{aligned}$$

since $\|\Phi\wedge\Psi\|\cap\|\neg\Phi\wedge\Psi\| = \emptyset$ and $\|\Phi\wedge\Psi\|\cup\|\neg\Phi\wedge\Psi\| = \|\Psi\|$, we have

$$\begin{aligned}&\mathrm{supp}_S(\Phi,\Psi)+\mathrm{supp}_S(\neg\Phi,\Psi)\\ &\quad=\mathrm{card}(\|\Phi\wedge\Psi\|)+\mathrm{card}(\|\neg\Phi\wedge\Psi\|)\\ &\quad=\mathrm{card}(\|\Psi\|),\end{aligned}$$

thus,

$$\frac{\mathrm{supp}_S(\Phi,\Psi)}{\mathrm{card}(U)}+\frac{\mathrm{supp}_S(\neg\Phi,\Psi)}{\mathrm{card}(U)} = \frac{\mathrm{card}(\|\Psi\|)}{\mathrm{card}(U)} = \mathrm{Fr}_S(\Psi).$$

(ii) says that $\mathrm{Fr}_S(\Psi)$ is a weighted average of $\mathrm{cer}_S(\Phi,\Psi)$ and $\mathrm{cer}_S(\neg\Phi,\Psi)$, with respective weights $\mathrm{Fr}_S(\Phi)$ and $\mathrm{Fr}_S(\neg\Phi)=[1-\mathrm{Fr}_S(\Phi)]$. Thus, if $\mathrm{cer}_S(\Phi,\Psi)>\mathrm{Fr}_S(\Psi)$, then necessarily $\mathrm{cer}_S(\neg\Phi,\Psi)<\mathrm{Fr}_S(\Psi)$.

On the basis of this observation, (i) can also be interpreted as follows:

- $c(\Phi,\Psi)>0$ means that property Ψ is satisfied more frequently when Φ is satisfied rather than when Φ is not satisfied,
- $c(\Phi,\Psi)=0$ means that property Ψ is satisfied with the same frequency when Φ is satisfied and when Φ is not satisfied,
- $c(\Phi,\Psi)<0$ means that property Ψ is satisfied more frequently when Φ is not satisfied rather than when Φ is satisfied.

The specific confirmation measures recalled in Section 2 can be rewritten in this context as follows:

$$d(\Phi,\Psi) = \mathrm{cer}_S(\Phi,\Psi)-\mathrm{Fr}_S(\Psi),$$

$$r(\Phi,\Psi) = \log\left[\frac{\mathrm{cer}_S(\Phi,\Psi)}{\mathrm{Fr}_S(\Psi)}\right],$$

$$l(\Phi,\Psi) = \log\left[\frac{\mathrm{cer}_S(\Psi,\Phi)}{\mathrm{cer}_S(\neg\Psi,\Phi)}\right],$$

$$f(\Phi,\Psi) = \frac{\mathrm{cer}_S(\Psi,\Phi)-\mathrm{cer}_S(\neg\Psi,\Phi)}{\mathrm{cer}_S(\Psi,\Phi)+\mathrm{cer}_S(\neg\Psi,\Phi)},$$

$$s(\Phi,\Psi) = \mathrm{cer}_S(\Phi,\Psi)-\mathrm{cer}_S(\neg\Phi,\Psi),$$

$$b(\Phi,\Psi) = \mathrm{cer}_S(\Phi,\Psi)-\mathrm{Fr}_S(\Phi)\,\mathrm{Fr}_S(\Psi).$$

Clearly, all the results about confirmation measures obtained within Bayesian confirmation theory are valid for the confirmation measures defined in the context of decision algorithms considered within rough set theory. Therefore, according to Fitelson's conclusions reminded in Section 2, we believe that $l(\Phi,\Psi)$ and $f(\Phi,\Psi)$ are the most convincing confirmation measures, which continue to be ordinally equivalent in this new context.

Let us remark that one of the most appreciated interestingness measures, proposed by Kamber and Shingal (1996), is strongly related to $l(\Phi,\Psi)$. It is called sufficiency measure and has the following formulation:

$$k(\Phi,\Psi) = \frac{\mathrm{cer}_S(\Psi,\Phi)}{\mathrm{cer}_S(\neg\Psi,\Phi)}.$$

Let us remark that $l(\Phi,\Psi) = \log[k(\Phi,\Psi)]$.

Below, we call the confirmation measures presented in the language of decision algorithms, the *rough set confirmation measures*.

5. Confirmation theory and the ravens' paradox

In view of logics, decision rules are often seen as implications. Hempel, in a series of articles, derived a theory of confirmation from logics. His starting point was the Nicod's criterion (Nicod, 1923). The Nicod's criterion says that an evidence confirms decision rule

"A implies B" if and only if it satisfies both the premise and the conclusion of the rule; it disconfirms the rule if and only if it satisfies the premise, but not the conclusion of the rule. Thus, according to the Nicod's criterion, an evidence is neutral, or irrelevant, with respect to the rule if it does not satisfy the premise.

To illustrate this point, Hempel introduced an example which became very well known in the specialized literature. The rule used for the illustration, denoted by (I1), is the following:

"if x is a raven, then x is black" or, in everyday language, "all ravens are black".

Remark that with respect to the considered rule, there are four possible evidences:

(a) black raven,
(b) black non-raven (for example, a black shoe),
(c) non-black raven (for example, a white raven),
(d) non-black non-raven (for example, a white shoe).

According to the Nicod's criterion, (a) is a positive instance of the rule, and so (a) confirms rule (I1). (c) is a negative instance, and so (c) disconfirms (I1). (b) and (d) do not satisfy the premise of the rule "all ravens are black" (i.e., neither (b) or (d) is a raven), and so they are non-instances and are irrelevant to (I1).

With the aim of discussing the Nicod's criterion, Hempel introduced the equivalence condition which says "*Whatever confirms* (*disconfirms*) *one of two equivalent sentences, also confirms* (*disconfirms*) *the other*" (Hempel, 1945, p. 12).

It seems that the truth of this condition is quite uncontestable. As Hempel claimed, the equivalence condition is "*a necessary condition*" and "*fulfillment of this condition makes the confirmation of a rule independent of the way in which it is formulated*" (Hempel, 1945, p. 12).

Even if the Nicod's criterion seems so natural and the equivalence condition so necessary in the theory of confirmation, when we put the two together, some problems arise.

To illustrate these problems, let us come back to rule (I1). Now, consider another rule, denoted by (I2)

"if x is non-black, then x is not a raven" or "all non-black things are not ravens"

which is logically equivalent to (I1). The equivalence can easily be observed from the truth tables of these rules seen as implications: "A implies B" and "$\neg B$ implies $\neg A$". It can also be observed that (I1) corresponds to *modus ponens* and (I2) to *modus tollens* reasoning patterns.

Using the logic of the Nicod's criterion, we see that evidence (d) confirms (I2), (c) disconfirms (I2) while (a) and (b) are irrelevant to (I2). The equivalence condition wants that whatever confirms (I1) also confirms (I2), however, (a) confirms (I1) but not (I2) and (d) confirms (I2) but not (I1). In this case, we see that application of the Nicod's criterion violates the equivalence condition. Hempel (1945, p. 11) concludes: "*This means that Nicod's Criterion makes confirmation depend not only on the content of the hypothesis, but also on its formulation*".

Thus, instead of the Nicod's criterion, Hempel proposes the positive instance criterion. Hempel claims that since (a) confirms (I1) and (d) confirms (I2), and (I1) is logically equivalent to (I2), so both (a) and (d) confirm (I1) and (I2). Moreover, (a) and (d) are not the only evidence that can confirm (I1) and (I2). In fact, let us consider the following rule (I3):

"if x is a raven or not, then x is black or not a raven" or "anything which is or is not a raven is either not a raven or black".

According to the Nicod's criterion, anything which is not a raven or black can confirm (I3). Obviously, (I3) is logically equivalent to (I1) and (I2). Therefore, according to the positive instance criterion, anything which is not a raven or a black raven confirms (I1) (and (I2)) as well. In other words, (a), (b) and (d) would confirm (I1), while only (c) disconfirms (I1). In Table 2, we summarize the application of the Nicod's criterion to (I1), (I2) and (I3).

