# Constructive Representation of Trust: Single Rule Paradigm

Arthur Ramer
*School of Computer Science & Engineering*
*University of New South Wales*
*Sydney 2052, Australia*
*Email: ramer@cse.unsw.edu.au*

Robert E Marks
*Australian School of Business*
*University of New South Wales*
*Sydney 2052, Australia*
*Email: r.marks@agsm.edu.au*

***Abstract*—A constructive computational framework for trust and reputation assessments is presented. It is proven free from any inconsistent or contradictory assessments under any scenarios of its application. A prototype implementation has been developed.**

**The framework focuses on a single information-theoretical rule as inference mechanism, thus avoiding any biases or spurious constraints in the solutions.**

**The users of our model will find its results intuitively plausible, free from clustering or drift to the extrema. The entire framework is suited for a direct use in economic, financial and intelligence analyses.**

***Keywords*-trust; belief revision; maximum entropy; reputation.**

## I. Motivation

Almost all the economic activities, save for the purely adversarial scenarios, involve some significant element of trust among the participants. Open markets, the main framework of modern economies worldwide, are particularly dependent on the ability of assessing the reliability and trustworthiness of counterparties.

Trust is pervasive, often taken for granted as a market "lubricant", until it disappears [22]. Lack of trust between banks led to interbank credit markets freezing last October, when the TED spread exceeded 450 points. Trust is hard won, but easily lost, as governments have found as they struggle to kick-start interbank lending, and hence the world financial system. All the major players in this crisis tended to assume nearly perfect trustworthiness of their counterparts when estimating risk and utility of transactions. They live to regret it and would like, in the future, to be able to quantify, in a systematic and consistent fashion, *trust* due to a wide range of financial partners [5], [17].

The concept of trust has been discussed in religion, philosophy, and psychology since antiquity, but invariably in a qualitative fashion. Social and economic interactions used to be circumscribed, with counterparties being 'known quantities'. The advent of electronic exchanges and the globalisation of finance call for establishing a quantitative framework that can function over large, anonymous gatherings over extended periods, whether measured in time or number of transactions.

*Trust* should express reliability and truthfulness of counterparties, in particular their willingness to act to *our* best advantage if need be. Thus it is materially different from the simple probability of successful transaction - the other party may work in a very adverse environment. It is also different from its game-theoretic namesake. There trust is a strategic element that can be enforced through correct mechanism design.

As life- and business-critical decisions may be based on trust assessments, these assessments should be computed in the most consistent manner. It is especially important that the relative rank and ordering are preserved whenever feasible because a majority of decisions centre on simple choices in favour of the most trustworthy, most secure option. The actual magnitude of the defining rank is often less germane.

## II. Work to Date

Other proposals either use numbers in a purely ordinal way—say three-valued $(+, -, 0)$ to mean *trust, trust not, neutral)*, or as near-probabilities—they allow arithmetic, but either avoid stating its precise rules or do not assure the rules are indeed from any standard framework (say, stochastic or fuzzy).

An important sub-stream are purely axiomatic approaches. Some of them aim firstly at social choice such as an impossibility' lemma—a sufficiently strong collection of otherwise plausible postulates that lead to contradiction or triviality. Others posit fairly strong analytical rules (say, linearity or convexity) and thence derive the numerical results. Unfortunately, these numbers are usually obtained either through approximating procedures [13] or through complex, computationally demanding algorithms [3]. Tractable closed-form formulae are conspicuous through their absence.

Earlier publications have suggested a range of ad-hoc methods for linking probability values to trust. They are usually driven by a convenient link to some specific form of probability revisions; they often suffered from lack of completeness and absence of provably verified consistency.

A significant step forward was to ask whether some general postulates for assessing and revising trust values would lead to a consistent, preferably unique model. Such approaches have been discussed by several authors - Bhattacharya, Devinney and Pillutla [4], Debenham and Sierra [8], Neilson [25], Foo and Renz [9], Kwok, Foo and Nayak [21], Sandbu [32], Jøsang [18], and A Ignjatovic, N Foo, CT Lee [13]. The latter derived a unique numerical scheme under the assumption of linearity; however, it i quite strong and may introduce bias into solutions.

All the earlier proposals only consider adding new reports from which to compute trust. However, in all real life applications one often needs to *withdraw* incorrect or falsified reports. Any proper, consistent approach must permit both increases and decreases in the pool of valid source data.

There is one other significant lacuna in these models. While the results they compute may be perfect for machines, they often are not suitably plausible for humans. Decisions may be perceived quite differently if based on inequalities such as $0.5001 > 0.5$ rather than $0.6 > 0.5$. Likewise, probabilistic values, if not carefully controlled, may exhibit 0–1 drift (values clustering near the extrema of the range). We have not noted any consideration given to intuitive plausibility anywhere in the literature.

