Evidence-Based Trust

A Mathematical Model Geared for Multiagent Systems

Yonghong Wang and Munindar P. Singh
North Carolina State University

An evidence-based account of trust is essential for an appropriate treatment of application-level interactions among autonomous and adaptive parties. Key examples include social networks and service-oriented computing. Existing approaches either ignore evidence or only partially address the twin challenges of mapping evidence to trustworthiness and combining trust reports from imperfectly trusted sources. This paper develops a mathematically well-formulated approach that naturally supports discounting and combining evidence-based trust reports.

This paper understands an agent Alice's trust in an agent Bob in terms of Alice's certainty in her belief that Bob is trustworthy. Unlike previous approaches, this paper formulates certainty in terms of evidence based on a statistical measure defined over a probability distribution of the probability of positive outcomes. This definition supports important mathematical properties ensuring correct results despite conflicting evidence: (1) for a fixed amount of evidence, certainty increases as conflict in the evidence decreases and (2) for a fixed level of conflict, certainty increases as the amount of evidence increases. Moreover, despite a subtle definition of certainty, this paper (3) establishes a bijection between evidence and trust spaces, enabling robust combination of trust reports and (4) provides an efficient algorithm for computing this bijection.

Categories and Subject Descriptors: I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms: Theory, Algorithms

Additional Key Words and Phrases: Application-level trust, evidence-based trust

Contents

Authors' addresses:
Yonghong Wang, Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206, USA. yhwang_y2k@yahoo.com
Munindar P. Singh, Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206, USA. singh@ncsu.edu

1. INTRODUCTION

Trust is a broad concept with many connotations. This paper concentrates on trust as it relates to beliefs about future actions and not, for example, to emotions. The target applications for this paper involve settings wherein independent (i.e., autonomous and adaptive) parties interact with one another, and each party may choose with whom to interact based on how much trust it places in the other. Examples of such applications are social networks, webs of information sources, and online marketplaces. We can cast each party as providing and seeking services, and the problem as one of service selection in a distributed environment.

1.1 What is Trust?

A key intuition about trust as it is applied in the above kinds of settings is that reflects the trusting party's belief that the trusted party will support its plans [Castelfranchi and Falcone 1998]. For example, if Alice trusts Bob to get her to the airport, then this means that Alice is putting part of her plans in Bob's hands. In other words, Alice believes that there will be a good outcome from Bob providing her with the specific service. In a social setting, a similar question would be whether Alice trusts Bob to give her a recommendation to a movie that she will enjoy watching or whether Alice trusts Bob to introduce her to a new friend, Charlie, with whom she will have pleasant interactions. In scientific computing on the cloud, Alice may trust a service provider such as Amazon that she will receive adequate compute resources for her analysis tool to complete on schedule.

Trust makes sense as a coherent concept for computing only to the extent that we confine ourselves to settings where it would affect the decisions made by one or more participants. Specifically, this places two constraints. One, the participants ought to have the possibility of predicting each other's future behavior. For example, if all interactions were random (in the sense of a uniform distribution), no benefit would accrue to any participant who attempts to model the trustworthiness of another. Two, if the setting ensured perfect anonymity for all concerned, trust would not be a useful concept because none of the participants would be able to apply trust.

Except in settings where we have full access to how all the participants involved are reasoning and where we can apply strict constraints on their reasoning and their capabilities, we cannot make any guarantees of success. More importantly, in complex settings, the circumstances can change drastically in unanticipated ways. When that happens, all bets are off. Even the most trustworthy and predictable party may fail—our placement of trust in such a party may not appear wise in retrospect. Taleb [2007] highlights unanticipated situations and shows the difficulties such situations have caused for humans. We do not claim that a computational approach would fare any better than humans in such situations. However, computational approaches can provide better bookkeeping than humans and thus facilitate activities in the applications of interest.

1.2 Applications: Online Markets and Social Networks

Of the many computer science applications of trust, our approach emphasizes two in particular. These applications, online markets and social networks, are among the most popular practical applications of large-scale distributed computing (involving tens of millions of users) and involve trust as a key feature.

Online markets provide a setting where people and businesses buy and sell goods and services. Companies such as eBay and Amazon host markets where buyers and sellers can register to obtain accounts. Such online markets host a facility where sellers can post their items for sale and buyers can find them. The markets provide a means to determine the price for the item—by direct announcement or via an auction. However, in general, key aspects of an item being traded are imperfectly specified, such as the condition of a used book. Thus commerce relies upon the parties trusting each other. Because an online market cannot readily ensure that buyer and seller accounts reflect real-world identities, each party needs to build up its reputation (based on which others would find it trustworthy) through interactions in the market itself. In other words, traditional ways to project trust, such as the quality of a storefront or one's attire, are not applicable. And trust is based to a large extent on the positive and negative experiences obtained by others.

To this end, marketplaces such as eBay and Amazon provide a means by which each participant in an interaction can rate the other participant. The marketplace aggregates the ratings received by each participant to compute the participant's reputation, and publishes it for others to see. The idea is that a participant's reputation would predict the behavior one would expect from it. Current approaches carry out a simplistic aggregation. As Section 4.5 shows, our proposed approach equals for exceeds current approaches in terms of predicting subsequent behavior.

Social networks provide another significant application area for trust. Existing social network approaches, such as Facebook or LinkedIn, provide a logically centralized notion of identity. Users then interact with others, potentially listing them as friends (including professional contacts here). Users may also state opinions about others. The above approaches treat friendship as a symmetric relationship. They enable users to introduce their friends as a way to expand the friends' social circles and help with tasks such as looking for a job or a contract. The idea is that trust can propagate, and can provide a valid basis for interaction between parties who were not previously directly acquainted with each other. The existing popular approaches do not compute the propagated trust explicitly, although the situation could change. Several have observed the intuitive similarity of social networks and the web, and developed trust propagation techniques (several of which we review in Section 5.1).

In terms of modeling, when we think of real-life social networks, we find it more natural to think of friendship and trust as potentially asymmetric. Alice may admire Bob but Bob may not admire Alice. This in addition maintains a stronger analogy with the web: Alice's home page may point to Bob's but not the other way around. For this reason, we think of a social network as a weighted directed graph in a natural manner. Each vertex of the graph is a person, each edge means that source is acquainted with the target, and the weight on an edge represents the level of trust placed by the source in the target. Symmetric situations can be readily captured by having two equally weighted edges, the source and target of one being the target and source of the other.

The directed graph representation is commonly used for several approaches including the Pretty Good Privacy (PGP) web of trust [Zimmermann 1995; WoT] and the FilmTrust [Kuter and Golbeck 2007] network for movie ratings. The PGP web of trust is based on the keyrings of different users—or, rather, of different identities. The idea is that each key owner user may apply his key to certify zero or more other keys. The certifying key owner's expresses his level of trust as an integer from 1 to 4. The intended use of the web of trust is to help a user Alice verify that a key she encounters is legitimate: if the key is signed by several keys that Alice trusts then it presumably is trustworthy. FilmTrust is a social network where users rate other users on the presumed quality of their movie ratings. An intended use of FilmTrust is to help a user Alice find users whose movie recommendations Alice would find trustworthy. Both these networks rely upon the propagation of trust.

Although the trust propagation is not the theme of this paper, it is a major motivation for the approach here. In intuitive terms, propagation relies upon an ability to discount and aggregate trust reports. What this paper offers are the underpinnings of approaches that propagate trust based on evidence. Hang *et al.* [2009] and Wang and Singh [2006] propose propagation operators that are based on the approach described in this paper. Importantly, Hang *et al.* evaluate these operators on existing PGP web of trust and FilmTrust datasets. As Section 4.5 shows, Hang *et al.* find that operators based on our approach yield superior predictions of propagated trust than some conventional approaches.

1.3 Modeling Trust

Let us briefly consider the pros and cons of the existing approaches in broad terms. (Section 5 discusses the relevant literature in some detail.)

Trust as an object of intellectual study has drawn attention from a variety of disciplines. Basing loosely on McCarthy's observation about three ways to approach computing [Filman 2009], we think of four main ways to approach the study of trust. The *logical* approaches develop models based on mathematical logic that describe how one party trusts another. The *cognitive* approaches develop models that seek to achieve realism in the sense of human psychology. The *socioeconomic* approaches characterize trust in terms of the personal or business relationships among the parties involved, taking inspiration from human relationships. The *statistical* approaches understand trust in terms of probabilistic and statistical measures.

Each family of approaches has advantages for different computer science applications. The logical approaches are nicely suited to the challenges of specifying policies such as for determining identity and authorization. The cognitive approaches describe the human experience and would yield natural benefits where human interfaces are involved. The socioeconomic approaches apply in settings such as marketplaces and social networks. The statistical approaches work best where the account of trust is naturally based on evidence, which can be used to assess the trust one party places in another. The approach we propose falls in the intersection of statistical and socioeconomic approaches, with an emphasis on the treatment of evidence in a way that can be discounted and aggregated as some socioeconomic approaches require. This approach relies upon logical approaches for identity and provide an input into decision-making about authorization. It should be clear that we make no claims about the purely cognitive aspects of trust.

The currently dominant computer science approaches for trust emphasize identity and generally take a qualitative stance in determining if a party is to be deemed trustworthy or not. Because of their prominence, we compare these approaches to our approach.

—*Identity*. Traditional approaches address trust primarily with respect to identity. A party attempts to establish its trustworthiness to another party by presenting a certificate. The certificate is typically obtained from a certificate authority or (as in webs of trust) from another party. The presumption is that the certificate issuer would have performed some offline verification. The best case for such an engagement is that a party truly has the identity that it professes to have.

Although establishing identity is crucial to enabling trust, identity by itself is inadequate for the problems we discuss here. In particular, identity does not yield a basis for determining if a given party will serve a desired purpose appropriately. For example, if Amazon presents a valid certificate obtained from Verisign, the most it means is that the presenter of the certificate is indeed Amazon. The certificate does not mean that Alice would have a pleasant shopping experience at Amazon. After all, Verisign's certificate is not based upon any relevant experience: simply put, the certificate does not mean that Verisign purchased goods from Amazon and had a pleasant experience. From the traditional standpoint, this example might sound outlandish, but ultimately if trust is to mean that one party can place its plans in the hands of another, the expected experience is no less relevant than the identity of the provider.

—*All or none*. Traditional approaches model trust qualitatively. This is based on an intuition of hard security. If one cannot definitely determine that a particular party has the stated identity, then that is sufficient reason not to deal with it at all.

Yet in many cases, requiring an all-or-none decision about trust can be too much to ask for, especially when we think not of identity but more broadly of whether a given party would support one's plans. When we factor in the complexity of the real world and the task to be performed, virtually no one would be able to make a hard guarantee about success. Following the above example, it would be impossible for Bob to guarantee that he will get Alice to the airport on time, recommend only the perfect movies, or introduce her to none other than her potential soul mate.

Approaches based on reputation management seek to address this challenge. They usually accommodate shades of trust numerically based on ratings acquired from users. However, these approaches are typically formulated in a heuristic, somewhat ad hoc manner. The meaning assigned to the aggregated ratings is not clear from a probabilistic (or some other mathematical) standpoint.

For the reasons adduced above, although the traditional approaches to trust are valuable, they are not adequate for dealing with the kinds of interactive applications that arise in settings such as social networks and service-oriented computing. This paper develops a mathematical approach that addresses such challenges.

