i-lex. Scienze Giuridiche, Scienze Cognitive e Intelligenza artificiale
Rivista quadrimestrale on-line: www.i-lex.it
Novembre 2012, numero 17

A MATHEMATICAL MODEL OF RELIABILITY OF ARGUMENTS AND STORIES

Rainhard Z. Bengez*

Abstract. In this contribution we introduce a measure of reliability of arguments. The concept behind is that we combine the power of a mere logical analysis and structure with the power of probability theory. In this paper we start with the basic description of the measure. In a forthcoming one we will combine it with probability theory and statistic. This approach can be used to support blended learning in legal education, legal machine learning and legal informatics.

Key-words. Logic, logic and law, legal theory, law and probability theory; AI and law, machine learning, legal education.

1. Introduction

A lawyer's weapons are arguments because, first, she has to try to convince the opposing party and, finally, the decision makers. Judges as decision makers are committed to justify their decisions according to the general conventions rationally[1]. In a criminal court at a local court with lay assessors or in criminal courts with a grand jury the lawyer has to convince a group of decision makers of her viewpoint and interpretation of the entire situation and evidences, artifacts, and rules which leads compulsorily to the same interpretation of rights and duties. And finally, this is what legal argumentation is all about, to convince the decision makers of someone's interpretation of rights and duties in a given context and situation of a specific case.

At this point we have to ask about the meaning of a preferred perception of facts of a case and statements of affairs. Is it about the best consequences for society or for all parties thereto?

---

* TU München.

[1] Of course, there are many different opinions on how and esp. when such a rational formulation of arguments starts. But all of them could agree that at least the proclamation and written grounds are following certain conventions and rational practice.

Suggested citation: Rainhard Z. Bengez, A Mathematical Model of Reliability of Arguments and Stories in: i-lex, 17, 2012, pp.191-204 (www.i-lex.it)
Invited paper