From the above reasoning, Hempel (1945, p. 14) concludes: "*This implies that any non-raven represents confirming evidence for the hypothesis that all ravens are black*". In this case, any red pencil, yellow chalk, blackboard, white shoe, etc. become evidence to confirm (I1). In other words, excluding "non-black ravens", all things in the world confirm (I1). It entails that if we want to test the hypothesis "all ravens are black", we do not need to find any black raven, or even any raven, to support the hypothesis. In other words, most things in the world can confirm any implication. Hempel (1945, p. 14) writes: "*We shall refer to these implications of the equivalence criterion and of the above sufficient condition of confirmation as the paradoxes of confirmation*", i.e., the ravens' paradox. How to solve the ravens' paradox? Hempel proposed some solutions to the paradox. We think that the most interesting solution is based on a misunderstanding of logic, as presented below.

Hempel claims that the rule "A implies B" is not only about the class of objects with property A, but about all objects. So, for (I1), it is not only a rule about ravens. It is a rule concerning all objects in the world, whether they are ravens or not. In other words, it is "for all x, if x is a raven, then x is black," and we are focusing on "all x" instead of the class of ravens only. If it is so, we should not be surprised that the objects like a yellow chalk or a blue book confirm (I1). In the following section we

Table 2
Application of the Nicod's criterion to (I1), (I2) and (I3)

	(I1): "if x is a raven, then x is black"	(I2): "if x is non-black, then x is not a raven"	(I3): "if x is a raven or not, then x is black or not a raven"
(a) Black ravens	Confirmatory	Irrelevant	Confirmatory
(b) Black non-ravens	Irrelevant	Irrelevant	Confirmatory
(c) Non-black ravens	Disconfirmatory	Disconfirmatory	Disconfirmatory
(d) Non-black non-ravens	Irrelevant	Confirmatory	Confirmatory

propose a new criterion relative to confirmation, which is different from both the Nicod's criterion and the Hempel's positive instance criterion: in the rough set context, it is called *property of monotonicity* of the confirmation measure.

6. Desirable properties for rough set confirmation measures

Even if all the formal properties of the Bayesian confirmation measures hold also for the corresponding rough set confirmation measures, we think that there is a new property which would be desirable for the latter measures.

To introduce this new property, let us remark that for each formula $\Phi \to \Psi$ in L, one can express the rough set confirmation measures in terms of the following four values:

- $a = \mathrm{supp}_S(\Phi, \Psi)$, i.e. the number of objects in U for which Φ and Ψ hold together,
- $b = \mathrm{supp}_S(\neg\Phi, \Psi)$, i.e. the number of objects in U for which Φ does not hold while Ψ holds,
- $c = \mathrm{supp}_S(\Phi, \neg\Psi)$, i.e. the number of objects in U for which Φ holds while Ψ does not hold,
- $d = \mathrm{supp}_S(\neg\Phi, \neg\Psi)$, i.e. the number of objects in U for which both Φ and Ψ do not hold.

Using the terms of the ravens' paradox, one can notice that if Φ is the property "to be a raven" and Ψ is the property "to be black", then

- $a = \mathrm{supp}_S(\Phi, \Psi)$ is the number of objects of U which are black ravens,
- $b = \mathrm{supp}_S(\neg\Phi, \Psi)$ is the number of objects of U which are black non-ravens,
- $c = \mathrm{supp}_S(\Phi, \neg\Psi)$ is the number of objects of U which are non-black ravens,
- $d = \mathrm{supp}_S(\neg\Phi, \neg\Psi)$ is the number of objects of U which are non-black non-ravens.

Therefore, the rough set confirmation measures can be expressed as follows:

$$d(\Phi, \Psi) = \frac{a}{a+c} - \frac{a+b}{a+b+c+d} = \frac{ad-bc}{(a+c)(a+b+c+d)},$$

$$r(\Phi, \Psi) = \log\left[\frac{(a/(a+c))}{((a+b)/(a+b+c+d))}\right],$$

$$l(\Phi, \Psi) = \log\left[\frac{(a/(a+b))}{(c/(c+d))}\right],$$

$$f(\Phi, \Psi) = \frac{\dfrac{a}{a+b} - \dfrac{c}{c+d}}{\dfrac{a}{a+b} + \dfrac{c}{c+d}} = \frac{ad-bc}{ad+bc+2ac},$$

$$s(\Phi, \Psi) = \frac{a}{a+c} - \frac{b}{b+d} = \frac{(ad-bc)}{(a+c)(b+d)},$$

$$b(\Phi, \Psi) = \frac{a}{a+b+c+d} - \frac{a+c}{a+b+c+d}\,\frac{a+b}{a+b+c+d} = \frac{ad-bc}{(a+b+c+d)^2}.$$

In this context, we propose the following property of monotonicity:

(M) $c(\Phi, \Psi) = F\,[\mathrm{supp}_S(\Phi, \Psi),\ \mathrm{supp}_S(\neg\Phi, \Psi),\ \mathrm{supp}_S(\Phi, \neg\Psi),\ \mathrm{supp}_S(\neg\Phi, \neg\Psi)]$ is a function non-decreasing with respect to $\mathrm{supp}_S(\Phi, \Psi)$ and $\mathrm{supp}_S(\neg\Phi, \neg\Psi)$ and non-increasing with respect to $\mathrm{supp}_S(\neg\Phi, \Psi)$ and $\mathrm{supp}_S(\Phi, \neg\Psi)$.

The monotonicity property (M) has the following interpretation. Monotonicity of $c(\Phi, \Psi)$ with respect to $\mathrm{supp}_S(\Phi,\Psi)$ means that any evidence in which Φ and Ψ hold together increases (or at least does not decrease) the credibility of the decision rule $\Phi \to \Psi$. Considering the example of black ravens, this means that the more black ravens we observe, the more credible becomes the decision rule "if x is a raven, then x is black" denoted by (I1) in Section 5. Monotonicity of $c(\Phi,\Psi)$ with respect to $\mathrm{supp}_S(\Phi,\neg\Psi)$ means that any evidence in which Φ holds and Ψ does not hold decreases (or at least does not increase) the credibility of the decision rule $\Phi \to \Psi$. In the example of black ravens, this means that the more non-black ravens we observe, the less credible becomes the decision rule (I1). Analogously, with respect to $\mathrm{supp}_S(\neg\Phi,\Psi)$, any evidence in which Φ does not hold and Ψ holds decreases (or at least does not increase) the credibility of the decision rule $\Phi \to \Psi$, and with respect to $\mathrm{supp}_S(\neg\Phi,\neg\Psi)$, any evidence in which both Φ and Ψ do not hold increases (or at least does not decrease) the

Table 3
Application of the Nicod's criterion to (I1), (I2) and (I3), compared with monotonicity property of the confirmation measure

	(I1): "if x is a raven, then x is black"	(I2): "if x is non-black, then x is not a raven"	(I3): "if x is a raven or not, then x is black or not a raven"	Monotonicty property of the confirmation measure
(a) Black-ravens	Confirmatory	Irrelevant	Confirmatory	Confirmatory
(b) Black non-ravens	Irrelevant	Irrelevant	Confirmatory	Disconfirmatory
(c) Non-black ravens	Disconfirmatory	Disconfirmatory	Disconfirmatory	Disconfirmatory
(d) Non-black non-ravens	Irrelevant	Confirmatory	Confirmatory	Confirmatory

credibility of the decision rule $\Phi \to \Psi$. In the example of black ravens, this means that the more black non-ravens we observe, the less credible becomes the decision rules (I1), while the more non-black non-ravens we observe the more credible becomes the decision rule (I1).

Let us remark that the monotonicity with respect to $\text{supp}_S(\Phi,\Psi)$ and $\text{supp}_S(\Phi,\neg\Psi)$ amount to the Nicod's criterion and are unquestionable. The monotonicity with respect to $\text{supp}_S(\neg\Phi,\Psi)$ and $\text{supp}_S(\neg\Phi,\neg\Psi)$ are more debatable. To put the Nicod's criterion, the Hempel's positive instance criterion and our monotonicity property in the same context, we propose Table 3.

Table 3 clearly shows that our monotonicity property is different not only from the Nicod's criterion but also from the Hempel's positive instance criterion represented in the last but one column. Let us note that in the context of automated analysis of data, Hajek and Havranek (1978) suggest the same monotonicity property M, however, not for a confirmation measure of rule $\Phi \to \Psi$ (for which they suggest the increasing monotonicity with respect to $\text{supp}_S(\Phi,\Psi)$ and the decreasing monotonicity with respect to $\text{supp}_S(\Phi,\neg\Psi)$ only), but for a more specific association measure. In other words, according to Hajek and Havranek (1978), within data analysis, a decision rule is confirmed according to the Nicod's criterion.

