Trust Transitivity and Conditional Belief Reasoning

Audun Josang, Tanja Aždderska, Stephen Marsh

Trust Transitivity and Conditional Belief Reasoning*

Audun Jøsang[1], Tanja Azderska[2], and Stephen Marsh[3]

[1] University of Oslo, Norway ♯
josang@mn.uio.no
[2] Jozef Stefan Institute, Slovenia
atanja@e5.ijs.si
[3] Communications Research Centre, Canada
steve.marsh@crc.gc.ca

Abstract. Trust transitivity is a common phenomenon embedded in human reasoning about trust. Given a specific context or purpose, trust transitivity is often manifested through the humans' intuition to rely on the recommendations of a trustworthy advisor about another entity that the advisor recommends. Although this simple principle has been formalised in various ways for many trust and reputation systems, there is no real or physical basis for trust transitivity to be directly translated into a mathematical model. In that sense, all mathematical operators for trust transitivity proposed in the literature must be considered *ad hoc*; they represent attempts to model a very complex human phenomenon as if it were lendable to analysis by the laws of physics. Considering this nature of human trust transitivity in reality, any simple mathematical model will essentially have rather poor predictive power. In this paper, we propose a new interpretation of trust transitivity that is radically different from those described in the literature so far. More specifically, we consider recommendations from an advisor as evidence that the relying party will use as input arguments in conditional reasoning models for assessing hypotheses about the trust target. The proposed model of conditional trust transitivity is based on the framework of subjective logic.

Keywords: Trust, Transitivity, Deduction, Abduction, Bayesian, Conditional

1 Introduction

Trust transitivity based on recommendations is a concept that can have different meanings. It can, for example, mean that, if Alice trusts Bob, and Bob trusts Claire, then by transitivity, Alice will also trust Claire. This is expressed in Eq.(1).

$$\text{Indirect(Alice} \longrightarrow \text{Claire)} \quad := \quad \text{Direct(Alice} \longrightarrow \text{Bob} \longrightarrow \text{Claire)} \tag{1}$$

Alice is here the originator relying party, Bob is the recommender (i.e., the advisor), and Claire is the trust target that Alice indirectly trusts as a result of this process. This transitive process assumes that Bob recommends Claire to Alice, i.e., there must be some communication from Bob to Alice about Claire's trustworthiness. This kind of reasoning can also be observed among animals. For example, when bees signal to

* 6th IFIPTM International Conference on Trust Management (IFIPTM 2012), Surat, May 2012.

each other where to find pollen, the other bees can derive trust in a specific pollen harvesting area; when animals give warnings about danger, it can be interpreted as a recommendation about distrust, as in the case of a presence of potential predator. Trust is a phenomenon that emerges naturally among living species equipped with advanced cognitive faculties. When assuming that software agents can be equipped with capabilities to reason about trust and risk, and to make decisions based on that, one can talk about artificial trust, as described by rapidly growing growing literature [1,2,6,13].

It is also common, at least among humans, that the relying party receives recommendations about the same target entity from multiple recommenders whose recommendations express different and possibly conflicting trust. The relying party then needs to fuse the different trust recommendations and form a single trust opinion about the target entity. These are principles of analytical reasoning that people handle more or less unconsciously in everyday life. In the computational trust literature many formal models have been proposed for the same purpose, such as in [5,9]. A distinction can be made between interpreting trust as a belief about the reliability of the target, and as a decision to depend on the target [7]. In this paper, trust is interpreted in the former sense, i.e. as a belief about reliability of a target.

A fundamental problem about modelling trust transitivity is that there is no benchmark for comparison and validation in nature, since practical trust transitivity seems to be idiosyncratic for humans and animals, with no true analog 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 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. However, it is the actual nature of trust that represents the relying party's subjective estimate of the reliability of the target entity for a purpose on which the relying party's welfare depends.

In contrast, when analysing the reliability of physical systems, there are mathematical models that can be easily validated through observation. For example, the correct operation of a serial system depends on the correct operation of each component, which translates into a conjunctive model where the system reliability can be predicted as the product of the reliabilities of each component in the series. The correct operation of a parallel system depends on the correct operation of at least one of its components, which translates into a disjunctive model where the system reliability can be predicted as the copoduct of the reliabilities of each component.

Let $p(x_i)$ express the reliability of component x_i, then the respective reliabilities of a serial system **Ser** and a parallel system **Par** are expressed as:

$$
\begin{aligned}
&\text{Serial system reliability:} && p(\text{Ser}) = \prod p(x_i)\ ,\ \text{i.e. product of reliabilities} \\
&\text{Parallel system reliability:} && p(\text{Par}) = \coprod p(x_i)\ ,\ \text{i.e. coproduct of reliabilities}
\end{aligned}
\tag{2}
$$

Addressing and computing trust based on chained recommendations, and on recommendations from multiple parties in parallel can not be modelled in the same manner as in Eq.(2), although some proposed trust models in the literature do precisely that. The product rule gives a very good estimate of serial system reliability, but is a very poor model for trust transitivity. To illustrate why this is so, assume a transitive trust path of n nodes where each node trusts the next with value p. The product rule dictates a derived trust $T = p^n$ which converges relatively quickly towards zero whenever $p < 1$, meaning that a trust value that is derived from a relatively long transitive trust path will be close to zero. Interpreting zero trust as distrust, as some models do, would clearly be wrong because there is no reason to think that an entity which is known only indirectly through a long transitive path should be distrusted for that reason. Interpreting zero trust as 'no information', as other models do, would be more intuitive in this case. Methods of computing trust transitivity that have been proposed for subjective logic [9] are analoguous to the latter interpretation, and we show that these methods are special cases of the conditional trust model for analysing trust transitivity proposed in this paper.