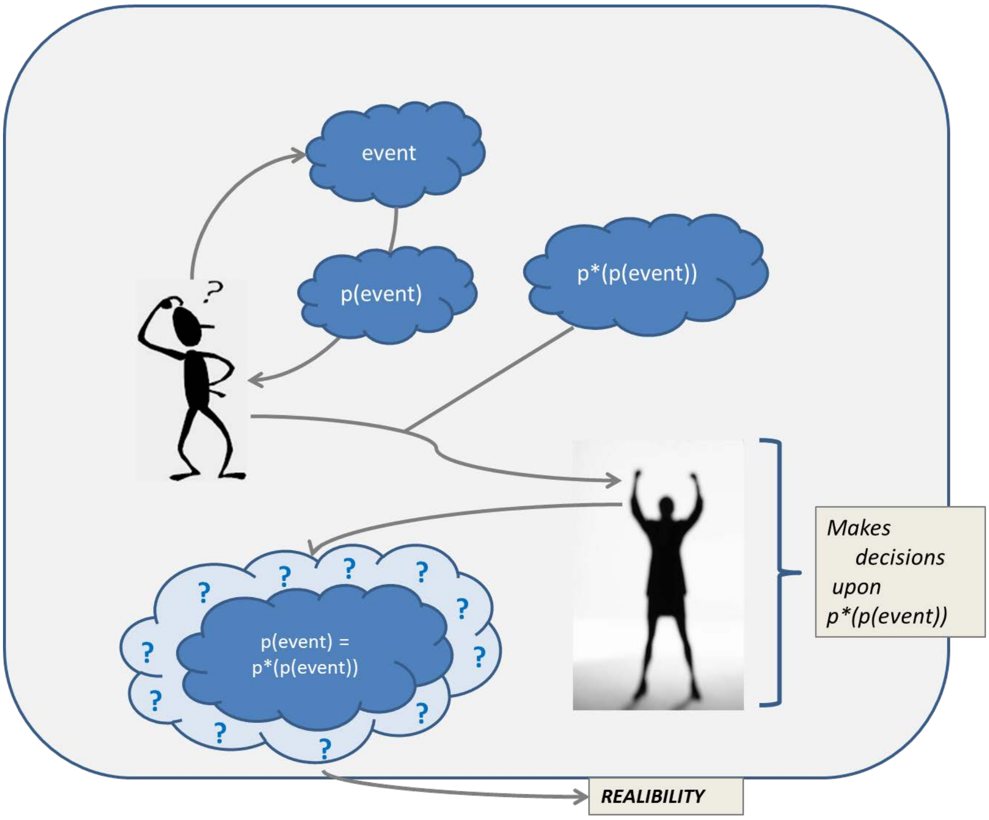

Fig. 1: Process of evaluation and decision making

2. Reliability of Arguments and Stories

Behind the general concepts of arguments and stories there is the mere question of trustworthiness and reliability (cf. figure 1); the first is matter of how we rate a person and the last of how we interpret given data and its representation. Among others arguments and stories there are components we evaluate in respect to its reliability. In this section we will introduce a mathematical model of reliability of arguments and stories. As both are using verbal communication and interpretation of data by using a natural language, we will embrace them as propositions. Thus, instead of speaking of arguments and/or stories, from now on we will speak about propositions and its evaluations. *By such an evaluation we mean the assignment of a truth value.* This can be of course a two-valued logic or also an assessment based on real numbers and associated metrics. In this paper we will restrict ourselves to two-valued logic. But the concept can be extended naturally to any kind of n-valued discrete or continuous logic and metrics. Our evaluation of a given proposition can be wrong. That means, the reliability we previously gave to statements has not been proven as maintainable.

We will call an evaluation of reliability of a proposition reliable, if (and only if) it is true under all circumstances, i.e. if the proposition is covering with what it states. In other words, if the truth value we assign to a statement is identical to that we have assigned to the state[2]. A statement is always given in a specific context and at a certain time. It is not necessarily time independent. During the remaining part of the paper we will introduce a stochastically based measure of the degree of reliability of statements, i.e. arguments and stories. The proposed measure of reliability will take into account the context of the proposition and its temporary character as well. Actually we have:

Definition 1: For us *Propositions* are arguments, statements, and stories, i.e. everything we can evaluate by a two-valued ore multi-valued logic or similar metrics. The associated logic or metrics[3] are reflecting a proposition's underlying truth-value concept. In this paper we rely on the standard two-valued truth value concept, but our approach is more general and not restricted to classical logic.

Definition 2: Evaluation & Reliability. Each proposition has two realities or domains, world0 and world1. In world1 a subject[4] interprets an event, i.e. she assigns truth-values[5] to a given proposition according to their subjective level of reliability. Our decision maker does not know what really happened in world0 where the proposition has been born as a coded reflection of a previous event. Now the crucial question arises:

[2] Not to come into philosophical or fundamental discussions on the nature of logic I will avoid calling it the real value. But if the reader do not make a philosophical plot out of it, we can say that an evaluation of a proposition is reliable if (and only if) the value we assign to the statement is the same as the real value the statements is based on.

[3] Fuzzy logic, for example would allow a different and more powerful approach to truth-value concepts, or according to some philosophers like Carnap a general stochastic framework. Such frameworks do not make sharp jumps between different truth values. They are offering ways to express transitions in a smoother way.

[4] For us this could be a human being or a bot making decisions based on several algorithms.

[5] Of course, this assignment presupposes a hidden concept of truth and belief.

Are the two different assignments of truth values similar? Our activity in world1 is called reliability and an associated proposition is reliable if both assignments in world0 and world1 are similar.

Observation 1: Reliability is time dependent and not necessarily time invariant. Any proposition evaluated in world1 depends on time and available data and our interpretation of it. In other words, it is very unlikely that we can make a time invariant decision about the reliability of a given task, because complexity of a task and the need of making a decision in a short period of time enforce us reducing complexity. Therefore, we have to select given data and to interpret it. On the other hand, each time we get new data we will start a reinterpretation. And this reinterpretation is based on individuals and society.

2.1 Construction of the Space of Possible Relations

To find an appropriate measure of the degrees of reliability of a proposition we have to start constructing a base of the relations between arguments and stories upon which we can derive and apply the measure. This base is what we call the space of possibilities or universe of discourse (cf. figure xx). This universe of discourse contains all possible propositions, i.e. arguments and/or stories of given context (case, situation, etc.). In a second step we will identify necessary and helpful substructures within this universe. These categories will divide the domain of relations among propositions (or statements) in seven sections (classes or categories).