1.4 Trust Management

In essence, trust management [Ioannidis and Keromytis 2005] refers to the approaches by which trust judgments are reached, including how trust information is maintained, propagated, and used. Approaches for trust management vary depending on the trust model being considered. The logic-based approaches lead to trust management approaches that in simplified terms are centered on the maintenance, propagation, and use of identity credentials expressed as values of attributes needed to make authorization decisions based on the stated policies. Other important elements of trust management involve architectural assumptions such as the existence of certificate authorities and the creation and evaluation of certificate chains. To our knowledge, trust management has not been explicitly addressed for the cognitive approaches.

The socioeconomic approaches have received a lot of interest lately. In the case of marketplaces and social networks maintained as web-sites, many such approaches postulate the existence of an authority that provides the identity for each of the participants. In some cases, an "enforcer" can eliminate participants that misbehave and can attempt to litigate against them in the real world, but the connection between a virtual identity and a real-world identity can be tenuous except in cases where a user has to provide some real-world credential such as a credit card number. Other networks, such as the Pretty Good Privacy (PGP) web of trust, postulate no central authority at all, and rely on direct relationships between pairs of participants. Most recent research in socioeconomic approaches takes a conceptually distributed stance, which is well-aligned with multiagent systems. Here the participants are modeled as peers who continually interact with and rate each other. The peers exchange their ratings of others as a way to help each other identify the best peers with whom to interact. Where the approaches differ is in how they represent trust, how they exchange trust reports, and how they aggregate trust reports. Sections 5.1 and 5.2 review the most relevant of these approaches.

1.5 Scope and Organization

This paper takes the view that a probabilistic account of trust that considers the interactions among parties is crucial for supporting the above kinds of applications.

The rest of this paper is organized as follows. Section 2 motivates an evidential treatment of trust. Section 3 proposes a new notion of certainty in evidence by which we can map evidence into trust effectively. Section 4 shows that this approach satisfies some important properties, and shows how to apply it computationally. Section 5 reviews some of the most relevant literature. Section 6 summarizes our contributions and brings forth some directions for future work.

2. MOTIVATING EVIDENCE-BASED TRUST

Subtle relationships underlie trust in social and organizational settings [Castelfranchi and Falcone 1998]. Without detracting from such principles, this paper takes a narrower view of trust. In simple terms, although our intuitions are similar to those of Castelfranchi and Falcone, we approach the topic from a detailed analysis of the probabilistic aspects of trust, whereas they approach the topic from a conceptual analysis of the broader conception of trust.

We model each party computationally as an agent. Each agent seeks to establish a belief or disbelief that another agent's behavior is good (thus abstracting out details of the agent's own plans as well as the social and organizational relationships between the two agents). The model we propose here, however, can in principle be used to capture as many dimensions of trust as needed, e.g., trust about timeliness, quality of service, and so on.

In broad terms, trust arises in two main settings studied in economics [Dellarocas 2005]. In the first, the agents adjust their behavior according to their payoffs. The corresponding approaches to trust seek to alter the payoffs by *sanctioning* bad agents so that all agents have an incentive to be good. In the second setting, which this paper considers, the agents are of (more or less) fixed types, meaning that they do not adjust whether their behavior is good or bad. The corresponding approaches to trust seek to distinguish good agents from bad agents, i.e., *signal* who the bad (or good) agents are. Of course, the payoffs of the agents would vary depending on whether other agents trust them or not. Thus, even in the second setting, agents may adjust their behavior. However, such incentive-driven adjustments would occur at a slower time scale.

The following are some examples of the signaling setting, which we study. An airline would treat all coach passengers alike. Its effectiveness in transporting passengers and caring for them in transit depends on its investments in aircraft, airport lounges, and staff training. Such investments can change the airline's trustworthiness for a passenger, but a typical passenger would do well to treat the airline's behavior as being relatively stable. In the same vein, a computational service provider's performance would depend on its investments in computing, storage, and networking infrastructure; a weather service's accuracy and timeliness on the quality of its available infrastructure (sensors, networks, and prediction tools).

Our approach doesn't inherently require that the agents' behavior be fixed. Common heuristic approaches for decaying trust values can be combined with our work. However, accommodating trust updates in a mathematically well-formulated manner is itself a challenging problem, and one we defer to future work.

The most prevalent trust models today are based on subjective ratings given by one party to another. Section 5.1 discusses a few such approaches. These ratings originate from subjective user assessments and may indicate how much one user liked another but without any corresponding precise relationship between such ratings and what is expected to transpire in a subsequent interaction.

By contrast, we understand a rational agent placing trust in another party based substantially on evidence consisting of positive and negative experiences with it. This evidence can be collected by an agent locally or via a reputation agency [Maximilien 2004] or by following a referral protocol [Sen and Sajja 2002]. In such cases, the evidence may be implicit in the trust reports obtained that somehow summarize the evidence being shared. This paper develops a mathematically well-formulated evidence-based approach for trust that supports the following two crucial requirements, which arise in multiagent systems applied in important settings such as electronic commerce or information fusion.

Dynamism. Practical agent systems face the challenge that trust evolves over time. This may happen because additional information is obtained, the parties being considered alter their behavior, or the needs of the rating party change.

Composition. It is clear that trust cannot be trivially propagated. For example, Alice may trust Bob who trusts Charlie, but Alice may not trust Charlie. However, as a practical matter, a party would not have the opportunity or be able to expend the cost, e.g., in money or time, to obtain direct experiences with every other party. This is the reason that a multiagent approach—wherein agents exchange trust reports—is plausible. Consequently, we need a way to combine trust reports that cannot themselves be perfectly trusted, possibly because of the source of such reports or the way in which such reports are obtained. And we do need to accommodate the requirement that trust is weakened when propagated through such chains.

Traditionally, mathematically well-formulated approaches to trust that satisfy the above requirements have been difficult to come by. With few exceptions, current approaches for combining trust reports tend to involve ad hoc formulas, which might be simple to implement but are difficult to understand and justify from a conceptual basis.

The common idea underlying solutions that satisfy the above requirements of dynamism and composition is the notion of *discounting*. Dynamism can be accommodated by discounting over time and composition by discounting over the space of sources (i.e., agents). Others have applied discounting before, but without adequate mathematical justification. For instance, Yu and Singh [2002] develop a heuristic discounting approach layered on their (otherwise mathematically well-formulated) Dempster-Shafer account.

Wang and Singh [2006] describe a multiagent application of the present approach. They develop an algebra for aggregating trust over graphs understood as webs of trust. Wang and Singh concentrate on their algebra and assume a separate, underlying trust model, which is a previous version of the one developed here. By contrast, the present paper is neutral about the discounting and aggregation mechanisms, and instead develops a mathematically well-formulated evidential trust model that would underlie any such agent system where trust reports are gathered from multiple sources.

Following Jøsang [2001], we understand trust in terms of the *probability of the probability* of outcomes, and adopt his idea of a trust space of triples of *belief* (in a good outcome), *disbelief* (or belief in a bad outcome), and *uncertainty*. Trust in this sense is neutral as to the outcome and is reflected in the *certainty* (i.e., one minus the uncertainty). Thus the following three situations are distinguished:

—Trust being placed in a party (i.e., regarding the party as being good): belief is high, disbelief is low, and uncertainty is low.
—Distrust being placed in a party (i.e., regarding the party as being bad): belief is low, disbelief is high, and uncertainty is low.
—Lack of trust being placed in a party (pro or con): belief is low, disbelief is low, and uncertainty is high.

However, whereas Jøsang defines certainty itself in a heuristic manner, we define certainty based on a well-known statistical measure over a probability distribution. Despite the increased subtlety of our definition, it preserves a bijection between trust and evidence spaces, enabling the combination of trust reports (via mapping them to evidence). Our definition captures the following key intuitions.

—*Effect of evidence*. Certainty *increases* as evidence increases (for a fixed ratio of positive and negative observations).
—*Effect of conflict*. Certainty *decreases* as the extent of conflict increases in the evidence.

Jøsang's approach satisfies the intuition about the effect of evidence but fails the intuition about the effect of conflict. It falsely predicts that mounting *conflicting* evidence increases certainty—and equally as much as mounting confirmatory evidence. Say Alice deals with Bob four times: in either case, her evidence would be between zero and four positive experiences. It should be uncontroversial that whereas Alice's certainty is greatest when the evidence is all in favor or all against, her certainty is least when the evidence is equally split. Section 4.2 shows that Jøsang, in contrast to our approach, assigns the same certainty in each case.

Yu and Singh [2002] model positive, negative, or neutral evidence, and apply Dempster-Shafer theory to compute trust. Neutral experiences yield uncertainty, but conflicting positive or negative evidence does not increase uncertainty. Further, for conflicting evidence, Dempster-Shafer theory can yield unintuitive results. The following is a well-known example about the Dempster-Shafer theory, and is not specific to Yu and Singh's use of it [Sentz and Ferson 2002; Zadeh 1979]. Say Pete sees two physicians, Dawn and Ed, for a headache. Dawn says Pete has meningitis, a brain tumor, or neither with probabilities 0.79, 0.20, and 0.01, respectively. Ed says Pete has a concussion, a brain tumor, or neither with probabilities 0.79, 0.20, and 0.01, respectively. Dempster-Shafer theory yields that the probability of a brain tumor is 0.725, which is highly counterintuitive and wrong, because neither Dawn nor Ed thought that a brain tumor was likely. Section 4.3 shows that our model of trust yields an intuitive result in this case: the probability of a brain tumor is 0.21, which is close to each individual physician's beliefs.

This paper makes the following contributions.

—A rigorous, probabilistic definition of certainty that satisfies the above key intuitions, especially with regard to accommodating conflicting information.
—The establishment of a bijection between trust reports and evidence, which enables the mathematically well-formulated combination of trust reports that supports discounting as motivated above.
—An efficient algorithm for computing the above-mentioned bijection.

It is worth briefly clarifying the scope of this paper. This paper deals with a numeric representation of trust that captures beliefs regarding the success of a prospective interaction between a trusting and a trusted party. This paper takes a rigorous probabilistic stance on trust. The novelty of this paper lies in the introduction of a measure of certainty, which naturally accommodates conflict in evidence. Thus the approach of this paper is suitable in a wide range of settings where autonomous parties interact. In particular, this approach applies where the parties share information about each other. However, the main focus of this paper is not the propagation of trust as an end in itself, but representing and reasoning about evidence-based trust.

3. MODELING CERTAINTY

The proposed approach is based on the fundamental intuition that an agent can model the behavior of another agent in probabilistic terms. Specifically, an agent can represent the probability of a positive experience with, i.e., good behavior by, another agent. This probability must lie in the real interval $[0, 1]$. The agent's trust corresponds to how strongly the agent believes that this probability is a specific value (whether large or small, we do not care). This strength of belief is also captured in probabilistic terms. To this end, we formulate a probability density function of the probability of a positive experience. Following [Jøsang 1998], we term this a *probability-certainty density function (PCDF)*. Crucially, in our approach, unlike in Jøsang's, certainty is a statistical measure defined on a PCDF, and thus naturally accommodates both the amount of evidence and the extent of the conflict in the evidence.