## III. Basic Design

We focus our introduction by giving a very basic scenario. We start with a simple situation where an agent receives a number of reports about the degree of trust he can place on a specific entity; that entity can be pictured, perhaps, as another agent acting in a constrained field of endeavour. We could picture an e-trading community who want to estimates their trading partners. They ask what the degree of *trust* can be placed on their business ethics and acumen, not just on their actual performance. For a change of pace, we could imagine that a security agency acquires a highly placed agent in the hostile organisation. The agent is not directly accessible; his trustworthiness and veracity need be assessed by the ordinary field sources.

All such and many similar scenarios require only that we abstract the notion of trust to just a handful of its properties that are computationally relevant.

We shall extrapolate from the suggestion of Jonker, Treur and Marx [15] to view trust as some numerical summary computed from (i) an initial trust value, (ii) a string of reported experiences, and (iii) the temporal frame of the reports. They had also presented a lengthy list of updating postulates. We replace all those with a simple probabilistic assumption: trust values are subjective probabilities and their values are revised through conditioning, whether direct or inverse. We term the reported values *experiences* and do not dwell on their nature. They simply fall in between 0 and 1, with a higher value meaning 'more trust'. We wish to combine the reported experiences into an overall evaluating of trust. We usually start from some initial default value of trust, with experiences serving to update it.

We report each experience $E_k$ as a probabilistic pair $(e_k, f_k), f_k = 1 - e_k$, with $e_k$ termed trust and $f_k$ distrust. Then all the reports, put together, should permit us to compute the composite trust and distrust. As the simple sums $e_1 + \ldots + e_k$ and $f_1 + \ldots + f_k$ would not from a probabilistic pair (in fact, their sum would be $k \gg 1$), we need a suitable scaling of $e_i$ and $f_i$. This is the most critical step of our methods - we want it to conform to the principle of minimum change and not rely on such simple-minded calculation like proportional scaling.

To derive the correct procedure let us first consider the obverse problem - that of *removing* a report. Thus, let us consider an already probabilistic family $P = \{e_1, \ldots, e_k, f_1, \ldots, f_k\}$ and remove the pair $e_k, f_k$. The minimal change solution $P' = \{e'_1, \ldots, e'_{k-1}, f'_1, \ldots, f'_{k-1}\}$ is the distribution closest in entropy to $P$, and this is well-know to be the conditional distribution of $P$. This points to the correct way of *inserting* a new report - we need to perform *inverse conditioning*.

## IV. AGM Probability revisions

Formalisations of belief change have been discussed, in various contexts, since 1970's. Notable specific applications are 'truth maintenance systems' [7] and 'database priorities' [10]. General, abstract protocols were introduced by philosophers Levi [23], [24], Harper [14], and then a series of works by Alchourron, Gardenfors and Makinson [1], [2], [12]. The last one gave the name to the system of postulates for belief revision as the *AGM framework*. Its basic design is founded on a revision scheme addressing needs of the finite propositional knowledge bases. In parallel with the purely logical framework there has been proposed a scheme for modifying *beliefs about probability* [12]. Here we consider a finite collection of *possible worlds* $X$ and a probability distribution thereupon. Any proposition $A$ may be held in some subcollection $X_A$ of these worlds, with its probability defined as the sum of probabilities of the worlds where it is held $P(A) := P(X_A)$. A proposition is *accepted* if its probability is 1; it is important to remember that it does not signify a universal acceptance, as there (usually) will be worlds, of probability 0, where the proposition may not hold.

*Expansion* of the state of beliefs wrt $A$ will mean adjusting the probability distribution to a such $P^+_A$ that $A$ is *accepted* $P +_A (A) = 1$. In keeping with the overall philosophy of such change, it is postulated that the passage from $P$ to $P^+_A$ should be effected with a *minimal* change. On a combination of philosophical and logical grounds it is strongly argued that such an expansion should be the *conditioning* wrt $A$, understood as

$$P^+_A(B) := P(A \wedge B)/P(A), \qquad P(A) > 0$$

and with a pseudo-distribution for the case of $P(A) = 0$. As a probabilistic operation it should be viewed as conditioning wrt the subset $A_X$; this subset may include some worlds of probability 0, but where $A$ is held. A group of four postulates is given to axiomatise it [12]

- $\boldsymbol{P^+1}$ For disjoint $A$ and $B$ ($\vdash \neg(A \wedge B)$), the distribution $P^+_{A\vee B}$ is a suitable convex combination of $P^+_A$ and $P^+_B$; taking $\alpha = P(A)/P(A \wedge B), \quad \beta = 1 - \alpha = P(B)/P(A \wedge B)$, one requires $P^+_{A\vee B} = \alpha P^+_A + \beta P^+_B$.
- $\boldsymbol{P^+2}$ $P^+_A(A) = 1$
- $\boldsymbol{P^+3}$ If $\vdash A$ then $P^+_A = P$
- $\boldsymbol{P^+4}$ If $P(A) = 0$ then $P^+_A = P_\perp$ - the 'absurd' pseudo-distribution[1]

