2010 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing

The Deficit and Dynamics of Trust

Abhaya C. Nayak
Department of Computing
Macquarie University
Sydney, Australia 2109
Email: abhaya.nayak@mq.edu.au

***Abstract*—Trust and Belief are two very closely connected notions. Hence one would expect that any mechanism that guides the management of one can efficaciously guide the other. In fact, devices such as Dempster-Shafer's theory of evidence have been successfully used in the management of both. This paper looks at a different mechanism devised for inductive inferencing, namely Spohn's *Ordinal Conditional Functions*, and shows that, when appropriately adapted and interpreted, it can fruitfully model the dynamics of trust via trust deficit.**

> *Do not trust people. They are capable of greatness.*
> Stanislaw Lem, 1963.

I. Introduction

Trust is important when multiple agencies are involved. In a typical situation, when there are other agents around, success of one in achieving its goals would require the cooperation – even if only passive cooperation – from others. Such cooperation cannot be taken for granted. Accordingly, the actions an agent would undertake aiming to achieve its goals would, to a significant extent, be constrained by who it trusts, and to what degree. For instance, to borrow an example from Yu and Singh [17], an agent's decision as to whether or not to buy a certain item from a particular seller in the eBay would be affected by whether or not the agent trusts the seller. Thus it is apparent that in any framework involving multiple agencies, each agent would maintain its estimate of how trustworthy other agents are. What measure should be used to represent trustworthiness of agents, how it should be evaluated, and how when required it should be updated are some of the seminal problems that must be satisfactorily dealt with in order to develop any multi-agent system.

Different approaches have been advocated to model trust in multi-agent systems. For instance, Yu and Singh [17] have used the Dempster-Shafer *belief functions* [13] in a rather simplified manner in their framework. Effectively, there is a belief function with respect to every agent a telling to what extent a is to be trusted, and to what extent distrusted. When recommendations are received, Dempster Rules of combination is used to combine such evidences. Colin Tan's approach to trust computation [15] is in many ways similar in the sense that it also uses belief functions; however it views trust as a weighted mixture of opinion and reputation. Liau [8] and Herzig, *et. al.* [6] have developed specialized modal logics of trust. Castelfranchi and Falcone [3], on the other hand, argue that the traditional way of modeling trust as a probability function, or using game theory is rather misguided, and that a more cognitively plausible account of trust needs to be developed.

Instead of looking at trust and its management through the coloured glasses of any particular theory, we will look at it as a rational phenomenon – as something people engage in in a principled manner, with the expectation that it will naturally lead us to develop a rational framework for representing trust and reasoning with it. Embracing political realism, we assume that it is not trust, but rather *distrust*, or in the politically more fashionable locution, *trust deficit*, that is the irreducible concept via which *trust* needs to be defined, and the rest of our framework is developed. In Section II we outline the rational principles that underpin the approach developed in this paper. This is followed by the development of the basic formal model in Section III that allows for determining how trustworthy a particular agent is. However, as the saying goes, the hand that feeds the hen eventually wrings its neck; hence it is only natural that the trustworthiness of an agent changes over time as new observations are made and new information is received. A rational account of how the trustworthiness of different agents are updated is developed in Section IV. Finally, we conclude in Section V with a brief outline of some related issues that we will take up in our future work.

II. Trust and its Deficit

Trust need not be innate to human nature. In realpolitik one does not take the actors to be working with a benevolent motive, but rather are assumed to be in a *state of nature* where, as Hobbes claimed way back in 1651, there is "continuall feare, and danger of violent death; And the life of man, solitary, poore, nasty, brutish, and short" [7]. Trust and cooperation are taken to emerge out of human behaviour in such a state of nature [2]. It then stands to reason to assume that it is not trust but rather its deficit, *suspicion* or *distrust* that should be taken to be a more basic concept, and the former should be defined in terms of it. Hence, we may define *trust* as follows:

Definition 1: An agent a trusts b just in case a is suspicious, not of b, but of the "enemies" of b.

In theory, there can be agents that are suspicious of every other agent. We would call such agents *paranoid* or *delusional*. Although the study of such agents is interesting from a psychological point of view, it is of little importance from a *rational* or *normative* perspective. Hence we will make the following assumption throughout:

978-0-7695-4322-2/10 $26.00 © 2010 IEEE
DOI 10.1109/EUC.2010.85

IEEE computer society

Assumption 1: No agent is paranoid; i.e., for every agent a there is an agent b (different from a) such that a is not suspicious of b.

