Trust Revision for Conflicting Sources [*]

Audun Jøsang
University of Oslo
Norway
josang@ifi.uio.no

Magdalena Ivanovska
University of Oslo
Norway
magdalei@ifi.uio.no

Tim Muller
Nanyang Technological University
Singapore
t.j.c.muller@gmail.com

***Abstract*—Sensors and information sources can produce conflicting evidence for various reasons, including errors and deception. When conflicting evidence is received, some sources produce more accurate evidence than others with regard to the ground truth. The reliability of sources can be expressed by assigning a level of trust to each source. In this situation, multiple fusion strategies can be applied: one strategy is to directly fuse the evidence based on *a priori* trust in each source, another strategy is to first revise *a priori* trust assignments as a function of the degree of conflict, before the evidence is fused. This paper focuses on the latter approach, and describes a method for applying trust revision in case of highly conflicting evidence. The trust revision method is expressed in the formalism of subjective logic.**

I. INTRODUCTION

Trust in an information source determines how information from the source is interpreted. Assume, for example, that Alice has trouble with her car and gets advice from her neighbour Bob about the issue. If she trusts Bob in matters of car mechanics, then she can take Bob's advice, as illustrated in Figure 1. If not she wold be reluctant to take Bob's advice. The general principle is that the relying party *discounts* the received advice as a function of the level of trust in the advisor.

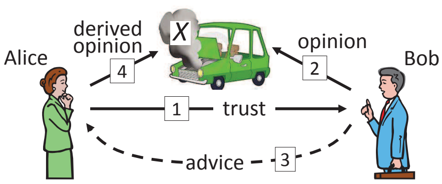

Figure 1. Deriving knowledge based on trust

A relying party may receive information about the same variable from multiple sources, where received information can be conflicting. The relying party then needs to fuse the different information elements and try to form a single opinion about the target variable. For example, if Alice has doubts about Bob's advice about car mechanics she might also ask her other neighbour Clark for a second opinion. In case the pieces of advice from Bob and Clark are highly conflicting, Alice might deduce that at least one of them is not to be trusted. As a result, Alice might revise her opinion about one or both of her neighbours, and consequently, derive a different conclusion from their advice. Humans handle this type of trust reasoning routinely.

[*] In the proceedings of the 18th International Conference on Information Fusion (FUSION 2015), Washington, July, 2015.

It is also possible for advisors to give deceptive advice on purpose in order to gain some advantage, so that relying parties need to beware and discount the advice they receive if there is a reason to believe that the source is not trustworthy. Trust is the compass that guides relying parties about whether advice from a specific source should be accepted or rejected. Trust is a phenomenon that emerges naturally among living species equipped with advanced cognitive faculties. Discounting of a received information or signal based on trust and trust revision happen almost instinctively and simultaneously. When assuming that software agents can be equipped with capabilities to reason about trust and risk, and to make decisions based on that, one talks about computational trust, as described by rapidly growing literature [1]–[5].

In the computational trust literature many formal models have been proposed for trust reasoning, such as the ones described in [6] and [7]. A distinction can be made between interpreting trust as 1) a belief about the reliability of the information source, and as 2) a decision to depend on the source [8]. In this paper, trust is interpreted as a belief about the reliability of the information source. A definition of this type of trust interpretation is provided by Gambetta [9]:

Trust is the subjective probability by which an individual A expects that another individual B performs a given action on which its welfare depends.

A fundamental problem for modelling trust discounting is the lack of benchmark for comparison and validation of models, since practical trust transitivity seems to be idiosyncratic for humans and animals, with no true analogue among non-living forms (and in the physical world for that matter). The efficacy of long chains of transitive trust in these circumstances is debatable, but nonetheless chains of trust can be observed in human trust. Human subjective trust is in reality a state of mind that results from the cognitive and affective predispositions of an entity to perceive and respond to external stimuli, and to combine them with the internal states and stimuli.

In this paper we present a model for fusing conflicting opinions coming from different information sources. In the model, the analyst assigns a trust opinion to each of the information sources, where a source considered reliable is assigned high trust level and vice versa. The respective trust opinions are considered to depend on the degree of conflict in the information provided by the information sources, and is updated based on that.

II. Subjective Logic

Subjective logic is a formalism that represents uncertain probabilistic information in the form of *subjective opinions*, and that defines a variety of operations for subjective opinions. In this section we present in detail the concept of subjective opinion, and, in particular, *binomial opinion*, which is used in the subsequent sections for representing agent's opinion arguments and, in particular, for representing trust.

A. Opinion Representations

In any modeled scenario, the objects of interest like the condition of the car, or the trustworthiness of a person or information source, can be represented as *variables* taking values from a certain *domain*. The condition of the car can be *good* or *bad*, a person can be *trusted* or *not*, etc. Domains are typically specified to reflect realistic situations for the purpose of being practically analysed in some way. The different values of the domain are assumed to be mutually exclusive and exhaustive, which means that the variable can take only one value at any time, and that all possible values of interest are included in the domain. For example, if the variable is the *WEATHER*, we can assume its domain to be the set {rainy, sunny, overcast}. The available information about the particular value of the variable is very often of a probabilistic type, in the sense that we don't now the particular value, but we might know its probability. Probabilities express likelihoods with which the variable takes the specific values and their sum over the whole domain is 1. A variable together with a probability distribution defined on its domain is a *random variable*.