The generalisation proposed by Jeffrey [16] was to combine conditioning wrt $A$ and its logical complement $\neg A$ as method of revision of probability functions. A revised function $P^{+J}$ should compute probability of an arbitrary assertion $B$ by conditioning wrt the property that $P^{+J}(A) = a$ for some $0 < a < 1$. He was led to the formula

$$P^{+J}(A) = aP(B|A) + (1 - a)P(B|\neg A).$$

It is convenient to express it in the language of expansions—Jeffrey conditionalisation becomes a linear combination of expansions wrt proposition $A$ and its complement $\neg A$. For the specified $0 < a < 1$

$$P^{+J}_A = aP^+_A + (1 - a)P^+_{\neg a}.$$

## V. Trust as Probabilistic Belief

Subjective probabilities are degrees of belief and their changes represent belief revision. The standard framework for modeling such changes is known as AGM, after the initials of its proposers [1], [12]. It is a system of postulates (revision scheme) for modifying probabilities under the control of propositional assertions. Objects of interest form a space $X$ of 'possible worlds', endowed with probability distribution $P$. An assertion $A$ defines a subset of operational worlds - admitting $A$ *restricts* the set of possible worlds to that $X_A \subset X$ where $A$ holds. This should lead to probability distribution $P^+_A$, supported on $X_A$, that obtains through some 'minimal' modification from $P$.

AGM state that $P^+_A$ should be $P$ conditioned onto $X_A$, and refer to the use of entropy for motivation. Namely, the conditional distribution $Q^* = P_{|A}$ can be found as the unique solution to minimising the entropy distance between the given $P$ and some $Q$ on $X_A$ [33]. Retraction of $A$ implies expanding the domain $X_A$; from $Q$ on $X_A$ we need to pass to $Q^-_A$ on the entire $X$. Direct use of entropy distance is unsatisfactory, insofar as it gives trivial results. The matter was left open by AGM, and later solved by us [28] - the correct method is to find $P^*$ of maximum entropy and such that it conditions back onto $Q$, that is $(P^*)^+_A = Q$. Such expansions and retractions are generalised by specifying constraints on the distributions on both $X_A$ and its complement $\bar{X}_A = X \setminus X_A$. Let $k$ be the cardinality of $\bar{X}_A$. Then the simple retraction of $A$ leads to $P^*$ described by

$$P^*(X_A) = \frac{2^{H(Q)}}{2^{H(Q)} + k}, \qquad P^*(y) = \frac{1}{2^{H(Q)} + k} \quad \text{for } y \in \bar{X}_A$$

If we specify both conditional distributions - one on $X_A$, the other on $\bar{X}_X$, we get the formulae from the previous section.

Suppose a probabilistic $P = \{e_1, \ldots, e_k, f_1, \ldots, f_k\}$ and try adding a new pair $\mathbf{p} = \{e_{k+1}, f_{k+1}\}$. An ensuing distribution $P''$ should satisfy

- conditioned to the first $k$ pairs it becomes $P$
- conditioned to the last pair it becomes $e_{k+1}, f_{k+1}$
- it has maximum possible entropy

A unique solution [29] obtains by 'compressing' the $P$-part and the $\mathbf{p}$-part to set their total probabilities to

$$\frac{2^{H(P)}}{2^{H(P)} + 2^{H(\mathbf{p})}} \quad \text{and} \quad \frac{2^{H(\mathbf{p})}}{2^{H(P)} + 2^{H(\mathbf{p})}}$$

More complex trust revision scenarios all use variants of this basic formula. While there is nothing difficult about those formulae, they cannot be just 'guessed' - their design and their properties need be meticulously proven.

## VI. Entropy and Max-Ent

Method of choice for revising probability assignments is *MaxEnt - maximum entropy* principle. It is usually introduced descriptively, based on very attractive properties of entropy function itself [11], the properties ranging from physics to coding theory. However, its use can also be justified axiomatically.

We seek a function $f(P, Q)$ of two probability distributions that can serve as an *objective* function for the passage from $P$ to $Q$. Given $P$ we want to find $Q$ that satisfies certain algebraic constraints and is closest to $P$ as measured by $f$

$$Q = \arg\min_R f(P, R)$$

Imposing just two consistency axioms on employing $f$ leads to 'inevitability' of entropy [26] - the only admissible function $f$ is either the entropy distance (aka cross-entropy, $I$-divergence, Kulback-Lieber metric)

$$f(P, Q) = D(P, Q) = \sum p_i \log \frac{p_i}{q_i}$$

or its monotonic transform.