The contribution of this paper is to propose a new computational model for trust transitivity based on conditional belief reasoning using the formalism of subjective logic. The idea is to model trust transitivity in similar manner as analysing competing hypotheses. Such models are applied, e.g., in the area of intelligence analysis [14]. Modelling trust transitivity as a simple evidence analysis problem removes the mysticism of trust transitivity, i.e. it does not assume that trust transitivity exists as a separate natural process by itself. Our model assumes that trust transitivity implies weighing the obtained evidence in order to draw conclusions about the trust target. In this way, the principle of trust propagation becomes general and flexible, and thus applicable to the online communities of people, organisations and software agents. The higher purpose, however, is providing decision support based on collaborative interpretation of evidence by the community members. Consequently, this leads to enhancing the ability of online communities to support the emergent properties that result from transitive trust-relations among their members.

2 Subjective Logic Basics

2.1 General Opinions

A subjective opinion expresses belief about one or multiple propositions from a state space of mutually exclusive states called a *"frame of discernment"* or *"frame"* for short. Let X be a frame of cardinality k. Belief mass is distributed over the reduced powerset of the frame denoted as $\mathcal{R}(X)$. More precisely, the reduced powerset is defined as:

$$\mathcal{R}(X) = 2^X \setminus \{X, \emptyset\} = \{x_i \mid i = 1 \dots k,\ x_i \subset X\}\ , \tag{3}$$

which means that all proper subsets of X are elements of $\mathcal{R}(X)$, but X itself is not in $\mathcal{R}(X)$. The emptyset $\emptyset$ is also not considered to be a proper element of $\mathcal{R}(X)$.

An opinion is a composite function that consists of a belief vector $\vec{b}$, an uncertainty parameter u and base rate vector $\vec{a}$. An opinion can also have as attributes the belief source/owner. Assigning belief mass to an element in the frame (i.e. to a singleton or a subset) is interpreted as positive belief that the proposition(s) represented by that element is/are true, and as negative belief in their complements. The belief vector can be additive (i.e. sum = 1) or sub-additive (i.e. sum $<$ 1).

Uncertainty is expressed by not assigning the totality of belief mass, where the level of uncertainty is equal to the amount of unassigned belief. Uncertainty is here interpreted as the perceived imprecision of probability estimates. In case of total uncertainty, i.e. when $u = 1$, then the probability estimates of elements in the frame are equal to their base rate probabilities. The sub-additivity of the belief vector and the complement property of the uncertainty are expressed by Eq.(4) and Eq.(5) below:

$$\text{Belief sub-additivity:} \sum_{x_i \in \mathcal{R}(X)} \vec{b}_X(x_i) \leq 1\ , \quad \vec{b}_X(x_i) \in [0,1] \tag{4}$$

$$\text{Belief and uncertainty additivity:}\ u_X + \sum_{x_i \in \mathcal{R}(X)} \vec{b}_X(x_i) = 1\ , \quad \vec{b}_X(x_i), u_X \in [0,1]\ . \tag{5}$$

An element $x_i \in \mathcal{R}(X)$ is a *focal element* when its belief mass is non-zero, i.e. when $\vec{b}_X(x_i) > 0$. The frame X can not be a focal element, even when $u_X > 0$. The base rate vector, denoted as $\vec{a}(x_i)$, expresses the base rates of elements $x_i \in X$, and is formally defined below.

Definition 1 (Base Rate Function). Let X be a frame of cardinality k, and let $\vec{a}_X$ be the function from X to $[0,1]^k$ satisfying:

$$\vec{a}_X(\emptyset) = 0, \quad \vec{a}_X(x_i) \in [0,1] \quad \text{and} \quad \sum_{i=1}^{k} \vec{a}_X(x_i) = 1\ . \tag{6}$$

Then $\vec{a}_X$ is a base rate distribution over X.

An opinion is normally denoted as $\omega_X^A = (\vec{b}, u, \vec{a})$ where A is the opinion owner, and X is the target frame to which the opinion applies [3].

Definition 2. General Opinion
Assume X to be a frame where $\mathcal{R}(X)$ denotes its reduced powerset. Let $\vec{b}_X$ be a belief vector over the elements of $\mathcal{R}(X)$, let u_X be the complementary uncertainty mass, and let $\vec{a}$ be a base rate vector over the frame X, all seen from the viewpoint of the opinion owner A. The composite function $\omega_X^A = (\vec{b}_X, u_X, \vec{a}_X)$ is then A's opinion over X.

The belief vector $\vec{b}_X$ has $(2^k - 2)$ parameters, whereas the base rate vector $\vec{a}_X$ only has k parameters. The uncertainty parameter u_X is a simple scalar. A general opinion thus contains $(2^k + k - 1)$ parameters. However, given Eq.(5) and Eq.(6), general opinions only have $(2^k + k - 3)$ degrees of freedom. The probability projection of hyper opinions is the vector $\vec{\mathrm{E}}_X$ from $\mathcal{R}(X)$ to $[0,1]^\kappa$ expressed as:

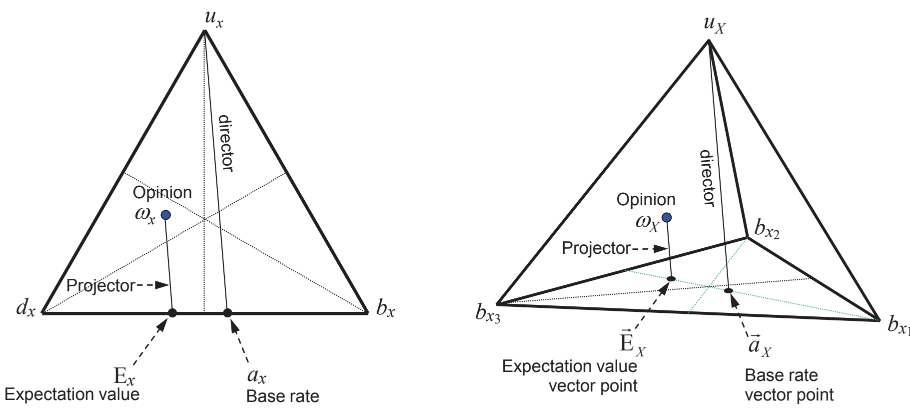
$$\vec{\mathrm{E}}_X(x_i) = \sum_{x_j \in \mathcal{R}(X)} \vec{a}_X(x_i/x_j)\, \vec{b}_X(x_j) \ + \ \vec{a}_X(x_i)\, u_X\ , \quad \forall\, x_i \in \mathcal{R}(X)\ . \tag{7}$$

Table 1 lists the different classes of opinions [10], of which *hyper opinions* represent the general case. Equivalent probabilistic representations of opinions, e.g. as Beta pdf (probability density function) or a Dirichlet pdf, offer an alternative interpretation of subjective opinions in terms of traditional statistics.