For a given variable of interest, the values of its domain are assumed to be the real possible states the variable can take in the situation to be analysed. In some cases, certain observations may indicate that the variable takes one of several possible states, but it is not clear which one in particular. For example, we might know for sure that the weather is either *rainy* or *sunny*, but do not know its exact state. For this reason it is very often more practical to consider subsets of the domain as possible values of the variable, i.e. instead of the original domain to consider a *hyperdomain*, which would contain all the singletons like {rainy}, but also composites like {rainy, sunny}; and assign beliefs to these values according to the available information, instead of providing a probability distributions on the original domain. In this case we are talking about a *hypervariable* in contrast to a random variable.

In the case of the *WEATHER* seen as a hypervariable, a possible value can be {rainy, sunny} which means that the actual weather is either rainy or sunny, but not both at the same time. Composites are only used as an artifact for assigning belief mass when the observer believes that one of several values is the case, but is confused about which one in particular is true. If the analyst wants to include the realistic possibility that there can be rain and sunshine simultaneously, then the domain would need to include a corresponding singleton value such as {rainy&sunny}. It is thus a question of interpretation how the analyst wants to separate between different types of weather, and thereby define the relevant domain.

A subjective opinion distributes a *belief mass* over the values of the hyperdomain. The sum of the belief masses is less than or equal to 1, and is complemented with an *uncertainty mass*. In addition to belief and uncertainty mass, a subjective opinion contains a *base rate* probability distribution expressing prior knowledge about the specific class of random variables, so that in case of significant uncertainty about a specific variable, the base rates provide a basis for default likelihoods. We give formal definitions of these concepts in what follows.

Let X be a variable over a domain $\mathbb{X} = \{x_1, x_2, \dots, x_k\}$ with cardinality k, where x_i $(1 \leq i \leq k)$ represents a specific value from the domain. The *hyperdomain* is a reduced powerset of $\mathbb{X}$, denoted by $\mathscr{R}(\mathbb{X})$, and defined as follows:

$$\mathscr{R}(\mathbb{X}) = \mathscr{P}(\mathbb{X}) \setminus \{\mathbb{X}, \emptyset\}. \tag{1}$$

All proper subsets of $\mathbb{X}$ are elements of $\mathscr{R}(\mathbb{X})$, but $\mathbb{X}$ and $\emptyset$ are not, since they are not considered possible observations to which we can assign beliefs. The hyperdomain has cardinality $2^k - 2$. We use the same notation for the elements of the domain and the hyperdomain, and consider X a *hypervariable* when it takes values from the hyperdomain.

Let A denote an *agent* which can be an individual, source, sensor, etc. A *subjective opinion* of the agent A on the variable X, ω_X^A, is a tuple

$$\omega_X^A = (b_X^A, u_X^A, a_X^A), \tag{2}$$

where $b_X^A : \mathscr{R}(\mathbb{X}) \rightarrow [0,1]$ is a belief mass distribution, the parameter $u_X^A \in [0,1]$ is an uncertainty mass, and $a_X^A : \mathbb{X} \rightarrow [0,1]$ is a base rate probability distribution satisfying the following additivity constrains:

$$u_X^A + \sum_{x \in \mathscr{R}(\mathbb{X})} b_X^A(x) = 1, \tag{3}$$

$$\sum_{x \in \mathbb{X}} a_X^A(x) = 1 \ . \tag{4}$$

In the notation of the subjective opinion ω_X^A, the subscript is the target variable X, the *object* of the opinion while the superscript is the opinion owner A, the *subject* of the opinion. Explicitly expressing subjective ownership of opinions makes is possible to express that different agents have different opinions on the same variable. Indication of opinion ownership can be omitted when the subject is clear or irrelevant, for example, when there is only one agent in the modelled scenario.

The belief mass distribution b_X^A has $2^k - 2$ parameters, whereas the base rate distribution a_X^A only has k parameters. The uncertainty parameter u_X^A is a simple scalar. A general opinion thus contains $2^k + k - 1$ parameters. However, given that Eq.(3) and Eq.(4) remove one degree of freedom each, opinions over a domain of cardinality k only have $2^k + k - 3$ degrees of freedom.

A subjective opinion in which $u_X = 0$, i.e. an opinion without uncertainty, is called a *dogmatic opinion*. A dogmatic opinion for which $b_X(x) = 1$, for some x, is called an *absolute opinion*. In contrast, an opinion for which $u_X = 1$, and consequently, $b_X(x) = 0$, for every $x \in \mathscr{R}(\mathbb{X})$, i.e. an opinion with complete uncertainty, is called a *vacuous opinion*.

Every subjective opinion "projects" to a probability distribution P_X over $\mathbb{X}$ defined through the following function:

$$\mathrm{P}_X(x_i) = \sum_{x_j \in \mathscr{R}(\mathbb{X})} a_X(x_i/x_j)\, b_X(x_j) + a_X(x_i)\, u_X\,, \tag{5}$$

where $a_X(x_i/x_j)$ is the *relative base rate* of $x_i \in \mathbb{X}$ with respect to $x_j \in \mathscr{R}(\mathbb{X})$ defined as follows:

$$a_X(x_i/x_j) = \frac{a_X(x_i \cap x_j)}{a_X(x_j)}\ , \tag{6}$$

where a_X is extended on $\mathscr{R}(\mathbb{X})$ additively. For the relative base rate to be always defined, it is enough to assume $a_X^A(x_i) > 0$, for every $x_i \in \mathbb{X}$. This means that everything we include in the domain has a non-zero probability of occurrence in general.