We can explain the monotonicity with respect to $\text{supp}_S(\neg\Phi,\Psi)$ and $\text{supp}_S(\neg\Phi,\neg\Psi)$, considering our interpretation of property (i) from Section 4: a positive value of a confirmation measure $c(\Phi,\Psi)$ means that property Ψ is satisfied more frequently when property Φ is satisfied rather than when Φ is not satisfied. From this viewpoint, an evidence in which Φ is not satisfied and Ψ is satisfied (objects $||\neg\Phi \wedge \Psi||$) increases the frequency of Ψ in the situations where Φ is not satisfied and thus it should decrease the value of the confirmation. Analogously, an evidence in which both Φ and Ψ are not satisfied (objects $||\neg\Phi \wedge \neg\Psi||$) decreases the frequency of Ψ in the situations where Φ is not satisfied and thus it should increase the value of the confirmation.

We want to give also a more formal justification to the monotonicty of confirmation measures with respect to $\text{supp}_S(\neg\Phi,\Psi)$ and $\text{supp}_S(\neg\Phi,\neg\Psi)$. Let us consider the following definition of confirmation: property Φ confirms property Ψ if

$$\text{cer}_S(\Phi,\Psi) > \text{Fr}_S(\Psi). \tag{iii}$$

Let us remark that definition (iii) corresponds to the definition of *incremental confirmation* introduced by Carnap (1962, new preface) under the name of "*confirmation as increase in firmness*" in the following form: evidence Φ confirms hypothesis Ψ if

$$\Pr(\Psi|\Phi) > \Pr(\Psi). \tag{iv}$$

The confirmation measures $d(\Phi,\Psi)$, $r(\Phi,\Psi)$, $l(\Phi,\Psi)$, $f(\Phi,\Psi)$, $s(\Phi,\Psi)$ and $b(\Phi,\Psi)$ can be seen as quantitative generalizations of the qualitative incremental confirmation (Fitelson, 2001).

Redefining (iii) in terms of $a = \text{supp}_S(\Phi,\Psi)$, $b = \text{supp}_S(\neg\Phi,\Psi)$, $c = \text{supp}_S(\Phi,\neg\Psi)$ and $d = \text{supp}_S(\neg\Phi,\neg\Psi)$, we get

$$\frac{a}{a+c} > \frac{a+b}{a+b+c+d}. \tag{iii$'$}$$

The following theorem is useful for justifying the property of monotonicity.

Theorem 1. *Let us consider case α in which*

$$a = \text{supp}_S(\Phi,\Psi), \quad b = \text{supp}_S(\neg\Phi,\Psi),$$

$$c = \text{supp}_S(\Phi,\neg\Psi), \quad d = \text{supp}_S(\neg\Phi,\neg\Psi)$$

and case α' in which

$$a' = \text{supp}_S(\Phi',\Psi'), \quad b' = \text{supp}_S(\neg\Phi',\Psi'),$$

$$c' = \text{supp}_S(\Phi',\neg\Psi'), \quad d' = \text{supp}_S(\neg\Phi',\neg\Psi').$$

Let us suppose, moreover, that

$$\text{cer}_S(\Phi,\Psi) < \text{Fr}_S(\Psi) \tag{v}$$

while

$$\text{cer}_S(\Phi',\Psi') > \text{Fr}_S(\Psi'). \tag{vi}$$

The following implications are satisfied:

(1) if $a' = a + \Delta, b' = b, c' = c$ and $d' = d$, then $\Delta > 0$,
(2) if $a' = a, b' = b + \Delta, c' = c$ and $d' = d$, then $\Delta < 0$,
(3) if $a' = a, b' = b, c' = c + \Delta$ and $d' = d$, then $\Delta < 0$,
(4) if $a' = a, b' = b, c' = c$ and $d' = d + \Delta$, then $\Delta > 0$.

Proof. (v) can be written as

$$\frac{a}{a+c}<\frac{a+b}{a+b+c+d}. \tag{vii}$$

If condition of implication (1) holds, (vi) can be rewritten as

$$\frac{a'}{a'+c'}>\frac{a'+b'}{a'+b'+c'+d'} \Leftrightarrow \frac{(a+\Delta)}{(a+\Delta)+c}>\frac{(a+\Delta)+b}{(a+\Delta)+b+c+d}. \tag{viii}$$

Using simple algebraic operations, we obtain from (viii)

$$\frac{(a+\Delta)}{(a+\Delta)+c}>\frac{(a+\Delta)+b}{(a+\Delta)+b+c+d} \Leftrightarrow ad-bc>-d\Delta. \tag{ix}$$

Due to (vii), we have

$$ad-bc<0. \tag{x}$$

Moreover, let us remark that

$$d>0 \tag{xi}$$

because if $d=0$, then (ix) would boil down to

$$-bc>0,$$

which is impossible because $b\geqslant 0$ and $c\geqslant 0$.For (x) and (xi), (ix) gives

$$\Delta>\frac{ad-bc}{-d}>0.$$

Thus, we proved implication (1). If condition of implication (2) holds, (vi) can be rewritten as

$$\frac{a'}{a'+c'}>\frac{a'+b'}{a'+b'+c'+d'} \Leftrightarrow \frac{a}{a+c}>\frac{a+(b+\Delta)}{a+(b+\Delta)+c+d}. \tag{xii}$$

From (vii) and (xii) we obtain

$$\frac{a+b}{a+b+c+d}>\frac{a}{a+c}>\frac{a+(b+\Delta)}{a+(b+\Delta)+c+d}. \tag{xiii}$$

Using simple algebraic operations, from (xiii) we obtain

$$\frac{a+(b+\Delta)}{a+(b+\Delta)+c+d}<\frac{a+b}{a+b+c+d} \Leftrightarrow (c+d)\Delta<0. \tag{xiv}$$

(xiv) holds if $\Delta<0$, because $c\geqslant 0$ and $d\geqslant 0$ by definition and, moreover, $c\neq 0$ or $d\neq 0$, because otherwise inequality (vii) would become

$$\frac{a}{a}<\frac{a+b}{a+b}$$

which, of course, is a contradiction. (xiv) says that (v) and (vi), or equivalently (vii) and (xii), can be satisfied together only if $\Delta<0$. This proves implication (2).

If condition of implication (3) holds, (vi) can be rewritten as

$$\frac{a'}{a'+c'}>\frac{a'+b'}{a'+b'+c'+d'} \Leftrightarrow \frac{a}{a+(c+\Delta)}>\frac{a+b}{a+b+(c+\Delta)+d}. \tag{xv}$$

From (xv), we obtain

$$\frac{a}{a+(c+\Delta)}>\frac{a+b}{a+b+(c+\Delta)+d} \Leftrightarrow ad-bc>b\Delta. \tag{xvi}$$

Due to (vii) and (x), and taking into account that $b\geqslant 0$, (xvi) can be satisfied only if $\Delta<0$. This proves implication (3). If condition of implication (4) holds, (vi) can be rewritten as

$$\frac{a'}{a'+c'}>\frac{a'+b'}{a'+b'+c'+d'} \Leftrightarrow \frac{a}{a+c}>\frac{a+b}{a+b+c+(d+\Delta)}. \tag{xvii}$$

Clearly, (vii) and (xvii) can be satisfied only if $\Delta>0$ and this proves implication (4). □

Theorem 1 has the following interpretation. Passing from case α to case α', we pass from a situation in which property Φ does not confirm property Ψ, to a situation in which property Φ' confirms property Ψ'. Theorem 1 says that this passage from non-confirmation to confirmation is permitted by an increase of $\mathrm{supp}_S(\Phi,\Psi)$ or $\mathrm{supp}_S(\neg\Phi,\neg\Psi)$, or by a decrease of $\mathrm{supp}_S(\neg\Phi,\Psi)$ or $\mathrm{supp}_S(\Phi,\neg\Psi)$. Thus, the theorem supports the claim that confirmation given by property Φ to property Ψ is positively related to $\mathrm{supp}_S(\Phi,\Psi)$ and $\mathrm{supp}_S(\neg\Phi,\neg\Psi)$, and negatively related to $\mathrm{supp}_S(\neg\Phi,\Psi)$ and $\mathrm{supp}_S(\Phi,\neg\Psi)$.