The implications of this fact are quite profound - they signify that conditioning, thus also inverse conditioning, are in a sense universal operations in the realm of belief revision. Furthermore, various schemes of bayesian revision

[1] It is defined as assigning probability 1 to all subsets.

of probabilities can be justified and derived as applications of MaxEnt principle.

Another option is to look for an alternative numerical framework that is rich enough to admit an entropy-like functions. If such a framework be preferred for quantification of beliefs, then that framework can serve to express trust, at least in its numerical aspects. Our earlier research developed such entropy structures in possibility theory, based on fuzzy sets [20]. It means that one can discuss *quantitatively* fuzzy trust value scheme.

## VII. Conditioning, inverse conditioning and entropy

The simplest expansion problem can be posed as a question about finding $\hat{Q}$ on $X$ where $\hat{Q}(A_X) = 1$ for $A_X \subset X$ - the subset where $A$ holds. This $\hat{Q}$ should be as close as possible to the given $P$ on $X$. A natural solution [27], [33] would be

$$\arg \min_{Q:Q(A_X)=1} D(Q\|P).$$

The solution is the familiar conditional distribution $P(\dot{|}A) : x_i \mapsto p_i/P(A_X)$ if $A(x_i)$, and $x_i \mapsto 0$ if $\neg A(x_i)$. The same result obtains if $D(P\|Q)$ is used in its place (hence also for the symmetric distance $D(Q\|P)+D(P\|Q)$). More significantly, use of Renyi entropy [19] (or many others) does not affect the result.

Jeffrey formula [16] intends an expansion where $P_A^+(A) = a$ for some $0 < a < 1$, leaving $P_A^+(\neg A) = 1-a$, and is defined through

$$\begin{aligned} P_A^{+J}(x_i) &= p_i/a \text{ if } A(x_i), \\ P_A^{+J}(x_i) &= p_i/(1-a) \text{ otherwise.} \end{aligned}$$

It is immediate that

$$\arg \min_{Q:Q(A_X)=a} D(Q\|P)$$

gives this conditionalisation. An easy extension is to specify a partition $X = A^{(1)} \cup \ldots \cup A^{(k)}$, a probability assignment on its elements $A^{(i)} \mapsto a_i, \quad \sum a_i = 1$ and require that $P^+(A^{(i)}) = a_i$. Such a generalised Jeffrey rule results from a like minimisation of information divergence. Again, change of entropy function turns out to be immaterial.

We should also note that we need to use $D(Q\|P)$ and cannot obtain a reasonable answer directly from $H(Q)$. The proximate cause seems to be that we must 'retain' the knowledge of $P$ and then add the fact that $P^+(A) = 1$. Using $H(P)$ would appear to recognise only the latter fact.

An attempt to replicate the previous method of direct minimisation of a distance between the distributions is bound to fail. The nearest $P_A^-$ which conditionalises back to $P$ is the very same $P$. If we insist that $P^-$ must be different, an $\epsilon$-change to $P^-(A) = 1-\epsilon$ would ensue. However, we observe that in case of conditionalisation the entropy $H(P^+) < H(P)$, therefore the minimum of $D(Q\|P)$ is somewhat related to minimising the distance from $Q$ to the most *uninformed* ie. uniform distribution. While this cannot be used for deciding on $P^+$ (as we would loose the knowledge of $P$), it suggests a useful approach to the $P^-$ problem.

To make the question specific, we assume that $P$ is supported on $A_X \subset X$ and that there are $m$ elements outside $A_X$. We shall seek distribution $\hat{Q}$, with *maximum* entropy, that conditionalises back to $P$. We need to compute

$$\arg \max_{Q:Q_A^+(A)=P} H(Q).$$

The answer has a very attractive form

$$\begin{aligned} P_A^-(x) &= \frac{1}{m + 2^{H(P)}}, \qquad x \notin A_X \\ P_A^-(A) &= \frac{2^{H(P)}}{m + 2^H(P)} \end{aligned}$$

Noting that $m = 2^{\log m}$, which is the entropy of the uniform distribution on $m$ elements, permits to anticipate the effect of *inverting* Jeffrey conditioning. We first compute $H(P_A^+)$ and $H(P_{\neg A}^+)$. Denoting $P^{-J}$ for the *inverse* Jeffrey rule

$$\begin{aligned} P^{-J}(A) &= \frac{2^{H(P_A^+)}}{2^{H(P_A^+)} + 2^{H(P_{\neg A}^+)}} \\ P^{-J}(\neg A) &= \frac{2^{H(P_{\neg A}^+)}}{2^{H(P_A^+)} + 2^{H(P_{\neg A}^+)}} \end{aligned}$$

The extension to an arbitrary partition is straightforward. Moreover, while simple inverse conditioning assigns to all the elements outside $A_X$ the same probability, one can adopt the inverse Jeffrey rule to recognise some specified proportions. The simplest, albeit somewhat informal method is to view $P(\neg A)$ as having an 'infinitesimal' value, which becomes 0 in actual computations, but permits retaining some meaningful proportions.