A natural question that arises at this point is with respect to the notion of suspicion. Is it possible to give it an operational definition? The following may be taken to be an operational (but very informal) definition of suspicion:

Definition 2: An agent a is suspicious of b just in case a expects to be surprised by b's behaviour at some point.

Agent a is suspicious of b just in case a expects that b will "spring a surprise" (on a) at any moment. Thus a's trust in b is reducible to b's potential of springing a surprise on a. This is quite interesting since the *degree of potential surprise* is a well established concept in social sciences, particularly in the context of theory choice, its origin going back to the work of Shackle [12]. Its role in the development of the *logic of belief change* [1], [4], [10] is well established through the works of Wolfgang Spohn [14]. Thus there appears to be a natural connection between the readily available formal framework of belief change and theory choice on one hand, and a formal framework for a rational account of trust on the other.

We will not digress into the details of different formal theories developed on the basis of a degree of potential surprise. Instead, we will start developing our account of trust maintenance on the basis of basic principles, and later on in Section V point out its connection to certain existing theories.

III. Trust in Equilibrium

In Section II we outlined the basic reasons why a theory of trust needs to be based on the concept of potential surprise. In this section, we develop that intuition in a more formal manner, effectively providing the "statics" of trust.

The first required ingredient for developing a multi-agent systems is a set of agents. We will assume a pre-theoretic notion of *agent*, without going to further elaboration of their properties. Let $\mathcal{A}^+ = \{a_0, a_1, \ldots, a_n\}$ be the set of all agents. We use a with or without decoration to denote individual agents. Presumably, members of $\mathcal{A}^+$ observe each other's behaviour, manipulate them to serve their own ends, trust them, distrust them, and what have you. We are not going to deal with most of those issues. We will, instead, assume a contextually fixed, prominent member, say a_0, and develop an account of how the trust of a_0 in *other members* of $\mathcal{A}^+$, namely $\mathcal{A} = \{a_1, \ldots, a_n\}$, is defined and maintained. So although a_0 is a member of the agent-system, and is subject to all the systemic constraints that other members of are subject to, and members of $\mathcal{A}$ trust/distrust a_0 just as a_0 does to them, we will abstract away most of those things. In particular, we will assume that a_0 has the ability to distinguish itself from the rest of $\mathcal{A}^+$, and primarily deals with the members of $\mathcal{A}$ in its "model of the world".

Suspicion directed toward an individual.

We assume that a_0 has views as to which members of $\mathcal{A}$ are friends (respectively foes) of which members.[1] Hence, a_0 assigns to each member a_i of $\mathcal{A}$ a set of members of $\mathcal{A}$ that it considers are friends of a_i, and another set of members of $\mathcal{A}$ that it considers are foes of a_i. This we accomplish by assuming unary functions: $friends_{a_0}(\cdot)$ and $foes_{a_0}(\cdot)$ as defined below. Since the agent a_0 is contextually fixed, we will drop the subscripts for readability.

Definition 3:

1) The function $friends : \mathcal{A} \longrightarrow 2^{\mathcal{A}}$ maps an agent $a \in \mathcal{A}$ to the set of agents $friends(a)$ that are considered (by a_0) to be a's friends.
2) The function $foes : \mathcal{A} \longrightarrow 2^{\mathcal{A}}$ maps an agent $a \in \mathcal{A}$ to the set of agents $foes(a)$ that are considered (by a_0) to be a's enemies.

How a_0 arrives at these two functions is an important issue, but we take it to be extraneous to the account we are developing, and will not explicitly deal with it. There are many other things that we will leave unspecified. For instance, we will leave it open whether, the *Friend* and *Foe* relations corresponding to the $friends(\cdot)$ and $foes(\cdot)$ functions are reflexive, symmetric or transitive. We can graphically illustrate who is whose friend (or enemy) by drawing different types of arrows. In the figure Fig. 1 below, we take $\mathcal{A}$ to be composed of agents $a_1 \ldots a_{10}$. We identify the sets $friends(a_1) = \{a_4, a_9\}$ and $foes(a_1) = \{a_7, a_{10}\}$ by drawing appropriate solid and broken arrows respectively. The friends and enemies of other agents are not being shown to avoid clutter. Apart from keeping information as to who is whose friend or enemy, agent a_0 also has its opinion on who is likely to spring on it a surprise. We do this by assigning another function $surp_{a_0}(\cdot)$ to it, which, for each agent a in $\mathcal{A}$, gives a non-negative integer indicating a's degree of potential surprise (trust deficit, or suspicion). As before, we will drop the subscript a_0 for readability.

