Exploring Different Types of Trust Propagation

Audun Jøsang[1], Stephen Marsh[2], and Simon Pope[3]

[1] Queensland University of Technology,
Brisbane, Australia
a.josang@qut.edu.au
[2] National Research Council,
Fredericton, Canada
Stephen.Marsh@nrc-cnrc.gc.ca
[3] Defence Science and Technology Organisation,
Adelaide, Australia
Simon.Pope@dsto.defence.gov.au

Abstract. Trust propagation is the principle by which new trust relationships can be derived from pre-existing trust relationship. Trust transitivity is the most explicit form of trust propagation, meaning for example that if Alice trusts Bob, and Bob trusts Claire, then by transitivity, Alice will also trust Claire. This assumes that Bob recommends Claire to Alice. Trust fusion is also an important element in trust propagation, meaning that Alice can combine Bob's recommendation with her own personal experience in dealing with Claire, or with other recommendations about Claire, in order to derive a more reliable measure of trust in Claire. These simple principles, which are essential for human interaction in business and everyday life, manifests itself in many different forms. This paper investigates possible formal models that can be implemented using belief reasoning based on subjective logic. With good formal models, the principles of trust propagation can be ported to online communities of people, organisations and software agents, with the purpose of enhancing the quality of those communities.

1 Introduction

Trust is a phenomenon that only exists among living species equipped with advanced cognitive faculties. One usually considers the appreciation of trust to be a purely human characteristic, but it would be arrogant to exclude animals. When assuming that software agents can equipped with capabilities to reason about trust, risk assessment and decision making, one can talk about artificial trust. There is a rapidly growing growing literature on this topic [2, 3, 12, 19].

What humans perceive through their senses is a more or less distorted version of a reality which they assume exists. A considerable part of human science consists of modelling aspects of the world for the purpose of understanding, prediction and control. When trying to make statements about the assumed world, we actually make statements about the subjective perceived world. However, most reasoning models are designed for the assumed reality, not for the perceived reality.

A quite different approach would be to design a reasoning model for the perceived world. A key component of such a model is to include uncertainty resulting from partial

K. Stølen et al. (Eds.): iTrust 2006, LNCS 3986, pp. 179–192, 2006.

ignorance. Several alternative calculi and logics which include degrees of uncertainty have been proposed and with some success applied to practical problems [4, 20]. The problem with many of the proposals has been that the calculi diverge considerably from standard probability calculus and therefore have received relatively little acceptance. A second key component of a model for the perceived world is to accept the fact that every belief is individual.

Subjective logic, which will be described here, takes both the uncertainty and individuality of beliefs into account while still being compatible with standard logic and probability calculus. The migration from the assumed towards the perceived world is achieved by adding an uncertainty dimension to the single valued probability measure, and by taking the individuality of beliefs into account.

A distinction can be made between interpreting trust as a belief about the reliability of an object, and as a decision to depend on an object [14]. In this paper, trust is interpreted in the former sense, as a belief about reliability. As a calculus of beliefs, subjective logic can therefore be used for trust reasoning. Although this model can never be perfect, and able to reflect all the nuances of trust, it can be shown to respect the main intuitive properties of trust and trust propagation.

As soon as one attempts to perform computations with input parameters in the form of subjective trust measures, parameter dependence becomes a major issue. If Alice for example wants to know whether tomorrow will be sunny, she can ask her friends, and if they all say it will be sunny she will start believing the same. However, her friends might all have based their opinions on the same weather-forecast, so their opinions are dependent, and in that case, asking only one of them would be sufficient. It would in fact be wrong of Alice to take all her friends' opinions into account as being independent, because it would strengthen her opinion without any good reason. Being able to identify cases of dependent opinions is therefore important, but alas difficult.

2 Trust Modeling with Subjective Logic

Subjective logic is a belief calculus specifically developed for modeling trust relationships. In subjective logic, beliefs are represented on binary state spaces, where each of the two possible states can consist of sub-states. Belief functions on binary state spaces are called *subjective opinions* and are formally expressed in the form of an ordered tuple $\omega_x^A = (b, d, u, a)$, where b, d, and u represent belief, disbelief and uncertainty respectively where $b, d, u \in [0, 1]$ and $b + d + u = 1$. The base rate parameter $a \in [0, 1]$ represents the base rate probability in the absence of evidence, and is used for computing an opinion's probability expectation value $\mathrm{E}(\omega_x^A) = b + au$, meaning that a determines how uncertainty shall contribute to $\mathrm{E}(\omega_x^A)$. A subjective opinion is interpreted as an agent A's belief in the truth of statement x. Ownership of an opinion is represented as a superscript so that for example A's opinion about x is denoted as ω_x^A.

Subjective opinions are equivalent to beta PDFs (probability density functions) denoted by $\mathrm{beta}\,(\alpha, \beta)$ [1]. The beta class of density functions express probability density over the same binary event spaces as for subjective opinions, and this is also the basis for their equivalence.

Let r and s express the number of positive and negative past observations respectively, and let a express the *a priori* or base rate, then α and β can be determined as:

$$\alpha = r + 2a\ , \qquad \beta = s + 2(1-a)\ . \tag{1}$$

The following bijective mapping between the opinion parameters and the beta PDF parameters can be determined analytically [5, 17].

$$\begin{cases} b_x = r/(r+s+2) \\ d_x = s/(r+s+2) \\ u_x = 2/(r+s+2) \\ a_x = \text{base rate of } x \end{cases} \Longleftrightarrow \begin{cases} r = 2b_x/u_x \\ s = 2d_x/u_x \\ 1 = b_x + d_x + u_x \\ a = \text{base rate of } x \end{cases} \tag{2}$$

Without evidence, the base rate alone determines the probability distribution. As more evidence becomes available, the influence of the base rate diminishes, until the evidence alone determines the probability distribution. In order to separate between base rate and evidence in the beta PDF, we define the *augmented beta PDF* notation below.