## VIII. Quantitative Trust Values

We assume here the basic structure where we are given some initial trust value $T_0$ and a series of inputs - experiences $e_1, \ldots, e_k$ that make us revise our initial opinion. In the electronic age the reports are likely to come closely spaced in real time, with their order being just a random occurrence depending on vagaries of the communication network. We consider it very important to have a model which permits updates $u(T_0; e_1, \ldots, e_k) = T_k$ which would be invariant wrt reordering of $(e_i)$

$$u(T_0; e_1, \ldots, e_k) = u(T_0; e_{\sigma(1)}, \ldots, e_{\sigma(k)})$$

for any permutation $\sigma$ of the indices.

We model trust as belief emanating from a family of reported experiences $e_i$. These experiences are trust values specific to such single acts as trades, intelligence submissions and like; if fully satisfied we put $e_i = 1$, while if

completely dissatisfied $e_i = 0$. The beliefs will apply to the domain comprising all the act instances $a_i$, and, if desired, also of some initial, ie. prior to $a_1$ instances.

In the simplest approach - too simple as it will be seen - we would just form a domain of instances $a_i$ , and give each the weight $e_i$. This will not form probability for the most obvious reason - they do not sum to 1. Rather more importantly, given that we already have a (normalised) probability distribution $(e'_1, \ldots, e'_{k-1})$, new experience $e_k$ can only be a part of some other distribution, and it is not at all clear how should $e_k$ serve to update the already present $(e'_i)$. Lastly, if we have several experiences, each should contribute only a little to our overall computation of trust. It would be reasonable to look at their sum as the total experience value, but that would be always one.

These problems are remedied if one models more carefully even a single reported experience. We present it as a two-point distribution $(e_i, d_i)$, with $d_i = 1 - e_1$ and termed *distrust*. Similarly, we form the space of the pairs $\{\langle a_i, b_i\rangle\}$, where each $a_i$ will carry a trust contribution and each $b_i$ the corresponding distrust. We arrive at the space

$$\{a_1, b_1, a_2, b_2, \ldots, a_k, b_k\}$$

representing the set of inputs available at epoch $k$, with probability distribution $\mathbf{q}^k$.

As the probabilities $\mathbf{q}(a_i)$ and $\mathbf{q}(b_i)$ arose from the reported pair of values $(e_i, d_i), e_i + d_i = 1$, we postulate that

$$\mathbf{q}(a_i) \div \mathbf{q}(b_i) = e_i \div d_i.$$

Now the trust, at epoch $k$, becomes

$$T_k = \sum_{i=1}^{k} \mathbf{q}(a_i)$$

and distrust

$$D_k = \sum_{i=1}^{k} \mathbf{q}(b_i).$$

We may write $\mathbf{q}(a_i)$ for short, although we often need a full notation $\mathbf{q}_k(a_i)$, as these probabilities evolve with the number of epochs considered.

To decide how to perform an update with a new report $(e_{k+1}, d_{k+1})$ arriving, let us first consider the case of report *removal*. Let us suppose that the report $(e_k, d_k)$ is deemed no longer valid. We would be left with a subspace $\{a_1, b_1, \ldots, a_{k-1}, b_{k-1}\}$. The obvious step would be to condition $\mathbf{q}^k$ onto a smaller space and get

$$T'_{k-1} = \frac{\sum_{i=1}^{k-1} \mathbf{q}^k(a_i)}{\sum_{i=1}^{k-1} \mathbf{q}^k(a_i) + \sum_{i=1}^{k-1} \mathbf{q}^k(b_i)}.$$

It is most reasonable, though never so far considered in the literature, to require that

$$T'_{k-1} = T_{k-1}.$$

In other words, trust $T_{k-1}$ ensuing from experiences $e_1, \ldots, e_{k-1}$ and trust $T'_{k-1}$ obtaining from first considering $e_1, \ldots, e_k$ and then eliminating $e_k$, should be identical.

If experience removal effects probability conditioning on the act space, then experience inclusion should be an inverse conditioning on the same space. Now we can apply the basic formulae from the previous sections. We shall treat including the pair $(e_{k+1}, d_{k+1})$ as inverse Jeffrey conditioning.