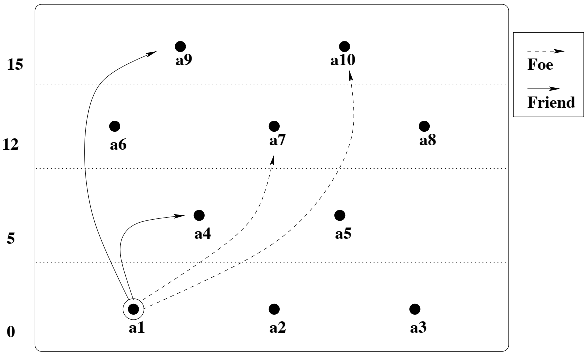

Fig. 1. The trustworthiness of agent a_1 is 12 since its surprise index is 0 and the least degree of suspicion directed toward its foes is 12.

Definition 4: The function $surp : \mathcal{A} \longrightarrow \mathcal{N}$ maps an agent $a \in \mathcal{A}$ to a non-negative integer $surp(a)$ which indicates how surprising a could be.

If $surp(a)$ is 0, it indicates that the agent a_0 is not suspicious of a at all. On the other hand, an inequality such as $surp(a_i) < surp(a_j)$ would indicate that the agent a_0 is more suspicious of a_j than of a_i. For all practical purpose, these

[1] Clearly it also has views as to which members of $\mathcal{A}$ are its own friends, and which are foes; but we are ignoring such issues for the sake of simplicity.

values are ordinal numbers, and we use them by and large for comparison. Keeping in mind Assumption 1, the following observation is relevant:

Observation 1: The agent a_0 is not paranoid if and only if there exists some agent $a \in \mathcal{A}$ such that $surp(a) = 0$.

Accordingly, in Fig. 1, there are some agents (namely, a_1, a_2, and a_3) that are assigned the degree of potential surprise 0. In fact, Assumption 1 leads to a more general constraint on the functions $surp(\cdot)$ as follows:

Observation 2: In presence of Assumption 1, given an arbitrary agent a_i, there exists some agent $a_j \in \mathcal{A}^+ \setminus \{a_i\}$ such that $surp_{a_i}(a_j) = 0$.
This observation states that since the agents we are dealing with are not paranoid, each of them places the minimum possible degree of distrust, 0, on at least one agent.

Suspicion directed toward a group.
Although an attitude like suspicion is primarily directed toward an individual, often, by extension it is directed toward a group of individuals. For instance, during an election, one might have strong suspicion against the working of particular interest groups. In principle, any subset $\mathcal{A}' \subseteq \mathcal{A}$ could constitute an interest group (or a group our eminent agent a_0 is interested in); and hence we should be able to discuss the degree to which a_0 suspects $\mathcal{A}'$. Accordingly we generalize the definition of the function $surp(\cdot)$ as follows:

Definition 5: $surp(\mathcal{A}') = min_{a \in \mathcal{A}'} surp(a)$
for any group of agents $\mathcal{A}' \subseteq \mathcal{A}$.
Note that this general definition of $surp(\cdot)$ gracefully deals with the special case when $\mathcal{A}'$ is a singleton set in that $surp(\{a\}) = surp(a)$ for any $a \in \mathcal{A}$. In this context, we define the notion of a $surp$-minimal agent as follows:

Definition 6: An agent $a \in \mathcal{A}$ is $surp$-minimal in $\mathcal{A}' \subseteq \mathcal{A}$ if and only if both $a \in \mathcal{A}'$ and $surp(a) = surp(\mathcal{A}')$.
The following observation is indicative of an alternative, equivalent formulation of this notion.

Observation 3: An agent a is $surp$-minimal in $\mathcal{A}' \subseteq \mathcal{A}$ if and only if both $a \in \mathcal{A}'$ and $surp(a) \leq surp(a')$ for every $a' \in \mathcal{A}'$.
Going back to Fig. 1, of particular interest would be the calculation of the degree to which agent a_0 would be suspicious of the friends of a_1 and of the enemies of a_1, that is, calculation of the values of $surp(friends(a_1))$ and of $surp(foes(a_1))$. Note that in this figure the set $\mathcal{A}$ is partitioned into equivalence classes modulo the $surp$-rank of its members, and the ranks (0, 5, 12 and 15) of different agents are indicated on the left. It is easily verified that these two values are respectively 5 and 12.