	Binomial opinion Binary frame Focal element $x \in X$	**Multinomial opinion** n-ary frame Focal elements $x \in X$	**General (Hyper) opinion** n-ary frame Focal elements $x \in \mathcal{R}(X)$
Uncertain $(u > 0)$ Probabilistic repr.:	**UB opinion** Beta pdf on x	**UM opinion** Dirichlet pdf over X	**UH opinion** Dirichlet pdf over $\mathcal{R}(X)$
Dogmatic $(u = 0)$ Probabilistic repr.:	**DB opinion** Probability of x	**DM opinion** Proba. distr. over X	**DH opinion** Proba. distr. over $\mathcal{R}(X)$

Table 1. Opinion classes with equivalent probabilistic representations

Specific opinion types can be visualised as a point inside a barycentric coordinate systems in the form of a regular simplex. In particular, Fig.1.a visualises a binomial opinion as a point inside an equal sided triangle, and Fig.1.b visualises a trinomial opinion as a point inside a tetrahedron. Hyper opinions or multinomial opinions larger than trinomial can not be visualised in the same way, and are in general challenging to visualise.



(a) Binomial opinion point in triangle (b) Trinomial opinion point in tetrahedron

Fig. 1. Visualisation of opinions in barycentric coordinate systems

In the triangle on the left hand side, the belief, disbelief and uncertainty axes run from one edge to the opposite vertex indicated by b_x, d_x and u_x. The base rate a_x of a binomial opinion are shown on the base line, and the probability expectation value E_x is determined by projecting the opinion point to the base line parallel to the base rate director. In the tetrahedron on the right hand side the belief, and uncertainty-axes run from one triangular plane to the opposite vertex indicated by the labels b_{x_i} and by u_X. A *vacuous opinion* is when the opinion point is at the top of the simplex $(u = 1)$, and a *dogmatic opinion* is when the opinion point is on the base line or plane $(u = 1)$. The base rate vector $\vec{a}_X$ of the trinomial opinion is shown as a point on the base plane, and the probability expectation vector $\vec{\mathrm{E}}_X$ is determined by projecting the opinion point onto the triangular base, parallel to the base rate director.

A special notation is used for representing binomial opinions over binary frames. A general n-ary frame X can be considered binary when seen as a binary partitioning consisting of one of its proper subsets x and the complement $\overline{x}$.

Definition 3 (Binomial Opinion). *Let $X = \{x, \overline{x}\}$ be either a binary frame or a binary partitioning of an n-ary frame. A binomial opinion about the truth of state x is the ordered quadruple $\omega_x = (b, d, u, a)$ where:*

b, belief: *belief mass in support of x being true,*
d, disbelief: *belief mass in support of $\overline{x}$ (i.e. NOT x),*
u, uncertainty: *uncommitted belief, uncertainty of probability expectation of x,*
a, base rate: *prior probability of x.*

We require $b + d + u = 1$ and $b, d, u, a \in [0, 1]$ as a special case of Eq.(5). The probability expectation value is computed with Eq.(8) as a special case of Eq.(7).

$$\mathrm{E}_x = b + au \ . \tag{8}$$

In case the point of a binomial opinion is located at the left or right base vertex in the triangle, i.e. with $d = 1$ or $b = 1$ and $u = 0$, the opinion is equivalent to boolean TRUE or FALSE, in which case subjective logic becomes equivalent with binary logic.

3 Previous Transitivity Operators of Subjective Logic

Several different trust transitivity operators for subjective logic have been proposed in the literature [9]. These are briefly described below.

Let A and B be two agents where A's trust in B's recommendations is expressed as $\omega^A_B = \{b^A_B, d^A_B, u^A_B, a^A_B\}$, and let C be an agent where B's trust in C is recommended to A with the opinion $\omega^B_C = \{b^B_C, d^B_C, u^B_C, a^B_C\}$. Let $\omega^{A:B}_C = \{b^{A:B}_C, d^{A:B}_C, u^{A:B}_C, a^{A:B}_C\}$ be A's derived trust in C as a result of the recommendation from B. Table 2 shows the derived opinion $\omega^{A:B}_C$ in case of uncertainty favouring transitivity, base rate sensitive transitivity, and opposite belief favouring transitivity.

Uncertainty favouring $\omega^{A:B}_C$	Base rate sensitive $\omega^{A:B}_C$	Opposite belief favouring $\omega^{A:B}_C$
$b^{A:B}_C = b^A_B b^B_C$ $d^{A:B}_C = b^A_B d^B_C$ $u^{A:B}_C = d^A_B + u^A_B + b^A_B u^B_C$ $a^{A:B}_C = a^B_C$	$b^{A:B}_C = \mathrm{E}(\omega^A_B) b^B_C$ $d^{A:B}_C = \mathrm{E}(\omega^A_B) d^B_C$ $u^{A:B}_C = 1 - \mathrm{E}(\omega^A_B)(b^B_C + d^B_C)$ $a^{A:B}_C = a^B_C$ $\mathrm{E}(\omega^A_B) = b^A_B + a^A_B u^A_B$	$b^{A:B}_C = b^A_B b^B_C + d^A_B d^B_C$ $d^{A:B}_C = b^A_B d^B_C + d^A_B b^B_C$ $u^{A:B}_C = u^A_B + (b^A_B + d^A_B) u^B_C$ $a^{A:B}_C = a^B_C$

Table 2. Trust transitivity operators of subjective logic proposed in the literature [9]

Uncertainty favouring transitivity means that A is uncertain about the trustworthiness of C not only to the extend that A is uncertain about the recommending agent B, but also to the extent that A distrusts B. *Base rate sensitive transitivity* means that that A's trust in C is a function of the expectation value of A's trust in the recommender B, which in turn is a function of the base rate. *Opposite belief favouring transitivity* means that *"the enemy of my enemy is my friend"*, i.e. that when A distrusts the recommending agent B, then A thinks that B consistently recommends the opposite of his real opinion, so that when C actually is trustworthy then B recommends distrust and vice versa.