3.1 Certainty from a PCDF

Because the cumulative probability of a probability lying within $[0, 1]$ must equal 1, all PCDFs must have the mean density of 1 over $[0, 1]$, and 0 elsewhere. Lacking additional knowledge, a PCDF would be a uniform distribution over $[0, 1]$. However, with additional knowledge, the PCDF would deviate from the uniform distribution. For example, knowing that the probability of good behavior is at least 0.5, we would obtain a distribution that is 0 over $[0, 0.5)$ and 2 over $[0.5, 1]$. Similarly, knowing that the probability of good behavior lies in $[0.5, 0.6]$, we would obtain a distribution that is 0 over $[0, 0.5)$ and $(0.6, 1]$, and 10 over $[0.5, 0.6]$. Notice that although a cumulative probability must equal 1, a probability density can be any nonnegative real number: densities are constrained only to ensure that cumulative probabilities equal 1.

In formal terms, let $p \in [0, 1]$ represent the probability of a positive outcome. Let the distribution of p be given as a function $f : [0, 1] \mapsto [0, \infty)$ such that $\int_0^1 f(p)dp = 1$. The probability that the probability of a positive outcome lies in $[p_1, p_2]$ can be calculated by $\int_{p_1}^{p_2} f(p)dp$. The mean value of f is $\frac{\int_0^1 f(p)dp}{1-0} = 1$. As explained above, when we know nothing else, f is a uniform distribution over probabilities p. That is, $f(p) = 1$ for $p \in [0, 1]$ and 0 elsewhere. This reflects the Bayesian intuition of assuming an equiprobable prior. The uniform distribution has a certainty of 0. As additional knowledge is acquired, the probability mass shifts so that $f(p)$ is above 1 for some values of p and below 1 for other values of p.

Our key intuition is that the agent's trust corresponds to increasing deviation from the uniform distribution. Two of the most established measures for deviation are standard deviation and mean absolute deviation (MAD) [Weisstein 2003]. MAD is more robust, because it does not involve squaring (which can increase standard deviation because of outliers or "heavy tail" distributions such as the Cauchy distribution). Absolute values can sometimes complicate the mathematics. But, in the present setting, MAD turns out to yield straightforward mathematics. In a discrete setting involving data points $x_1 \ldots x_n$ with mean $\hat{x}$, MAD is given by $\frac{1}{n}\Sigma_{i=1}^{n}|x_i - \hat{x}|$. In the present case, instead of summation we have an integral, so instead of dividing by n we divide by the size of the domain, i.e., 1. Because a PCDF has a mean value of 1, increase in some parts above 1 must yield a matching reduction below 1 elsewhere. Both increase and reduction from 1 are counted by $|f(p) - 1|$. Definition 1 scales the MAD for f by $\frac{1}{2}$ to remove this double counting; it also conveniently places certainty in the interval $[0, 1]$.

DEFINITION 1. *The certainty based on f, c_f, is given by $c_f = \frac{1}{2}\int_0^1 |f(p) - 1| dp$*

In informal terms, certainty captures the fraction of the knowledge that we do have. (Section 5.3 compares this approach to information theory.) For motivation, consider randomly picking a ball from a bin that contains N balls colored white or black. Suppose p is the probability that the ball randomly picked is white. If we have no knowledge about how many white balls there are in the bin, we cannot estimate p with any confidence. That is, certainty $c = 0$. If we know that exactly m balls are white, then we have perfect knowledge about the distribution. We can estimate $p = \frac{m}{N}$ with $c = 1$. However, if all we know is that at least m balls are white and at least n balls are black (thus $m + n \leq N$), then we have partial knowledge. Here $c = \frac{m+n}{N}$. The probability of drawing a white ball ranges from $\frac{m}{N}$ to $1 - \frac{n}{N}$. We have

$$f(p) = \begin{cases} 0, & [0, \frac{m}{N}) \\ \frac{N}{N-m-n} & p \in [\frac{m}{N}, 1 - \frac{n}{N}] \\ 0 & (1 - \frac{n}{N}, 1]. \end{cases}$$

Using Definition 1, we can confirm that certainty based on the function f as defined above, $c_f = \frac{m+n}{N}$:

$$\begin{aligned} c_f &= \tfrac{1}{2}\int_0^1 |f(p) - 1| dp \\ &= \tfrac{1}{2}(\int_0^{\frac{m}{N}} 1\ dp + \int_{\frac{m}{N}}^{1-\frac{n}{N}} (\tfrac{N}{N-m-n} - 1) dp + \int_{1-\frac{n}{N}}^1 1\ dp \\ &= \tfrac{1}{2}(\tfrac{m}{N} + \tfrac{N-m-n}{N}(\tfrac{N}{N-m-n} - 1) + \tfrac{n}{N}) \\ &= \tfrac{m+n}{N} \end{aligned}$$

3.2 Evidence Space

For simplicity, we begin by thinking of a (rating) agent's experience with a (rated) agent as a binary event: positive or negative. Evidence is conceptualized in terms of the numbers of positive and negative experiences. When an agent makes unambiguous direct observations of another, the corresponding evidence could be expressed as natural numbers (including zero). However, our motivation is to combine evidence in the context of trust. As Section 1 motivates, for reasons of dynamism or composition, the evidence may need to be discounted to reflect the weakening of the evidence source due to the effects of aging or the effects of imperfect trust having been placed in it. Intuitively, because of such discounting, the evidence is best understood as if there were real (i.e., not necessarily natural) numbers of experiences. Similarly, when a rating agent's observations are not clearcut positive or negative, we can capture the ratings via arbitrary nonnegative real numbers (as long as their sum is positive).

Accordingly, following [Jøsang 2001], we model the evidence space as $E = \mathbb{R}^+ \times \mathbb{R}^+ \setminus \{\langle 0, 0\rangle\}$, a two-dimensional space of nonnegative reals whose sum is strictly positive. (Here $\mathbb{R}^+$ is the set of nonnegative reals.) The members of E are pairs $\langle r, s\rangle$ corresponding to the numbers of positive and negative experiences, respectively.

DEFINITION 2. *Evidence space $E = \{\langle r, s\rangle | r \geq 0, s \geq 0, t = r + s > 0\}$*

Combining evidence as a result is a trivial operation: simply add up the positive and negative evidence separately.

Let x be the probability of a positive outcome. The posterior probability of evidence $\langle r, s\rangle$ is the conditional probability of x given $\langle r, s\rangle$ [Casella and Berger 1990, p. 298].

DEFINITION 3. *The conditional probability of x given $\langle r, s\rangle$ is*

$$\begin{aligned} f(x|\langle r,s\rangle) &= \frac{g(\langle r,s\rangle|x)f(x)}{\int_0^1 g(\langle r,s\rangle|x)f(x)dx} \\ &= \frac{x^r(1-x)^s}{\int_0^1 x^r(1-x)^s dx} \end{aligned}$$

where $g(\langle r,s\rangle|x) = \begin{pmatrix} r+s \\ r \end{pmatrix} x^r(1-x)^s$

Throughout this paper, r, s, and $t = r + s$ refer to positive, negative, and total evidence, respectively. The following development assumes that there is some evidence; i.e., $t > 0$.

Traditional probability theory models the event $\langle r, s\rangle$ by the pair $(p, 1-p)$, the expected probabilities of positive and negative outcomes, respectively, where $p = \frac{r+1}{r+s+2} = \frac{r+1}{t+2}$. The idea of adding 1 each to r and s (and thus 2 to $r + s$) follows Laplace's famous *rule of succession* for applying probability to inductive reasoning [Ristad 1995]. This rule in essence reflects the assumption of an equiprobable prior, which is common in Bayesian reasoning. Before any evidence, positive and negative outcomes are equally likely, and this prior biases the evidence obtained subsequently.

In practical terms, Laplace's rule of succession, alluded to above, reduces the impact of sparse evidence. It is sometimes termed *Laplace smoothing*. If you only made one observation and it was positive, you would not want to conclude that there would never be a negative observation. As the body of evidence increases, the increment of 1 has a negligible effect. More sophisticated formulations of rules of succession exist [Ristad 1995], but Laplace's rule is simple and reasonably effective for our present purposes. Laplace's rule is insensitive to the number of outcomes in that 1 is always added. The effect of this statistical "correction" (the added 1) decreases inversely as the number of outcomes being considered increases. More sophisticated approaches may be thought of as decreasing the effects of their corrections more rapidly.

Importantly, as explained above, total evidence in our approach is modeled as a nonnegative real number. Due to the effect of discounting, the total evidence can appear to be lower than 1. In such a case, the effect of any Laplace smoothing can become dominant. For this reason, this paper differs from Wang and Singh [2007] in defining a measure of the conflict in the evidence that is different from the probability to be inferred from the evidence.

3.3 Conflict in Evidence

The conflict in evidence simply refers to the relative weights of the negative and positive evidence. Conflict is highest when the negative and positive evidence are equal, and least when the evidence is unanimous one way or the other. Definition 4 characterizes the amount of *conflict* in the evidence. To this end, we define α as $\frac{r}{t}$. Clearly, $\alpha \in [0, 1]$: α being 0 or 1 indicates unanimity, whereas $\alpha = 0.5$ means $r = s$, i.e., maximal conflict in the body of evidence. Definition 4 captures this intuition.

DEFINITION 4. *conflict*$(r, s) = \min(\alpha, 1 - \alpha)$

3.4 Certainty in Evidence

In our approach, as Definition 1 shows, certainty depends on a PCDF. The particular PCDF we consider is the one of Definition 3, which generalizes over binary events. It helps in our analysis to combine these so as to define certainty based on evidence $\langle r, s\rangle$, where r and s are the positive and negative bodies of evidence, respectively. Definition 5 merely writes certainty as a function of r and s.

DEFINITION 5. $c(r, s) = \frac{1}{2}\int_0^1 |\frac{(x^r(1-x)^s}{\int_0^1 x^r(1-x)^s dx} - 1|dx$

Recall that $t = r + s$ is the total body of evidence. Thus $r = t\alpha$ and $s = t(1 - \alpha)$. We can thus write $c(r, s)$ as $c(t\alpha, t(1 - \alpha))$. When α is fixed, certainty is a function of t, and is written $c(t)$. When t is fixed, certainty is a function of α, and is written $c(\alpha)$. And, $c'(t)$ and $c'(\alpha)$ are the corresponding derivatives.

3.5 Trust Space

The traditional probability model outlined above ignores uncertainty. Thus it predicts the same probability whenever r and s have the same ratio (correcting for the effect of the Laplace smoothing) even though the total amount of evidence may differ significantly. For example, we would obtain $p = 0.70$ whether $r = 6$ and $s = 2$ or $r = 69$ and $s = 29$. However, the result would be intuitively much more certain in the second case because of the overwhelming evidence: the good outcomes hold up even after a large number of interactions. For this reason, we favor an approach that accommodates certainty.

Following [Jøsang 2001], we define a trust space as consisting of *trust reports* modeled in a three-dimensional space of reals in $[0, 1]$. Each point in this space is a triple $\langle b, d, u\rangle$, where $b+d+u = 1$, representing the weights assigned to belief, disbelief, and uncertainty, respectively. Certainty c is simply $1-u$. Thus $c = 1$ and $c = 0$ indicate perfect knowledge and ignorance, respectively. Definition 6 states this formally.

DEFINITION 6. *Trust space* $T = \{\langle b, d, u\rangle | b \geq 0, d \geq 0, b+d > 0, u > 0, b+d+u = 1\}$

Combining trust reports is nontrivial. Our proposed definition of certainty is key in accomplishing a bijection between evidence and trust reports. The problem of combining independent trust reports is reduced to the problem of combining the evidence underlying them. Section 3.6 further explains how evidence and trust space are used in this approach.