From suspicion to trust.
Trust is a positive attitude. It is more than simply the lack of suspicion. In order that agent a_0 positively trust some agent a, it is a necessary condition that a_0 bears no suspicion toward a; but that does not constitute a sufficient condition. As a rational agent, I have no reason to be suspicious of a perfect stranger; but that does not entitle me to trust them. This is the rationale behind Definition 1 which was couched in informal terms. Now we are in a position to offer a more formal and concrete definition of trust. As before, the reference to the agent a_0 is being dropped for readability.

Definition 7:

1) An agent a is trusted (by a_0) if and only if both $surp(a) = 0$ and $surp(foes(a)) > 0$ for all $a \in \mathcal{A}$.
2) Given an agent $a \in \mathcal{A}$ trusted by a_0, the degree of trust that a_0 places on a is given by $trust(a) = surp(foes(a))$.

Thus, going back to Fig. 1, the degree of trust that a_0 places on the agent a_1 is 12 since $surp(a_1) = 0$ and $surp(foes(a_1)) = surp(\{a_7, a_{10}\}) = 12$. On the other hand, since $surp(a_5) = 5$, it follows that a_0 does not trust a_5 at all. In fact, a_0 distrusts (is suspicious of) a_5 to the degree 5.

Now we are in a position to define *trust in a group*. Trusting a set of agents $\mathcal{A}'$ entails trusting every member of that group. The degree of that trust is the minimal trust earned by members of that group in the relevant manner.[2]

Definition 8:

1) Given any $\mathcal{A}' \subseteq \mathcal{A}$, agent a_0 trusts $\mathcal{A}'$ if and only if a_0 trusts every agent $a \in \mathcal{A}'$.
2) Given that agent a_0 trusts some $\mathcal{A}' \subseteq \mathcal{A}$, the degree of trust that a_0 places on $\mathcal{A}'$ is given by $trust(\mathcal{A}') = min_{a \in \mathcal{A}'} trust(a)$.

Referring back to Fig. 1 again, we can easily verify that $trust(\{a_1, a_5, a_6\})$ is undefined since the agents a_5 and a_6 are not trusted. On the other hand, $trust(\{a_1, a_2, a_3\})$ cannot be evaluated at this point since the values of $foes(a_2)$ and $foes(a_3)$ have not been specified.

IV. The Dynamics of Trust

So far, we have outlined the the basics of a framework to describe a *trust state* that is in equilibrium – the *statics of trust*, as it were. However trust is not constant. It changes over time subject to an agent making new observations, receiving new information, and even receiving recommendations from others. For instance, if agent a_0 were to receive a recommendation from a reputable source that a_5 is a very trustworthy person, a_0's evaulation of the trustworthiness of a_5 will change. How this change is effected – the *dynamics of trust* – is the topic of this section.

There could be many reasons that may initiate change in a_0's trust in other agents. Without any loss of generality, we can formulate it as the study of how the function $surp(\cdot)$ changes in response to relevant triggers. However, some such causes such as the observations a_0 makes of the behaviour of other agents cannot be dealt with within our simple framework without enriching it, and hence their discussion is beyond the scope of this paper. We will restrict the discussion of trust dynamics to how $surp(\cdot)$ changes in response to received recommendations of the appropriate sort.

[2] It is not always correct. I might trust a company because its CEO is my trusted friend, even if I don't trust many employees of this company. But such power relations cannot be captured in our framework without adding further structure to it.

There are two important things to consider while discussing how recommendations effect trust evaluation, namely, (1) Who makes the recommendation? In particular, what is their reputation? and (2) What form does the recommendation take? Let us discussion these two points at some length.

(1) Who makes the recommendation? The answer to this question makes a big difference as to how the recommendation effects a_0's trust function. Imagine that the recommendation comes from someone whom a_0 considers to be completely unreliable. In such eventuality a_0 is likely to ignore the recommendation. On the other hand, if the recommendation comes from someone whose judgment a_0 trusts, then it will take up the recommendation seriously and "update" its trust values in different agents. There are other relevant factors as well. If the recommendation regarding a_j comes from a_i, and a_0 "knows" that a_j is a friend of a_i, then chances are a_0 would take the recommendation not as seriously as it normally would; it might in fact "discount" the recommendation to some extent. On the other hand, if a_j is believed to be an enemy of a_i, then it would have the opposite effect. The framework developed here has the efficacy to deal with these variations in a judicious manner. In this paper we will primarily deal with the simplest case, namely when the recommendation comes from someone whom agent a_0 considers to be practically infallible in its judgment, has no conflict of interest with the agent the recommendation about, and so on. It is, as it were, the recommendation comes from the God, and agent a_0 does have to take the recommendation as is. As we mention in Section V, we will take up the other cases in our future work.