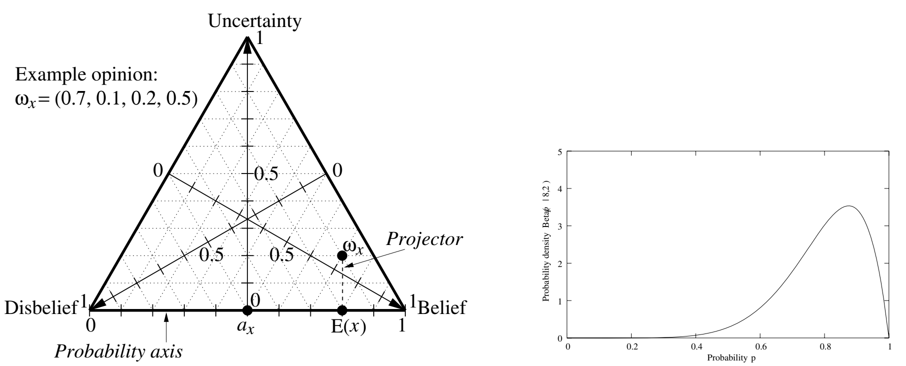
Definition 1 (Augmented Beta PDF Notation). *Let the* a priori *beta PDF as a function of the base rate* a*, without evidence, be expressed as* $\mathrm{beta}(2a,\ 2(1-a))$*. Let the* a posteriori *beta PDF with positive evidence* r *and negative evidence* s *be expressed as* $\mathrm{beta}(r+2a,\ s+2(1-a))$*. The augmented beta PDF with the 3 parameters* (r, s, a) *is then simply written as* $\varphi(r, s, a)$*, defined by:*

$$\varphi(r, s, a) = \text{beta}(r + 2a,\ s + 2(1-a))\ . \tag{3}$$

Opinions can be mapped into the interior of an equal-sided triangle, and augmented beta PDFs can be visualised as 2D plots, as illustrated in Fig.1.

Fig.1 illustrates the example of a subjective opinion $\omega_x = (0.7,\ 0.1,\ 0.2,\ 0.5)$, and the corresponding equivalent augmented beta PDF $\varphi(7, 1, \frac{1}{2})$.

The fact that subjective logic is compatible with binary logic and probability calculus means that whenever corresponding operators exist in probability calculus, the probability expectation value $\mathrm{E}(\omega)$ of an opinion ω that has been derived with subjective



(a) Opinion $\omega_x = (0.7,\ 0.1,\ 0.2,\ 0.5)$ (b) Augm. beta: $\varphi(7, 1, \frac{1}{2})$

Fig. 1. Example equivalent subjective opinion and beta PDF

logic, is always equal to the probability value that would have been derived had simple probability calculus been applied. Similarly, whenever corresponding binary logic operators exist, an absolute opinion (i.e. equivalent to binary logic TRUE or FALSE) derived with subjective logic, is always equal to the truth value that can be derived with binary logic.

Subjective logic has a sound mathematical basis and is compatible with binary logic and traditional Bayesian analysis. Subjective logic defines a rich set of operators for combining subjective opinions in various ways [5, 6, 7, 8, 9, 10, 11, 12, 15, 16, 18]. Some operators represent generalisations of binary logic and probability calculus, whereas others are unique to belief calculus because they depend on belief ownership. With belief ownership it is possible to explicitly express that different agents have different opinions about the same issue.

The advantage of subjective logic over probability calculus and binary logic is its ability to explicitly express and take advantage of ignorance and belief ownership. Subjective logic can be applied to all situations where probability calculus can be applied, and to many situations where probability calculus fails precisely because it can not capture degrees of ignorance. Subjective opinions can be interpreted as probability density functions, making subjective logic a simple and efficient calculus for probability density functions. An online demonstration of subjective logic can be accessed at: http://www.fit.qut.edu.au/∼josang/sl/.

3 Trust Fusion

3.1 Fusion of Independent Trust

This operator is most naturally expressed in the evidence space, so we will define it there first and subsequently map it over to the opinion space.

Definition 2 (Consensus Operator for Independent Beta PDFs). *Let $\varphi(r_x^A, s_x^A, a_x^A)$ and $\varphi(r_x^B, s_x^B, a_x^B)$ be two augmented beta PDFs respectively held by the agents A and B regarding the trustworthiness of x. The augmented beta PDF $\varphi(r_x^{A\diamond B}, s_x^{A\diamond B}, a_x^{A\diamond B})$ defined by*

$$\begin{cases} r_x^{A\diamond B} &= r_x^A + r_x^B \\ s_x^{A\diamond B} &= s_x^A + s_x^B \\ a_x^{A\diamond B} &= \frac{a_x^A(r_x^A+s_x^A)+a_x^B(r_x^B+s_x^B)}{r_x^A+r_x^B+s_x^A+s_x^B} \end{cases}$$

is then called the consensus of A's and B's estimates, as if it was an estimate held by an imaginary agent $[A, B]$. By using the symbol $\oplus$ to designate this operation, we get $\varphi(r_x^{A\diamond B}, s_x^{A\diamond B}, a_x^{A\diamond B}) = \varphi(r_x^A, s_x^A, a_x^A) \oplus \varphi(r_x^B, s_x^B, a_x^B)$.

The consensus rule for combining independent opinions is easily obtained by using Def.2 above and the evidence-opinion mapping of Eq.(2).