Observation 2: The universe of discourse around a specific argumentative context is called

$$U := \{x \mid x \text{ is an acceptable proposition within a specific context}\}$$

and, of course, with *proposition* we just embrace arguments and stories suitable to convince decision makers. The pure context specific universe *U* can be split into categories around a selected proposition, i.e. by fixing a proposition we can create a relative structure based on this selected proposition(s). For this we need

Definition 3: The above mentioned fixed proposition p is called the *root* or *focus of reliability*.

The first four categories are:

(1) $$P_{\to p} := P(p)_{\to p} := \{x \mid x \to p\}$$

(2) $$P_{p\to} := P(p)_{p\to} := \{x \mid p \to x\}$$

(3) $$P_{\to\neg p} := P(p)_{\to\neg p} := \{x \mid x \to \neg p\}$$

(4) $$P_{\neg p\to} := P(p)_{\neg p\to} := \{x \mid \neg p \to x\}$$

These categories are symmetric with respect to the (logical) negation and based on the logical implication which can be seen (cf. Brewer) as an always important structure in argumentation. At least ex post we can rewrite everything into if-then-form. Associated with a fixed proposition **p** we only have included those categories where we can conclude **p** or where **p** is the base from where we conclude to associated arguments. The next step is to include those structures allowing us to go into both directions simultaneously, i.e. we are putting together arguments which can be seen as similar to our fixed proposition **p**.

$$(5) \qquad P_{\leftrightarrow} := P(p)_{\leftrightarrow} := \{x \mid x \leftrightarrow p\}$$
$$(6) \qquad P_{\leftrightarrow\perp} := P(p)_{\leftrightarrow\perp} := \{x \mid x \leftrightarrow \neg p\}$$

Now we have included categories mapping similar arguments or stories to a given proposition and those associated as base or conclusion as well. What remains is to add a category of all arguments around a specific context independently on the fixed proposition **p**, in other words another or new world of thinking about the situation.

We will call this category *IND* for all propositions independently on the fixed proposition **p**:

$$(7) \qquad IND := \{x \mid x \notin (P_{\to p} \cup P_{p\to} \cup P_{\to\neg p} \cup P_{\neg p\to} \cup P_{\leftrightarrow} \cup P_{\leftrightarrow\perp})\}$$

Altogether these seven categories are forming our space of all possible argumentation based relations within a given context. And it does not matter with which argument **p** we are starting. Stated otherwise, the entire space would always be the same, but the single categories will differ. Thus we can make

Observation 3: The space of all argumentative relations (abbreviated as ***A(p)***) is invariant (or independent, i.e. well-defined) for a given proposition **p,** but the arrangement and structure of each of the seven categories will differ as they are dependent on the fixed proposition **p** (cf. figure 2).

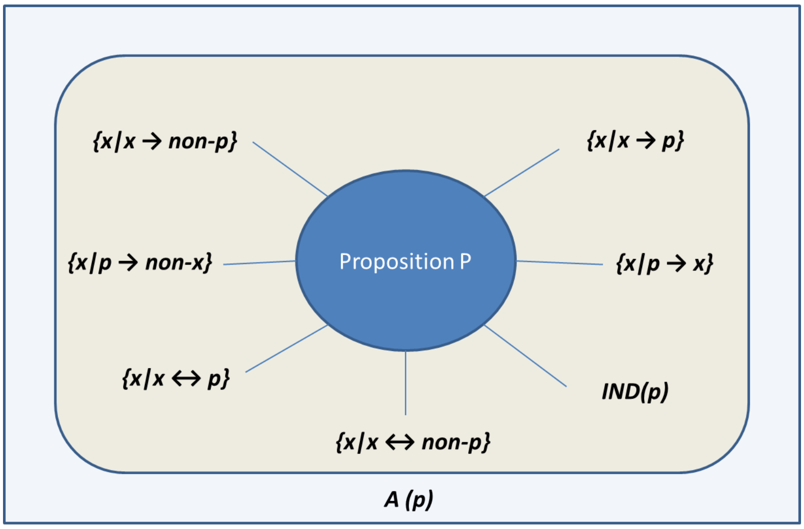

Fig. 2: The space A(p) of all propositions regarding a specific task

Now we have everything we need to answer the first question on the degree of reliability of a chosen proposition.