(2) What form does the recommendation take? This is the other issue with significant bearing upon how the trust function of a_0 gets updated. There are many aspects of the recommendations related to *speech acts* [11] and *conversational implicature* [5] that are clearly beyond the scope of this paper. For instance, the richness of a recommendation such as *"I cannot recommend agent a_{007} highly enough"* cannot be expressed without having a very sophisticated language in place. We will instead deal with recommendations expressed in a very limited way, such as: *I recommend you to trust agent* a_5, and *I recommend you to trust agent* a_5 *to degree* 3. In fact we will primarily deal with the second form of recommendation, and briefly mention in Section V how the former can be treated as a special case of it.

Why is trust update a problem?

So let us assume that agent a_0 has received a recommendation from an infallible source to trust agent a_5 to degree 3. This will clearly effect a change in the trust function of a_0 which is reducible to its surprise function $surp$. Thus, we will restrict our discussion to how agent a_0's function $surp(\cdot)$ gets updated in light of the received recommendation of the form $new\_trust(agent\ a, value\ x)$. In response to this recommendation, agent a_0 will update its function $surp(\cdot)$ to, say, $surp'(\cdot) = surp(\cdot \mid \langle a, x\rangle)$ such that when the trust value of agent a is computed modulo $surp'(\cdot)$, we will get the result x. Borrowing terminology from probability theory, we well term $surp(\cdot)$ to be the *prior surprise function* of the agent a_0, and $surp'(\cdot) = surp(\cdot \mid \langle a, x\rangle)$ to be its *posterior surprise function*. When no confusion is imminent, we will represent the posterior $surp'(\cdot) = surp(\cdot \mid \langle a, x\rangle)$ simply as $surp_{\langle a,x\rangle}(\cdot)$.

Definition 9:

1) A *trust recommendation* is denoted by a pair $\langle a, x\rangle$ where $a \in \mathcal{A}$ is an agent, and $x \in \mathcal{N}$ is the recommended trust-value of a.
2) $surp(\cdot \mid \langle a, x\rangle)$ represents the "posterior surprise function", also denoted $surp_{\langle a,x\rangle}(\cdot)$, after the prior $surp(\cdot)$ has been updated in a principled manner in light of the trust recommendation $\langle a, x\rangle$.

We are assuming that a recommendation is accepted at its face-value, that is, a_0 does not further negotiate or scale the recommended trust value x of a after receiving the trust recommendation $\langle a, x\rangle$. Accordingly, the posterior surprise function $surp_{\langle a,x\rangle}(\cdot)$ should be so constructed that the trust value for the agent a generated from it as per Definition 7 should be x. This naturally leads to the following two constraints:

Assumption 2:

1) $surp_{\langle a,x\rangle}(a) = 0$
2) $surp_{\langle a,x\rangle}(foes(a)) = x$

Consider for instance the prior surprise function $surp(\cdot)$ represented by the central rectangle in the figure Fig. 2 (labelled "A"). According to this scenario, the agent a_0 distrusts a_5 to degree 5; the friends of a_5 are two, namely a_8 and a_{10}; and a_5 also has two enemies: a_1 and a_6. Suppose also that the agent a_0 receives a recommendation $\langle a_5, 3\rangle$ that it must accept at face value. Given Assumption 2, it then follows that $surp_{\langle a_5,3\rangle}(a_5) = 0$, and $surp_{\langle a_5,3\rangle}(\{a_1, a_6\}) = 3$. Questions arising:

1) Clearly the posterior surprise function $surp'(\cdot) = surp_{\langle a_5,3\rangle}(\cdot)$ will assign a_5 the value 0. What would it do to the friends of a_5? For instance, should $surp(a_8)$ be left untouched in this updating process, or $surp_{\langle a_5,3\rangle}(a_8)$ would differ from it? If the latter, what would be the correct value?
2) The condition $surp_{\langle a_5,3\rangle}(\{a_1, a_6\}) = 3$ implies that the minimum of the values assigned by $surp_{\langle a_5,3\rangle}(\cdot)$ to a_1 and a_6 is 3. But there is no way of telling which of the two, a_1 and a_6, will receive the value 3.
3) As in the case (1) above, there is still the issue to be settled as to whether or not, given that in this trust updating process, the $surp(\cdot)$ value of a_1 is changed from 0 to 3, whether that of a_6 should be left unchanged.