Theorem 1 (Consensus Operator for Independent Opinions). *Let $\omega_x^A = (b_x^A, d_x^A, u_x^A, a_x^A)$ and $\omega_x^B = (b_x^B, d_x^B, u_x^B, a_x^B)$ be trust in x from A and B respectively. The opinion $\omega_x^{A\diamond B} = (b_x^{A\diamond B}, d_x^{A\diamond B}, u_x^{A\diamond B}, a_x^{A\diamond B})$ is then called the consensus between ω_x^A and ω_x^B, denoting the trust that an imaginary agent $[A, B]$ would have in x, as if* *that agent represented both A and B. In case of Bayesian (totally certain) opinions, their relative weight can be defined as $\gamma^{A/B} = \lim(u_x^B/u_x^A)$.*

Case I: $u_x^A + u_x^B - u_x^A u_x^B \neq 0$

$$\begin{cases} b_x^{A\diamond B} = \frac{b_x^A u_x^B + b_x^B u_x^A}{u_x^A + u_x^B - u_x^A u_x^B} \\ d_x^{A\diamond B} = \frac{d_x^A u_x^B + d_x^B u_x^A}{u_x^A + u_x^B - u_x^A u_x^B} \\ u_x^{A\diamond B} = \frac{u_x^A u_x^B}{u_x^A + u_x^B - u_x^A u_x^B} \\ a_x^{A\diamond B} = \frac{a_x^A u_x^B + a_x^B u_x^A - (a_x^A + a_x^B) u_x^A u_x^B}{u_x^A + u_x^B - 2u_x^A u_x^B} \end{cases}$$

Case II: $u_x^A + u_x^B - u_x^A u_x^B = 0$

$$\begin{cases} b_x^{A\diamond B} = \frac{(\gamma^{A/B} b_x^A + b_x^B)}{(\gamma^{A/B} + 1)} \\ d_x^{A\diamond B} = \frac{(\gamma^{A/B} d_x^A + d_x^B)}{(\gamma^{A/B} + 1)} \\ u_x^{A\diamond B} = 0 \\ a_x^{A\diamond B} = \frac{\gamma^{A/B} a_x^A + a_x^B}{\gamma^{A/B} + 1} \end{cases} .$$

By using the symbol '$\oplus$' to designate this operator, we can write $\omega_x^{A\diamond B} = \omega_x^A \oplus \omega_x^B$.

It can be shown that $\oplus$ is both commutative and associative which means that the order in which opinions are combined has no importance. Opinion independence must be assured, which obviously translates into not allowing an entity's opinion to be counted more than once.

The effect of independent consensus is to reduce uncertainty. For example the case where several witnesses give consistent testimony should amplify the judge's opinion, and that is exactly what the operator does. Consensus between an infinite number of not totally uncertain (i.e. $u < 1$) opinions would necessarily produce a consensus opinion with $u = 0$. Fig.2 illustrates an example of applying the consensus operator for independent opinions where $\omega_x^A = \{0.8, 0.1, 0.1, a\}$ and $\omega_x^B = \{0.1, 0.8, 0.1, a\}$, so that $\omega_x^{A\diamond B} = \omega_x^A \oplus \omega_x^B = \{0.47, 0.47, 0.06, a\}$.

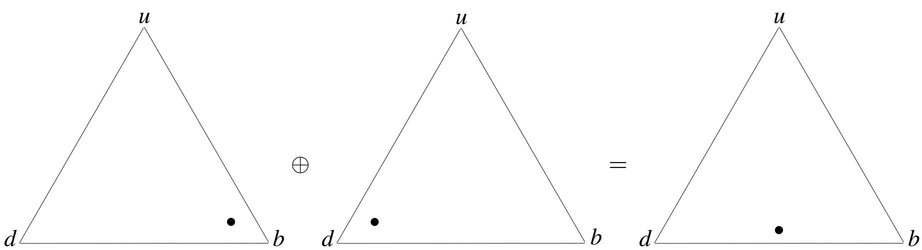


Fig. 2. Example of applying the consensus operator for fusing independent trust

3.2 Fusion of Dependent Trust

Assume two agents A and B having simultaneously observed the same process. Because their observations are identical, their respective opinions will necessarily be dependent, and a consensus according to Def.2 would be meaningless.

If the two observers have made exactly the same observations, and their estimates are equal, it is sufficient to take only one of the estimates into account. However, although two observers witness the same phenomenon, it is possible (indeed, likely) that they record and interpret it differently. The observers may have started and ended the observations at slightly different times, one of them may have missed or misinterpreted some of the events, resulting in varying, but still dependent opinions.

We will define a consensus rule for dependent beta PDFs based on the average of recorded positive and negative observations. Let two dependent augmented beta PDFs be $\varphi(r_x^A, s_x^A, a_x^A)$ and $\varphi(r_x^B, s_x^B, a_x^B)$, then we define the consensus estimate by the average of their parameters as $\varphi(\frac{r_x^A+r_x^B}{2}, \frac{s_x^A+s_x^B}{2}, \frac{a_x^A+a_x^B}{2})$. The general expression for the consensus between n dependent augmented beta PDFs can be defined as follows:

Definition 3 (Consensus Operator for Dependent Beta PDFs). *Let $\varphi(r_x^{A_i}, s_x^{A_i}, a_x^{A_i})$, where $i \in [1, n]$, be n dependent augmented beta PDFs respectively held by the agents $A_1, ..., A_n$ about the proposition x. The depended consensus beta PDF is then $\varphi(r_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}, s_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}, a_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n})$, where:*

$$\begin{cases} r_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \frac{\sum_1^n r_x^{A_i}}{n} \\ \\ s_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \frac{\sum_1^n s_x^{A_i}}{n} \\ \\ a_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \frac{\sum_1^n a_x^{A_i}}{n} \end{cases}$$

By using the symbol $\underline{\oplus}$ to designate this operation, we get
$\varphi(r_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}, s_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}, a_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}) = \varphi(r_x^{A_1}, s_x^{A_1}, a_x^{A_1}) \underline{\oplus} \ldots \underline{\oplus} \varphi(r_x^{A_n}, s_x^{A_n}, a_x^{A_n})$. □

The corresponding consensus operator is obtained by applying Eq.(2) to Def.3.