While these three interpretations of trust transitivity provide relatively intuitive results in specific situations, none of them are suitable as a model for trust transitivity in general. The next section describes a general approach to modelling trust transitivity, where the three interpretations of Table 2 are special cases.

4 Conditional Belief Reasoning

Conditional reasoning with subjective logic is defined for binomial [8] and multinomial [4] opinions, and can also be extended to general opinions. For binomial deduction and abduction the following notation is used:

$\omega_{y|x}$: conditional opinion on y given that x is TRUE
$\omega_{y|\overline{x}}$: conditional opinion on y given that x is FALSE
ω_{x} : opinion on the antecedent state x
$\omega_{y\|x}$: deduced opinion on the consequent state y
$\omega_{y\|\hat{x}}$: hypothetically deduced opinion on y given vacuous $\omega_{\hat{x}}$

The image space of the deduced opinion $\omega_{y\|x}$ is a sub-triangle with base vertices defined by the two conditionals $\omega_{y|x}$ and $\omega_{y|\overline{x}}$, and the top vertex defined by the consequent opinion $\omega_{y\|\hat{x}}$ of the vacuous antecedent $\omega_{\hat{x}}$. This mapping determines the position of the consequent opinion $\omega_{y\|x}$ within the child sub-triangle, as illustrated in Fig.2.

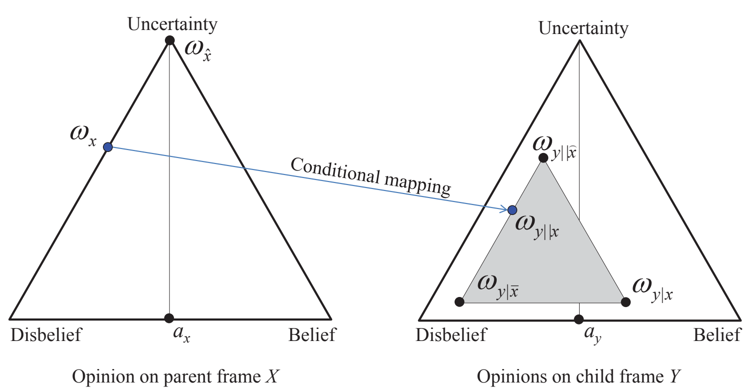


Fig. 2. Mapping from parent triangle to child subtriangle

It can be noticed that when $\omega_{y|x} = \omega_{y|\overline{x}}$, the child sub-triangle is reduced to a point, so that necessarily $\omega_{y\|x} = \omega_{y|x} = \omega_{y|\overline{x}} = \omega_{y\|\hat{x}}$ in this case. This means that there is no relevance relationship between antecedent and consequent.

In general, the child sub-triangle is not regular (equal-sided) as in the example of Fig.2. By setting base rates of x and y different from 0.5, and by defining conditionals with different uncertainty, the child image sub-triangle will be skewed (irregular), and it is even possible that the uncertainty of $\omega_{y\|\hat{x}}$ is less that that of $\omega_{x|y}$ or $\omega_{x|\overline{y}}$.

For multinomial opinions, let $X = \{x_i | i = 1 \dots k\}$ be the parent frame and $Y = \{y_j | j = 1 \dots l\}$ be the child frame. The general notation for conditionals is:

$\omega_{Y|X}$: set of conditional opinions on Y given that a specific state in X is TRUE
$\omega_{Y|\overline{X}}$: set of conditional opinions on Y given that a specific state in X is FALSE
ω_X : opinion on the antecedent (parent) frame X
$\omega_{Y\|X}$: deduced opinion on the consequent (child) frame Y
$\omega_{Y\|\hat{X}}$: hypothetically deduced opinion on Y given vacuous $\omega_{\hat{X}}$

Assume the antecedent opinion ω_X where $|X| = k$, and k conditional opinions $\omega_{Y|x_i}$. There is thus one conditional opinion for each element x_i, where a conditional opinion $\omega_{Y|x_i}$ expresses the subjective opinion on Y, given that x_i is TRUE. The subscript notation indicates not only the child frame Y it applies to, but also the element x_i in the parent frame it is conditioned on. The notation $\omega_{Y|X}$ expresses the set of $k = |X|$ different opinions conditioned on each $x_i \in X$ respectively.

Generalisation from the binomial case of Fig.2 to the multinomial case results in a parent regular simplex which is mapped to a sub-simplex inside the child simplex, defined by the set of conditional $\omega_{Y\|X}$. The ternary case is illustrated in Fig.3.

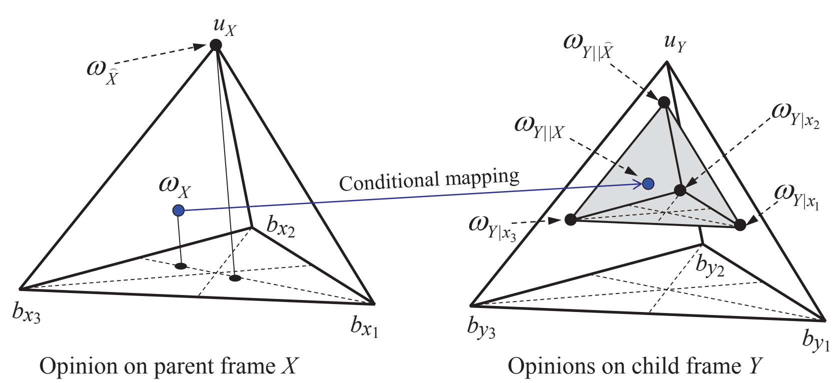


Fig. 3. Projection from parent opinion tetrahedron to child opinion sub-tetrahedron.