3.6 From Evidence to Trust Reports

As remarked above, it is easier to aggregate trust in the evidence space age and to discount it in trust space. As trust is propagated, each agent involved would map the evidence it obtains to trust space, discount it, map it back to evidence space, and aggregate it as evidence. We cannot accomplish the above merely by having the agents perform all their calculations in either the evidence space or the trust space. Therefore, we need a function to map evidence space to trust space. This function should be (uniquely) invertible.

Definition 7 shows how to map evidence to trust. This mapping relates positive and negative evidence to belief and disbelief, respectively, but with each having been discounted by the certainty. Definition 7 generalizes the pattern of [Jøsang 1998] by identifying the degree of conflict α and certainty $c(r, s)$. The development below describes two important differences with Jøsang's approach.

DEFINITION 7. *Let* $Z(r, s) = \langle b, d, u\rangle$ *be a transformation from* E *to* T *such that* $Z = \langle b(r, s), d(r, s), u(r, s)\rangle$*, where*

(*1*) $b(r, s) = \alpha c(r, s)$
(*2*) $d(r, s) = (1 - \alpha)c(r, s)$
(*3*) $u(r, s) = 1 - c(r, s)$

where $\alpha = \frac{r}{t}$ *and* $c(r, s)$ *is as given in Definition 5.*

One can easily verify that $c(0, 1) > 0$. In general, because $t = r + s > 0$, $c(r, s) > 0$. Moreover, $c(r, s) < 1$: thus, $1 - c(r, s) > 0$. This ensures that $b + d > 0$, and $u > 0$. Notice that $\alpha = \frac{b}{b+d}$.

Jøsang [1998] maps evidence $\langle r, s\rangle$ to a trust triple $(\frac{r}{t+1}, \frac{s}{t+1}, \frac{1}{t+1})$. Two main differences with our approach are:

—Our definition of certainty depends not only on the amount of evidence but also on the conflict, which Jøsang ignores.
—Our definition of certainty incorporates a subtle characterization of the probabilities whereas, in essence, Jøsang defines certainty as $\frac{t}{t+1}$. He offers no mathematical justification for doing so. The underlying intuition seems to be that certainty increases with increasing evidence. We finesse this intuition to capture that increasing evidence yields increasing certainty but *only if* the conflict does not increase.

Section 4.2 shows a counterintuitive consequence of Jøsang's definition.

In passing, we observe that discounting as defined by Jøsang [1998] and Wang and Singh [2006] reduces the certainty but does not affect the probability of a good outcome. Discounting in their manner involves multiplying the belief and disbelief components by the same constant, $\gamma \neq 0$. Thus a triple $\langle b, d, u\rangle$ is discounted by γ to yield $\langle b\gamma, d\gamma, 1-b\gamma-d\gamma\rangle$. Recall that the probability of a good outcome is given by $\alpha = \frac{b}{b+d}$. The probability of a good outcome from a discounted report is $\frac{b\gamma}{b\gamma+d\gamma} = \frac{b}{b+d}$, which is the same as α.

Let us consider a simple example. Suppose Alice has eight good and two bad transactions with a service provider, Charlie, yielding a trust triple of $\langle 0.42, 0.1, 0.48\rangle$. Suppose Bob has one good and four bad transactions with the Charlie, yielding a trust triple of $\langle 0.08, 0.33, 0.59\rangle$. Suppose Alice and Bob report their ratings of Charlie to Ralph. Suppose that Ralph's trust in Alice is $\langle 0.2, 0.3, 0.5\rangle$ and his trust in Bob is $\langle 0.9, 0.05, 0.05\rangle$. Ralph then carries out the following steps.

—Ralph discounts Alice's report by the trust he places in Alice (i.e., the belief component of his triple for Alice, 0.2), thus yielding $\langle 0.084, 0.02, 0.896\rangle$. Ralph discounts Bob's report in the same way by 0.9, thereby yielding $\langle 0.072, 0.297, 0.631\rangle$.
—Ralph transforms the above two discounted reports into the evidence space, thus obtaining $\langle 0.429, 0.107\rangle$ from Alice's report and $\langle 0.783, 3.13\rangle$ from Bob's report.
—Ralph combines these in evidence space, thus obtaining a total evidence of $\langle 1.212, 3.237\rangle$.

—Transforming these back to trust space, Ralph obtains that he trusts Charlie to $\langle 0.097, 0.256, 0.645 \rangle$.

Notice how, in the above, since Ralph places much greater credibility in Bob than in Alice, Ralph's overall assessment of Charlie is closer to Bob's than to Alice's.

4. IMPORTANT PROPERTIES AND COMPUTATION

We now show that the above definition yields important formal properties and how to compute with this definition.

4.1 Increasing Experiences with Fixed Conflict

Consider the scenario where the total number of experiences increases for fixed $\alpha = 0.50$. For example, compare observing 5 good episodes out of 10 with observing 50 good episodes out of 100. The expected value, α, is the same in both cases, but the certainty is clearly greater in the second. In general, we would expect certainty to increase as the amount of evidence increases. Definition 5 yields a certainty of 0.46 from $\langle r, s \rangle = \langle 5, 5 \rangle$, but a certainty of 0.70 for $\langle r, s \rangle = \langle 50, 50 \rangle$.

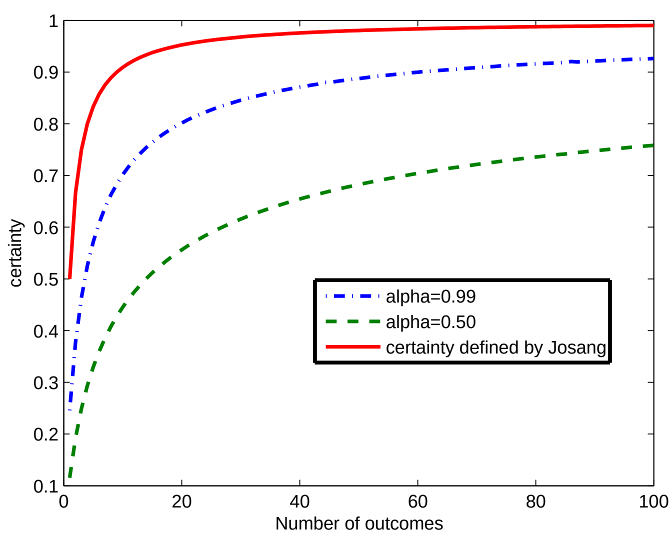

Fig. 1. Certainty increases with t both in Jøsang's approach and in our approach when the level of conflict is fixed; for our approach, we show $\alpha = 0.5$ and $\alpha = 0.99$; in Jøsang's approach, certainty is independent of the level of conflict; X-axis: t, the amount of evidence; Y-axis: $c(t)$, the corresponding certainty.

Figure 1 plots how certainty varies with t both in our approach and in Jøsang's approach. Notice that the specific numeric values of certainty in our approach should not be compared to those in Jøsang's approach. The trend is monotonic and asymptotic to 1 in both approaches. The important observation is that our approach yields a higher certainty curve when the conflict is lower.

Theorem 1 captures this property in general for our approach.

THEOREM 1. *Fix α. Then $c(t)$ increases with t for $t > 0$.*

Proof sketch: The proof of this theorem is built via a series of steps. The main idea is to show that $c'(t) > 0$ for $t > 0$. Here $f(r, s, x)$ is the function of Definition 3 viewed as a function of r, s, and x.

(1) Let $f(r,s,x) = \frac{(x^r(1-x)^s}{\int_0^1 x^r(1-x)^s dx}$. Then $c(r,s) = \frac{1}{2}\int_0^1 |f(r,s,x) - 1| dx$. We can write c and f as functions of t and α. That is, $c = c(t,\alpha)$ and $f = f(t, \alpha, x)$.

(2) Eliminate the absolute sign. By Lemma 9, we can define A and B where $f(A) = f(B) = 1$ so that $c(t,\alpha) = \frac{1}{2}\int_0^1 |f(t,\alpha,x) - 1| dx = \int_A^B (f(t,\alpha,x) - 1)dx$ A and B are also functions of t and α.

(3) When α is fixed, $c(t,\alpha)$ is a function of t and we can differentiate it by t. Notice that: $\frac{d}{dt}\int_{A(t)}^{B(t)}(f(t,x) - 1)dx = B'(t)(f(t,B) - 1) - A'(t)(f(t,A) - 1) + \int_{A(t)}^{B(t)}(\frac{\partial}{\partial t}f(t,x) - 1)dx$. The first two terms are 0 by the definition of A and B.

(4) Using the formula, $\frac{d}{dx}a^{f(x)} = \ln a f'(x) a^{f(x)}$ we can calculate $\frac{\partial}{\partial t}f(t,\alpha,x)$.

(5) Then we break the result into two parts. Prove the first part to be positive by Lemma 9, and the second part to be 0 by exploiting the symmetry of the terms.

Hence, $c'(t) > 0$, as desired. □

The appendix includes full proofs of this and the other theorems.

4.2 Increasing Conflict with Fixed Experience

Another important scenario is when the total number of experiences is fixed, but the evidence varies to reflect different levels of conflict by using different values of α. Clearly, certainty should increase as r or s dominates the other (i.e., α approaches 0 or 1) but should reduce as r and s are in balance (i.e., α approaches 0.5). Figure 2 plots certainty for fixed t and varying conflict.

Table I. Certainty computed by different approaches for varying levels of conflict.

	$\langle 0,4\rangle$	$\langle 1,3\rangle$	$\langle 2,2\rangle$	$\langle 3,1\rangle$	$\langle 4,0\rangle$
Our approach	0.54	0.35	0.29	0.35	0.54
Jøsang	0.80	0.80	0.80	0.80	0.80
Yu & Singh	1.00	1.00	1.00	1.00	1.00

More specifically, consider Alice's example from Section 1. Table I shows the effect of conflict where $t = 4$. Briefly, Yu and Singh [2002] base uncertainty not on conflict, but on intermediate (neither positive not negative) outcomes. If there is no intermediate value, the certainty is at its maximum.

Let's revisit Pete's example of Section 1. In our approach, Dawn and Ed's diagnoses would correspond to two b, d, u triples (where b means "tumor" and d means "not a tumor"): $\langle 0.20, 0.79, 0.01\rangle$ and $\langle 0.20, 0.79, 0.01\rangle$, respectively. Combining these we obtain the b, d, u triple of $\langle 0.21, 0.78, 0.01\rangle$. That is, the weight assigned to a tumor is 0.21 as

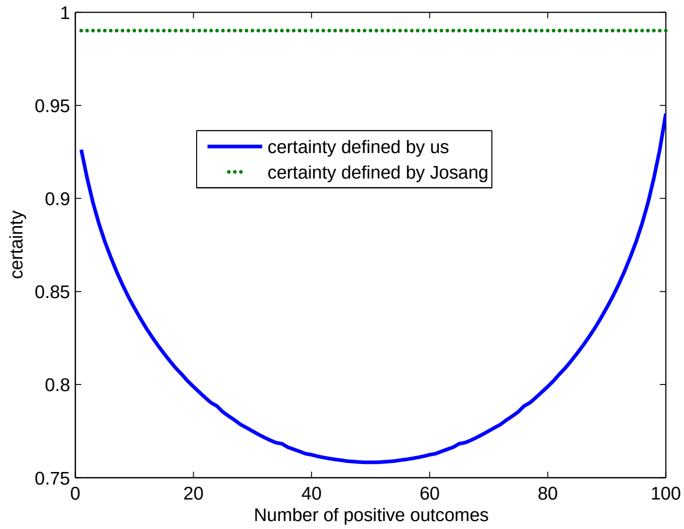

Fig. 2. Certainty is concave when t is fixed at 100; X-axis: $r + 1$; Y-axis: $c(\alpha)$; minimum occurs at $r = s = 5$; certainty according to Jøsang is constant and is shown for contrast.

opposed to 0.725 by Dempster-Shafer theory—which is unintuitive, because a tumor is Dawn and Ed's least likely prediction.