2.2 A metrics of reliability

In the following we will abbreviate the degree of reliability as *deg(p)*. Thus, we have

Definition 3: The degree of reliability of a given proposition (i.e. argument or story or statement) will be denoted by *deg(p).*

Based on this definition and the above introduced space of possibilities or events, we can derive a metrics of reliability. Our intention now is to take the logical structure – in this paper the two-dimensional or binary structure of two-valued logic as a base for a metrical measure. This measure can be extended to stochastics (beta distribution and Dirichlet distribution and Dirichlet process) with the two-valued truth functional concept. And, as stated above, this approach is not restricted to two-valued logic. It can be extended in a natural way to many valued and continuous logic. Such an approach would lead to a different family of probability distributions and processes. But this is a topic for other papers going deeper into AI and law and machine learning then the actual one.

But for now we are taking the above constructed space of all relations **A(p)** which has been fixed around a specific task and a selected proposition. For this emphasized proposition **p** we are interested in its degree of reliability. Stated otherwise, we try to find a number which expresses how good our decision was about the trustworthiness and acceptance of a given argument or story. We try to answer the question whether a given proposition does reflect the previous event in a rational way. For this we assign to each proposition a number based on its relation to other arguments in the given context.

Observation 4: The measure or metrics of reliability of arguments, stories, or statements (proposition for short) based on the above created space of events has the following formula:

$$\deg(p) = \frac{P_{\rightarrow p}}{P_{\rightarrow p} + P_{p \rightarrow} + IND}$$

Here we get a number expressing a propositions degree of reliability. Now, we can put all the arguments we have into this formula and assign a number. These numbers can be ordered according to their cardinality. For this task the following formulas are helpful.

2.3 Some derivations and further observations

Observation 5: $0 \leq \deg(p) \leq 1$

Observation 6: $\deg(p) + \deg(\neg p) \leq 1$

Observation 7: $\deg(p \vee q) + \deg(p \wedge q) = \deg(p) + \deg(q)$

Observation 8: $(p \rightarrow q) \rightarrow (\deg(p) \leq \deg(q))$

Observation 9: $\deg(p \wedge q) = \min\{0, \deg(p), \deg(q)\}$

Certainly, there are more than just these few observations, but we have restricted ourselves to them demonstrating the mere principle behind and what is possible to gain by using our approach. But what is it good for? Based on these observations we can try to find in a rational way the most reliable (or most promising) argument, story, or statement out of a given collection of specific and relevant propositions to support our (rational) decisions.

2.4 Example

First we just want to give a pure numerical example to get more familiar with the space **A(p)** and its relative categorical structure.

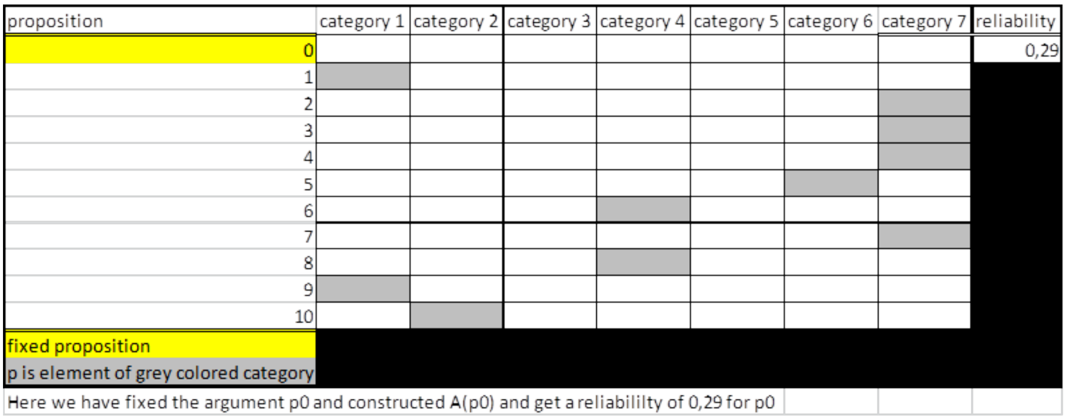

Fig.3: A(p0) and deg(p0).