Binomial opinions apply to binary random variables where the belief mass is distributed over two elements. Multinomial opinions apply to random variables in n-ary domains, and where the belief mass is distributed over the elements of the domain. General opinions, also called *hyper-opinions*, apply to hypervariables where belief mass is distributed over elements in hyperdomains obtained from n-ary domains. A binomial opinion is equivalent to a Beta probability density function, a multinomial opinion is equivalent to a Dirichlet probability density function, and a hyper-opinion is equivalent to a Dirichlet hyper-probability density function [10]. Binomial opinions thus represent the simplest opinion type, which can be generalised to multinomial opinions, which in turn can be generalised to hyper-opinions. Simple visualisations for binomial and trinomial opinions are based on barycentric coordinate systems as illustrated in Figures 2 & 3 below.

B. Binomial Opinions

Binomial opinions have a special notation that is used for modelling trust in the subsequent sections.

Let X be a random variable on domain $\mathbb{X} = \{x, \overline{x}\}$. The binomial opinion of agent A about variable X can be seen as an opinion about the truth of the statement "X is x" (denoted by $X = x$, or just x) and given as an ordered quadruple:

$$\omega_x^A = (b_x^A, d_x^A, u_x^A, a_x^A)\ , \tag{7}$$

b_x^A (*belief*)	belief mass in support of x,
d_x^A (*disbelief*)	belief mass in support of $\overline{x}$ (NOT x),
u_x^A (*uncertainty*)	uncertainty about probability of x,
a_x^A (*base rate*)	non-informative prior probability of x.

In case of binomial opinions Eq.(3) is simplified to Eq.(8).

$$b_x^A + d_x^A + u_x^A = 1\ . \tag{8}$$

Similarly, in the special case of binomial opinions the projected probability of Eq.(5) is simplified to:

$$\mathrm{P}_x^A = b_x^A + a_x^A u_x^A\ . \tag{9}$$

A binomial opinion can be represented as a point inside an equilateral triangle, which is a 3-axis barycentric coordinate system representing belief, disbelief, and uncertainty masses, with a point on the baseline representing base rate probability, as shown in Fig. 2. The axes run through the vertices along the altitudes of the triangle. The belief, disbelief, and uncertainty axes run through the vertices denoted by x, $\overline{x}$, and u, correspondingly, which have coordinates coordinates $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$, correspondingly. In Fig. 2, $\omega_x = (0.20,\ 0.40,\ 0.40,\ 0.75)$, with projected probability $\mathrm{P}_x = 0.50$, is shown as an example. A strong positive opinion, for example, would be represented by a point towards the bottom right belief vertex.

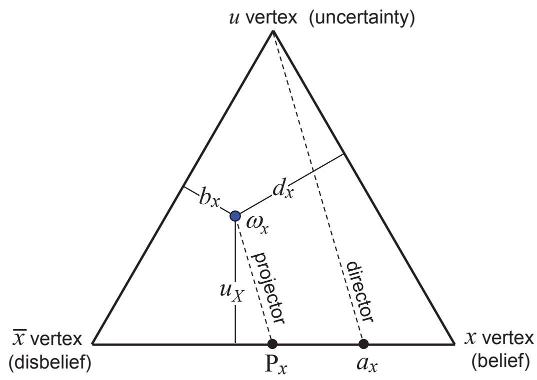

Figure 2. Visualisation of a binomial opinion

In case the opinion point is located at the left or right vertex of the triangle, i.e. has $d_x = 1$ or $b_x = 1$ (and $u_x = 0$), the opinion is equivalent to boolean TRUE or FALSE, in which case subjective logic is reduced to binary logic. In case the opinion point is located on the base line of the triangle, i.e. has $u_x = 0$, then the opinion is equivalent to a probability distribution, in which case subjective logic is reduced to probability calculus.

In general, a multinomial opinion can be represented as a point inside a regular simplex. In particular, a trinomial opinion can be represented inside a tetrahedron (a 4-axis barycentric system), as shown in Figure 3.

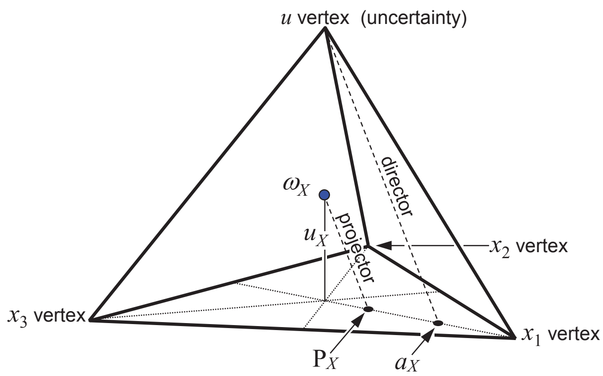

Figure 3. Visualisation of a trinomial opinion

Assume the random variable X on domain $\mathbb{X} = \{x_1, x_2, x_3\}$. Figure 3 shows multinomial opinion ω_X with belief mass distribution $b_X = (0.20, 0.20, 0.20)$, uncertainty mass $u_X = 0.40$ and base rate distribution $a_X = (0.750, 0.125, 0.125)$.

III. Subjective Opinion Fusion

In many cases subjective opinions from multiple sources are available and one needs to fuse them in some way and produce a single opinion. The purpose of opinion fusion is to produce a new opinion that would be a relevant representative of the original opinions in the given context. It can be challenging to determine the most appropriate fusion operator for a specific setting. A discussion on this topic is provided in [11]. Here we use the averaging fusion operator as an example.