Theorem 2 captures the property that certainty increases with increasing unanimity.

THEOREM 2. *The function $c(\alpha)$ is decreasing when $0 < \alpha \leq \frac{1}{2}$, and increasing when $\frac{1}{2} \leq \alpha < 1$. Thus $c(\alpha)$ is minimized at $\alpha = \frac{1}{2}$.*

Proof sketch: The main idea is to show that $c'(\alpha) < 0$ when $\alpha \in (0, 0.5)$ and $c'(\alpha) > 0$ when $\alpha \in [0.5, 1.0)$. This is accomplished via steps similar to those in the proof of Theorem 1. First remove the absolute sign, then differentiate, then prove the derivative is negative in the interval $(0, 0.5)$ and positive in the interval $(0.5, 1)$. $\square$

Putting the above results together suggests that the relationship between certainty on the one hand and positive and negative evidence on the other hand is nontrivial. Figure 3 confirms this intuition by plotting certainty against r and s as a surface. The surface rises on the left and right corresponding to increasing unanimity of negative and positive evidence, respectively, and falls in the middle as the positive and negative evidence approach parity. The surface trends upward going from front to back corresponding to the increasing evidence at a fixed level of conflict.

It is worth emphasizing that certainty does *not* necessarily increase even as the evidence grows. When additional evidence conflicts with the previous evidence, a growth in evidence can possibly yield a loss in certainty. This accords with intuition because the arrival of conflicting evidence can shake one's beliefs, thus lowering one's certainty.

Figure 4 demonstrates a case where we first acquire negative evidence, thereby increasing certainty. Next we acquire positive evidence, which conflicts with the previous evidence, thereby lowering certainty. In Figure 4, the first ten transactions are all negative; the next ten transactions are all positive. Certainty grows monotonically with unanimous

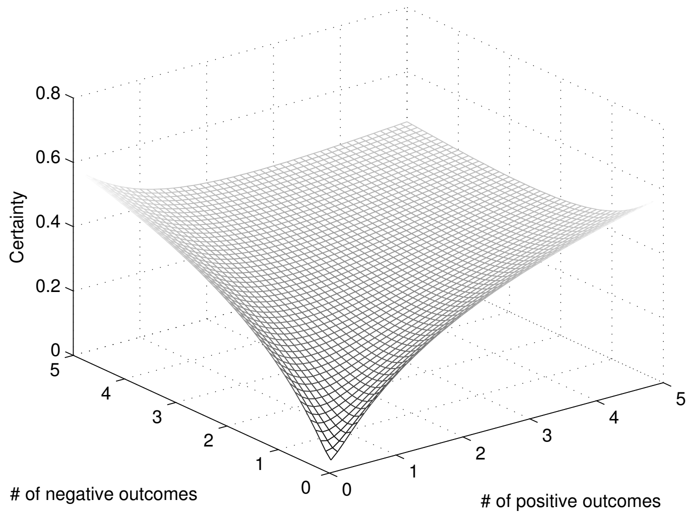

Fig. 3. X-axis: r, number of positive outcomes; Y-axis: s, number of positive outcomes; Z-axis: certainty $c(r, s)$, the corresponding certainty.

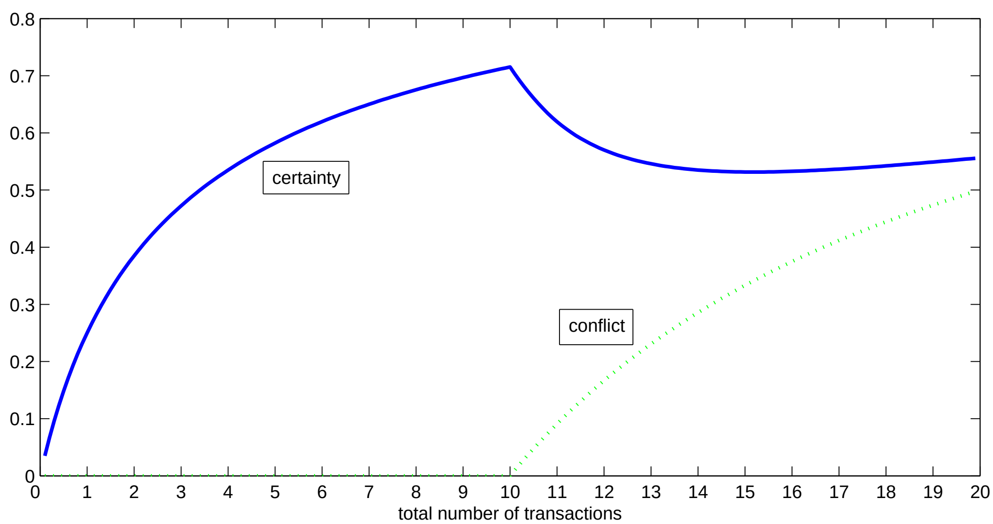

Fig. 4. Certainty increases as unanimous evidence increases; the addition of conflicting evidence lowers certainty; X-axis: number of transactions. Y-axis: c certainty.

evidence and falls as we introduce conflicting evidence. Because of the dependence of certainty on the size of the total body of evidence, it doesn't fall as sharply as it rises, and levels off as additional evidence is accrued.

4.3 Bijection Between Evidence and Trust Reports

A major motivation for modeling trust and evidence spaces is that each space facilitates computations of different kinds. Discounting trust is simple in the trust space whereas aggregating trust is simple in the evidence space.

Recall that, as Theorem 1 shows, we associate greater certainty with larger bodies of evidence (assuming conflict is fixed). Thus the certainty of trust reports to be combined clearly matters: we should place additional credence where the certainty is higher (generally meaning the underlying evidence is stronger). Consequently, we need a way to map a trust report to its corresponding evidence in such a manner that higher certainty yields a larger body of evidence.

The ability to combine trust reports effectively relies on being able to map between the evidence and the trust spaces. With such a mapping in hand, to combine two trust reports, we would simply perform the following steps:

(1) Map trust reports to evidence.
(2) Combine the evidence.
(3) Transform the combined evidence to a trust report.

The following theorem establishes that Z has a unique inverse, Z^{-1}.

THEOREM 3. *The transformation Z is a bijection.*

Proof sketch: Given $\langle b, d, u \rangle \in T$, we need $(r, s) \in E$ such that $Z(r, s) = \langle b, d, u \rangle$. As explained in Section 3.6, $\alpha = \frac{b}{b+d}$. Thus, we only need to find t such that $c(t) = 1 - u$. The existence and uniqueness of t is proved by showing that

(1) $c(t)$ is increasing when $t > 0$ (Theorem 1)
(2) $\lim_{t\to\infty} c(t) = 1$ (Lemma 11)
(3) $\lim_{t\to 0} c(t) = 0$ (Lemma 12)

Thus there is a unique t that corresponds to the desired level of certainty. $\square$

4.4 Algorithm and Complexity

Definition 5, which provides the basis for Definition 7, lies at the heart of our algorithm. We calculate the integral of Definition 5 via an application of the well-known trapezoidal rule. To further improve performance, we cache a matrix of certainty values for different values of positive and negative evidence.

Theorem 3 shows the existence of Z^{-1}. However, no closed form is known for Z^{-1}. For this reason, we develop an iterative, approximate algorithm for computing Z^{-1}.

As explained in Section 3.6, the ratio α depends solely on b and d. Thus given $\langle b, d, u \rangle$, we can determine α immediately as $\frac{b}{b+d}$. Since $r = t\alpha$ and $s = t(1-\alpha)$, in this manner, we know the relationships between r and t, and between s and t. But we do not immediately know t. In essence, no closed form for Z^{-1} is known because no closed form is known for its t component.

The intuition behind our algorithm for computing t is that after fixing α, the correct value of t is one that would yield the desired certainty of $(1 - u)$. This works because, as remarked in the proof sketch for Theorem 3, $c(t)$ ranges between 0 and 1. Further, Theorem 1 shows that for fixed α, $c(t)$ is monotonically increasing with t. In general, t being the size of the body of evidence is not bounded. However, as a practical matter, an upper bound can be placed on t. Thus, a binary search is an obvious approach. (When no bound is known, a simple approach would be to (1) guess exponentially increasing values for t until a value is found for which the desired certainty is exceeded; and then (2) conduct binary search between that and the previously guessed value.)

For binary search, since we are dealing with real numbers, it is necessary to specify $\epsilon > 0$, the desired precision to which the answer is to be computed. (In our experiments we set $\epsilon = 10^{-4}$.)

Algorithm 1 calculates Z^{-1} via binary search on $c(t)$ to a specified precision, $\epsilon > 0$. Here $t_{max} > 0$ is the maximum size of the body of evidence considered. (Recall that $\lg$ means logarithm to base 2.)

```
α = b/(b+d);
c = 1 − u;
t1 = 0;
t2 = tmax;
while t2 − t1 ≥ ε do
    t = (t1+t2)/2;
    if c(t) < c then
        t1 = t;
    else
        t2 = t;
r = tα;
s = t − r;
return r, s
```

Algorithm 1: Calculating $(r, s) = Z^{-1}(b, d, u)$.

THEOREM 4. *The complexity of Algorithm 1 is* $\Omega(-\lg \epsilon)$.

Proof: After the **while** loop iterates i times, $t_2 - t_1 = t_{max}2^{-i}$. Eventually, $t_2 - t_1$ falls below ϵ, thus terminating the loop. Assume the loop terminates in n iterations. Then, $t_2 - t_1 = t_{max}2^{-n} < \epsilon \leq t_{max}2^{-n+1}$. This implies $2^n > \frac{t_{max}}{\epsilon} \geq 2^{n-1}$. That is, $n > (\lg t_{max} - \lg \epsilon) \geq n - 1$.

4.5 Empirical Evaluation

The experimental validation of this work is made difficult by the lack of established datasets and testbeds, especially those that would support more than a scalar representation of trust. The situation is improving in this regard [Fullam et al. 2005], but current testbeds do not support exchanging trust reports of two dimensions (as in $\langle b, d, u \rangle$ because $b + d + u = 1$).

We have evaluated this approach on two datasets. The first dataset includes ratings received by five sellers (of *Slumdog Millionaire Soundtracks*) on Amazon Marketplace. Amazon supports integer ratings from 1 to 5. Amazon summarizes the information on each seller as an average score along with the total number of ratings received. However, it also makes the raw ratings available—these are what we use. We map the ratings to evidence $\langle r, s \rangle$, where $r + s = 1$. Specifically, we map 1 to $\langle 0.0, 1.0 \rangle$, 2 to $\langle 0.25, 0.75 \rangle$, and so on, increasing r in increments of 0.25 and decreasing s by the same amount to maintain $r + s = 1$.