In fact, Theorem 1 supports the monotonicity property (M) because, if the passage from a situation of non-confirmation to a situation of confirmation implies a specific sign of modifications of the four values $\mathrm{supp}_S(\Phi,\Psi)$, $\mathrm{supp}_S(\neg\Phi,\neg\Psi)$, $\mathrm{supp}_S(\neg\Phi,\Psi)$ and $\mathrm{supp}_S(\Phi,\neg\Psi)$, it is natural to expect that confirmation measures will react analogously to modifications of the above values.

Considering Hempel's theory of confirmation, a possible objection to our monotonicity property (M) could be the following: is the equivalence condition respected by the monotonicity property (M)? To answer this question we can reconsider the three equivalent implications (I1), (I2) and (I3). Let us denote again by Φ the property "to be a raven" and by Ψ the property "to be black". With respect to (I1), "if x is a raven, then x is black", using (iii) we have that Φ confirms Ψ if

$$\mathrm{cer}_S(\Phi,\Psi)>\mathrm{Fr}_S(\Psi)$$

which can be rewritten as

$$\frac{\mathrm{supp}_S(\Phi,\Psi)}{\mathrm{supp}_S(\Phi,\Psi)+\mathrm{supp}_S(\Phi,\neg\Psi)} > \frac{\mathrm{supp}_S(\Phi,\Psi)+\mathrm{supp}_S(\neg\Phi,\Psi)}{\mathrm{supp}_S(\Phi,\Psi)+\mathrm{supp}_S(\Phi,\neg\Psi)+\mathrm{supp}_S(\neg\Phi,\Psi)+\mathrm{supp}_S(\neg\Phi,\neg\Psi)}. \tag{xviii}$$

With respect to (I2), "if x is non-black, then x is not a raven", using (iii) we have that $\neg\Psi$ confirms $\neg\Phi$ if

$$\mathrm{cer}_S(\neg\Psi,\neg\Phi) > \mathrm{Fr}_S(\neg\Phi)$$

which can be rewritten as

$$\frac{\mathrm{supp}_S(\neg\Phi,\neg\Psi)}{\mathrm{supp}_S(\Phi,\neg\Psi)+\mathrm{supp}_S(\neg\Phi,\neg\Psi)} > \frac{\mathrm{supp}_S(\neg\Phi,\Psi)+\mathrm{supp}_S(\neg\Phi,\neg\Psi)}{\mathrm{supp}_S(\Phi,\Psi)+\mathrm{supp}_S(\Phi,\neg\Psi)+\mathrm{supp}_S(\neg\Phi,\Psi)+\mathrm{supp}_S(\neg\Phi,\neg\Psi)}. \tag{xix}$$

Using elementary algebra, we can prove that, both, (xviii) and (xix) are equivalent to

$$\mathrm{supp}_S(\Phi,\Psi)\times\mathrm{supp}_S(\neg\Phi,\neg\Psi) > \mathrm{supp}_S(\neg\Phi,\Psi)\,\mathrm{supp}_S(\Phi,\neg\Psi). \tag{xx}$$

The equivalence of (xviii) and (xix) says that the incremental confirmation of the implication $\Phi\to\Psi$ is equivalent to the incremental confirmation of the implication $\neg\Psi\to\neg\Phi$. Thus, with respect to (I1) and (I2) Hempel's equivalence condition holds.

Let us remark, however, that equivalent incremental confirmation does not mean equivalent degree of confirmation. In fact, considering, for example, confirmation measure l we can see that with respect to implication $\Phi\to\Psi$, the confirmation measure is $l(\Phi,\Psi)=\log[\mathrm{cer}_S(\Psi,\Phi)/\mathrm{cer}_S(\neg\Psi,\Phi)]$, while with respect to implication $\neg\Psi\to\neg\Phi$, the confirmation measure is $l(\neg\Phi,\neg\Psi)=\log[\mathrm{cer}_S(\neg\Phi,\neg\Psi)/\mathrm{cer}_S(\Phi,\neg\Psi)]$ and, in general, $l(\Phi,\Psi)\neq l(\neg\Psi,\neg\Phi)$.

In terms of confirmation measure $c(\Phi,\Psi)$, equivalent incremental confirmation of $\Phi\to\Psi$ and $\neg\Psi\to\neg\Phi$ means that for each Φ and Ψ:

$$c(\Phi,\Psi)>0 \Leftrightarrow c(\neg\Psi,\neg\Phi)>0, \tag{xxi}$$

$$c(\Phi,\Psi)<0 \Leftrightarrow c(\neg\Psi,\neg\Phi)<0, \tag{xxii}$$

$$c(\Phi,\Psi)=0 \Leftrightarrow c(\neg\Psi,\neg\Phi)=0. \tag{xxiii}$$

It is worth noting that equivalent incremental confirmation holds for $\Phi\to\Psi, \neg\Psi\to\neg\Phi$, $\Psi\to\Phi$ and $\neg\Phi\to\neg\Psi$, thus (xxi), (xxii) and (xxiii) can be generalized as follows:

$$c(\Phi,\Psi)>0 \Leftrightarrow c(\neg\Psi,\neg\Phi)>0 \Leftrightarrow c(\Psi,\Phi)>0 \Leftrightarrow c(\neg\Phi,\neg\Psi)>0, \tag{xxiv}$$

$$c(\Phi,\Psi)<0 \Leftrightarrow c(\neg\Psi,\neg\Phi)<0 \Leftrightarrow c(\Psi,\Phi)<0 \Leftrightarrow c(\neg\Phi,\neg\Psi)<0, \tag{xxv}$$

$$c(\Phi,\Psi)=0 \Leftrightarrow c(\neg\Psi,\neg\Phi)=0 \Leftrightarrow c(\Psi,\Phi)=0 \Leftrightarrow c(\neg\Phi,\neg\Psi)=0. \tag{xxvi}$$

With respect to (I3), "if x is a raven or not, then x is black or not a raven", the discussion is a little more complex. Formally, (I3) can be written as $(\Phi\vee\neg\Phi)\to(\neg\Phi\vee\Psi)$. Thus, with respect to (I3), using (iii) we have that $(\Phi\vee\neg\Phi)$ confirms $(\neg\Phi\vee\Psi)$ if

$$\mathrm{cer}_S(\Phi\vee\neg\Phi,\neg\Phi\vee\Psi) > \mathrm{Fr}_S(\neg\Phi\vee\Psi). \tag{xxvii}$$

Let us observe that $\|\Phi\vee\neg\Phi\|=U$, i.e. "all x being ravens or not ravens" means all objects in the universe of discourse U. Therefore, we have

$$\mathrm{cer}_S(\Phi\vee\neg\Phi,\neg\Phi\vee\Psi) = \mathrm{Fr}_S(\neg\Phi\vee\Psi). \tag{xxviii}$$

Of course, (xxvii) and (xxviii) are incompatible. Remark, however, that $\Phi\to\Psi$ can be considered as equivalent to $(\Phi\vee\neg\Phi)\to(\neg\Phi\vee\Psi)$ only if $\to$ is seen as a typical logical implication, but this is not the case if $\to$ is seen as a decision rule. According to the definition of a typical logical implication, $\Phi\to\Psi$ is false when Φ is true and Ψ is false. In all other cases $\Phi\to\Psi$ is true. The implication is thus equivalent to $\neg\Phi\vee\Psi$, so it is also equivalent to $(\Phi\vee\neg\Phi)\to(\neg\Phi\vee\Psi)$. In the literature, this interpretation of an implication corresponds precisely to so-called material implication. It has been considered by Frege, Łukasiewicz, Russell, Wittgenstein and logical positivists (for an exhaustive reconstruction see Edgington, 1995). We do not wish to enter the discussion about different interpretations of an implication, however, we would like to show that the material implication is not the best interpretation for decision rules resulting from the rough set approach and, more generally, from data analysis.

The best known objections to the material implication are the paradoxes of implication. For example, an implication like, "if Rome is in Poland, then 2 + 2 = 5" is true as material implication because the premise is false (if Rome is in Poland), so, whatever the conclusion is, the material implication is true. One of the most natural alternative interpretations of the implication, which agrees with decision rules, has been given by Edgington (1995) (see also Stalnaker, 1968); this non-truth-functional interpretation is the following:

(a) $\Phi\to\Psi$ is false when Φ is true and Ψ is false,
(b) $\Phi\to\Psi$ is true when Φ and Ψ are both true,
(c) $\Phi\to\Psi$ may be false or true when Φ is false.