Theorem 2 (Consensus Operator for Dependent Opinions). *Let $\omega_x^{A_i} = \{b_x^{A_i}, d_x^{A_i}, u_x^{A_i}, a_x^{A_i}\}$ where $i \in [1, n]$, be n dependent opinions respectively held by agents $A_1, ..., A_n$ about the same proposition x. The depended consensus is then $\omega_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \{b_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}, d_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}, u_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}, a_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n}\}$, where:*

$$\begin{cases} b_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \frac{\sum_1^n (b_x^{A_i}/u_x^{A_i})}{\sum_1^n (b_x^{A_i}/u_x^{A_i}) + \sum_1^n (d_x^{A_i}/u_x^{A_i}) + n} \\ \\ d_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \frac{\sum_1^n (d_x^{A_i}/u_x^{A_i})}{\sum_1^n (b_x^{A_i}/u_x^{A_i}) + \sum_1^n (d_x^{A_i}/u_x^{A_i}) + n} \\ \\ u_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \frac{n}{\sum_1^n (b_x^{A_i}/u_x^{A_i}) + \sum_1^n (d_x^{A_i}/u_x^{A_i}) + n} \\ \\ a_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \frac{\sum_1^n a_x^{A_i}}{n} \end{cases}$$

where all the $u_x^{A_i}$ are different from zero. By using the symbol $\underline{\oplus}$ to designate this operation, we get $\omega_x^{A_1 \underline{\diamond} \ldots \underline{\diamond} A_n} = \omega_x^{A_1} \underline{\oplus} \ldots \underline{\oplus} \omega_x^{A_n}$.

The $\underline{\oplus}$ operator is both commutative and associative. The effect of the dependent consensus operator is to produce an opinion which is based on an average of positive and an average of negative evidence. Fig.3 illustrates an example of applying the consensus operator for dependent opinions where $\omega_x^A = \{0.8, 0.1, 0.1, a\}$ and $\omega_x^B = \{0.1, 0.8, 0.1, a\}$, so that $\omega_x^{A\underline{\diamond}B} = \omega_x^A \underline{\oplus} \omega_x^B = \{0.45,\ 0.45,\ 0.10,\ a\}$.

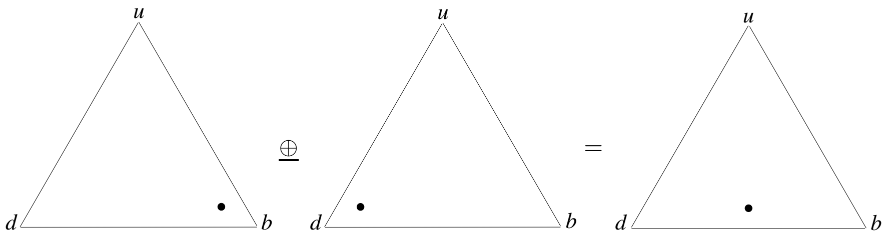


Fig. 3. Example of applying the consensus operator for dependent opinions

3.3 Fusion of Trust Under Partial Dependence

Let two agents A and B observed the same process during two partially overlapping periods. If it is known exactly which events were observed by both, one of the agents can simply dismiss these observations, and their opinions will be independent. However, it may not always be possible to determine which observations are identical.

Fig.4 illustrates a situation of partly dependent observations. Assuming that the fraction of overlapping observations is known, the dependent and the independent parts of their observations can be estimated, so that a consensus operator can be defined [13].

In the figure, $\omega_x^{A\mathrm{i}(B)}$ and $\omega_x^{B\mathrm{i}(A)}$ represent the independent parts of A and B's opinions, whereas $\omega_x^{A\mathrm{d}(B)}$ and $\omega_x^{B\mathrm{d}(A)}$ represent their dependent parts.

Let φ_x^A's fraction of dependence with φ_x^B and vice versa be represented by the dependence factors $\lambda_x^{A\mathrm{d}(B)}$ and $\lambda_x^{B\mathrm{d}(A)}$. The dependent and independent augmented betas can then be defined as a function of the dependence factors.

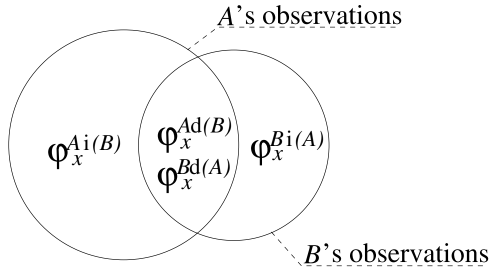


Fig. 4. Beta PDFs based on partly dependent observations

$$\varphi_x^{\mathrm{A}i(B)}:\begin{cases} r_x^{\mathrm{A}i(B)} = r_x^A(1-\lambda_x^{\mathrm{A}d(B)}) \\ s_x^{\mathrm{A}i(B)} = s_x^A(1-\lambda_x^{\mathrm{A}d(B)}) \end{cases} \qquad \varphi_x^{B\mathrm{i}(A)}:\begin{cases} r_x^{B\mathrm{i}(A)} = r_x^B(1-\lambda_x^{B\mathrm{d}(A)}) \\ s_x^{B\mathrm{i}(A)} = s_x^B(1-\lambda_x^{B\mathrm{d}(A)}) \end{cases}$$

$$\varphi_x^{A\mathrm{d}(B)}:\begin{cases} r_x^{A\mathrm{d}(B)} = r_x^A\lambda_x^{A\mathrm{d}(B)} \\ s_x^{A\mathrm{d}(B)} = s_x^A\lambda_x^{A\mathrm{d}(B)} \end{cases} \qquad \varphi_x^{B\mathrm{d}(A)}:\begin{cases} r_x^{B\mathrm{d}(A)} = r_x^B\lambda_x^{B\mathrm{d}(A)} \\ s_x^{B\mathrm{d}(A)} = s_x^B\lambda_x^{B\mathrm{d}(A)} \end{cases} \tag{4}$$

The cumulative fusion of partially dependent beta PDFs can then be defined as a function of the dependent and independent parts.