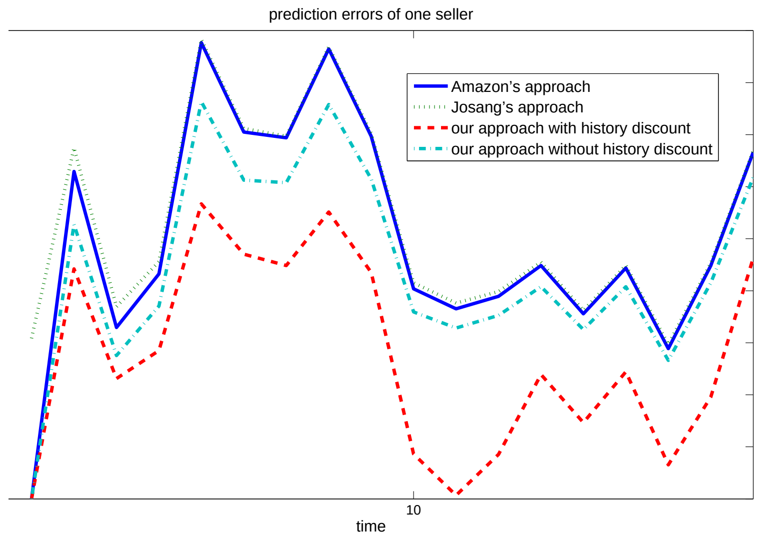

Fig. 5. Prediction errors based on ratings received by a seller on Amazon using different methods. Here, one timestep is 25 transactions, errors are the average of the 25 prediction errors, based on ratings in the range [1, 5].

For each seller, we consider the ratings that it received in the order in which it received them. The idea is that the party who carried out the $(i+1)^{\text{st}}$ transaction with the seller would generally have had access to the previous i ratings received by that seller. Therefore, for each i we map the first i ratings to a $\langle b, d, u \rangle$ triple and use this triple to predict the $(i+1)^{\text{st}}$ rating.

Figure 5 shows the prediction errors that result by applying different methods on the ratings received by one of the sellers. The Amazon approach refers to treating the average current rating as the predictor of the next rating. In the other approaches shown, the prediction is the b value computed from the ratings till the present rating. Jøsang's approach and our approach calculate b as already discussed. Our approach with discounting involves discounting the past ratings as a way to place additional weight on the more recent ratings. Specifically, we discount each rating by 10% for each time unit.

Table II. Average prediction errors for trustworthiness of five Amazon sellers based on their ratings, based on ratings in the range [1, 5].

	Seller A	Seller B	Seller C	Seller D	Seller E
Amazon's approach	0.473	0.287	0.233	0.135	0.502
Jøsang's approach	0.557	0.333	0.375	0.195	0.530
Our approach	0.388	0.244	0.186	0.122	0.445
Our approach with discounting	0.303	0.186	0.159	0.095	0.276

Figure 5 shows that our approach yields a lower prediction error than Amazon and Jøsang's approaches. Jøsang's approach is worse than Amazon's whereas ours is better. Moreover, our approach coupled with discounting yields even better results. The results for the other sellers are similar, and we omit them for brevity. Table II summarizes the results for all five sellers and shows that the same pattern holds for them.

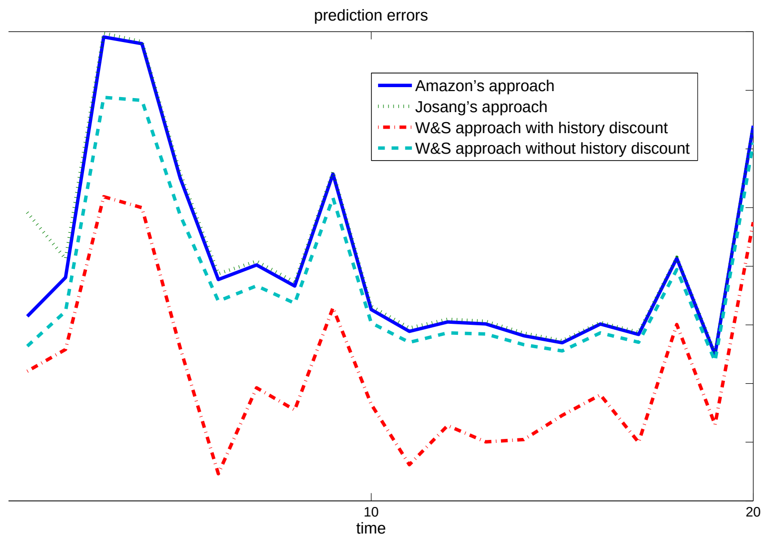

Fig. 6. Prediction errors based on ratings received by an artificial "multiple personality" seller using different methods. This seller's list of ratings is a concatenation of the ratings of the actual sellers. Here, one timestep is 50 transactions and shows the average prediction errors, based on ratings in the range [1, 5].

We next evaluate our approach with respect to its ability to track a changing behavior pattern. To develop this test-case while staying in the realm of actual data, we artificially construct a seller whose ratings are a concatenation of the ratings obtained by different sellers. In this way, this seller models a seller who changes his strategy arbitrarily. Figure 6 shows the results of applying the above approaches to this artificial seller. It finds that the same pattern of results holds as in Figure 5. In Figure 6 too, Jøsang's approach yields worse results than Amazon whereas our approach yields superior results. Further, with the benefit of history discounting, our approach does even better.

A possible way to understand these results is the following. Amazon calculates the average rating as the prediction whereas Jøsang incorporates Laplace smoothing (recall the discussion in Section 3.2). Thus Jøsang ends up with higher error in many cases. Further, Jøsang's definition of certainty ignore conflict and thus increases monotonically with evidence. Thus his predictions are the worst. Our approach takes a more nuanced approach than Amazon's but without the pitfalls of Jøsang's approach, and thus produces better results. With the benefit of discounting, it reduces the effect of old evidence, thus improving the outcome further.

The second evaluation of the proposed approach is based on its use within trust propagation operators. Recently, Hang *et al.* [2009] proposed trust propagation operators based on the approach described in this paper. They evaluated their operators using two network datasets, namely, FilmTrust (538 vertices representing users; 1,234 weighted directed edges representing ratings) [Kuter and Golbeck 2007] and the PGP web of trust (39,246 vertices representing users (or rather keys) and 317,979 weighted directed edges representing the strength of an endorsement) [WoT]. Hang *et al.* report how the operators based on our approach perform better than other approaches applied on those datasets. The definitions of operators and the various path search strategies are nuanced and cannot be readily repeated here, so we refer the reader to Hang *et al.*'s paper for additional details.

5. LITERATURE

A huge amount of research has been conducted on trust. We now review some of the most relevant literature from our perspective of an evidential approach.

5.1 Literature on Distributed Trust

In general, the works on distributed trust emphasize techniques for propagating trust. In this sense, they are not closely related to the present approach, which emphasizes evidence and only indirectly considers propagation. Many of the existing approaches rely on subjective assessments of trust. Potentially, one could develop variants of these propagation algorithms that apply on evidence-based trust reports instead of subjective assessments. However, two challenges would be (1) accommodating $\langle b, d, u \rangle$ triples instead of scalars; and (2) conceptually making sense of the propagated results in terms of evidence. Hang *et al.* [2009], discussed above, address both of these challenges.

Carbone *et al.* [2003] study trust formally in a setting based on distributed computing systems. They propose a two-dimensional representation of trust consisting of (1) trustworthiness and (2) certainty placed in the trustworthiness. Carbone *et al.*'s notion of trustworthiness is abstract and they do not discuss how the trust originates. In particular, they do not relate trustworthiness with evidence. Carbone *et al.* partially order trustworthiness and certainty. The level of trustworthiness for them is the extent to which, e.g., the amount of money loaned, an agent will fully trust another. There is no probabilistic interpretation, and they do not specify how to calculate certainty or any properties of it.

Weeks [2001] introduces a mathematical framework for distributed trust management systems. He uses the least fixed point in a lattice to define the semantics of trust computations. Importantly, Weeks only deals with the so-called hard trust among agents that underlies traditional authorization and access control approaches. He doesn't deal with evidential trust, as studied in this paper.

Several approaches understand trust in terms of aggregate properties of graphs, such as can be described via matrix operations [Guha et al. 2004; Kamvar et al. 2003; Richardson et al. 2003]. The propagation of trust corresponds to matrix multiplication. Such aggregate methods can be attractive and have a history of success when applied to link analysis of web pages. Such link analysis is inspired by the random browser model. However, it is not immediately apparent why trust should map to the random browser model, or whether it is even fair to expect trust ratings to be public the way links on web pages are. A further unintuitive consequence is that to ensure convergence these approaches split trustworthiness. For example, if Alice trusts two people but Bob only trusts only one person, Alice's trustworthiness is split between two people but Bob's trustworthiness propagates fully to his sole contact. There is no conceptual reason for this discrepancy.

Ziegler and Lausen [2004] develop an approach based on spreading activation for propagating trust. They too model trust itself as an arbitrary subjective rating: based on the agents' opinions, not on evidence. Ziegler and Lausen model trust as energy to be propagated through spreading activation, but do not justify their energy interpretation adequately on any mathematical or conceptual grounds. It simply seems like an approach that they think might work. Their notion of trust is global in that each party ends up with an energy level that describes its trustworthiness. Thus the relational aspect of trust is lost. The above remark about splitting trustworthiness among multiple parties applies to energy based models as well.

Quercia *et al.* [2007] design a system to propagate trust among mobile users. They relate nodes corresponding to users based on the similarity of their ratings. Quercia *et al.* apply a graph-based learning technique by which a node may compute its rating of another node. Thus their method is similar to collaborative filtering applied by each node. A fundamental difference with our approach is that Quercia *et al.*'s model is based on subjective ratings, not on evidence. Thus it makes no attempt to relate the ratings to expected behavior. However, our approach could potentially be combined with the prediction part of Quercia *et al.*'s model.

Schweitzer *et al.* [2006] propose an approach for propagating trust in ad hoc networks that builds on the Jøsang's representation. So it could potentially benefit from the present approach. Schweitzer *et al.* do not explicitly accommodate the certainty of the reports. Their approach is heuristic and in this sense differs from Hang *et al.* [2009]. However, Schweitzer *et al.* include a nice feature where a participant can warn those to whom it previously sent a referral if it finds that it no longer trusts the party it previously recommended.

Kuter and Golbeck [2007] propose a trust propagation algorithm that computes all paths from a source to a sink vertex in a graph, and combines trust ratings from those paths along with a confidence measure. The underlying data in their approach are subjective ratings given by one user to another. In this way, this work fits into the above family of trust propagation research and not into the evidential approaches, which this paper emphasizes.

5.2 Literature on Trust and Evidence

Abdul-Rahman and Hailes [2000] present an early model for computing trust. However, their approach is highly ad hoc and limited. Specifically, various weights are simply added up without any mathematical justification. Likewise, the term *uncertainty* is described but without being given any mathematical foundation.

Sierra and Debenham [2007] define an agent strategy by combining the three dimensions of utility, information, and semantic views. Their framework defines trust, reputation, and uncertainty. Their definition is rather complex. It is justified based on the authors' intuitions and is experimentally evaluated. Thus it is plausible in a conceptual way. However, it lacks an explicit mathematical justification, such as we have sought to develop in this work.

The Regret system combines several aspects of trust, notably the social aspects [Sabater and Sierra 2002]. It involves a number of formulas, which are given intuitive, but not mathematical, justification. A lot of other work, e.g., [Huynh et al. 2006], involves heuristics that combine multiple information sources to judge trust. It would be an interesting direction to combine a rigorous approach such as ours with the above heuristic approaches to capture a rich variety of practical criteria well.