Admittedly there is no hard-an-fast rule as to how these three issues are to be settled. Each of them admit of multiple solutions. For instance, unless further reasonable constraints are imposed, the following, admittedly bizarre, solution will fit the bill as well as any other:

Proposal 1 (a silly proposal): For all $a' \in \mathcal{A}$,

$$surp_{\langle a_5,3\rangle}(a') = \begin{cases} 0 & \text{if either } a' = a_5 \\ & \quad \text{or } a' \in friends(a_5) \\ 3 & \text{otherwise} \end{cases}$$

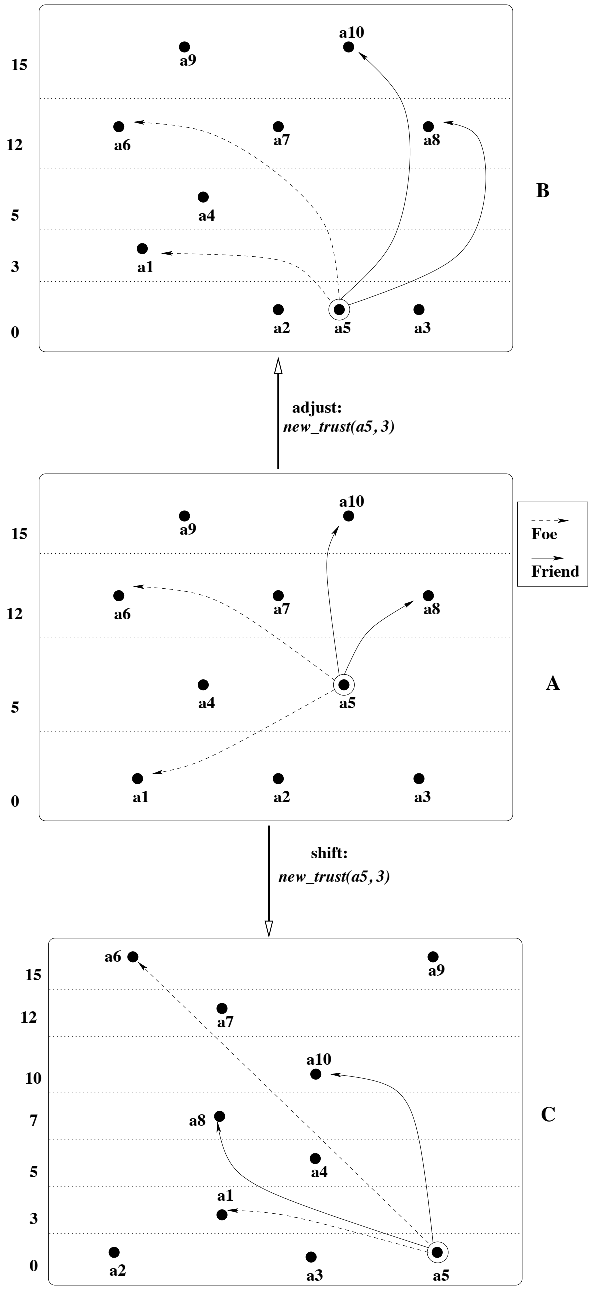

Fig. 2. The trustworthiness of agent a_5 who was distrusted to degree 5, upon recommendation, is now trusted to degree 3. In the *trust adjustment* process (labelled "B"), its friends a_8 and a_{10} are not distrusted any less; only its foe a_1 is now distrusted more. In the *trust shift* process (labelled "C"), its friends a_8 and a_{10} are distrusted less; and its foes a_1 and a_6 are distrusted more.

Such a solution is clearly not defensible. Further principled constraints are required to prevent such whimsical trust modification. We outline below two proposals, called *trust adjustment* and *trust shift*, that incorporate such constraints.

Trust Adjustment

One way of addressing this problem is to follow the *principle of minimal repair*. We may assume that agent a_0 has developed its $surp(\cdot)$ function over time, and in a certain sense, this function summarizes a lot of historical information. For instance it carries the information that while a_2 may possibly be trusted, a_9 should not be trusted at all. Such information is valuable, and it should not be lost without very good reason. However, Proposal 1 will obliterate such distinction, even when the trust recommendation, $\langle a_5, 3\rangle$ has nothing to do, directly or indirectly, either with a_2 or with a_9. It is guilty of *over repair*. We should not rectify the $surp(\cdot)$ function more than necessary.