Definition 4 (Consensus Operator for Partially Dependent Beta PDFs). *Let φ_x^A and φ_x^B be two augmented beta PDFs respectively held by the agents A and B regarding the trustworthiness of x. We will use the symbol $\widetilde{\oplus}$ to designate consensus between partially dependent augmented betas. As before $\underline{\oplus}$ is the operator for entirely dependent augmented betas. The consensus of A and B's augmented betas can then be written as:*

$$\begin{aligned} \varphi_x^A \widetilde{\oplus} \varphi_x^B &= \varphi_x^{A\widetilde{\diamond}B} \\ &= \varphi_x^{(A\mathrm{d}(B)\underline{\diamond}B\mathrm{d}(A))\diamond A\mathrm{i}(B)\diamond B\mathrm{i}(A)} \\ &= (\varphi_x^{A\mathrm{d}(B)} \underline{\oplus} \varphi_x^{B\mathrm{d}(A)}) \oplus \varphi_x^{A\mathrm{i}(B)} \oplus \varphi_x^{B\mathrm{i}(A)} \end{aligned} \tag{5}$$

The equivalent representation of dependent and independent opinions can be obtained by using Eq.(4) and the evidence-opinion mapping Eq.(2). The reciprocal dependence factors are as before denoted by $\lambda^{A\mathrm{d}(B)}$ and $\lambda^{B\mathrm{d}(A)}$.

$$\omega_x^{A\mathrm{i}(B)}:\begin{cases} b_x^{A\mathrm{i}(B)} = b_x^A\mu_x^{A\mathrm{i}(B)} \\ d_x^{A\mathrm{i}(B)} = d_x^A\mu_x^{A\mathrm{i}(B)} \\ u_x^{A\mathrm{i}(B)} = u_x^A\mu_x^{A\mathrm{i}(B)}/(1-\lambda_x^{A\mathrm{d}(B)}) \end{cases}, \qquad \mu_x^{A\mathrm{i}(B)} = \frac{1-\lambda_x^{A\mathrm{d}(B)}}{(1-\lambda_x^{A\mathrm{d}(B)})(b_x^A+d_x^A)+u_x^A}$$

$$\omega_x^{A\mathrm{d}(B)}:\begin{cases} b_x^{A\mathrm{d}(B)} = b_x^A\mu_x^{A\mathrm{d}(B)} \\ d_x^{A\mathrm{d}(B)} = d_x^A\mu_x^{A\mathrm{d}(B)} \\ u_x^{A\mathrm{d}(B)} = u_x^A\mu_x^{A\mathrm{d}(B)}/\lambda_x^{A\mathrm{d}(B)} \end{cases}, \qquad \mu_x^{A\mathrm{d}(B)} = \frac{\lambda_x^{A\mathrm{d}(B)}}{\lambda_x^{A\mathrm{d}(B)}(b_x^A+d_x^A)+u_x^A}$$

$$\omega_x^{B\mathrm{i}(A)}:\begin{cases} b_x^{B\mathrm{i}(A)} = b_x^B\mu_x^{B\mathrm{i}(A)} \\ d_x^{B\mathrm{i}(A)} = d_x^B\mu_x^{B\mathrm{i}(A)} \\ u_x^{B\mathrm{i}(A)} = u_x^B\mu_x^{B\mathrm{i}(A)}/(1-\lambda_x^{B\mathrm{d}(A)}) \end{cases}, \qquad \mu_x^{B\mathrm{i}(A)} = \frac{1-\lambda_x^{B\mathrm{d}(A)}}{(1-\lambda_x^{B\mathrm{d}(A)})(b_x^B+d_x^B)+u_x^B}$$

$$\omega_x^{B\mathrm{d}(A)}:\begin{cases} b_x^{B\mathrm{d}(A)} = b_x^B\mu_x^{B\mathrm{d}(A)} \\ d_x^{B\mathrm{d}(A)} = d_x^B\mu_x^{B\mathrm{d}(A)} \\ u_x^{B\mathrm{d}(A)} = u_x^B\mu_x^{B\mathrm{d}(A)}/\lambda_x^{B\mathrm{d}(A)} \end{cases}, \qquad \mu_x^{B\mathrm{d}(A)} = \frac{\lambda_x^{B\mathrm{d}(A)}}{\lambda_x^{B\mathrm{d}(A)}(b_x^B+d_x^B)+u_x^B} \tag{6}$$

Having specified the separate dependent and independent parts of two partially dependent opinions, we can now define the consensus operator for partially dependent opinions.

Theorem 3 (Consensus Operator for Partially Dependent Opinions). *Let A and B have the partially dependent opinions ω_x^A and ω_x^B respectively, about the same* *proposition x, and let their dependent and independent parts be expressed according to Eq.(6). We will use the symbol $\widetilde{\oplus}$ to designate consensus between partially dependent opinions. As before $\underline{\oplus}$ is the operator for entirely dependent opinions. The consensus of A and B's opinions can then be written as:*

$$\begin{aligned} \omega_x^A \widetilde{\oplus} \omega_x^B &= \omega_x^{A \widetilde{\diamond} B} \\ &= \omega_x^{(A\mathrm{d}(B) \underline{\diamond} B\mathrm{d}(A)) \diamond A\mathrm{i}(B) \diamond B\mathrm{i}(A)} \\ &= (\omega_x^{A\mathrm{d}(B)} \underline{\oplus} \omega_x^{B\mathrm{d}(A)}) \oplus \omega_x^{A\mathrm{i}(B)} \oplus \omega_x^{B\mathrm{i}(A)} \end{aligned} \tag{7}$$

It is easy to prove that for any opinion ω_x^A with a dependence factor $\lambda_x^{A\mathrm{d}(B)}$ to any other opinion ω_x^B the following equality holds:

$$\omega_x^A = \omega_x^{A\mathrm{i}(B)} \oplus \omega_x^{A\mathrm{d}(B)} \tag{8}$$

4 Trust Transitivity

Assume two agents A and B where A trusts B, and B believes that proposition x is true. Then by transitivity, agent A will also believe that proposition x is true. This assumes that B recommends x to A. In our approach, trust and belief are formally expressed as opinions. The transitive linking of these two opinions consists of discounting B's opinion about x by A's opinion about B, in order to derive A's opinion about x. This principle is illustrated in Fig.5 below. The solid arrows represent initial direct trust, and the dotted arrow represents derived indirect trust.