While situations (a) and (b) are quite natural, situation (c) needs some clarification. Suppose that Φ = "you touch that wire" and Ψ = "you will get an electric shock"; then $\Phi\to\Psi$ is read as "if you touch that wire, then you will get an electric shock". Remark that if you will not touch that wire, you will not know if $\Phi\to\Psi$ is true or false. This corresponds to situation (c) in which we are not able to say anything about the truth or the falsity of $\Phi\to\Psi$.

Therefore, according to the non-truth-functional interpretation of an implication identified with a decision rule, we are interested by the cases where Φ and Ψ are both true (because they support the decision rule) or where Φ is true and Ψ is false (because they discard the decision rule), while we are not interested by the cases where Φ is false. To justify our lack of interest by the last type of cases, let us consider a decision table with medical data where Φ ="presence of symptom s" and Ψ ="presence of disease d" (consequently, $\neg\Phi$ ="absence of symptom s" and $\neg\Psi$ ="absence of disease d"). In the table there are cases with properties $(\neg\Phi \wedge \Psi)$ and $(\neg\Phi \wedge \neg\Psi)$, while there is no case with properties $(\Phi \wedge \Psi)$ or $(\Phi \wedge \neg\Psi)$. In this situation, the decision rule $\Phi \rightarrow \Psi \equiv$ "if there is symptom s, then there is disease d" will *not* be induced. If, however, we would accept the interpretation of the material implication, decision rule $\Phi \rightarrow \Psi$ would be induced, because, for this interpretation, the only case permitting to block the induction of $\Phi \rightarrow \Psi$ is the case with properties $(\Phi \wedge \neg\Psi)$ and there is no case of this type in the considered decision table.

Thus, coming back to Hempel's equivalence condition, $\Phi \rightarrow \Psi$ and $(\Phi \vee \neg\Phi) \rightarrow (\neg\Phi \vee \Psi)$ are equivalent if we consider the material implication, but they are not equivalent if we consider the non-truth-functional interpretation of implication $\Phi \rightarrow \Psi$. This explains why the Hempel's equivalence condition and our monotonicity condition do not agree with respect to formulation (I3), "if x is a raven or not, then x is black or not a raven".

Let us end this section, with a remark about our approach to the ravens' paradox and the so-called Bayesian solution (Mackie, 1963; Horwich, 1982; Earman, 1992). We applied the incremental confirmation to the ravens' paradox on the basis of the rule $\Phi \rightarrow \Psi \equiv$ "if x is a raven, then x is black". Thus, we considered Φ= "x is a raven" as an evidence for the hypothesis Ψ = "x is black". The confirmation measure $c(\Phi, \Psi)$ calculated for this rule says how much evidence Φ supports (when $c(\Phi, \Psi) > 0$), discards (when $c(\Phi, \Psi) < 0$) or is indifferent to (when $c(\Phi, \Psi) = 0$) hypothesisΨ. Within the Bayesian solution of the ravens' paradox, the interpretation of evidence and hypothesis is different. There are four evidences: Φ_{BR} = "x is a black raven", Φ_{NR}= "x is a non-black raven", Φ_{BN}= "x is a black non-raven", Φ_{NN}= "x is a non-black non-raven". The hypothesis is Ψ_{BR}= "all ravens are black". Bayesian approach to the ravens' paradox considers the following measures of confirmation:

- $c(\Phi_{\mathrm{BR}}, \Psi_{\mathrm{BR}})$, which says how much the observation of a black raven (Φ_{BR}), supports (when $c(\Phi_{\mathrm{BR}}, \Psi_{\mathrm{BR}}) > 0$), discards (when $c(\Phi_{\mathrm{BR}}, \Psi_{\mathrm{BR}}) < 0$) or is indifferent to (when $c(\Phi_{\mathrm{BR}}, \Psi_{\mathrm{BR}}) = 0$) the hypothesis that all ravens are black (Ψ_{BR});
- $c(\Phi_{\mathrm{NR}}, \Psi_{\mathrm{BR}})$, which says how much the observation of a non-black raven (Φ_{NR}), supports (when $c(\Phi_{\mathrm{NR}}, \Psi_{\mathrm{BR}}) > 0$), discards (when $c(\Phi_{\mathrm{NR}}, \Psi_{\mathrm{BR}}) < 0$) or is indifferent to (when $c(\Phi_{\mathrm{NR}}, \Psi_{\mathrm{BR}}) = 0$) the hypothesis that all ravens are black (Ψ_{BR});
- $c(\Phi_{\mathrm{BN}}, \Psi_{\mathrm{BR}})$, which says how much the observation of a black non-raven (Φ_{BN}), supports (when $c(\Phi_{\mathrm{BN}}, \Psi_{\mathrm{BR}}) > 0$), discards (when $c(\Phi_{\mathrm{BN}}, \Psi_{\mathrm{BR}}) < 0$) or is indifferent to (when $c(\Phi_{\mathrm{BN}}, \Psi_{\mathrm{BR}}) = 0$) the hypothesis that all ravens are black (Ψ_{BR});
- $c(\Phi_{\mathrm{NN}}, \Psi_{\mathrm{BR}})$, which says how much the observation of a non-black non-raven (Φ_{BR}), supports (when $c(\Phi_{\mathrm{NN}}, \Psi_{\mathrm{BR}}) > 0$), discards (when $c(\Phi_{\mathrm{NN}}, \Psi_{\mathrm{BR}}) < 0$) or is indifferent to (when $c(\Phi_{\mathrm{NN}}, \Psi_{\mathrm{BR}}) = 0$) the hypothesis that all ravens are black (Ψ_{BR}).

Bayesian solution of ravens' paradox says that the evidence "black raven" has a higher degree of confirmation of the hypothesis "all ravens are black" than the evidence "non-black non-raven" (for a first attempt of this solution see (Hosiasson-Lindenbaum, 1940); let us mention that Hosiasson-Lindenbaum was the first to formulate the ravens' paradox in print; she attributed it to Hempel but gave no reference; Hempel (1945) referred to "discussions" with her). This can solve the paradox because the observation of "non-black non-ravens" supports the hypothesis "all ravens are black" so weakly that people will always think that "non-black non-ravens" do not confirm the hypothesis.

Without entering into details of a complex discussion about interpretation of probabilities involved in the Bayesian solution, we will give an idea, of the complexity of the Bayesian approach to the ravens' paradox, using an example proposed by Good (1967), who was one of the first to show that the Nicod's criterion is not always acceptable. Good considered the case where exactly one of the following hypotheses is true:

- (H): there are 100 black ravens, no non-black raven and 1 million of other birds,
- ($\neg H$): there are 1000 black ravens, 1 white raven and 1 million of other birds.

Suppose that an evidence is a bird selected at random and that it is a black raven. Let us now calculate the probability of selecting a black raven under the hypothesis H, i.e. $P(E|H) = 100/1000100$, and the probability of selecting a black raven under the hypothesis $\neg H$, i.e. $P(E|\neg H) = 1000/1001001$. Clearly,

$$P(E|H) < P(E|\neg H). \qquad \text{(xxix)}$$

(xxix) means that selection of a black raven (E) is supporting more hypothesis ($\neg H$) "not all ravens are black" than hypothesis (H) "all ravens are black".

This example permits to understand that the Bayesian approach to the ravens' paradox is based on the estimation of probabilities of quite complex events, such as: the probability that a randomly selected raven is black under the hypothesis that all ravens are black, or the probability that a randomly selected raven is black under the hypothesis that not all ravens are black. This approach is completely not appropriate to characterization of decision rules induced from observations contained in a decision table where we have nothing similar to the two possible distributions of the populations of birds in hypothesis H and $\neg H$. To be more precise, within Bayesian approach, H and $\neg H$ are two different universes. This is simply meaningless within data analysis where we have only one universe represented by the available data.