How much repair in $surp(\cdot)$ is really necessary? Strictly speaking, the really necessary changes are explicitly identified in Assumption 2. If we follow this recipe, then, in the context of Fig.2, the following modification in $surp(\cdot)$ would appear to be appropriate:

Proposal 2: For all $a' \in \mathcal{A}$,

$$surp_{\langle a_5,3\rangle}(a') = \begin{cases} 0 & \text{if } a' = a_5 \\ 3 & \text{if } a' \text{ is } surp\text{-minimal} \\ & \qquad \text{in } foes(a_5) \\ surp(a') & \text{otherwise.} \end{cases}$$

According to this proposal, then, the $surp$-value of a_5 will be reduced to 0, and that of a_1 will be increased to 3. The $surp$-value of every other agent in $\mathcal{A}$ will be left as is. This appears to be the approach that ensures that the a_5 will be trusted to degree 3 with minimal repair to the function $surp(\cdot)$. However, in general this approach will not work. In particular, supposethat the agent a_5 had one more enemy, say a_4 with $surp(a_4) = 2$. In such a case, after $surp(\cdot)$ was updated to $surp_{\langle a_5,3\rangle}(\cdot)$ in response to the received trust-recommendation $\langle a_5, 3\rangle$ following the procedure specified in Proposal 2, the agent will be trusted to the degree 2 instead of the desired degree 3. Hence a slight correction in Proposal 2 is due. The *trust adjustment* function is accordingly defined:

Definition 10 (Trust Adjustment): For $a,\ a' \in \mathcal{A},\ x \in \mathcal{N}$,

$$surp^{adjust}_{\langle a,x\rangle}(a') = \begin{cases} 0 & \text{if } a = a' \\ x & \text{if both } a' \in foes(a), \\ & \qquad \text{and } surp(a') < x \\ surp(a') & \text{otherwise.} \end{cases}$$

The figure Fig. 2 (B) graphically illustrates how the $surp$-values of different agents get adjusted in this process.

Trust Shift

Although *trust adjustment* satisfies both Assumption 2, and the principle of minimal repair, one might take issues with it on account of the way it distorts the relationship between different friends (or enemies) of the agent a, the focus of the recommendation $\langle a, x\rangle$. Consider again Fig. 2. All the three friends, a_5, a_8 and a_{10} were distrusted to various degrees, ranging from 5 to 15. A recommendation was received that the agent a_5 is fairly trust worthy, and should be trusted to degree 3, and a_0 accordingly trusted a_5. It is only rational to expect that agent a_0 would now distrust a_8 and a_{10} less than it did before. But that does not happen. Similarly, this recommendation "taints" a_5's enemy a_1, who was not distrusted before. However, there is no corresponding loss of trust in a_6, the other enemy of a_5. To this extent, the *trust adjustment* approach is rather unsatisfactory. It will be more appropriate to shift *en bloc* the surprise values of a_5's friends (or enemies), up or down as the case may be, by the same magnitude. This is captured as follows:

Proposal 3: For all $a' \in \mathcal{A}$,

$$surp_{\langle a_5,3\rangle}(a') = \begin{cases} 0 & \text{if } a' = a_5 \\ max(0,(surp(a') & \\ \qquad\qquad -5)) & \text{if } a' \in friends(a_5) \\ surp(a') + 3 & \text{if } a' \in foes(a_5) \\ surp(a') & \text{otherwise.} \end{cases}$$

The result of such uniform shift is illustrated in the figure Fig.2 (C). It is easily verified that the relative *trust deficit* between different friends (as between different enemies) of a_5 is not affected via this trust-update process. This proposal, when generalized, leads to the following definition:

Definition 11 (Trust Shift): Let $a,\ a' \in \mathcal{A},\ x \in \mathcal{N}$, and y denote $surp(foes(a))$.

$$surp^{shift}_{\langle a,x\rangle}(a') = \begin{cases} 0 & \text{if } a = a' \\ max(0,(surp(a')- & \\ \qquad\qquad surp(a))) & \text{if } a' \in friends(a) \\ surp(a') - y + x & \text{if both } a' \in foes(a) \\ & \qquad\qquad \text{and } x \neq y \\ surp(a') & \text{otherwise.} \end{cases}$$

The first clause in this definition ensures that a is not distrusted. Clearly this reduced the $surp$-value of a by $surp(a)$; and that is the magnitude by which we would like to reduce the $surp$-value of all its friends. This is done in the second clause, with the proviso that the revised $surp$-value has a floor value of 0. The more interesting case is captured by the third clause. Note that y, defined as $surp(foes(a))$, is really, as per Definition 5, the minimal $surp$-value associated with any enemy of a. Since a' is an enemy of a, it follows that y will never exceed $surp(a')$, and hence $surp(a') - y + x$ will always be non-negative, and the $surp$-value of all enemies of a will be uniformly shifted, up or down, by the same magnitude of $|x - y|$, as desired. The last clause effectly says that agents that are neither friends nor enemies of a are not affected.