Teacy *et al.* [2005] propose TRAVOS, the Trust and Reputation model for Agent-based Virtual OrganisationS. TRAVOS uses a probabilistic treatment of trust. Teacy *et al.* model trust in terms of confidence that the expected value lies within a specified error tolerance. An agent's confidence increases with the error tolerance. Teacy *et al.* study combinations of probability distributions to which the evaluations given by different agents might correspond. They do not formally study certainty. Further, Teacy *et al.*'s approach does not yield a probabilistically valid method for combining trust reports, which is the focus of this paper.

Despotovic and Aberer [2005] propose a simple probabilistic method, maximum likelihood estimation, to aggregate ratings. This method dramatically reduces the calculation overhead of propagating and aggregating trust information. Further, the aggregated trust admits a clear statistical interpretation. However, Despotovic and Aberer's model is overly simplified: the agents rate a target as either good or bad. Thus their approach cannot be used where the agents are required to give more accurate ratings, e.g., even a scalar (as a real value) from 0 to 1. Further, Despotovic and Aberer's method does not consider the uncertainty of a rating or equivalently the number of transactions on which a rating might be based. Since witnesses can be any agents, in order to estimate the maximum likelihood, each agent needs to record the trustworthiness of all possible witnesses, thus increasing the complexity of scaling up. More importantly, since only a small number of witnesses are chosen and each agent only knows a small number of all agents, most of the time, the agent cannot compute how much trust to place in the necessary witnesses.

Reece *et al.* [2007] develop a method to consolidate an agent's direct experience with a service provider and trust reports about that service provider received from other agents. Reece *et al.* calculate a covariance matrix based on the Dirichlet distribution that describes the uncertainty and correlations between different dimensional probabilities. The covariance matrix can be used to communicate and fuse ratings. The Dirichlet distribution considers only the ratio of positive and negative transactions. It does not depend on the number of transactions, so the resulting uncertainty or certainty estimates are independent of the total number of transactions. As we explained above, this is contrary to intuition because certainty does increase with mounting evidence if the ratio of positive and negative transactions is kept constant. Lastly, Reece *et al.* neglect the trustworthiness of the agent who provides the information. The present paper provides a basis for accommodating the trustworthiness of agents who provide information: this aspect is studied by Wang and Singh [2006], which applies the present approach to specify operators for propagating trust.

Halpern and Pucella [2006] consider evidence as an operator that maps prior beliefs to posterior beliefs. Similar to our certainty function, they use a confirmation function to measure the strength of the evidence. However, there are many kinds of confirmation functions available, and it is not clear which one to use. Halpern and Pucella use the log-likelihood ratio. They do not give a mathematical justification for its optimality, only that it avoids requiring a prior distribution on hypotheses.

Fullam and Barber [2006] describe trust-related decisions based on agent role (trustee or truster) and transactions (fundamental transaction or reputation transaction). They propose applying Q-learning and explain why the learning is complicated when reputation transaction is used. Fullam and Barber use the Agent Reputation and Trust (ART) test-bed to evaluate their learning techniques. Fullam and Barber [2007] study different sources of trust information: direct experience, referrals from peers, and reports from third parties. They propose a dynamical learning technique to identify the best sources of trust, based on some parameters. Both the above works do not consider the uncertainty of trust information and do not offer any mathematical justification for their approach.

The following work is not directly related to our work but is worth discussing because it deals with service discovery based on uncertainty. Li *et al.* [2008] describe ROSSE, a search engine for grid service discovery. They introduce the notion of "property uncertainty" when matching services. A property is uncertain when it is used in some but not all advertisements for services in the same category. Thus, for Li *et al.*, uncertainty means how unlikely a service has the property in question. This is quite different from our meaning based on the probability of the probability of a particular outcome. Li *et al.* use rough set theory to deal with property uncertainty and select the best matched services.

Other approaches study systems in which agents alter their relationships or their behaviors based on calculations of each other's trustworthiness. Jurca and Faltings [2007] describe a mechanism that uses side-payment schemes to provide incentives for agents so that it becomes rational for the agents to report ratings of other agents truthfully. Jurca and Faltings use a simplistic trust model. They express trust as a scalar from 0 to 1, and do not consider uncertainty. As a result, one bad experience out of ten yields the same level of trust as would 1,000 bad experiences out of 10,000. By contrast, our approach finds the two cases to yield different measures of certainty. Overall, though, our approach is complementary to theirs. Jurca and Faltings are interested in obtaining individual ratings; we are interested in aggregating the ratings into measures of belief of certainty, which can then be propagated [Wang and Singh 2006].

Sen and Sajja [2002] also address the problem of deceptive agents. They study reputation-based trust with an emphasis on the problem of estimating the number of raters (some of whom may be liars) to query in order to obtain a desired likelihood threshold about the quality of a service provider. They model the agents' performance as drawn from two distributions (high and low); agents use reputation to determine if some of the raters are liars. Sen and Sajja experimentally study the effect on performance of systems wherein some of the raters are liars with a view to identifying thresholds beyond which the number of liars can disrupt a system. Yu and Singh [2003] show how agents may adaptively detect deceptive agents. Yolum and Singh [2003] study the emergent graph-theoretic properties of referral systems. This paper complements such works because it provides an analytical treatment of trust that they do not have whereas they address system concerns that this paper does not study.

5.3 Literature on Information Theory

Shannon entropy [1948] is the best known information-theoretic measure of uncertainty. It is based on a discrete probability distribution $p = \langle p(x) | x \in X \rangle$ over a finite set X of alternatives (elementary events). Shannon's formula encodes the number of bits required to obtain certainty: $S(p) = -\sum_{x \in X} p(x) \lg p(x)$. Here $S(p)$ can be viewed as the weighted average of the conflict among the evidential claims expressed by p. Jaynes [2003] provides examples, intuitions, and precise mathematical treatment of the entropy principle. More complex, but less well-established, definitions of entropy have been proposed for continuous distributions as well, e.g., [Smith 2001].

Entropy, however, is not suitable for the present purposes of modeling evidential trust. Entropy captures (bits of) missing information and ranges from 0 to ∞. At one level, this disagrees with our intuition that, for the purposes of trust, we need to model the confidence placed in a probability estimation. Moreover, the above definitions cannot be used in measuring the uncertainty of the probability estimation based on past positive and negative experiences.

6. DISCUSSION

This paper offers a theoretical development of trust that would underlie a variety of situations where trust reports based on evidence are combined. In particular, it contributes to a mathematical understanding of trust, especially as it underlies a variety of multiagent applications. These include social networks understood via referral systems and webs of trust, in studying which we identified the need for this research. Such applications require a natural treatment of composition and discounting in an evidence-based framework. As Section 1 shows, these applications broaden to service-oriented computing in general.

Further, an evidence-based notion of trust must support important properties regarding the effects of increasing evidence (for constant conflict) and of increasing conflict (for constant evidence). Theoretical validation, as provided here, is valuable for a general-purpose conceptually driven mathematical approach such as ours. The main technical insight of this paper is how to manage the duality between trust and evidence spaces in a manner that provides a rigorous basis for combining trust reports.

A payoff of this approach is that an agent who wishes to achieve a specific level of certainty can compute how much evidence would be needed at different levels of conflict. Or, the agent can iteratively compute certainty to see if the certainty of its beliefs or disbeliefs has reached an acceptably high level.

It is worth considering briefly the benefits of treating trust as a well-defined concept. Potentially, instead of exchanging trust reports, the agents could exchange probability distributions based upon the evidence. However, discounting such evidence would require going through the trust report representation that we described. Because of the bijection that Theorem 3 establishes, the trust and evidence representations are equivalent, so the choice between them is arbitrary. However, trust is an important concept for both conceptual and practical reasons. In conceptual terms, trust represents a form of *relational* capital [Castelfranchi et al. 2006] among the agents. From a practical standpoint, trust summarizes the prospects for an interaction in a way that makes intuitive sense to people and fits into practical agent architectures. Making intuitive sense to people is crucial from the standpoint of effective requirements elicitation and for explaining outcomes to users, which are crucial for improving credibility and predictability. Patrick *et al.* [2005] and Yan *et al.* [2008] discuss additional ramifications of trust and usability on each other. Moreover, in open architectures where the agents are implemented heterogeneously, a numeric treatment of trust can provide a simple means of facilitating interoperation without requiring that the implementations agree on their internal representations.

6.1 Conclusions

The broad family of applications that this approach targets includes social networks and service-oriented computing. These applications rely on the parties concerned acquiring evidence in order to make reasoned judgments about interacting with others. As a practical matter, it is inevitable that there will be conflicts in the trust reports received from others. There is agreement that certainty is crucial in combining trust reports. However, previous approaches calculate certainty naïvely in a manner that disregards conflict. Thus our results are a significant advance even though our approach begins from the same PCDF framework as applied by Jøsang in his treatment of trust.

We now summarize our technical contributions.

—This paper offers a theoretical development of trust based on certainty that would underlie a variety of situations where trust reports based on evidence are combined. Specifically, in this approach, for a fixed amount of evidence, certainty increases as conflict in the evidence decreases. And, for a fixed level of conflict, certainty increases as the amount of evidence increases.

—Moreover, despite a more subtle definition of certainty than in the literature, this paper establishes a bijection between evidence and trust spaces, enabling robust combination of trust reports. Further, it provides an efficient algorithm for computing this bijection.

6.2 Directions

This work has opened up some important directions for future work. First, the above work treats all past transactions equally and simply adds up all positive transactions and negative transactions. We might give more weight to most recent transactions, i.e., discount the evidence by its age. The foregoing showed how trust evolves with respect to increasing evidence under different conditions. The same properties apply to the evolution of trust over time, that is, as time passes and more evidence is obtained. A crucial observation is that because of the bijection established in Theorem 3, the historical evidence at any point can be summarized in a belief-disbelief-uncertainty triple. New evidence can then be added as explained above. Moreover, we can discount the value of evidence over time if necessary. For example, we may discount the evidence at every time step (chosen based on the domain: e.g., one hour or one day, or after each transaction). As a result of such discounting, new evidence would have a greater impact than older evidence.

Second, our work assumes a uniform prior probability distribution. Other prior probability distributions such as the Gaussian distribution may be useful in different settings. We conjecture that certainty defined on other probability distributions would support the mathematical properties (monotonicity with increasing evidence for fixed conflict and for decreasing conflict for fixed evidence) just as well as the certainty formalized here.

Third, an important technical challenge is to extend the above work from binary to multi-valued events. Such an extension would enable us to handle a larger variety of interactions among people and services.

Fourth, we can imagine new models that encompass all the challenging aspects of the beta model, which can analyze the model and provide with algorithms for computing the various probabilities in this model.

Acknowledgments

This is a revised and extended version of [Wang and Singh 2007]. We thank Dennis Bahler, Greg Byrd, Ting Yu, Chris Hazard, and Chung-Wei Hang for useful discussions and the anonymous reviewers for helpful comments. Chris provided the Amazon dataset that we used for our evaluation. This research was partially supported by the National Science Foundation under grant ITR-0081742.

REFERENCES

ABDUL-RAHMAN, A. AND HAILES, S. 2000. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on Systems Science*. IEEE Computer Society Press, Los Alamitos.

CARBONE, M., NIELSEN, M., AND SASSONE, V. 2003. Formal model for trust in dynamic networks. In *Proceedings of 1st International Conference on Software Engineering and Formal Methods (SEFM)*. IEEE Computer Society, Los Alamitos, 54–63.