Averaging fusion of opinions takes as input arguments two or more subjective opinions about the same variable, as illustrated in Figure 4, where agents B and C have separate opinions about the variable X. The resulting opinion obtained by merging the input ones is held by an imaginary combined agent, which in case of averaging fusion is denoted by $B\underline{\diamond}C$.

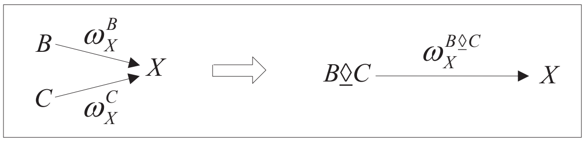

Figure 4. Averaging fusion of opinions

Averaging opinion fusion is used when the fused opinions are assumed to be dependent. In this case, including more opinions does not necessarily mean that more evidence is supporting the conclusion. An example of this type of situation is when a jury tries to reach a verdict after having observed the court proceedings. Because the evidence is limited to what was presented in court, the certainty about the verdict does not increase by having more jury members expressing their beliefs, since they all received the same evidence.

More formally, we assume a variable X with a domain $\mathbb{X}$ of cardinality k and corresponding hyperdomain $\mathscr{R}(\mathbb{X})$. Let the opinions ω_X^B and ω_X^C apply to the hypervariable X which takes its values from the hyperdomain. The superscripts B and C identify the respective opinion sources or opinion owners. The expression for *averaging belief fusion* of these two opinions is the following:

$$\omega_X^{B\underline{\diamond}C} = \omega_X^B \underline{\oplus} \omega_X^C \ . \tag{10}$$

In the above expression, the averaging combination of agents denoted by the symbol '$\underline{\diamond}$', corresponds to averaging opinion fusion operator denoted by '$\underline{\oplus}$'.

The definition of averaging belief fusion operator is obtained from averaging opinions represented as evidence through the bijective mapping between evidence and belief in subjective logic [10]. The expressions for the beliefs and uncertainty of the resulting opinion $\omega_X^{B\underline{\diamond}C}$ are provided separately for the two different cases given below.

Case I: For $u_X^B \neq 0 \ \vee \ u_X^C \neq 0$:

$$\begin{cases} b_X^{B\underline{\diamond}C}(x) &= \frac{b_X^B(x)u_X^C + b_X^C(x)u_X^B}{u_X^B + u_X^C} \\ \\ u_X^{B\underline{\diamond}C} &= \frac{2u_X^B u_X^C}{u_X^B + u_X^C} \end{cases} \tag{11}$$

Case II: For $u_X^B = 0 \ \wedge \ u_X^C = 0$:

$$\begin{cases} b_X^{B\underline{\diamond}C}(x) &= \gamma_X^B b_X^B(x) + \gamma_X^C b_X^C(x) \\ \\ u_X^{B\underline{\diamond}C} &= 0 \end{cases} \tag{12}$$

$$\text{where} \quad \begin{cases} \gamma_X^B = \lim\limits_{\substack{u_X^B \to 0 \\ u_X^C \to 0}} \frac{u_X^C}{u_X^B + u_X^C} \\ \\ \gamma_X^C = \lim\limits_{\substack{u_X^B \to 0 \\ u_X^C \to 0}} \frac{u_X^B}{u_X^B + u_X^C} \end{cases}$$

It is usually assumed that the two opinions to be fused have equal base rates, which leads to the same base rate for the resulting opinion as well. In the case when $a_X^B \neq a_X^C$, the base rate of the fused opinion, $a_X^{B\underline{\diamond}C}$, is defined simply as an average of the functions a_X^B and a_X^C.

It can be verified that the averaging fusion rule is commutative and idempotent; but it is *not* associative.

IV. Trust Discounting

The general idea behind trust discounting is to express degrees of trust in an information source and then to discount information provided by the source as a function of the trust in the source. We represent both the trust and the provided information in the form of subjective opinions, and then define an appropriate operation on these opinions to find the trust discounted opinion.

Let agent A denote the relying party and agent B denote an information source. Assume that agent B provides information to agent A about the state of a variable X expressed as a subjective opinion on X. Assume further that agent A has an opinion on the trustworthiness of B with regard to providing information about X. Based on the combination of A's trust in B and on B's opinion about X given as an advice to A, it is possible for A to derive an opinion about X. This process is illustrated in Figure 6.