Using Shannon entropy, let $H(T_k) = H(\mathbf{q}(a_1), \mathbf{q}(b_1), \ldots, \mathbf{q}(b_k))$ and $H(E_{k+1}) = H(e_{k+1}, d_{k+1})$. We require that

$$\mathbf{q}^{k+1}(\{a_1, \ldots, b_k\}) = \frac{2^{H(T_k)}}{2^{H(T_k)} + 2^{H(E_{k+1})}}$$

and

$$\mathbf{q}_{k+1}(\{a_{k+1}, b_{k+1}\}) = \frac{2^{H(E_{k+1})}}{2^{H(T_k)} + 2^{H(E_{k+1})}}$$

with probabilities of individual elements assigned proportionally. This defines $\mathbf{q}^{k+1}$ uniquely, thus determines $T_{k+1}$ and $D_{k+1}$ unambiguously.

It can be verified that any combination of inverse conditioning steps and conditioning steps is consistent. This leads to consistent updates of trust values for any conceivable scenarios of experience reports. Numerical computations are quite easy. It is known that entropy and graph entropy computations are tractable. For the basic trust computations they are simply linear. Perhaps because the proofs of compositionality of trust updated might be difficult, this property was never required by other researchers. Other basic properties are all easily proven to hold for our model.

## IX. Examples of Trust Updating

We recall two very simple examples from [29]. First, let us consider first an almost trivial situation

$$T_0 = t_0 = T' = t'_{-1} + t'_0, D_0 = d_0 = D'_0 = d'_{-1} + d'_0$$

where

$$t_0 = d_0 = \frac{1}{2}, \quad t'_{-1} = t'_0 = d'_{-1} = d'_0 = \frac{1}{4}$$

Let the new experience be $E_1 = (1, 0)$, ie. a totally positive trust report arrives. Following the rules of inverse conditioning wrt Jeffrey rule [16], we get for $(T_1, D_1)$ (evolved form $(T_0, D_0)$)

$$t_0, t_1 : \frac{1}{3}, \frac{1}{3}, \quad d_0, d_1 : \frac{1}{3}, 0$$

and $T_1 = \frac{2}{3} = 67\%$, $D_1 = \frac{1}{3} = 33\%$. Starting from $(T'_0, D'_0)$, we get at epoch 1

$$t_{-1}, t_0, t_1 : \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \quad d_{-1}, d_0, d_1 : \frac{1}{5}, \frac{1}{5}, 0$$

thus $T'_1 = 60\%$, $D'_1 = 40\%$.

Let us redo the same pair of examples, using $E_1 = (\frac{2}{3}, \frac{1}{3})$. Applying the entropy-based inverse conditioning gives, for $t_0 = d_0 = \frac{1}{2}$

$$\begin{aligned} t_0^{(1)} + d_0^{(1)} &= \frac{2^{H(\frac{1}{2},\frac{1}{2})}}{2^{H(\frac{1}{2},\frac{1}{2})} + 2^{H(\frac{2}{3},\frac{1}{3})}} = \frac{2}{2 + 3/2^{2/3}} \\ &= \frac{2}{2 + 1.5\sqrt[3]{2}} \\ t_1^{(1)} + d_1^{(1)} &= \frac{2^{H(\frac{2}{3},\frac{1}{3})}}{2^{H(\frac{1}{2},\frac{1}{2})} + 2^{H(\frac{2}{3},\frac{1}{3})}} = \frac{1.5\sqrt[3]{2}}{2 + 1.5\sqrt[3]{2}} \end{aligned}$$

Remembering that $t_0^{(1)} = d_0^{(1)}$ and $t_1^{(1)}/d_1^{(1)} = 2$, we find

$$T_1 = t_0^{(1)} + t_1^{(1)} = \frac{1 + \sqrt[3]{2}}{2 + 1.5\sqrt[3]{2}} \approx \frac{18}{31}$$

$$D_1 = d_0^{(1)} + d_1^{(1)} = \frac{1 + 0.5\sqrt[3]{2}}{2 + 1.5\sqrt[3]{2}} \approx \frac{13}{31}$$

If we start from $t_{-1}^{(0)} = t_0^{(0)} = d_{-1}^{(0)} = d_0^{(0)} = \frac{1}{4}$, we need use $H(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) = \log 4 = 2$ instead of $H(\frac{1}{2}, \frac{1}{2}) = 1$. We get

$$t_{-1}^{(1)} + d_{-1}^{(1)} + t_0^{(1)} + d_0^{(1)} = \frac{4}{4 + 1.5\sqrt[3]{2}}$$

$$t_1^{(1)} + d^{(1)} = \frac{1.5\sqrt[3]{2}}{4 + 1.5\sqrt[3]{2}}$$

As $t_{-1}^{(1)} = t_0^{(0)} = d_{-1}^{(1)} = d_0^{(0)}$, while $t_1^{(1)}/d_1^{(1)} = 2$

$$T_1' = t_{-1}^{(1)} + t_0^{(1)} + t_1^{(1)} = \frac{2 + \sqrt[3]{2}}{4 + 1.5\sqrt[3]{2}} \approx \frac{26}{47}$$

$$D_1' = d_{-1}^{(1)} + d_0^{(1)} + d_1^{(1)} = \frac{2 + 0.5\sqrt[3]{2}}{4 + 1.5\sqrt[3]{2}} \approx \frac{21}{47}$$