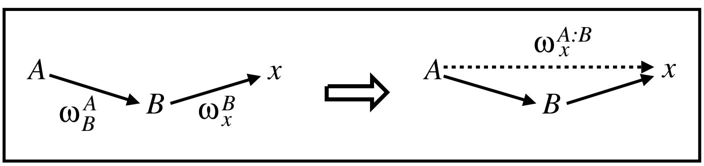


Fig. 5. Principle of the discounting operator

Trust transitivity, as trust itself, is a human mental phenomenon, so there is no such thing as objective transitivity, and trust transitivity therefore lends itself to different interpretations. We see two main difficulties. The first is related to the effect of A disbelieving that B will give a good advice. What does this exactly mean? We will give two different interpretations and definitions. The second difficulty relates to the effect of base rate trust in a transitive path. We will briefly examine this, and provide the definition of a base rate sensitive discounting operator as an alternative to the two previous which are base rate insensitive.

4.1 Uncertainty Favouring Trust Transitivity

A's disbelief in the recommending agent B means that A thinks that B ignores the truth value of x. As a result A also ignores the truth value of x.

Definition 5 (Uncertainty Favouring Discounting). *Let A, B and be two agents where A's opinion about B's recommendations is expressed as $\omega_B^A = \{b_B^A, d_B^A, u_B^A, a_B^A\}$, and let x be a proposition where B's opinion about x is recommended to A with the opinion $\omega_x^B = \{b_x^B, d_x^B, u_x^B, a_x^B\}$. Let $\omega_x^{A:B} = \{b_x^{A:B}, d_x^{A:B}, u_x^{A:B}, a_x^{A:B}\}$ be the opinion such that:*

$$\begin{cases} b_x^{A:B} = b_B^A b_x^B \\ d_x^{A:B} = b_B^A d_x^B \\ u_x^{A:B} = d_B^A + u_B^A + b_B^A u_x^B \\ a_x^{A:B} = a_x^B \end{cases}$$

then $\omega_x^{A:B}$ is called the uncertainty favouring discounted opinion of A. By using the symbol $\otimes$ to designate this operation, we get $\omega_x^{A:B} = \omega_B^A \otimes \omega_x^B$. □

It is easy to prove that this operator is associative but not commutative. This means that the combination of opinions can start in either end of the path, and that the order in which opinions are combined is significant. In a path with more than one recommending entity, opinion independence must be assumed, which for example translates into not allowing the same entity to appear more than once in a transitive path. Fig.6 illustrates an example of applying the discounting operator for independent opinions, where $\omega_B^A = \{0.1,\ 0.8,\ 0.1\}$ discounts $\omega_x^B = \{0.8,\ 0.1,\ 0.1\}$ to produce $\omega_x^{A:B} = \{0.08,\ 0.01,\ 0.91\}$.

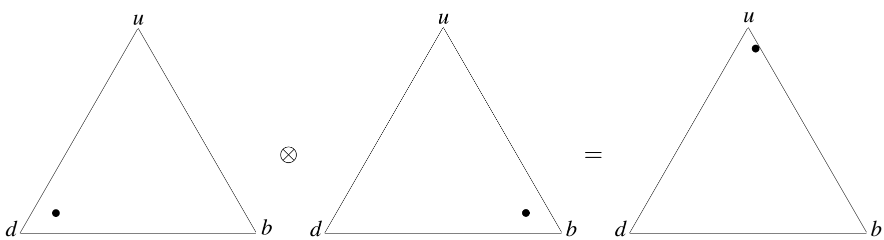


Fig. 6. Example of applying the discounting operator for independent opinions

4.2 Opposite Belief Favouring

A's disbelief in the recommending agent B means that A thinks that B consistently recommends the opposite of his real opinion about the truth value of x. As a result, A not only disbelieves in x to the degree that B recommends belief, but she also believes in x to the degree that B recommends disbelief in x, because the combination of two disbeliefs results in belief in this case.

Definition 6 (Opposite Belief Favouring Discounting). *Let A, B and be two agents where A's opinion about B's recommendations is expressed as $\omega_B^A = \{b_B^A, d_B^A, u_B^A, a_B^A\}$, and let x be a proposition where B's opinion about x is recommended to A as the opinion $\omega_x^B = \{b_x^B, d_x^B, u_x^B, a_x^B\}$. Let $\omega_x^{A:B} = \{b_x^{A:B}, d_x^{A:B}, u_x^{A:B}, a_x^{A:B}\}$ be the opinion such that:*

$$\begin{cases} b_x^{A:B} = b_B^A b_x^B + d_B^A d_x^B \\ d_x^{A:B} = b_B^A d_x^B + d_B^A b_x^B \\ u_x^{A:B} = u_B^A + (b_B^A + d_B^A) u_x^B \\ a_x^{A:B} = a_x^B \end{cases}$$

then $\omega_x^{A:B}$ *is called the opposite belief favouring discounted recommendation from B to A. By using the symbol* $\otimes$ *to designate this operation, we get* $\omega_x^{A:B} = \omega_B^A \otimes \omega_x^B$. $\square$

This operator models the principle that *"your enemy's enemy is your friend"*. That might be the case in some situations, and the operator should only be applied when the situation makes it plausible. It is doubtful whether it is meaningful to model more than two arcs in a transitive path with this principle. In other words, it is doubtful whether the enemy of your enemy's enemy necessarily is your enemy too.