Finally, from the viewpoint of the incremental confirmation considered within the rough set approach or, in general, within data analysis, the confirmation measure $c(\Phi,\Psi)$ depends on the number of "black ravens", "black non-ravens", "non-black ravens" and "non-black non-ravens" contained in the universe of discourse described in the decision table. Thus, within rough set approach, "black ravens", "black non-ravens", "non-black ravens" and "non black non-ravens" are confirmatory, disconfirmatory or irrelevant with respect to the decision rule $\Phi \to \Psi \equiv$"if x is a raven, then x is black" to the extent to which an increment of the respective number increases, decreases or does not influence $c(\Phi,\Psi)$. We think that this incremental view of the confirmation measure is very convincing because it does not need probability estimation (and thus avoids complex discussions about interpretation of probability; see for example Kyburg (1970)) and, instead, it makes use of some elementary parameters of the considered data set (numbers of objects satisfying some properties). We believe that for all these reasons the incremental view is the only acceptable interpretation of the confirmation measures for data analysis.

7. Rough set confirmation measures satisfying monotonicity

Theorem 2, presented in this section, answers the following question: which of the above confirmation measures do satisfy monotonicity property (M)?

Theorem 2. *$l(\Phi,\Psi)$, $f(\Phi,\Psi)$ and $s(\Phi,\Psi)$ satisfy* (M), *while $d(\Phi,\Psi)$, $r(\Phi,\Psi)$ and $b(\Phi,\Psi)$ do not satisfy* (M).

Proof. In the following we shall assume that modification Δ is positive. Let us start with $l(\Phi,\Psi)$ and modification of a. In this situation we have

$$\frac{a+\Delta}{a+b+\Delta} \geqslant \frac{a}{a+b} \Leftrightarrow (a+\Delta)(a+b) \geqslant a(a+b+\Delta) \Leftrightarrow b\Delta \geqslant 0. \quad \text{(xxx)}$$

Since, by hypothesis, $b\Delta \geqslant 0$ is always true, then (xxx) proves that $l(\Phi,\Psi)$ is not decreasing with respect to $a = \text{supp}_S(\Phi,\Psi)$. For modification of c, we have

$$\frac{c+\Delta}{c+d+\Delta} \geqslant \frac{c}{c+d} \Leftrightarrow (c+\Delta)(c+d) \geqslant c(c+d+\Delta) \Leftrightarrow d\Delta \geqslant 0. \quad \text{(xxxi)}$$

Since, by hypothesis, $d\Delta \geqslant 0$ is always true, then (xxxi) proves that $l(\Phi,\Psi)$ is not increasing with respect to $c = \text{supp}_S(\Phi,\neg\Psi)$. It is obvious, moreover, that $s(\Phi,\Psi)$ is not increasing (more exactly decreasing) with respect to b and not decreasing (more exactly increasing) with respect to d. Due to the fact that $l(\Phi,\Psi)$ and $f(\Phi,\Psi)$ are ordinally equivalent, the proof that (M) is satisfied by $l(\Phi,\Psi)$ is sufficient to conclude that (M) is satisfied also by $f(\Phi,\Psi)$. Let us remark that using arguments similar to those used in (xxx), one can prove that $s(\Phi,\Psi)$ is not decreasing with respect to $a = \text{supp}_S(\Phi,\Psi)$. For modification of b, we have

$$\frac{b+\Delta}{b+d+\Delta} \geqslant \frac{b}{b+d} \Leftrightarrow (b+\Delta)(b+d) \geqslant b(b+d+\Delta) \Leftrightarrow d\Delta \geqslant 0. \quad \text{(xxxii)}$$

Since, by hypothesis, $d\Delta \geqslant 0$ is always true, then (xxxii) proves that $s(\Phi,\Psi)$ is not increasing with respect to $b = \text{supp}_S(\Phi,\neg\Psi)$. It is obvious, moreover, that $s(\Phi,\Psi)$ is not increasing (more exactly decreasing) with respect to c and not decreasing (more exactly increasing) with respect to d. Now, we will prove by a counterexample that $d(\Phi,\Psi)$, $r(\Phi,\Psi)$ and $b(\Phi,\Psi)$ do not satisfy (M). Let us consider the case α in which

$$a = \text{supp}_S(\Phi,\Psi) = 100, \quad b = \text{supp}_S(\neg\Phi,\Psi) = 0,$$

$$c = \text{supp}_S(\Phi,\neg\Psi) = 0, \quad d = \text{supp}_S(\neg\Phi,\neg\Psi) = 1$$

and the case α' in which

$$a' = \text{supp}_S(\Phi',\Psi') = 101, \quad b' = \text{supp}_S(\neg\Phi',\Psi') = 0,$$

$$c' = \text{supp}_S(\Phi',\neg\Psi') = 0, \quad d' = \text{supp}_S(\neg\Phi',\neg\Psi') = 1.$$

We can easily verify that

$$d(\Phi,\Psi) = \frac{100}{100+0} - \frac{100+0}{100+0+0+1} > \frac{101}{101+0} - \frac{101+0}{101+0+0+1} = d(\Phi',\Psi')$$

which proves that confirmation measure $d(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $a = \text{supp}_S(\Phi,\Psi)$. We can also verify that

$$r(\Phi,\Psi) = \log\left[\frac{\left(\frac{100}{100+0}\right)}{\left(\frac{100+0}{100+0+0+1}\right)}\right] > \log\left[\frac{\left(\frac{101}{101+0}\right)}{\left(\frac{101+0}{101+0+0+1}\right)}\right] = r(\Phi',\Psi')$$

which proves that also confirmation measure $r(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $a = \mathrm{supp}_S(\Phi,\Psi)$. Finally, we can verify that

$$b(\Phi,\Psi) = \frac{100*1 - 0*0}{(100+0+0+1)^2} > \frac{101*1 - 0*0}{(101+0+0+1)^2} = b(\Phi',\Psi')$$

which proves that also confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $a = \mathrm{supp}_S(\Phi,\Psi)$. For the sake of completeness, we shall prove that while confirmation measure $d(\Phi,\Psi)$ and $r(\Phi,\Psi)$ satisfy the monotonicity property with respect to $b = \mathrm{supp}_S(\neg\Phi,\Psi)$, $c = \mathrm{supp}_S(\Phi,\neg\Psi)$ and $d = \mathrm{supp}_S(\neg\Phi,\neg\Psi)$, the confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $b = \mathrm{supp}_S(\neg\Phi,\Psi)$, $c = \mathrm{supp}_S(\Phi,\neg\Psi)$ and $d = \mathrm{supp}_S(\neg\Phi,\neg\Psi)$. For modification of b in $d(\Phi,\Psi)$, we have

$$\frac{a+(b+\Delta)}{a+(b+\Delta)+c+d} \geqslant \frac{a+b}{a+b+c+d} \Leftrightarrow (a+b+\Delta)(a+b+c+d) \geqslant (a+b)(a+b+\Delta+c+d) \Leftrightarrow (c+d)\Delta \geqslant 0. \quad \text{(xxxiii)}$$

Since, by hypothesis, $(c+d)\Delta \geqslant 0$ is always true, then (xxxiii) proves that $d(\Phi,\Psi)$ is not increasing with respect to $b = \mathrm{supp}_S(\neg\Phi,\Psi)$. For modification of c in $d(\Phi,\Psi)$, we also have

$$\frac{a}{a+(c+\Delta)} - \frac{a+b}{a+b+(c+\Delta)+d} \leqslant \frac{a}{a+c} - \frac{a+b}{a+b+c+d} \Leftrightarrow -b\Delta \leqslant 0. \quad \text{(xxxiv)}$$