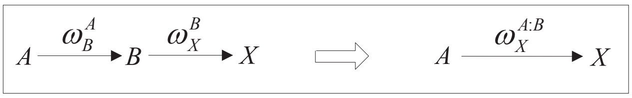

Figure 6. Trust discounting of opinions

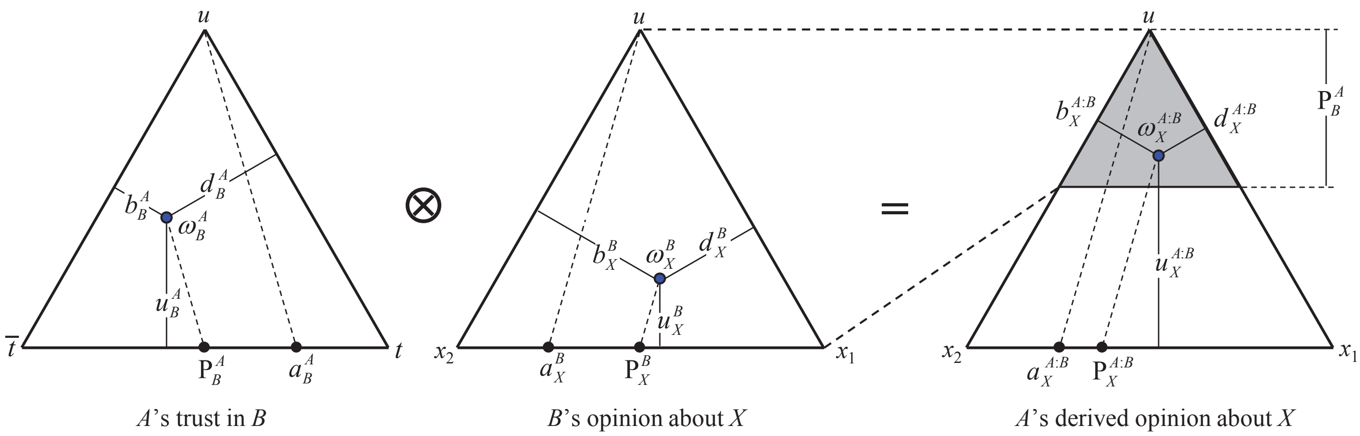

Figure 5. Uncertainty-favouring trust discounting

Several trust discounting operators for subjective logic are described in the literature [4], [7]. The general representation of trust discounting is through conditionals [4], while special cases can be expressed with specific trust discounting operators. In this paper we use the specific case of *uncertainty-favouring trust discounting* which enables the uncertainty in A's derived opinion about X to increase as a function of the projected *distrust* in the recommender B. The uncertainty-favouring trust discounting operator is described below.

Agent A's trust in B is formally expressed as a binomial opinion on domain $\mathbb{T} = \{t, \bar{t}\}$ where the values t and $\bar{t}$ denote *trusted* and *distrusted* respectively. We denote this opinion by $\omega^A_B = (b^A_B, d^A_B, u^A_B, a^A_B)$ [1]. The values b^A_B, d^A_B, and u^A_B represent the degrees to which A trusts, does not trust, or is uncertain about the trustworthiness of B in the current situation, while a^A_B is a base rate probability that A would assign to the trustworthiness of B *a priori*, before receiving the advice.

Assume variable X on domain $\mathbb{X}$, and let $\omega^B_X = (b^B_X, u^B_X, a^B_X)$ be B's general opinion on X as recommended to A. Trust discounting is expressed with the following notation:

$$\omega^{A:B}_X = \omega^A_B \otimes \omega^B_X \ . \tag{13}$$

Trust discounted agent obtained by discounting agent B by the trust of agent A to it, denoted by $A : B$, corresponds to transitive discounting of opinions with the operator denoted by $\otimes$. $\omega^{A:B}_X$ denotes A's subjective opinion on X derived as a function of A's trust in B and B's recommended opinion about X. There are multiple variants of this operator given in [4], where the specific case of uncertainty-favouring trust discounting is defined in the following way:

$$\omega^{A:B}_X : \begin{cases} b^{A:B}_X(x) & = \ \mathrm{P}^A_B\, b^B_X(x) \\ u^{A:B}_X & = 1 - \mathrm{P}^A_B \sum\limits_{x \in \mathscr{R}(\mathbb{X})} b^B_X(x) \\ a^{A:B}_X(x) & = \ a^B_X(x) \end{cases} \tag{14}$$

[1]According to the notation introduced in Section II-B, a more correct notation for this opinion would be $\omega^A_{\mathbb{T}_B} = (b^A_t, d^A_t, u^A_t, a^A_t)$. For practical reasons, especially for the case where there can be more different agents in the role of B, we choose to use this modified notation here.

The effect of this operator is illustrated in Figure 5 with the following example. Let $\omega^A_B = (0.20, 0.40, 0.40, 0.75)$ be A's trust opinion on B, with projected probability $\mathrm{P}^A_B = 0.50$. Let further $\omega^B_x = (0.45, 0.35, 0.20, 0.25)$ be B's opinion about the state of variable X, with projected probability $\mathrm{P}^B_x = 0.50$. According to Eq.(14) we can compute A's derived opinion about X as $\omega^{A:B}_x = (0.225, 0.175, 0.60, 0.25)$ which has projected probability $\mathrm{P}^{A:B}_x = 0.375$. The trust-discounted opinion $\omega^{A:B}_x$ typically has increased uncertainty, compared to the original opinion given by B, where the decree of discounting is dictated by the projected probability of the trust opinion.

Figure 5 illustrates the general behaviour of the uncertainty-favouring trust discounting operator, where the derived opinion is constrained to the shaded sub-triangle at the top of the rightmost triangle. The size of the shaded sub-triangle corresponds to the projected probability of trust in the trust opinion. The effect of this is that the barycentric representation of ω^B_x is shrunk proportionally to P^A_B to become a barycentric opinion representation inside the shaded sub-triangle.

Some special cases are worth mentioning. In case the projected trust probability equals one, which means complete trust, the relying party accepts the recommended opinion as it is. In case the projected trust probability equals zero, which means complete distrust, the recommended opinion is reduced to a vacuous opinion, meaning that the recommended opinion is completely discarded.

The following example illustrates how this kind of trust discounting is applied intuitively in real situations. While visiting a foreign country Alice is looking for a restaurant where the locals go, because she would like to avoid places overrun by tourists. She meets a local called Bob who tells her that restaurant Xylo is the favourite place for locals. Assume that Bob is a stranger to Alice. Then *a priori* her trust in Bob is affected by high uncertainty. However, it is enough for Alice to assume that locals in general give good advice, which results in a high base rate for her trust. Even if her trust in Bob is vacuous, a high base rate will result in a projected probability of trust close to one, so that Alice will derive a strong opinion about the restaurant Xylo based on Bob's advice.