The sub-simplex formed by the conditional projection of the parent simplex into the child simplex is shown as the shaded tetrahedron on the right hand side in Fig.3. The position of the derived opinion $\omega_{Y\|X}$ is geometrically determined by the point inside the sub-simplex that linearly corresponds to the opinion ω_X in the parent simplex.

In general, a sub-simplex will not be regular as in the example of Fig.3, and can be skewed in all possible ways. The dimensionality of the sub-simplex is equal to the smallest cardinality of X and Y. Visualising a simplex larger than ternary (tetrahedron) is difficult. Subjective logic conditional deduction is expressed as in Eq.(9).

$$\omega_{Y\|X} = \omega_X \circledcirc \omega_{Y|X} \tag{9}$$

where $\circledcirc$ is the general conditional deduction operator for subjective opinions.

Reasoning in the opposite direction is called derivative reasoning, or abduction. Let $\omega_{X|Y} = \{\omega_{X|y_j} | j = 1 \ldots l\}$ be a set of $l = |Y|$ different multinomial opinions conditioned on each $y_j \in Y$ respectively. Conditional abduction is expressed as:

$$\omega_{Y\overline{\|}X} = \omega_X \overline{\circledcirc}\, \omega_{X|Y} \tag{10}$$

where $\overline{\circledcirc}$ is the general conditional abduction operator. Detailed expressions for evaluating Eq.(9) and Eq.(10) are provided in [4, 8, 14].

5 Conditional Reasoning about Trust Transitivity

The basic idea of conditional trust transitivity is to express trust in the target entity in terms of conditional opinions. Assume the recommendation frame $X = \{x, \overline{x}\}$ interpreted as x: *"Positive"* and $\overline{x}$: *"Negative"*. Trust in the recommender can then be expressed as the conditional propositions $y|x$: *"Target is trusted in case of positive recommendation"* and $y|\overline{x}$: *"Target is trusted in case of negative recommendation"*, where the conditional opinions $\omega_{y|x}$ and $\omega_{y|\overline{x}}$ are the actual trust values. Note that $\omega_{y|\overline{x}}$ typically contains disbelief, meaning that the target is distrusted in case of negative recommendation. As an example, trust in the recommender B can be expressed as the conditional opinions $\omega_{y|x} = (0.9,\ 0.0,\ 0.1,\ a)$ and $\omega_{y|\overline{x}} = (0.0,\ 0.9,\ 0.1,\ a)$, illustrated in Fig.4.

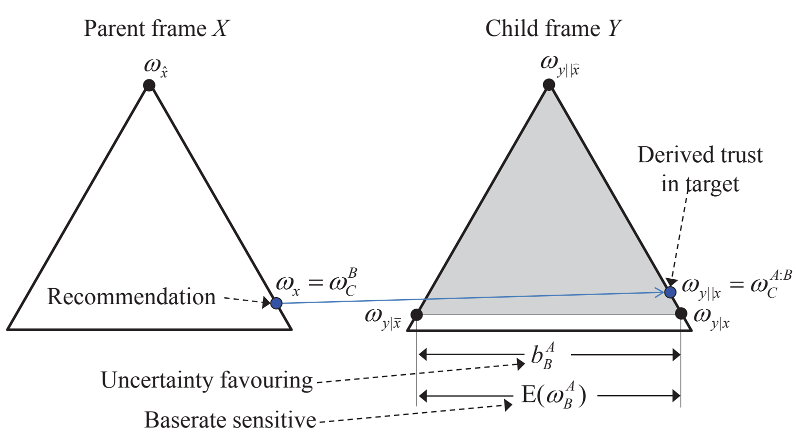


Fig. 4. Trusted recommender, derived transitive trust equals recommended trust.

In case of absolute trust the child sub-triangle is equal to the child triangle itself. The consequence of this is that the recommendation opinion in the parent triangle on the left hand side will be mapped to the same position in the child triangle on the right hand side. In other words, the relying party will believe whatever the recommender says.

Strong distrust in the recommender can be specified by placing the conditional trust opinions at, or close to, the uncertainty vertex, e.g. as $\omega_{y|x} = (0.3,\, 0.0,\, 0.7,\, a)$ and $\omega_{y|\overline{x}} = (0.0,\, 0.3,\, 0.7,\, a)$. The result is a small sub-triangle, or a point in case of total distrust, at the top of the child triangle, as illustrated in Fig.5.

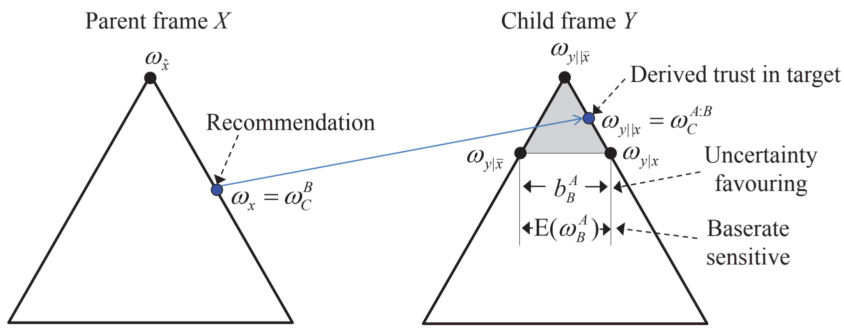


Fig. 5. Distrusted recommender, derived transitive conditional trust equals uncertainty.

The examples of Fig.4 and Fig.5 are specific instances of conditional trust transitivity that is equivalent to both the uncertainty favouring and the base rate sensitive transitivity operators of Table 2. Conditional trust transitivity is equivalent to the uncertainty favouring and the baserate sensitive transitivity operators whenever 1) the child subtriangle is equal-sided, 2) the positive conditional $\omega_{y|x}$ is the rightmost conditional, and 3) the uncertainty of the sub-triangle is maximised. The latter requirement means that the child sub-triangle must be positioned as high as possible inside the child triangle, as in Fig.4 and Fig.5. In case of the uncertainty favouring transitivity operator the base of the sub-triangle equals b_B^A. In case of the baserate sensitive transitivity operator, the base of the sub-triangle equals $\mathrm{E}(\omega_B^A)$.