Finally $T_0' = T_0 < T_1' < T_1$, in agreement with two facts

- new experience $E_1$ is more positive (towards the trust) than the available trust value $T_0 = T_0' = \frac{1}{2}$;
- earlier history for $T_0'$ is more entrenched (two periods, at $k = -1, 0$) than that for $T_0$ (one period, at $k = 0$).

These examples emphasise a recurring theme - computations are not difficult conceptually, but even for the toy examples use of a computer becomes essential - we computed all the numbers above entirely 'by hand', but would not attempt it for a longer sequence of updates.

## X. Timed Experiences

We can observe that as the new trust data becomes more entrenched the effect of the initial experience reports should become slighter. The implication is that the experience which are more entrenched should have a stronger impact. A simple experience report at epoch $k$ is simply a pair of numbers $(e_k, f_k)$; to reinforce its impact, it would be most natural to repeat it, perhaps several times. (It would be akin to making one's point in a conversation by reiterating it several times.) However, to repeat a report would require bringing it up at the next several time instances. There is a better method - instead of reporting the pair $E_k = (e_k, f_k), e_k + f_k = 1$, we present a pair of $n$-sequences $E_k^n = \left(\frac{e_k}{n}, \ldots, \frac{e_k}{n}; \frac{f_k}{n}, \ldots, \frac{f_k}{n}\right)$. This will, of course, increase the entropy of the experience, thus its influence in the computation of the updated trust. This even though the totals of the $e$-values and $f$-values remain the same. To make it concrete, let the initial (default) trust be $t_0 = 0.5$ ($e = f = 0.5$) and the report be $t_r = \frac{2}{3}$ ($e = \frac{2}{3}, f = \frac{1}{3}$). Entrenchment over $n$ periods gives, after simplification, $t_n = \frac{1+\sqrt[3]{2}n}{2+1.5\sqrt[3]{2}n} \approx \frac{1+1.26n}{2+1.89n}$. It depends on $n$ as it should: with $n \to \infty$ it approaches $\frac{2}{3}$ - only the reported experience counts in the limit; while if $n = 0$ it is $\frac{1}{2}$ - with no experience reported it remains at the initial value. The effect of such replication has the desired effect of making the experience more credible. In fact, the formula matches the standard *credibility* formula as used in property-liability insurance [6] - not an unexpected link to actuarial science.

Just as new reports may carry higher relative weight, the older reported experiences may be accorded lower weight. We first generalise the entropy formulae to permit fractional multiplicities of the entries. We effect the time discounting by assigning decaying multiplicities to earlier trust and distrusts. A very reasonable method is to give the report that is $n$ epochs old the multiplicity $\frac{1}{n}$; assigning zero multiplicity would eliminate such an earlier report altogether.

## XI. Conclusion

Our entire approach is based on just two universal principles

- use of probability to express subjective judgements
- use of Occam's razor - the minimal change principle

to derive the full computational scheme. Proceeding from these principles only, and introducing no additional assumptions gives our scheme a high degree of logical consistency. It guarantees no unexpected reversals of comparative trust values, thus avoiding potential paradoxes in trust-based decision making.

Our design is very modular - we could, at various stages, have replaced use of probability for computations with another numerical framework. Equally, we could replace use of entropy [19] with another 'objective function' whose optimisation would determine the computed trust values. There are other paradigms of reasoning about beliefs besides probability; one better known relies on fuzzy sets and numbers. Without debating their relative merits, we simply point out that our entire design can be directly ported to such an alternative scheme.

## Acknowledgment

This research was supported by the Australia Research Council grant DP0210999 "Asynchronous Continuous Time

Conditioning" and by the FRGP funding in 2008 from the Faculty of Engineering, UNSW.

Technical preparation of this article was assisted by Christopher Petrov, Computing Support Officer of the School of Computer Science and Engineering.

## REFERENCES

[1] CE Alchourron, P Gardenfors, D Makinson. *On the logic of theory change: partial meet contractions and revision functions.* J. Symbolic Logic 50(1985), 510–530.

[2] CE Alchourron, D Makinson. *Hierarchies of regulations and their logic.* In R Hilpinen (ed) New Studies in Deontic Logic., 125–148. D Reidel Publ. Co, Dordrecht 1981.