V. Trust Revision

We continue the example from the previous section by assuming that Alice stays in a hostel where she meets another traveler named Clark who tells her that he already tried the Xylo restaurant, and that it actually was very bad, and that there were no locals there. Even if Clark is also a stranger to Alice which means that she is uncertain about his trustworthiness, she assumes that fellow travelers are trustworthy in general which translates into a high base rate trust for Clark. Now Alice has a second advice which gives her reason to revise her initial trust in Bob, which could translate into distrusting Bob.

Trust discounting and fusion can be combined for fusing information from different sources with different trust levels, as illustrated in Figure 7.

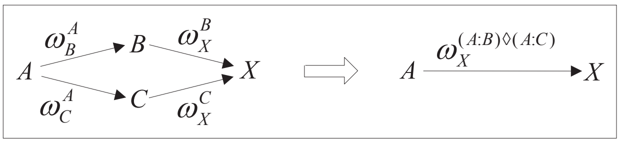

Figure 7. Fusion of trust-discounted opinions

A complicating element in this scenario is when multiple sources provide highly conflicting advice, which might indicate that one or both sources are unreliable. In this case a strategy is needed for dealing with the conflict. The chosen strategy must be suitable for the specific situation.

The most straightforward strategy would be to consider the trust opinions as static, and not to revise trust at all. With this strategy the relying party only needs to determine the most suitable fusion operator for the situation to be analysed. For example, if averaging fusion is considered suitable, then a simple model would be to derive A's opinion about X as follows:

$$\omega_X^{(A:B)\underline{\diamond}(A:C)} = (\omega^A_B \otimes \omega^B_X)\underline{\oplus}(\omega^A_C \otimes \omega^C_X) \quad (15)$$

However, there are several situations where simple fusion might be considered inadequate, and where it would be natural to revise one or both trust opinions.

One such situation is when the respective opinions provided by B and C are highly conflicting in terms of their projected probability distributions on X. Note here that in some situations this might be natural, such as in case of short samples of random processes where a specific type of events might be observed in clusters. Another situation where highly conflicting beliefs might occur naturally is when the observed system can change characteristics over time and the observed projected probability distributions refer to different time periods. However, if the sources A and B observe exactly the same situation or event at the same time, but still produce different opinions, then trust revision should be considered.

Another situation that calls for a trust revision is when the relying party A learns that the ground truth about X is radically different from the recommended opinions.

The most basic measure of conflict is the *projected distance*, denoted PD, between the projected probability distributions of two trust discounted opinions.

$$\mathrm{PD}(\omega_X^{A:B}, \omega_X^{A:C}) = \frac{\sum_{x\in\mathbb{X}} |\mathrm{P}_X^{A:B}(x) - \mathrm{P}_X^{A:C}(x)|}{2} \quad (16)$$

Obviously $\mathrm{PD} \geq 0$. Using basic absolute value inequalities, it can be proven that the numerator in Eq.(16) is not greater than 2, independently of the cardinality of $\mathbb{X}$, so $\mathrm{PD} \in [0,1]$. We obtain $\mathrm{PD} = 0$ when the two opinions have equal projected probability distributions, in which case the opinions are non-conflicting (even though they might be different). The maximum value $\mathrm{PD} = 1$ occurs e.g. in case of two absolute binomial opinions with opposite projected probability values.

A large PD does not necessarily indicate a problem, because conflict is diffused in case one (or both) opinions have high uncertainty. The more uncertain the opinions, the more a large PD should be tolerated. This corresponds to the fact that uncertain opinions carry little weight in the fusion process.

A natural measure of simultaneous certainty of two opinions is the *conjunctive certainty* denoted by CC:

$$\mathrm{CC}(\omega_X^{A:B}, \omega_X^{A:C}) = (1 - u_X^{A:B})(1 - u_X^{A:C}) \quad (17)$$

It can be seen that $\mathrm{CC} \in [0,1]$ where $\mathrm{CC} = 0$ means that one or both opinions are vacuous, and $\mathrm{CC} = 1$ means that both opinions are dogmatic, i.e. have zero uncertainty mass.

The *degree of conflict* (DC) is the product of PD and CC.

$$\mathrm{DC}(\omega_X^{A:B}, \omega_X^{A:C}) = \mathrm{PD} \cdot \mathrm{CC} \quad (18)$$

It is natural to let the degree of trust revision be a function of DC, but the most uncertain opinion should be revised the most. We define *uncertainty difference* (UD) to be a measure for comparing uncertainty between two opinions:

$$\mathrm{UD}(\omega_X^{A:B}, \omega_X^{A:C}) = \frac{u_X^{A:B} - u_X^{A:C}}{u_X^{A:B} + u_X^{A:C}} \quad (19)$$

It can be seen that $\mathrm{UD} \in [-1,1]$, where $\mathrm{UD} = 0$ means that both opinions have equal uncertainty, $\mathrm{UD} = 1$ means that the trust opinion $\omega_X^{A:B}$ is infinitely more uncertain than the other opinion, and $\mathrm{UD} = -1$ means that the trust opinion $\omega_X^{A:C}$ is infinitely more uncertain than the other opinion.

We can use UD to define how much revision each opinion needs. The revision factors are denoted RF_B and RF_C.

$$\mathrm{RF}_B = \frac{1+\mathrm{UD}}{2}, \qquad \mathrm{RF}_C = \frac{1-\mathrm{UD}}{2}. \quad (20)$$

It can be seen that $\mathrm{RF} \in [0,1]$, where $\mathrm{RF}_B + \mathrm{RF}_C = 1$. The case when $\mathrm{RF}_B = 1$ means that only $\omega_X^{A:B}$ is revised, the case $\mathrm{RF}_B = \mathrm{RF}_C = 0.5$ means that both opinion are equally revised, and the case $\mathrm{RF}_C = 1$ means that only $\omega_X^{A:C}$ is revised.