The opposite belief favouring trust transitivity operator can be modelled by allowing the negative conditional $\omega_{y|\overline{x}}$ to be the rightmost trust opinion, e.g. expressed as $\omega_{y|x} = (0.2,\, 0.8,\, 0.0,\, a)$ (positive conditional trust opinion) and $\omega_{y|\overline{x}} = (0.8\,,0.2\,,0.0,\, a)$ (negative conditional trust opinion). The resulting sub-triangle is then flipped around, as illustrated with the check-pattern of Fig.6. However, the negative conditional is not necessarily the rightmost trust opinion, it can also be the leftmost trust opinion, as in the case of Fig.4. However, the situation where *"the enemy of my enemy is my friend"* and *"the friend of my enemy is also my enemy"* must be expressed with the negative conditional as the rightmost trust opinion, as in Fig.6 below.

In case of opposite belief favouring trust transitivity it can be seen that the derived trust opinion inside the child sub-triangle is the opposite of the recommendation in the parent triangle. The opposite belief favouring operator is similar to the uncertainty

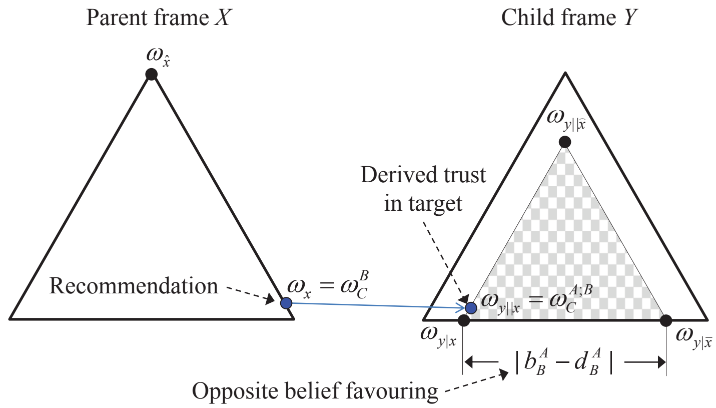


Fig. 6. Distrusted recommender, with the assumption that "the friend of my enemy is my enemy".

favouring operator in the sense that it does not consider the base rate of ω_B^A; the difference is that the sub-triangle produces minimal uncertainty, i.e. that it is located at the base of the child triangle. The criteria for the equivalence between the transitivity operators of Table.2 and conditional trust transitivity are expressed in Table (3) .

Criteria for equivalence between conditional trust transitivity and $\omega_C^{A:B}$		
Uncertainty favouring $\omega_C^{A:B}$	Base rate sensitive $\omega_C^{A:B}$	Opposite belief favouring $\omega_C^{A:B}$
$b_{y\|x} - b_{y\|\overline{x}} = b_B^A$ $b_{y\|x} = d_{y\|\overline{x}}$ $d_{y\|x} = b_{y\|\overline{x}}$ $u_{y\|x} = u_{y\|\overline{x}} = 1 - b_{y\|x} - d_{y\|x}$	$b_{y\|x} - b_{y\|\overline{x}} = \mathrm{E}(\omega_B^A)$ $b_{y\|x} = d_{y\|\overline{x}}$ $d_{y\|x} = b_{y\|\overline{x}}$ $u_{y\|x} = u_{y\|\overline{x}} = 1 - b_{y\|x} - d_{y\|x}$	$\|b_{y\|x} - b_{y\|\overline{x}}\| = \|b_B^A - d_B^A\|$ $b_{y\|x} = b_B^A$ $b_{y\|\overline{x}} = d_B^A$ $u_{y\|x} = u_{y\|\overline{x}} = 0$

Table 3. Equivalence criteria between transitivity operators [9] and conditional trust transitivity

With the conditional transitivity model it is possible to specify any form of trust transitivity. As an example, assume a recommender whom the relying party finds unreliable in the sense that positive recommendations can only be relied upon by 50%, expressed as $\omega_{y|x} = (0.5, 0.5, 0, a)$, and where negative recommendations are taken at face value, in order to be on the safe side, as expressed by $\omega_{y|\overline{x}} = (0, 1, 0, a)$. The resulting child sub-triangle is then the triangle illustrated on Fig.7.

The conditional trust transitivity model of Fig.7 is different from those specified in Table 2. In fact, conditional belief reasoning allows the specification of arbitrary trust transitivity models to suit any specific situation.

Table 3 demonstrates the formal side of the advantage of generalisation of conditional belief reasoning over the rest of the models in the current literature we stated at

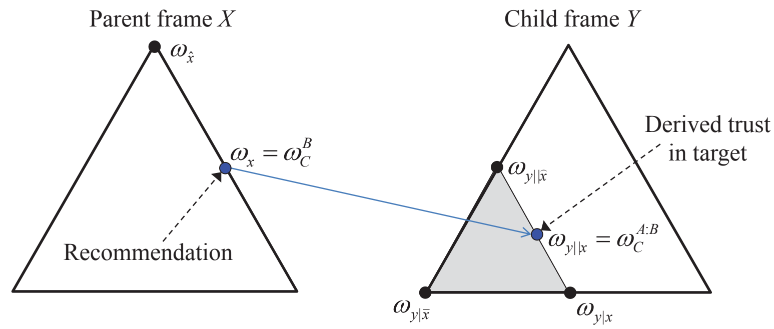


Fig. 7. Unreliable recommender, where positive recommendations are half trusted, and negative recommendations are fully trusted.

the beginning. Moreover, the provided example further clarifies the generalisation over the existing transitivity operators in Subjective Logic, showing the applicability of the approach in the context of trust modelling.

It is meaningful, however, to separate between the honesty and the ability to do something in the sense that an honest person will do their best to deliver a service, and the quality of the service then only depends on ability. A dishonest person who is able to deliver a quality service might on purpose deliver a low quality service.