4.3 Base Rate Sensitive Transitivity

In the transitivity operators defined in Sec.4.1 and Sec.4.2 above, a_B^A had no influence on the discounting of of the recommended (b_x^B, d_x^B, u_x^B) parameters. This can seem counterintuitive in many cases such as in the example described next.

Imagine a stranger coming to a town which is know for its citizens being honest. The stranger is looking for a car mechanic, and asks the first person he meets to direct him to a good car mechanic. The stranger receives the reply that there are two car mechanics in town, David and Eric, where David is cheap but does not always do quality work, and Eric might be a bit more expensive, but he always does a perfect job.

Translated into the formalism of subjective logic, the stranger has no other info about the person he asks than the base rate that the citizens in the town are honest. The stranger is thus ignorant, but the expectation value of a good advice is still very high. Without taking a_B^A into account, the result of the definitions above would be that the stranger is completely ignorant about which if the mechanics is the best.

An intuitive approach would then be to let the expectation value of the stranger's trust in the recommender be the discounting factor for the recommended (b_x^B, d_x^B) parameters.

Definition 7 (Base Rate Sensitive Discounting). *The base rate sensitive discounting of a belief* $\omega_x^B = (b_x^B, d_x^B, u_x^B, a_x^B)$ *by a belief* $\omega_B^A = (b_B^A, d_B^A, u_B^A, a_B^A)$ *produces the transitive belief* $\omega_x^{A\diamond B} = (b_x^{A\diamond B}, d_x^{A\diamond B}, u_x^{A\diamond B}, a_x^{A\diamond B})$ *where*

$$\begin{cases} b_x^{A\diamond B} = \mathrm{E}(\omega_B^A) b_x^B \\ d_x^{A\diamond B} = \mathrm{E}(\omega_B^A) d_x^B \\ u_x^{A\diamond B} = 1 + \mathrm{E}(\omega_B^A) u_x^B - \mathrm{E}(\omega_B^A) \\ a_x^{A\diamond B} = a_x^B \end{cases} \tag{9}$$

where the probability expectation value $\mathrm{E}(\omega_B^A) = b_B^A + a_B^A u_B^A$.

However this operator must be applied with care. Assume again the town of honest citizens, and let let the stranger A have the opinion $\omega_B^A = (0,\ 0,\ 1,\ 0.99)$ about the first person B she meets, i.e. the opinion has no basis in evidence other than a very high base rate defined by $a_B^A = 0.99$. If the person B now recommends to A the opinion $\omega_x^B = (1,\ 0,\ 0,\ a)$, then, according to the base rate sensitive discounting operator of Def.7, A will have the belief $\omega_x^{A:B} = (0.99,\ 0,\ 0.01,\ a)$ in x. In other words, the highly certain belief $\omega_x^{A:B}$ is derived on the basis of the highly uncertain belief ω_B^A, which can seem counterintuitive. This potential problem could be amplified as the trust path gets longer. A safety principle could therefore be to only apply the base rate sensitive discounting to the last transitive link.

There might be other principles that better reflect human intuition for trust transitivity, but we will leave this question to future research. It would be fair to say that the base rate insensitive discounting operator of Def.5 is safe and conservative, and that the base rate sensitive discounting operator of Def.7 can be more intuitive in some situations, but must be applied with care.

5 Mass Hysteria

One of the strengths of this work is in its analytical capabilities. As an example, consider how mass hysteria can be caused by people not being aware of dependence between opinions. Let for example person A recommend an opinion about a particular statement x to a group of other persons. Without being aware of the fact that the opinion came from the same origin, these persons can recommend their opinions to each other as illustrated in Fig.7.

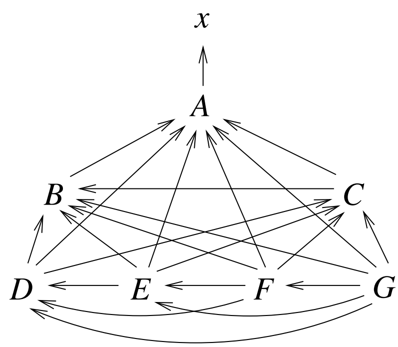


Fig. 7. The principle of mass hysteria

The arrows represent trust so that for example $B \longrightarrow A$ can be interpreted as saying that B trusts A to recommend an opinion about statement x. The actual recommendation goes, of course, in the opposite direction to the arrows in Fig.7.

It can be seen that A recommends an opinion about x to 6 other agents, and that G receives 6 recommendations in all. If G assumes the recommended opinions to be independent and takes the consensus between them, his opinion can become abnormally strong and in fact even stronger than A's opinion.

As a numerical example, let A's opinion ω_x^A about x as well as the agents' opinions about each other ($\omega_A^B, \omega_A^C, \omega_B^C, \omega_A^D, \omega_B^D, \omega_C^D, \omega_A^E, \omega_B^E, \omega_C^E, \omega_D^E, \omega_A^F, \omega_B^F, \omega_C^F, \omega_D^F, \omega_E^F,$ $\omega_A^G, \omega_B^G, \omega_C^G, \omega_D^G, \omega_E^G, \omega_F^G$) all have the same value given by $(0.7,\ 0.1,\ 0.2,\ a)$.

In this example, we will apply the consensus operator for *independent* beliefs to illustrate the effect of unknown dependence. We also apply the uncertainty favouring discounting operator which does not take base rates into account.