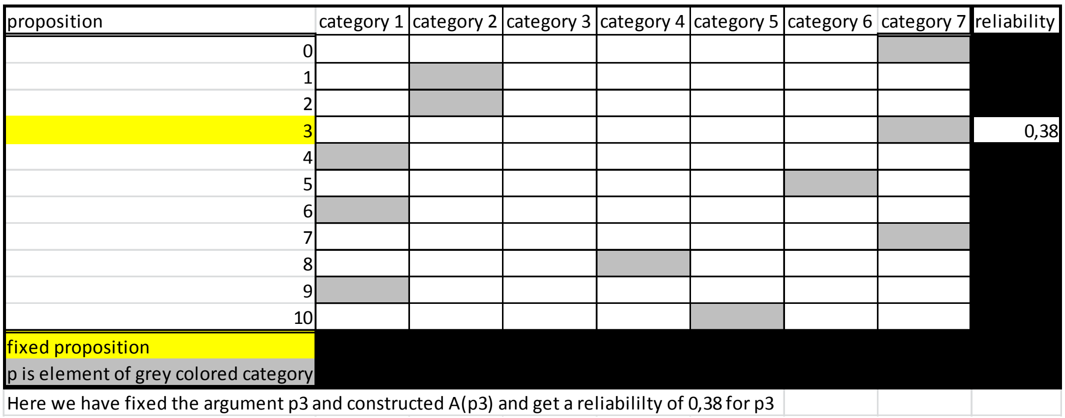

Fig.4: A(p3) and deg(p3).

From both tables we get reliability values deg(p0) and deg(p3). Because of deg(p0)< deg(p3) we can state that p3 is a more reliable proposition. Another application of this graphical method could be to evaluate the degree of a single proposition several times or by asking several groups of people. Both would give us different insights into sociological structures and learning effects of groups and/or machines.

3. Conclusion and prospect

What we have done was the introduction of a measure of reliability. But what is it good for? During the development we had three issues in mind; the first was legal education, the second machine learning and the third legal sociology and the research of the change of argumentation and its structures and acceptance. For all these issues we need a model which synthesizes logical approaches and approaches based on probability theory because of the underlying complexity. In the legal education we had in mind to combine it with approaches from Hajime Yoshino from Japan and/or Scott Brewer's logocratic method. In this context our measure can be helpful to find the most appropriate argument in a rational way. In the context of AI and law or, more generally, of machine learning this measure can be helpful in practicing abduction and finding a good measure to separate and combine new information. In the sociological domain this measure can be helpful to investigate in the changing acceptance of arguments and stories and find reasonable explanations or to unmask discrimination. Another field could be to apply it or its extended version to Vern R. Walker's logically structured case data base.

This sounds good, but can this small system really assist the bigger systems? Yes, we think so, because what makes our approach a little bit unique is that we start with simple 2-valued logic and we combine it with real numbers and later in a natural way with probability theory and the stochastic processes. Thus it is a very powerful framework paying attention to the structure of arguments and used logic without any need of a different system like Fuzzy logic or continuous logic, but it can also include it. The system we have just introduced here has to show how it can be integrated into a stochastic framework. This will happen by using beta distributions, Dirichlet distributions and Dirichlet processes. This means for the machine learning domain it can be easily combined with Bayes networks or Markov Logic. If one decides to use a different set of logic the associated distributions and processes can be different. In the next paper we will show how this approach fits together with the above mentioned distributions and how to apply it.

4. Proofs of observations

In this section we will proof the above made unproven observations.

4.1 Observation 3:

This is straightforward. Let's say that **A(p)** and **A(q)** are both spaces of all relations regarding to a specific context from where we have fixed two propositions **p** and **q**. If **p** is connected with **q** by implication or equivalence, then it would be in one of the collections (1), (2) or (5), otherwise we could find such a connection with **non-q** or **non-q**. If **q** and **p** are independent from each other, but within the same context, then category number seven "*IND*" would collect them. This is the way we can show that **A(p) = A(q),** if we neglect the categorical structure and think of **A** as a mere collection of propositions without any structure, and, therefore, that there is a unique space of relations **A** for each specific context.

Now, what remains open is to show that generally **A(p)** is not identical to **A(q)**, if we are taking the seven categories into account, i.e. if we are focusing on the specific content of the categories of both spaces **A(p)** and **A(q)**. If **p** and **q** are independent propositions, then **p** would be an element in *IND(q)* and vice versa. Hence, we have at least identified one category different in both spaces **A(p)** and **A(q)**. If **p** and **q** are not independent, then they could be associated equivalently or by implication. In the first case **A(p)** and **A(q)** are having the same categorical structure; in the second case we can find them in the antecedence or succedent category either. Therefore, they have different categories. Thus we have shown that in general **A(p)** and **A(q)** are not having the same categorical structure.