V. Future Outlook

In this paper we developed a simple framework for representing the statics of trust, and provided two reasonable mechanisms for managing its dynamics. In the framework, ironically, what is directly represented is not trust, but rather *trust deficit* (suspicion or surprise), and trust is defined in terms of it. The format of the recommendation we assumed is a pair $\langle agent, new\_trust\rangle$. The mechanisms proposed ensures that after the recommendation is processed, the trust in $agent$ gets updated to new_trust, and the trust value of other agents are modified, if required, in a rational fashion.

There are many issues related to this proposal that we could not discuss due to space limitation. Some of these issues are:

1) It can be easily proved that an agent's *friends* and *enemies* constitute two mutually exclusive sets.
2) If we also assume that every agent is its own best friend, and whoever is not a friend is an enemy, this framework will resemble Spohn's *Ordinal Conditional Functions* [14]; *trust adjustment* and *trust shift* correspond to similar concepts studied in belief change [9], [16].
3) The definitions 10 and 11 can be generalized so that they can handle more precise recommendations of the form, say, *"Trust agent* a *to degree* 5 *in matters of money, and to degree* 12 *in matters of politics"*.
4) Less informative recommendations such as *"trust* a*"* can be dealt with by re-interpreting this recommendation as: $\langle a, surp(foes(a))\rangle$.
5) An $\langle a, 0\rangle$ trust-update can model *trust suspension*.
6) A recommendation may concern a group of agents ("Don't trust any secondhand car salesman"), or may be coming from a group of agents (Senator Obama receiving endorsement from the National Association of Letter Carriers). This framework can be enhanced to deal with such recommendation.
7) The current trust-value (*reputation*) of the recommender can be used to further fine-tune this framework.
8) The proposed framework can potentially be used to manage trust in communication networks.

We will take up these issues on a future occasion.

References

[1] C E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] Robert M. Axelrod. *Evolution of Cooperation*. Basic Books, 1984.

[3] C. Castelfranchi and R. Falcone. *Trust Theory: A Socio-Cognitive and Computational Model*. Wiley, 2010.

[4] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge Massachusetts, 1988.

[5] Herbert Paul Grice. Utterer's meaning and intentions. *Philosophical Review*, 78:14777, 1969.

[6] A. Herzig, E. Lorini, J F Hübner, and L. Vercouter. A logic of trust and reputation. *Logic Journal of the IGPL*, 18(1):214–244, 2010.

[7] Thomas Hobbes. *Leviathan*. The Green Dragon in St. Pauls Churchyard, 1651. (The Project Gutenberg EBook #3207, 2002).

[8] C-J Liau. Belief, information acquisition, and trust in multi-agent systems–a modal logic formulation. *Artif. Intell.*, 149(1):31–60, 2003.

[9] Abhaya C. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.

[10] Abhaya C. Nayak, Maurice Pagnucco, and Pavlos Peppas. Dynamic belief revision operators. *Artif. Intell.*, 146(2):193–228, 2003.

[11] John Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

[12] G. L. S. Shackle. The logic of surprise. *Econometrica*, new series, 149(78):112–117, 1953.

[13] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton, 1976.

[14] Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W. Harper and B. Skryms, eds, *Causation in Decision, Belief Change, and Statistics, II*, pp. 105–34. Kluwer, 1988.

[15] Colin K-Y Tan. A belief augmented frame computational trust model. In I. Russell *et. al.*, eds, *FLAIRS Conf.*, pp. 647–52. AAAI, 2005.

[16] Mary-Anne Williams. Transmutations of knowledge systems. In *Proceedings of the KR-94*, pp. 619–29. Morgan Kaufmann, 1994.

[17] Bin Yu and Munindar P. Singh. Distributed reputation management for electronic commerce. *Computational Intelligence*, 18(4):535–549, 2002.