Taking all the possible recommendations of Fig.7 into account creates a relatively complex trust graph, and a rather long notation. In order to reduce the size of the notation, the transitivity symbol ":" will simply be omitted, and the cumulative fusion symbol $\diamond$ will simply be written as ",". Analysing the whole graph of dependent paths, as if they were independent, will then produce:

$$\omega_x^{\begin{pmatrix} GA,\ GBA,\ GCA,\ GCBA,\ GDA,\ GDBA,\ GDCA,\ GDCBA,\ GEA, GEBA, GECA, \\ GECBA,\ GEDA, GEDBA,\ GEDCA,\ GEDCBA,\ GFA,\ GFBA,\ GFCA,\ GFCBA, \\ GFDA,\ GFDBA,\ GFDCA,\ GFDCBA,\ GFEA,\ GFEBA,\ GFECA,\ GFECBA, \\ GFEDA,\ GFEDBA,\ GFEDCA,\ GFEDCBA \end{pmatrix}} = (0.76, 0.11, 0.13, a)$$

For comparison, if G only took the recommendation from A into account (as he should), his derived opinion would be $\omega_x^{G:A} = \{0.49,\ 0.07,\ 0.44,\ a\}$.

In real situations it is possible for recommended opinions to return to their originator through feedback loops, resulting in even more exaggerated beliefs. When this process continues, an environment of self amplifying opinions, and thereby hysteria, is created.

6 Conclusion

Subjective logic is a belief calculus which takes into account the fact that perceptions about the world always are subjective. This translates into using a belief model that can express degrees of uncertainty about probability estimates, and we use the term *opinion* to denote such subjective beliefs. In addition, ownership of opinions is assigned to particular agents in order to reflect the fact that opinions always are individual. The operators of subjective logic use opinions about the truth of propositions as input parameters, and produce an opinion about the truth of a proposition as output parameter.

In this paper, trust is interpreted as a belief about reliability, and we have shown how subjective logic can be used for trust reasoning. Although this model can never be perfect, and able to reflect all the nuances of trust, it can be shown to respect the main and intuitive properties of trust and trust propagation.

One difficulty with applying subjective logic is that trust and beliefs can be dependent without people being aware of it, in which case the calculus will produce "wrong" results. Our example illustrated how dependent opinions can influence peoples opinions without any objective reason, and even cause hysteria. In order to avoid this problem we introduced operators for belief and trust fusion that explicitly take dependence into account. This makes it possible to models real world situations involving dependent beliefs more realistically.

Another difficulty is to find a sound and intuitive operator for trust transitivity. This problem comes from the fact that trust transitivity is a psychosocial phenomenon that can not be objectively observed and modelled in traditional statistical or probabilistic terms. We have proposed possible alternative models to the traditional and conservative uncertainty favouring transitivity operator of subjective logic. However, we feel that more research and experiments are needed in order to determine optimal principles of modelling trust transitivity. It might also be the case that no single transitivity operator is suitable for all situations, and that particular situations will require specially designed transitivity operators.

References

1. M.H. DeGroot and M.J. Schervish. *Probability and Statistics (3rd Edition)*. Addison-Wesley, 2001.
2. Li Ding and Timothy Finin. Weaving the Web of Belief into the Semantic Web. In *Proceedings of the 13th International World Wide Web Conference*, New York, May 2004.
3. K.K Fullam et al. The Agent Reputation and Trust (ART) Testbed Architecture. In *Proceedings of the 8th Int. Workshop on Trust in Agent Societies (at AAMAS'05)*. ACM, 2005.
4. A. Hunter. *Uncertainty in information systems*. McGraw-Hill, London, 1996.
5. A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.
6. A. Jøsang. Subjective Evidential Reasoning. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty (IPMU2002)*, Annecy, France, July 2002.
7. A. Jøsang. The Consensus Operator for Combining Beliefs. *Artificial Intelligence Journal*, 142(1–2):157–170, October 2002.
8. A. Jøsang, D. Bradley, and S.J. Knapskog. Belief-Based Risk Analysis. In *Proceedings of the Australasian Information Security Workshop (AISW)*, Dunedin, January 2004.
9. A. Jøsang, M. Daniel, and P. Vannoorenberghe. Strategies for Combining Conflicting Dogmatic Beliefs. In Xuezhi Wang, editor, *Proceedings of the 6th International Conference on Information Fusion*, 2003.
10. A. Jøsang, E. Gray, and M. Kinateder. Simplification and Analysis of Transitive Trust Networks (in press, accepted May 2005). *Web Intelligence and Agent Systems*, 00(00):00–00, 2006.
11. A. Jøsang, R. Hayward, and S. Pope. Trust Network Analysis with Subjective Logic. In *Proceedings of the* 29$^{\text{th}}$ *Australasian Computer Science Conference (ACSC2006), CRPIT Volume 48*, Hobart, Australia, January 2006.
12. A. Jøsang, R. Ismail, and C. Boyd. A Survey of Trust and Reputation Systems for Online Service Provision (in press, accepted June 2005). *Decision Support Systems*, 00(00):00–00, 2006.
13. A. Jøsang and S.J. Knapskog. A Metric for Trusted Systems (full paper). In *Proceedings of the 21st National Information Systems Security Conference*. NSA, October 1998.
14. A. Jøsang and S. Lo Presti. Analysing the Relationship Between Risk and Trust. In T. Dimitrakos, editor, *Proceedings of the Second International Conference on Trust Management (iTrust)*, Oxford, March 2004.
15. A. Jøsang and D. McAnally. Multiplication and Comultiplication of Beliefs. *International Journal of Approximate Reasoning*, 38(1):19–51, 2004.
16. A. Jøsang and S. Pope. Semantic Constraints for Trust Tansitivity. In S. Hartmann and M. Stumptner, editors, *Proceedings of the Asia-Pacific Conference of Conceptual Modelling (APCCM) (Volume 43 of Conferences in Research and Practice in Information Technology)*, Newcastle, Australia, February 2005.
17. A. Jøsang and S. Pope. Normalising the Consensus Operator for Belief Fusion. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty (IPMU2006)*, Paris, July 2006.
18. A. Jøsang, S. Pope, and M. Daniel. Conditional deduction under uncertainty. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, 2005.
19. S. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, 1994.
20. A. Motro and Ph. Smets. *Uncertainty management in information systems: from needs to solutions*. Kluwer, Boston, 1997.