4.2 Observation 4:

What do we need to decide how reliable a given proposition is? If our proposition is reliable, then it should be a consequence of other accepted evidence based and rational decisions. This is what we mean if we are claiming that something makes sense in a logical way. It is nothing more than just a result of previous steps of reasoning and explanations. Thus we would look at ***P(→p)*** and counting how many elements are in this collection to say how many useful propositions we have. That means we are getting the number of the good and reasonable arguments.

We need a second number of all possible and useful arguments. Together with this number and the above constructed number of reasonable propositions we can set up our measure. What are useful propositions? Because we have fixed a proposition **p,** we can answer this question regarding to **p**. Now, it is helpful to take the space of all possible propositions concerning a specific context and mapping associated events **A(p)** which consists of seven categories. As we have fixed **p** we do not need **non-p**. Thus we can cancel category (3) and (4), and also category (5) and (6) as equivalence would mean a logical substitution, a tautology without any reasonable increase of information. What remains is category (1), (2) and (7) and they seem to be reasonable as category (7) collects all propositions (good and bad ones) logically independent on the fixed **p**, i.e. arguments based on different evidences, etc**; c**ategory (1) and (2) are conserving the **p** in the sense that it plays a role as cause or consequence. Therefore, our second number consists of these three categories.

Now we have the number of good and reliable propositions and the number of reasonable propositions. As the first one is part of the second it is very natural to build its proportion and that is how we get the measure of reliability. It is just the ratio of reliable arguments of the good and reasonable arguments.

But why do we need such a sophisticated stochastic-logic theory? Because it is a better mapping of what happened in reality than just analyzing the structure of arguments and persisting in the logical structure.

4.3 Observation 5:

It is quite clear that **deg(p)** is a positive number as it is defined over the cardinality of sets, and because the nominator is a part of the denominator it is straightforward to see that **deg(p)** must be less or equal to 1.

4.4 Observation 6:

For this we have to go back to the definition of **deg(p)** and to apply it to **non-p**. By combining both results by using set operation we will get the wished result.

a) Definition of

$$\deg(p) = \frac{P_{\to p}}{P_{\to p} + P_{p \to} + IND}$$

b) Definition of

$$\deg(\neg p) = \frac{P_{\to \neg p}}{P_{\to \neg p} + P_{\neg p \to} + IND}$$

c) Combining both definitions by using set operation

$$\begin{aligned}
\deg(p) + \deg(\neg p) &= \frac{P_{\to p}}{P_{\to p} + P_{p \to} + IND} + \frac{P_{\to \neg p}}{P_{\to \neg p} + P_{\neg p \to} + IND} \\
&= \frac{\{x \mid x \to p \vee x \to \neg p\}}{\{x \mid x \to p \vee x \to \neg p \vee p \to x \vee \neg p \to x \vee x \notin \bigcup P\}} \\
&= \frac{P_{\to p} + P_{\to \neg p}}{P_{\to p} + P_{p \to} + P_{\to \neg p} + P_{\neg p \to} + IND} \\
&\le 1
\end{aligned}$$

The last step (less or equal to 1) uses the same argument as in the proof of observation 5 that the nominator is part of the denominator and, therefore, the fraction can never be greater than 1.

4.5 Observation 7:

Here we will use again the definition of the degree of a proposition **p** and a second proposition **q**, combining both with logical connectors and by using set operation we can split them back into its original form.

a) $\deg(p \vee q)$

$$\deg(p \vee q) = \frac{\{x \mid x \to p \vee x \to q\}}{\{x \mid x \to p \vee x \to q \vee p \to x \vee q \to x \vee x \notin \bigcup P\}}$$

b) $\deg(p \wedge q)$

$$\deg(p \wedge q) = \frac{\{x \mid x \to p \wedge x \to q\}}{\{x \mid x \to p \vee x \to q \vee p \to x \vee q \to x \vee x \notin \bigcup P\}}$$

c) By combining both definitions we get:

$$\begin{aligned}\deg(p \vee q) + \deg(p \wedge q) &= \frac{\{x \mid x \to p \vee x \to q \vee (x \to p \wedge x \to q)\}}{\{x \mid x \to p \vee x \to q \vee p \to x \vee q \to x \vee x \notin \bigcup P\}} \\ &= \frac{\{x \mid x \to p\} \cup \{x \mid x \to q)\}}{\{x \mid x \to p \vee p \to x \vee x \notin \bigcup P\} \cup \{x \mid x \to q \vee q \to x \vee x \notin \bigcup P\}} \\ &= \frac{\{x \mid x \to p\}}{\{x \mid x \to p \vee p \to x \vee x \notin \bigcup P\}} + \frac{\{x \mid x \to q)}{\{x \mid x \to q \vee q \to x \vee x \notin \bigcup P\}} \\ &= \deg(p) + \deg(q)\end{aligned}$$

In the second step we used the logical argument that if we have **(A or (A and B))** that **A** is the dominant argument and, therefore, **(A or (A and B))** is logically equal to **A**. In the third step we used that if the denominator consists of both domains *domain(p*) and *domain(q),* we can neglect the non-corresponding set, e.g. if in the dominator there is **p,** then we can neglect the domain of **q** in the denominator of the fraction. Why? Because if **q** and **p** are independent, then the *domain(q)* would already be in *IND* of *domain(p),* otherwise **q** could be an antecedence or a succedent of **p** and, therefore, already in the domain of **p**. If **q** is neither independent on **p** nor its antecedence nor its succedent, then it would be equal to **p** and, therefore, it would be covered by *domain(p)*.

4.6 Observation 8:

Here we will proof the monotony of the degree of reliability. For this let **p** and **q** be two fixed propositions of specified context under the assumption that **p** is the antecedence of **q**. This means that **p** is part of the set {x| x → q} and **q** is in the set {x| p → x} respectively. The first set belongs to nominator and denominator of the definition of *deg(q),* whereas the second set belongs to the denominator of the definition of *deg(p)*.

We can hold, that if **p** and **q** would be logically equivalent, then ***deg(p) = deg(q).*** Thus we have to show *deg(p) < deg(q), if p → q*.

By using logical transformations and applying the definition to the new proposition p→q as well as observation 5 we get: $0 \le \deg(p \to q) = \deg(\neg p \vee q)$. Now we can use observation 7 to extend this result to:

$\deg(\neg p \vee q) + \deg(\neg p \wedge q) - \deg(\neg p \wedge q) = \deg(\neg p) + \deg(q) - \deg(\neg p \wedge q).$

By using observation 6 we get: $\deg(\neg p) + \deg(q) - \deg(\neg p \wedge q) \le 1 - \deg(p) + \deg(q) - \deg(\neg p \wedge q)$.

Now we can apply observation 1 to the last term and get

$1 - \deg(p) + \deg(q) - \deg(\neg p \wedge q) \le 1 - \deg(p) + \deg(q) - 1 = -\deg(p) + \deg(q)$

Thus we have 0 ≤ - deg(p) + deg(q) and, therefore, deg(p) ≤deq(q) and $(p \to q) \to (\deg(p) \le \deg(q))$ hence.

4.7 Observation 9:

Here we justify a simple formula based on the order of (real) numbers to calculate the degree of two logically connected arguments.

If **p** and **q** are propositions without having anything in common, then it is straightforward that its logical and-connection must be false and therefore the associated set empty. The empty set in the nominator of the definition means 0 dived by a number and that is zero, again. If the domains of **p** and **q** have arguments, i.e. elements, in common and we connect them by using the and-connection, then we omit all the remaining arguments, i.e. in the nominator of the joint degree the nominator of the smallest set dominates, whereas the denominator is the same – thus the smallest degree wins. In other words, if proposition **p** and **q** have nothing in common, then the joint degree is 0 and otherwise the joint degree is identical to the smallest one of both.

Thus we have: $\deg(p \wedge q) = \min\{0, \deg(p), \deg(q)\}$

References

BREWER, S. (2011), *Logocratic method and the analysis of arguments in evidence*, in: Law, Probability and Risk 10(3): pp. 175-202

YOSHINO, H. *Creative Legal Mind Project*, in: http://www.meijigakuin.ac.jp/~yoshino/en/less/abstract.html

VERN R. WALKER, *Law, Logic, and Technology Research Laboratory*, in: http://www.lltlab.org/