Since, by hypothesis, $b\Delta \geqslant 0$ is always true, then (xxxiv) proves that $d(\Phi,\Psi)$ is not increasing with respect to $c = \mathrm{supp}_S(\Phi,\neg\Psi)$. It is obvious, moreover, that $d(\Phi,\Psi)$ is not decreasing (more exactly increasing) with respect to $d = \mathrm{supp}_S(\neg\Phi,\neg\Psi)$. Let us remark that (xxxiii) proves also that $r(\Phi,\Psi)$ is not increasing with respect to $b = \mathrm{supp}_S(\neg\Phi,\Psi)$. Analogously, (xxxiv) proves that $r(\Phi,\Psi)$ is not increasing with respect to $c = \mathrm{supp}_S(\Phi,\neg\Psi)$. It is obvious, moreover, that $r(\Phi,\Psi)$ is not decreasing (more exactly increasing) with respect to $d = \mathrm{supp}_S(\neg\Phi,\neg\Psi)$. To prove that confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $b = \mathit{supp}_S(\neg\Phi,\Psi)$, let us consider the case α^* in which

$$a^* = \mathrm{supp}_S(\Phi^*,\Psi^*) = 0, \quad b^* = \mathrm{supp}_S(\neg\Phi^*,\Psi^*) = 100,$$

$$c^* = \mathrm{supp}_S(\Phi^*,\neg\Psi^*) = 1, \quad d^* = \mathrm{supp}_S(\neg\Phi^*,\neg\Psi^*) = 0$$

and the case α^{**} in which

$$a^{**} = \mathrm{supp}_S(\Phi^{**},\Psi^{**}) = 0,$$

$$b^{**} = \mathrm{supp}_S(\neg\Phi^{**},\Psi^{**}) = 101,$$

$$c^{**} = \mathrm{supp}_S(\Phi^{**},\neg\Psi^{**}) = 1,$$

$$d^{**} = \mathrm{supp}_S(\neg\Phi^{**},\neg\Psi^{**}) = 0.$$

We can easily verify that

$$b(\Phi^*,\Psi^*) = \frac{0*0 - 100*1}{(0+100+1+0)^2} < \frac{0*0 - 101*1}{(0+101+1+0)^2} = b(\Phi^{**},\Psi^{**})$$

which proves that confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $b = \mathrm{supp}_S(\neg\Phi,\Psi)$. To prove that confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $c = \mathrm{supp}_S(\Phi,\neg\Psi)$, let us consider the case α° in which

$$a^\circ = \mathrm{supp}_S(\Phi^\circ,\Psi^\circ) = 0, \quad b^\circ = \mathrm{supp}_S(\neg\Phi^\circ,\Psi^\circ) = 1,$$

$$c^\circ = \mathrm{supp}_S(\Phi^\circ,\neg\Psi^\circ) = 100, \quad d^\circ = \mathrm{supp}_S(\neg\Phi^\circ,\neg\Psi^\circ) = 0$$

and the case $\alpha^{\circ\circ}$ in which

$$a^{\circ\circ} = \mathrm{supp}_S(\Phi^{\circ\circ},\Psi^{\circ\circ}) = 0,$$

$$b^{\circ\circ} = \mathrm{supp}_S(\neg\Phi^{\circ\circ},\Psi^{\circ\circ}) = 1,$$

$$c^{\circ\circ} = \mathrm{supp}_S(\Phi^{\circ\circ},\neg\Psi^{\circ\circ}) = 101,$$

$$d^{\circ\circ} = \mathrm{supp}_S(\neg\Phi^{\circ\circ},\neg\Psi^{\circ\circ}) = 0.$$

We can easily verify that

$$b(\Phi^\circ,\Psi^\circ) = \frac{0*0 - 1*100}{(0+1+100+0)^2} < \frac{0*0 - 1*101}{(0+1+101+0)^2} = b(\Phi^{\circ\circ},\Psi^{\circ\circ})$$

which proves that confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $c = \mathrm{supp}_S(\Phi,\neg\Psi)$. To prove that confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $d = \mathrm{supp}_S(\neg\Phi,\neg\Psi)$, let us consider the case $\alpha^\#$ in which

$$a^\# = \mathrm{supp}_S(\Phi^\#,\Psi^\#) = 1,$$

$$b^\# = \mathrm{supp}_S(\neg\Phi^\#,\Psi^\#) = 0,$$

$$c^\# = \mathrm{supp}_S(\Phi^\#,\neg\Psi^\#) = 0,$$

$$d^\# = \mathrm{supp}_S(\neg\Phi^\#,\neg\Psi^\#) = 100$$

and the case $\alpha^{\#\#}$ in which

$$a^{\#\#} = \mathrm{supp}_S(\Phi^{\#\#}, \Psi^{\#\#}) = 1,$$

$$b^{\#\#} = \mathrm{supp}_S(\neg\Phi^{\#\#}, \Psi^{\#\#}) = 0,$$

$$c^{\#\#} = \mathrm{supp}_S(\Phi^{\#\#}, \neg\Psi^{\#\#}) = 0,$$

$$d^{\#\#} = \mathrm{supp}_S(\neg\Phi^{\#\#}, \neg\Psi^{\#\#}) = 101.$$

We can easily verify that

$$b(\Phi^{\#}, \Psi^{\#}) = \frac{1*100 - 0*0}{(1+0+0+100)^2} > \frac{1*101 - 0*0}{(1+0+0+101)^2} = b(\Phi^{\#\#}, \Psi^{\#\#})$$

which proves that confirmation measure $b(\Phi,\Psi)$ does not satisfy the monotonicity property with respect to $d = \mathrm{supp}_S(\neg\Phi, \neg\Psi)$. □

The content of the above Theorem 2 is quite clear and immediate: among the rough set confirmation measures considered in the paper $l(\Phi, \Psi)$, $f(\Phi, \Psi)$ and $s(\Phi, \Psi)$ satisfy the monotonicity property (M), while $d(\Phi, \Psi)$, $r(\Phi, \Psi)$ and $b(\Phi, \Psi)$ do not satisfy property (M). However, a more detailed comment may be useful. From our viewpoint, the most important discovery coming from Theorem 2 is that the confirmation measure $d(\Phi, \Psi)$ does not satisfy the monotonicity property. The importance of this result is threefold:

(1) $d(\Phi, \Psi)$ is a very simple rough set confirmation measure, coherent with the definition of incremental confirmation; it is rather counterintuitive that $d(\Phi, \Psi)$ does not satisfy monotonicity, while other confirmation measures having as complex formulation as $l(\Phi, \Psi)$ and $f(\Phi, \Psi)$ do;
(2) $d(\Phi, \Psi)$ does not satisfy monotonicity with respect to $\mathrm{supp}_S(\Phi, \Psi)$; in this case the monotonicity property is indeed an uncontestable principle; using the terms of the Hempel's example, the property says that the more black ravens we see, the more the rule "all ravens are black" is confirmed;
(3) $d(\Phi, \Psi)$ is not ruled out by the symmetry/asymmetry test performed by Eells and Fitelson (2002); this means that the contribution of monotonicity property (M) in reducing the field of "coherent" confirmation measures is very relevant; in fact, the only confirmation measures which satisfy both symmetry/asymmetry properties of Eells and Fitelson and monotonicity property (M) are the two ordinally equivalent confirmation measures $l(\Phi, \Psi)$ and $f(\Phi, \Psi)$.

8. Conclusions

The answer to the question put in the title is positive. The main result of this paper states that, among the confirmation measures considered in the literature and recalled in Section 2, there are two confirmation measures satisfying the desirable properties of symmetry/asymmetry of Eells and Fitelson (2002), as well as our new monotonicity property (M): these are the two ordinally equivalent measures $l(\Phi, \Psi)$ and $f(\Phi, \Psi)$. In particular, our property (M) rules out $d(\Phi, \Psi)$. Let us remark that using the symmetry/asymmetry properties, it is not possible to discard $d(\Phi, \Psi)$, while using our monotonicity property, it is not possible to discard $s(\Phi, \Psi)$. This can be interpreted in the sense that the symmetry/asymmetry properties together with our monotonicity property (M) can be considered as complementary basic principles on which a sound theory of confirmation measures can be founded. A special attention merits, moreover, the violation of the monotonicity property by confirmation measure $b(\Phi, \Psi)$ which is the corroboration measure proposed by Carnap (1962). From our point of view, the violation of the monotonicity property by this confirmation measure is more troubling than its violation of the symmetry/asymmetry property proposed by Eells and Fitelson (2002). In fact Carnap (1962) liked that his corroboration measure b satisfies all four symmetry properties ES, HS, CS and TS, because he was interested in representing quantitatively a completely symmetric relevance relation. This means that in terms of Hempel's paradox, b measures the "correlation" between "black-ness" and "raven-ness". However, let us observe that this "correlation" between "black-ness" and "raven-ness" is very particular – in some situations it increases when one observes non-black ravens and black non-ravens, and it decreases when one observes black ravens and non-black non-ravens!

In addition to the new desirable monotonicity property proposed for confirmation measures in the rough set context, we compared the confirmation measures to several related issues, like independence (dependence) of logical formulas, interestingness measures in data mining and Bayesian solutions of raven's paradox. Justification of the monotonicity property required, moreover, to clarify relationships between logical (material) implications and decision rules. Hopefully, all this can contribute to better understanding of what decision rules mean and what quantitative measures are the most suitable for them.