[3] R Andersen, C Borgs, J Chayes, U Feige, A Flaxman, A Kalai, V Mirrokni, M Tennenholtz. *Trust-based recommendation systems: an axiomatic approach,* WWW 2008: Internet Monetization, Recommendation and Security, Beijing, April 2008.

[4] R Bhattacharya, TM Devinney, MM Pillutla. *A Formal Model of Trust Based on Outcomes,* Academy of Management Review, 23:3(1998), 459–472.

[5] G Brennan. *Democratic trust: a rational-choice theory view,* Ch. 8 in Trust and Governance, ed. VA Braithwaite, M Levi. Russell Sage Foundation, NY 1998.

[6] H Bühlmann. Mathematical Methods in Risk Theory. Springer Verlag, Berlin 1970.

[7] J Doyle. *A truth maintenance system.* Artificial Intelligence 12(1979), 231–272.

[8] JK Debenham, C Sierra. *Agents, Information and Trust,* AI'2005 - 18th Australian Conf. Artificial Intelligence, Sydney, Australia, December 2005.

[9] N Foo, J Renz. *Experience, Trust and Reputation,* Trends in Logic IV - Studia Logica Int. Conf. Torun, Poland, September 2006.

[10] R Fagin, J Ullman, M Vardi. *Updating logical databases.* Adv. Computing Research 3(1986), 1–18.

[11] S Guiasu. Information Theory with Applications. McGraw-Hill, New York 1977.

[12] P Garderfors. Knowledge in Flux. The MIT Press, Cambridge MA 1988.

[13] A Ignjatovic, N Foo, CT Lee. *An analytic approach to reputation ranking of participants in online transactions,* WI2008 - IEEE Int Conf Web Intelligence, Sydney, Australia Dec 2008.

[14] W Harper. *Rational conceptual change.* PSA'76 - Philosophy of Science Asooc. Biennial Meeting, East Lansing, Mich. 1976, 462–494.

[15] CM Jonker, J Treur. *Formal analysis of models for the dynamics of trust based on experiences,* in Multi-Agent System Engineering. Lecture Notes in AI 1647, Springer Verlag, Berlin, 1999. (Proc. 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World.)

[16] RC Jeffrey. The Logic of Decision. McGraw-Hill, New York 1965.

[17] A Jøsang, R Ismailb, C Boyd. *A survey of trust and reputation systems for online service provision,* Decision Support Systems, 43: 2, 2007, 618–644.

[18] A Jøsang. *Prospectives for Online Trust Management,* Working Paper 2008 (submitted to IEEE Trans. Knowledge and Data Eng.) http://persons.unik.no/josang/publications.html

[19] JN Kapur, HK Kesavan. Entropy Optimization Principles. Academic Press, New York 1992.

[20] G Klir, T Folger. Fuzzy Sets, Uncertainty, and Information. Prentice Hall, Englewood Cliffs, NJ 1988.

[21] R Kwok, N Foo, A Nayak. *Experience and Trust: A Sytems-Theoretic Approach,* UNSW Comp. Sci.&Eng. Technical Report UNSW-CSE-TR-0719, 2007.

[22] A Leigh. *Trust, Inequality and Ethnic Heterogeneity,* Economic Record, 82: 258, 2006, 268–280.

[23] I Levi. *Subjunctives, dispositions, and chances.* Synthese 34(1977), 423–455.

[24] I Levi. The Enterprise of Knowledge. The MIT Press, Cambridge, MA 1980.

[25] WS Neilson. *Axiomatic reference-dependence in behavior toward others and toward risk,* Economic Theory, 28(3), 2006.

[26] JB Paris. The Uncertain Reasoner's Companion: A Mathematical Perspective. Cambridge University Press, London 1995.

[27] A Ramer. *Note on defining conditional probability,* Amer. Math. Monthly, 97(1990), 336–337.

[28] A Ramer. *Belief revision as combinatorial optimisation,* IPMU-2002 - 9th Int. Conf. Information Processing and Management of Uncertainty, Annecy, France, July 2002.

[29] A Ramer. *Trust updating as belief revision,* IPMU-2006 - 11th Int. Conf. Information Processing and Management of Uncertainty, Paris, France, July 2006.

[30] A Ramer. *Computational quantification of trust updates,* AIDM2006 - 1st Int. Workshop on Integrating AI and Data Mining, Hobart, Tasmania, December 2006.

[31] A Ramer. *GraphMaxEnt,* in A Mohammad-Djafari, ed. Bayesian Inference and Maximum Entropy Methods. AIP Conference Proceedings, Vol. 872, Melville, New York 2006.

[32] MS Sandbu. *Axiomatic foundations for fairness-motivated preferences,* Social Choice and Welfare, 31(4), 2008.

[33] PM Williams. *Bayesian conditionalisation and the principle of minimum information,* Brit. J. Phil. Science 31, 1980, 131–144.