In the previous examples it was implicitly assumed that trust in the recommender reflected the relying party's trust in honesty and ability simultaneously, except perhaps in case of the opposite belief favouring operator. When the relying party is able to assess honesty and ability of the recommender separately, it is possible to build a model which can express explicitly when a recommender does not provide their honest opinion. Trust in the presence of this assumption can be modelled as two separate conditional relationships, first between the recommendation and the recommender's internal trust opinion, and then between the recommender's internal trust opinion and the target trusted entity. The formal expression for trust transitivity from Eq.(1) can then be extended as in Eq.(11)

$$\text{Alice} \rightarrow \text{Claire} \;:=\; \text{Alice} \rightarrow \text{Bob's recomm.} \rightarrow \text{Bob's opinion} \rightarrow \text{Claire} \qquad (11)$$

Assume, for example, that the recommender Bob is a financial advisor and that Alice asks him about an investment product called C. Bob's recommendations are represented as a ternary frame X consisting of x_1: *"Says C is good "*, x_2: *"Says C is bad"* and x_3: *"Says don't know"*. Let further Bob's genuine opinion be represented as the binary frame Y consisting of y: *"Bob judges C to be good"* and $\overline{y}$: *"Bob judges C to be bad"*. Alice makes the following assumptions about the advisor Bob: If the advice is x_1 (Says C is good) it is probably a product that he gets a commission on, but he does not necessarily judge it to be good, so for Alice it is uncertain what he really thinks is good, formally expressed as $\omega_{y|x_1} = (0, 0, 1, a)$. If the advice is x_2 (Says C is bad) the Bob probably judges it to be a bad product, formally expressed as $\omega_{y|x_2} = (0, 1, 0, a)$. If the advice is x_3 (Says don't know) then it is possible that he genuinely is uncertain about the quality, but it is also possible that he judges it to be good but because he does not get a commission he does not want to recommend it, but at the same time does not want to be caught lying about a product which objectively is good; hence, he does not want to give a recommendation against it either, leaving him the option of x_3 (Says don't know). It is therefore possible that Bob is either uncertain, or that he judges the product to be good i.c.o. x_3 (Says don't know), formally expressed as $\omega_{y|x_3} = (0.5,\ 0.0,\ 0.5,\ a)$.

Assume in addition that Alice does not have full trust in Bob's ability to objectively assess investment products. Let the possible qualities of the investment product C be represented as the binary frame $Z = \{z, \overline{z}\}$ expressed as z: *"Good product"* and $\overline{z}$: *"Bad product"*. Then, Alice's doubt in Bob's ability can be expressed as $\omega_{z|y} = (0.5,\ 0.0,\ 0.5,\ a)$ and $\omega_{z|\overline{y}} = (0.0,\ 0.5,\ 0.5,\ a)$. The conditional connection between Bob's recommendations and Alice's derived trust in the investment product *Claire* is visualised in Fig.8.

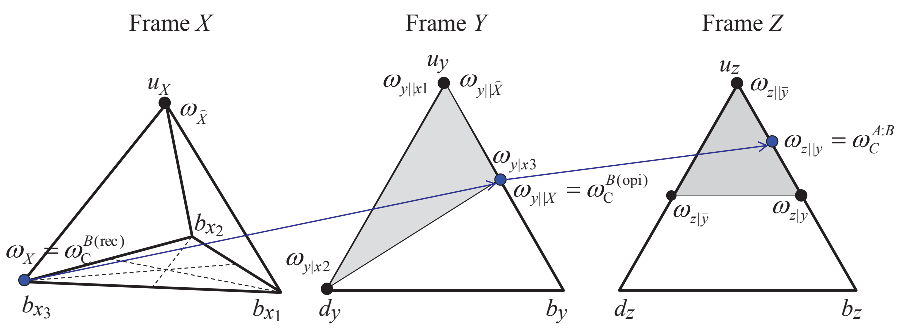


Fig. 8. Bayesian network for analysing recommendations from financial advisor.

The visual analysis of Fig.8 can be done mathematically using the methods for conditional reasoning described in [8,4]

6 Determining Conditionals

So far we have not discussed methods for determining the actual trust conditionals, and for this purpose it is important to be aware of the base rate fallacy [11]. As implicitly assumed in the analysis of Sec.5, the relying party A could simply express its subjective trust in B as opinions on the following conditional hypotheses:

$$\begin{array}{ll} \omega^A_{C|(B:C)} & : \text{"In case } B \text{ says } C \text{ has good quality, then } C \text{ has good quality"} \\ \omega^A_{C|(B:\overline{C})} & : \text{"In case } B \text{ says } C \text{ has bad quality, then } C \text{ has good quality"} \end{array} \qquad (12)$$

These are called *derivative conditionals* because they are expressed in anti-causal form, meaning that the act of saying that C has good or bad quality does not cause C to have good or bad quality. In reality, the opposite is the case, i.e. the fact that C has good or bad quality causes B to get an opinion on C's quality and to express recommendations about it, which thereby can be expressed as causal conditionals. The important point is whether B's opinion correctly reflects C's quality. The values of $\omega^A_{C|(B:C)}$ and $\omega^A_{C|(B:\overline{C})}$ thus rely on B's capability to correctly detect C's quality, which must be expressed as opinions on the corresponding causal conditionals:

$$\begin{array}{l} \omega^A_{(B:C)|C} : \text{"In case } C \text{ has good quality, then } B \text{ will say } C \text{ has good quality"} \\ \omega^A_{(B:C)|\overline{C}} : \text{"In case } C \text{ has bad quality, then } B \text{ will say } C \text{ has good quality"} \end{array} \quad (13)$$

Trust in B is naturally expressed in terms of the causal conditionals of Eq.(13), because they express the reliability of B as a sensor for detecting C's quality. B's assessments can of course be wrong sometimes, which represent cases of false positive or false negative. When interpreting *"C has quality"* as the positive case, then Eq.(13) express the TPR (True Positive Rate) and FPR (False Positive Rate). However, the opposite conditionals are needed (i.e. those of Eq.(12)) for assessing C's quality.