Trust revision consists of increasing distrust at the cost of trust and uncertainty. The idea is that sources found to be unreliable should be distrusted more. A source found to be completely unreliable should be absolutely distrusted.

In terms of the opinion triangle, trust revision consists of moving the opinion point towards the $\bar{t}$ vertex, as shown in Figure 8. Given a trust opinion $\omega_B^A = (b_B^A, d_B^A, u_B^A, a_B^A)$ the revised trust opinion denoted $\tilde{\omega}_B^A = (\tilde{b}_B^A, \tilde{d}_B^A, \tilde{u}_B^A, \tilde{a}_B^A)$ is :

$$\tilde{\omega}_B^A : \begin{cases} \tilde{b}_B^A = b_B^A - b_B^A \cdot \mathrm{RF}_B \cdot \mathrm{DC} \\ \tilde{d}_B^A = d_B^A + (1 - d_B^A) \cdot \mathrm{RF}_B \cdot \mathrm{DC} \\ \tilde{u}_B^A = u_B^A - u_B^A \cdot \mathrm{RF}_B \cdot \mathrm{DC} \\ \tilde{a}_B^A = a_B^A \end{cases} \tag{21}$$

Figure 8 illustrates the effect of trust revision which consists of making a trust opinion more distrusting.

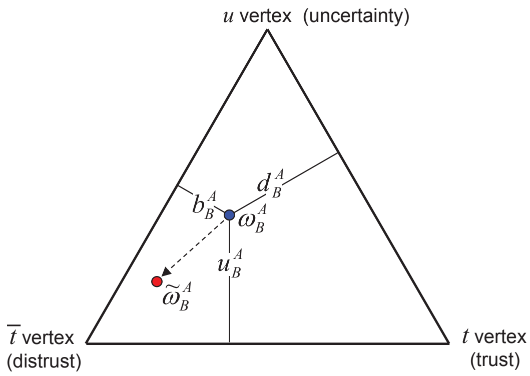

Figure 8. Revision of trust opinion ω_B^A

After trust revision has been applied, the opinion fusion according to Eq.(15) can be repeated, with reduced conflict.

VI. Example

As an example we apply the method to a scenario where advisors provide conflicting recommendations.

The target domain $\mathbb{X}$ is binary, i.e. we assume, for example, that the two advisors express their opinion about X being *good* or *bad*. Furthermore, we assume that the base rate for trustworthiness and for the variable X are both 0.5.

Assume two advisors B and C, each with their own hidden integrity parameter. The integrities of the advisors are the parameters of the model. User A forms opinions ω_B^A and ω_C^A about the trustworthiness of the two advisors, as an estimate of their integrity. The two advisors provide recommendations about X in the form of opinions ω_X^B and ω_X^C.

The integrity of a recommender is represented by the probability that its recommendation reflects the truth of X. Unreliability is represented by the complement probability, i.e. by the probability that its recommendation is misleading.

User A derives opinions $\omega_X^{A:B}$, $\omega_X^{A:C}$ based on the advisors' recommendations ω_X^B and ω_X^C, and the trust opinions ω_B^A, ω_C^A, applying trust discounting and averaging fusion.

User A can revise the trust in the advisors, based on the obtained opinions $\omega_X^{A:B}$, $\omega_X^{A:C}$, and their degree of conflict. The graphs below displays the different aspects of trust revision introduced in Section V, parametrized with advisors.

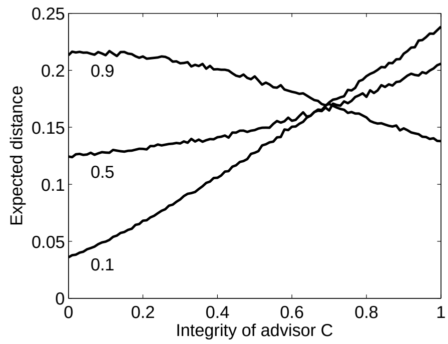

Figure 9. Projected probability distance

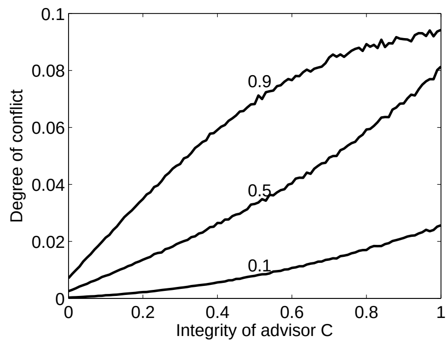

Figure 10. Degree of conflict

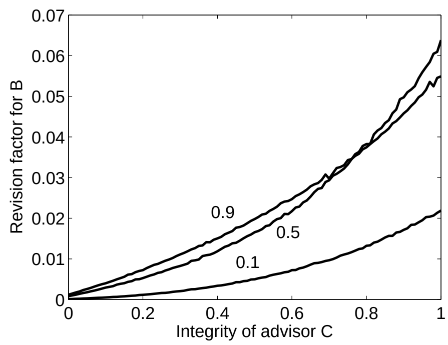

Figure 11. Revision factor for B

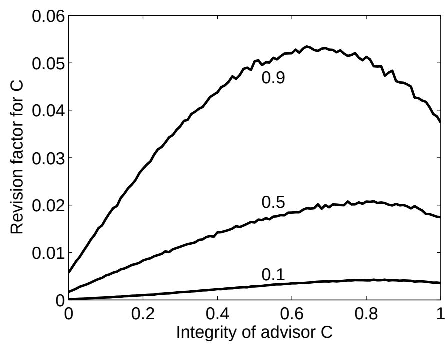

Figure 12. Revision factor for C

Figures 9–12 show three plots each. On the x-axis, we have the integrity of the advisor *C*. The integrity of the advisor *B* is constant for each graph (the shown curves correspond to values 0.1, 0.5, and 0.9). In each of the four figures, we measure a different aspect of trust evaluation.