We think that the quite theoretical results presented in this paper can be the basis for important operational development within rough set theory and, in general, within data analysis. Only to give some idea of interesting issues for future researches consider the use of measures $l(\Phi, \Psi)$ and $f(\Phi, \Psi)$ for assessing the interest of "*if* …, *then* …" decision rules induced from a data table, as well as classification with these rules. Considering the huge number of decision rules which can be induced from a data set, and the necessity of presenting only the most interesting rules to the users, this is a problem of primary importance for data analysis. We believe that the contribution of rough set confirmation measures to solving this problem is very important.

Acknowledgements

The research of the first author has been supported by the Italian Ministry of Education, University and Scientific Research (MIUR). The third author wishes to acknowledge financial support from the State Committee for Scientific Research (KBN).

References

Agrawall, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I., 1996. Fast discovery of association rules. In: Fayyad, U.M., Piatetsky–Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining. AAAI Press and MIT Press, Menlo Park and Cambridge, MA, pp. 307–328.

Carnap, R., 1962. Logical Foundations of Probability, 2nd Edition. University of Chicago Press, Chicago.

Christensen, D., 1999. Measuring confirmation. Journal of Philosophy XCVI, 437–461.

Earman, J., 1992. Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory. MIT Press, Cambridge, MA.

Edgington, D., 1995. On conditionals. Mind 104, 235–329.

Eells, E., 1982. Rational Decision and Causality. Cambridge University Press, Cambridge.

Eells, E., Fitelson, B., 2002. Symmetries and assymmetries in evidential support. Philosophical Studies 107 (2), 129–142.

Fayyad, U.M., Piatetsky–Shapiro, G., Smyth, P., Uthurusamy, R., 1996. Advances in Knowledge Discovery and Data Mining. AAAI Press, MIT Press, Menlo Park, Cambridge, MA.

Fitelson, B., 2001. Studies in Bayesian confirmation theory. Ph.D. Thesis, University of Wisconsin, Madison.

Fitelson, B., 2003. A probabilistic theory of coherence. Analysis 63 (3), 194–199.

Flach, P.A., 1995. Conjectures: an inquiry concerning the logic of induction. Ph.D. Thesis, Tilburg University.

Flach, P.A., Lachiche, N., 2001. Confirmation-guided discovery of first-order rules with tertius. Machine Learning 42, 61–95.

Gillies, D., 1986. In defense of the Popper–Miller argument. Philosophy of Science 53, 110–113.

Good, I.J., 1967. The white shoe is a red herring. British Journal for the Philosophy of Science 17, 322.

Good, I.J., 1984. The best explicatum for weight of evidence. Journal of Statistical Computation and Simulation 19, 294–299.

Greco, S., Pawlak, Z., Słowiński, R., 2002. Generalized decision algorithms, rough inference rules and flow graphs. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (Eds.), Rough Sets and Current Trends in Computing (RSCTC-2002) Lecture Notes in Artificial Intelligence, Vol. 2475. Springer, Berlin, pp. 93–104.

Hajek, P., Havranek, T., 1978. Mechanising Hypothesis Formulation: Mathematical Foundation for a General Theory. Springer, Berlin.

Heckerman, D., 1988. An axiomatic framework for belief updates. In: Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence 2. Elsevier Science Publishers, New York, pp. 11–22.

Hempel, C.G., 1945. Studies in the logic of confirmation (I). Mind 54, 1–26.

Hilderman, R.J., Hamilton, H.J., 2001. Knowledge Discovery and Measures of Interest. Kluwer Academic Publishers, Boston.

Horvitz, E., Heckerman, D., 1986. The inconsistent use of certainty measures in artificial intelligence research. In: Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence (1). Elsevier Science Publishers, New York, pp. 137–151.

Horwich, P., 1982. Probability and Evidence. Cambridge University Press, Cambridge.

Hosiasson-Lindenbaum, J., 1940. On confirmation. Journal of Symbolic Logic 5, 133–148.

Jeffrey, R., 1992. Probability and the Art of Judgment. Cambridge University Press, Cambridge.

Joyce, J., 1999. The Foundations of Causal Decision Theory. Cambridge University Press, Cambridge.

Kamber, M., Shingal, R., 1996. Evaluating the interestingness of characteristic rules. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, August, pp. 263–266.

Kemeny, J., Oppenheim, P., 1952. Degrees of factual support. Philosophy of Science 19, 307–324.

Keynes, J., 1921. A Treatise on Probability. Macmillan, London.

Kyburg, H., 1970. Probability and Inductive Logic. Macmillan, New York.

Kyburg, H., 1983. Recent work in inductive logic. In: Lucey, K.G., Machan, T.R. (Eds.), Recent Work in Philosophy. Rowman and Allanheld, Totowa, NJ, pp. 89–150.

Łukasiewicz, J., 1913. Die Logischen Grundlagen der Wahrscheinlichkeitsrechnung, Cracow; English translation. In: Borkowski, L. (Ed.), Jan Łukasiewicz—Selected Works. North-Holland, Polish Scientific Publishers, Amsterdam, Warsaw, pp. 16–63.

Mackie, J.L., 1963. The paradox of confirmation. The British Journal for the Philosophy of Science 13, 265–277.

Mackie, J., 1969. The relevance criterion of confirmation. The British Journal for the Philosophy of Science 20, 27–40.

Michalski, R.S., Bratko, I., Kubat, M., 1998. Machine Learning and Data Mining—Methods and Applications. Wiley, New York.

Milne, P., 1995. A Bayesian defence of Popperian science? Analysis 55, 213–215.

Milne, P., 1996. Log [P(h|e)/P(h|b)] is the one true measure of confirmation. Philosophy of Science 63, 21–26.

Nicod, J., 1923. Le problème de la logique de l'induction. Alcan, Paris.

Pawlak, Z., 1982. Rough sets. International Journal of Computer and Information Sciences 11, 341–356.

Pawlak, Z., 1991. Rough Sets—Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Boston.

Pawlak, Z., 2002. Rough sets, decision algorithms and Bayes' theorem. European Journal of Operational Research 136, 181–189.

Pawlak, Z., 2003. Flow graphs and decision algorithms. In: Wang, G., Liu, Q., Yao, Y.Y., Skowron, A. (Eds.), Proceedings of the Ninth International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC-2003) Lecture Notes in Artificial Intelligence, Vol. 22639. Springer, Berlin, pp. 1–11.

Pawlak, Z., 2004. Decision networks. In: Grzymala-Busse, J.W., Komorowski, J., Słowiński, R., Tsumoto, S. (Eds.), Rough Sets and Current Trends in Computing (RSCTC-2004), Lecture Notes in Artificial Intelligence, Vol. 3066. Springer, Berlin, pp. 1–8.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Francisco.

Pollard, S., 1999. Milne's measure of confirmation. Analysis 59, 335–337.

Popper, K.R., 1959. The Logic of Scientific Discovery. Hutchinson, London.

Rosenkrantz, R., 1994. Bayesian confirmation: paradise regained. The British Journal for the Philosophy of Science 45, 467–476.

Schlesinger, G., 1995. Measuring degrees of confirmation. Analysis 55, 208–212.

Schum, D., 1994. The Evidential Foundations of Probabilistic Reasoning. Wiley, New York.

Stalnaker, R., 1968. A theory of conditionals. In: Monograph Series, Vol. 2. Studies in Logical Theory, American Philosophical Quarterly. Blackwell, Oxford, pp. 98–112.

Tsumoto, S., 2002. Statistical test for rough set approximation based on Fisher's exact test. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (Eds.), Rough Sets and Current Trends in Computing (RSCTC-2002), Lecture Notes in Artificial Intelligence, Vol. 2475. Springer, Berlin, pp. 381–388.

Yao, Y.Y., Zhong, N., 1999. An analysis of quantitative measures associated with rules. In: Zhong, N., Zhou, L. (Eds.), Methodologies for Knowledge Discovery and Data Mining, Lecture Notes in Artificial Intelligence, Vol. 1574. Springer, Berlin, pp. 479–488.

Zembowicz, R., Zytkow, J., 1996. From contingency tables to various forms of knowledge in databases. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining. AAAI Press, The MIT Press, Menlo Park, Cambridge, MA, pp. 329–349.