The conditionals of Eq.(12) are influenced by the base rate of quality in the population of C. A problem can arise when the derivative conditionals, such as Eq.(12), have been determined in a population with a specific base rate, using the same conditionals in a population with a different base rate, producing untruthful conclusions. As an example, consider a medical test for a specific disease, where the test gives mostly true positive and true negative results in a population where the disease is common. When the same test gives a positive result for a person in a population where the disease is extremely rare, then it is most likely a false positive result. Ignoring this fact is called the base rate fallacy in medicine. In the context of trust systems, a well-known problem is the one of exaggerated positive evaluations in systems like Amazon and eBay [15, 12], also known as positive bias. Failing to foresee the potential of base rate fallacy may lead to the conclusion that the system performs well just because of the low amount of false positives.

The base rate fallacy can be avoided by first determining the causal conditionals of Eq.(13), e.g. on a statistical or subjective basis, and subsequently inverting them into the form of Eq.(12) which takes into account the base rate of quality in target C's population. This is called derivative reasoning because of its nature, and can be done with the abduction operator of subjective logic, expressed in Eq.(10).

7 Discussion and Conclusion

Most of the current approaches that represent opinions of trust/distrust as a binary value of 1/0, without assigning any other attributes, dismiss all the shades of belief that exist between trust and distrust and capture only the two extremes of the state of trust met in the human perception and reasoning. Other approaches do account for the possibility to represent trust in a multi-valued manner, but assign maximum certainty to each of the values. Conditional belief reasoning, on the other hand, offers a framework for exploiting all the shades of trust-opinions that can be subject to the human perception and reasoning. By joining belief mass, uncertainty, and base rates into a single trust opinion, the model completely satisfies both the statistical and the subjective properties of trust inference and propagation that exist in the real world. It is reasonable to assume that, since trust in artificial systems is to an extent different, although derived from and modelled on, human subjective trust, the rules of transitivity can be more properly defined. Thus, when analysing trust transitivity through conditional belief reasoning, not only the frequentist nature of probability is captured (i.e., the count of positive and negative outcomes), but the ascribed subjectivity of trust also accounts for the impact of an entity's opinion on the outcome of a transaction.

The practical implications of employing conditional belief reasoning to address trust transitivity are much deeper than just providing the formal apparatus to reason about trust relationships in an intuitive way. One advantage of this approach is that the model easily accounts for situations where trust is not transitive, as well as situations where trust us transitive. Another advantage is that it successfully disentangles the notions of trust and reputation, in addition to acknowledging them as community values:

- it recognizes the influence of the subjective opinions of a single entity on the trust relationships established among community members – in this case, of a single recommender – on the perception of the relying entity about the target entity;
- it also implements the idea that the reputation, as a more general opinion about an entity's trustworthiness, results from the established trust relationships among all the community members.
- the introduction of conditionals in addition to the existing transitivity operators allows to capture the intuitive causality between trust and reputation that exists in the human reasoning, but is not always followed by rational decisions: e.g. "You are said to reputable, therefore i trust you", but also "You are said to be reputable, therefore i do not trust you";

Due to the formal framework for aggregating conflicting and non-independent opinions offered by subjective logic, a third advantage of the model is its power to reduce the complexity that arises from conflict resolution of differing (competing) trust-opinions. Moreover, by employing conditionals to infer an entity's trustworthiness based on presented evidence, the model accounts for the expectations of the relying party in the light of available evidence.

Although tacitly implied, it is worth pointing out the advantage of interoperability of conditional belief reasoning with subjective logic with the rest of the formal apparatus of statistics and probability theory used for modelling situations based on observed and statistical evidence. This makes conditional belief reasoning easily employable for the purpose of replacing, enhancing and adding functionality, or correcting some of the inefficiencies in the current models.

The main disadvantage of the model in its current state is that, although it accounts for the subjectivity of perceptions, the reasoning/inference phase still assigns a great deal of rationality on the side of the decision-maker. Despite of the fact that uncertainty is taken into consideration, its value results from a sound calculative model, rather than being an ad-hoc representation of the unpredictable nature of the transaction outcome. Further work will examine the potential for incorporating different aspects of this unpredictability, including via context, into the reasoning process, where changes in context (including, for instance, location or mobility) may change the subjective opinions of recommenders. This is interesting because, all other things being equal, changes in context may affect the final values even for the same recommenders.

References

1. 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.
2. 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.
3. A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.
4. A. Jøsang. Conditional Reasoning with Subjective Logic. *Journal of Multiple-Valued Logic and Soft Computing*, 15(1):5–38, 2008.
5. A. Jøsang, R. Hayward, and S. Pope. Trust Network Analysis with Subjective Logic. In *Proceedings of the* 29^{th} *Australasian Computer Science Conference (ACSC2006), CRPIT Volume 48*, Hobart, Australia, January 2006.
6. A. Jøsang, R. Ismail, and C. Boyd. A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems*, 43(2):618–644, 2007.
7. 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.
8. 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.
9. 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.
10. Audun Jøsang. Multi-Agent Preference Combination using Subjective Logic. In *International Workshop on Preferences and Soft Constraints (Soft'11)*, Perugia, Italy, 2011.
11. Jonathan Koehler. The Base Rate Fallacy Reconsidered: Descriptive, Normative and Methodological Challenges. *Behavioral and Brain Sciences*, 19, 1996.
12. Jo Mackiewicz. Reviewer Motivations, Bias, and Credibility in Online Reviews. In Sigrid Kelsey and Kirk St. Amant, editors, *Handbook of Research on Computer Mediated Communication*, pages 252–266. IGI Global, 2008.
13. Stephen Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, 1994.
14. Simon Pope and Audun Jøsang. Analsysis of Competing Hypotheses using Subjective Logic. In *Proceedings of the 10th International Command and Control Research and Technology Symposium (ICCRTS)*. United States Department of Defense Command and Control Research Program (DoDCCRP), 2005.
15. P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The Value of Reputation on eBay: A Controlled Experiment. *Experimental Economics*, 9(2):79–101, 2006. Avaliable from http://www.si.umich.edu/~presnick/papers/postcards/PostcardsFinalPrePub.pdf.

(♯) The work reported in this paper has been partially funded by UNIK.