In Figure 9, we measure projected distance (PD). The PD correlates negatively with uncertainty. This may be somewhat counter-intuitive, since opinions with high uncertainty should be less accurate. However, the PD is not measuring accuracy. An opinion with high uncertainty will provide a projected probability for belief close to 0.5. The absolute distance to a central value, such as 0.5, tends to be small.

In Figure 10, we measure the degree of conflict (DC). The DC correlates much stronger negatively with uncertainty. This is an obvious consequence of the fact that $\mathrm{DC} = \mathrm{CC} \cdot \mathrm{PD}$, and CC correlates negatively with uncertainty too (by definition).

In Figures 11 and 12, we measure the actual revision for the first (fixed) and the second (variable) advisor, respectively. The former graph is increasing, since the total uncertainty decreases (making DC increase), whereas the relative uncertainty of the fixed advisor increases. The latter graph is first increases, then decreases. The initial increase is caused by the total decrease of uncertainty, and the final decline is caused by the relative decrease of uncertainty.

Table I
CORRELATION BETWEEN ERROR AND REVISION OF *B*

$B \backslash C$	Random	0.1	0.5	0.9
Random	0.3259	0.4100	0.3474	0.2208
0.1	0.1885	0.2955	0.1677	−0.0475
0.5	0.2945	0.4509	0.3234	0.1060
0.9	0.6415	0.7681	0.6621	0.4594

Table I uses the same set-up as the graphs. We measure the correlation between the error in the opinion resulting from *C*'s recommendation and the amount of revision that *C* receives. There are four types of advisors: random advisors, bad advisors (10% of recommendations are true), average advisors (50%) and good advisors (90%). The top-left cell value shows the overall correlation (for a pair of random advisors), which is strongly positive. This means that revisions tend to be provided when the opinions tend to be bad. We see a general trend that if the quality of the advisors goes up, they tend to get less useful revisions. In one extreme case, the revisions are even slightly counterproductive. The best revisions are revisions of opinions given by a bad advisor, with a good other advisor to compare to.

VII. CONCLUSIONS

We have described a model for trust revision based on the degree of conflict in the fusion between opinions. The model is formalised in terms of subjective logic which explicitly handles uncertainty in a probabilistic information, which is necessary to adequately model trust. Our model closely corresponds to intuitive human reasoning which includes uncertainty in the beliefs and default assumptions through base rates. This model provides the basis for a sound analysis of situations where the analyst receives information from sources that are considered to have varying levels of trustworthiness. Application areas are for example intelligence analysis, sensor network pruning, and social networks trust management.

REFERENCES

[1] L. Ding and T. Finin, "Weaving the Web of Belief into the Semantic Web," in *Proceedings of the 13th International World Wide Web Conference*, New York, May 2004.

[2] 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.

[3] A. Jøsang, R. Ismail, and C. Boyd, "A Survey of Trust and Reputation Systems for Online Service Provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.

[4] A. Jøsang, T. Aždeska, and S. Marsh, "Trust Transitivity and Conditional Belief Reasoning," in *Proceedings of the 6th IFIP International Conference on Trust Management (IFIPTM 2012)*, Surat, India, May 2012.

[5] S. Marsh, "Formalising Trust as a Computational Concept," Ph.D. dissertation, University of Stirling, 1994.

[6] A. Jøsang, R. Hayward, and S. Pope, "Trust Network Analysis with Subjective Logic," in *Proceedings of the* 29th *Australasian Computer Science Conference (ACSC2006), CRPIT Volume 48*, Hobart, Australia, January 2006.

[7] A. Jøsang, S. Pope, and S. Marsh, "Exploring Different Types of Trust Propagation," in *Proceedings of the 4th International Conference on Trust Management (iTrust)*, Pisa, May 2006.

[8] A. Jøsang and S. Lo Presti, "Analysing the Relationship Between Risk and Trust," in *Proceedings of the Second International Conference on Trust Management (iTrust)*, T. Dimitrakos, Ed., Oxford, March 2004.

[9] D. Gambetta, "Can We Trust Trust?" in *Trust: Making and Breaking Cooperative Relations*, D. Gambetta, Ed. Basil Blackwell. Oxford, 1990, pp. 213–238.

[10] A. Jøsang and R. Hankin, "Interpretation and Fusion of Hyper Opinions in Subjective Logic." in *Proceedings of the 15th International Conference on Information Fusion (FUSION 2012)*, Singapore, July 2012.

[11] A. Jøsang, P. C. Costa, and E. Blash, "Determining Model Correctness for Situations of Belief Fusion." in *Proceedings of the 16th International Conference on Information Fusion (FUSION 2013)*, Istanbul, July 2013.

ACKNOWLEDGEMENTS

The work reported in this paper has been partially funded by the US Army Research Program Activity R&D 1712-IS-01.

Partial funding has also been provided by UNIK.