CASELLA, G. AND BERGER, R. L. 1990. *Statistical Inference*. Duxbury Press, Pacific Grove, CA.

CASTELFRANCHI, C. AND FALCONE, R. 1998. Principles of trust for MAS: cognitive anatomy, social importance, and quantification. In *Proceedings of the 3rd International Conference on Multiagent Systems*. IEEE Computer Society Press, Los Alamitos, 72–79.

CASTELFRANCHI, C., FALCONE, R., AND MARZO, F. 2006. Being trusted in a social network: Trust as relational capital. In *Trust Management: Proceedings of the iTrust Workshop*. LNCS, vol. 3986. Springer, Berlin, 19–32.

CRANOR, L. F. AND GARFINKEL, S. 2005. *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly, Sebastopol, CA.

DELLAROCAS, C. 2005. Online reputation mechanisms. In *Practical Handbook of Internet Computing*, M. P. Singh, Ed. Chapman Hall & CRC Press, Baton Rouge, Chapter 20.

DESPOTOVIC, Z. AND ABERER, K. 2005. Probabilistic prediction of peers' performances in P2P networks. *International Journal of Engineering Applications of Artificial Intelligence 18,* 7, 771–780.

FILMAN, R. 2009. Personal communication about Filman's doctoral adviser McCarthy's views on approaches to artificial intelligence.

FULLAM, K. AND BARBER:, K. S. 2006. Learning trust strategies in reputation exchange networks. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. ACM Press, New York, 1241–1248.

FULLAM, K. AND BARBER, K. S. 2007. Dynamically learning sources of trust information: Experience vs. reputation. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. IFAAMAS, Columbia, SC, 1062–1069.

FULLAM, K., KLOS, T. B., MULLER, G., SABATER, J., SCHLOSSER, A., TOPOL, Z., BARBER, K. S., ROSENSCHEIN, J. S., VERCOUTER, L., AND VOSS, M. 2005. A specification of the agent reputation and trust (ART) testbed: Experimentation and competition for trust in agent societies. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. ACM Press, New York, 512–518.

GUHA, R., KUMAR, R., RAGHAVAN, P., AND TOMKINS, A. 2004. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*. ACM Press, New York, 403–412.

HALPERN, J. Y. AND PUCELLA, R. 2006. A logic for reasoning about evidence. *Journal of AI Research 26*, 1–34.

HANG, C.-W., WANG, Y., AND SINGH, M. P. 2009. Operators for propagating trust and their evaluation in social networks. In *Proceedings of the 8th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. IFAAMAS, Columbia, SC.

HUYNH, T. D., JENNINGS, N. R., AND SHADBOLT, N. R. 2006. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and MultiAgent Systems 13,* 2 (Sept.), 119–154.

IOANNIDIS, J. AND KEROMYTIS, A. D. 2005. Distributed trust. In *Practical Handbook of Internet Computing*, M. P. Singh, Ed. Chapman Hall & CRC Press, Baton Rouge, Chapter 20.

JAYNES, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK.

JØSANG, A. 1998. A subjective metric of authentication. In *Proceedings of the 5th European Symposium on Research in Computer Security (ESORICS)*. LNCS, vol. 1485. Springer-Verlag, Heidelberg, 329–344.

JØSANG, A. 2001. A logic for uncertain probabilities. *Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9*, 279–311.

JURCA, R. AND FALTINGS, B. 2007. Obtaining reliable feedback for sanctioning reputation mechanisms. *Journal of Artificial Intelligence Research (JAIR) 29*, 391–419.

KAMVAR, S. D., SCHLOSSER, M. T., AND GARCIA-MOLINA, H. 2003. The EigenTrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th International World Wide Web Conference*. ACM Press, New York, 640–651.

KUTER, U. AND GOLBECK, J. 2007. SUNNY: A new algorithm for trust inference in social networks using probabilistic confidence models. In *Proceedings of the 22st National Conference on Artificial Intelligence (AAAI)*. AAAI Press, Menlo Park, 1377–1382.

LI, M., YU, B., RANA, O., AND WANG, Z. 2008. Grid service discovery with rough sets. *IEEE Transactions on Knowledge and Data Engineering 20,* 6, 851–862.

MAXIMILIEN, E. M. 2004. Toward autonomic web services trust and selection. Ph.D. thesis, Department of Computer Science, North Carolina State University.

PATRICK, A. S., BRIGGS, P., AND MARSH, S. 2005. Designing systems that people will trust. In *[Cranor and Garfinkel 2005]*. Chapter 5, 75–100.

QUERCIA, D., HAILES, S., AND CAPRA, L. 2007. Lightweight distributed trust propagation. In *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM)*. IEEE Computer Society, Los Alamitos, 282–291.

REECE, S., ROGERS, A., ROBERTS, S., AND JENNINGS, N. R. 2007. Rumours and reputation: Evaluating multi-dimensional trust within a decentralised reputation system. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. IFAAMAS, Columbia, SC, 1063–1070.

RICHARDSON, M., AGRAWAL, R., AND DOMINGOS, P. 2003. Trust management for the semantic Web. In *The Semantic Web: Proceedings of the 2nd International Semantic Web Conference (ISWC)*. LNCS, vol. 2870. Springer-Verlag, 351–368.

RISTAD, E. S. 1995. A natural law of succession. TR 495-95, Department of Computer Science, Princeton University. July.

SABATER, J. AND SIERRA, C. 2002. Reputation and social network analysis in multi-agent systems. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. ACM Press, New York, 475–482.

SCHWEITZER, C. M., CARVALHO, T. C. M. B., AND RUGGIERO, W. V. 2006. A distributed mechanism for trust propagation and consolidation in ad hoc networks. In *Proceedings of the International Conference on Information Networking, Advances in Data Communications and Wireless Networks (ICOIN)*. LNCS, vol. 3961. Springer, Berlin, 156–165.

SEN, S. AND SAJJA, N. 2002. Robustness of reputation-based trust: Boolean case. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. ACM Press, New York, 288–293.

SENTZ, K. AND FERSON, S. 2002. Combination of evidence in Dempster Shafer theory. TR 0835, Sandia National Laboratories, Albuquerque, New Mexico.

SHANNON, C. E. 1948. The mathematical theory of communication. *Bell System Technical Journal 27,* 3, 379–423.

SIERRA, C. AND DEBENHAM, J. K. 2007. Information-based agency. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, Detroit, 1513–1518.

SMITH, J. D. H. 2001. Some observations on the concepts of information-theoretic entropy and randomness. *Entropy 3*, 1–11.

TALEB, N. N. 2007. *The Black Swan: The Impact of the Highly Probable*. Random House, New York.

TEACY, L., PATEL, J., JENNINGS, N., AND LUCK, M. 2005. Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. ACM Press, New York, 997–1004.

WANG, Y. AND SINGH, M. P. 2006. Trust representation and aggregation in a distributed agent system. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*. AAAI Press, Menlo Park, 1425–1430.

WANG, Y. AND SINGH, M. P. 2007. Formal trust model for multiagent systems. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, Detroit, 1551–1556.

WEEKS, S. 2001. Understanding trust management systems. In *Proceedings of the IEEE Symposium on Security and Privacy*. IEEE Computer Society, Los Alamitos, 94–105.

WEISSTEIN, E. W. 2003. Mean deviation. http://mathworld.wolfram.com/MeanDeviation.html.

WoT. Web of trust. http://www.lysator.liu.se/ jc/wotsap/wots2/.

YAN, Z., NIEMI, V., DONG, Y., AND YU, G. 2008. A user behavior based trust model for mobile applications. In *Proceedings of the 5th International Conference Autonomic and Trusted Computing (ATC)*. LNCS, vol. 5060. Springer, Berlin, 455–469.

YOLUM, P. AND SINGH, M. P. 2003. Emergent properties of referral systems. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. ACM Press, New York, 592–599.

YU, B. AND SINGH, M. P. 2002. Distributed reputation management for electronic commerce. *Computational Intelligence 18,* 4 (Nov.), 535–549.

YU, B. AND SINGH, M. P. 2003. Detecting deception in reputation management. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. ACM Press, New York, 73–80.

ZADEH, L. A. 1979. On the validity of Dempster's rule of combination. TR 79/24, Department of Computer Science, University of California, Berkeley.

ZIEGLER, C.-N. AND LAUSEN, G. 2004. Spreading activation models for trust propagation. In *IEEE International Conference on e-Technology, e-Commerce, and e-Services (EEE)*. IEEE Computer Society, Los Alamitos, 83–97.

ZIMMERMANN, P. 1995. *PGP Source Code and Internals*. MIT Press, Cambridge, MA.

A. PROOFS OF THEOREMS AND AUXILIARY LEMMAS

LEMMA 5. *$f_{r,s}(x)$ is increasing when $x \in [0, \frac{r}{r+s})$ and decreasing when $x \in (\frac{r}{r+s}, 1]$ $f_{r,s}(x)$ is maximized at $x = \frac{r}{r+s}$.*

Proof: To show monotonicity, it is adequate to assume r and s are integers and $r + s > 0$. The derivative

$$\begin{aligned} \frac{df_{r,s}(x)}{dx} &= \frac{x^{r-1}(1-x)^{s-1}}{\int_0^1 x^r(1-x)^s dx}(r(1-x) - sx) \\ &= \frac{x^{r-1}(1-x)^{s-1}}{\int_0^1 x^r(1-x)^s dx}(r - (r+s)x) \end{aligned}$$

Since $r - (r + s)x > 0$ when $x \in [0, \frac{r}{r+s})$ and $r - (r + s)x < 0$ when $x \in (\frac{r}{r+s}, 1]$, we have $\frac{df_{r,s}(x)}{dx} > 0$ when $x \in [0, \frac{r}{r+s})$ and $\frac{df_{r,s}(x)}{dx} < 0$ when $x \in (\frac{r}{r+s}, 1]$. Then $f_{r,s}(x)$ is increasing when $x \in [0, \frac{r}{r+s})$ and $f_{r,s}(x)$ is decreasing when $x \in (\frac{r}{r+s}, 1]$ $f_{r,s}(x)$ has maximum at $x = \frac{r}{r+s}$. □

The motivation behind Lemma 6 is, in essence, to remove the absolute value function that occurs in the definition of certainty. Doing so enables differentiation.

LEMMA 6. *Given A and B defined by $f_{r,s}(A) = f_{r,s}(B) = 1$, $0 < A < \frac{r}{r+s} < B < 1$, we have $c_f = \int_A^B (f_{r,s}(x) - 1)dx$*

Proof: As in Definition 2, $r + s > 0$ throughout this paper. By Lemma 5, $f_{r,s}(x)$ is strictly increasing for $x \in [0, \frac{r}{r+s})$, strictly decreasing for $x \in (\frac{r}{r+s}, 1]$ and maximized at $x = \frac{r}{r+s}$. Since the average of $f_{r,s}(x)$ in $[0, 1]$ is 1.0, we have that $f_{r,s}(\frac{r}{r+s}) > 1.0$. Since $f_{r,s}(0) = f_{r,s}(1) = 0$, there are A and B such that $f_{r,s}(A) = f_{r,s}(B) = 1$ and $0 < A < \frac{r}{r+s} < B < 1$
From Lemma 5, we have $f_{r,s}(x) < 1$ when $x \in [0, A)$ or $x \in (B, 1]$ and $f_{r,s}(x) > 1$ when $x \in (A, B)$. By Definition 3, we have $\int_0^1 (f_{r,s}(x) - 1)dx = 0$.
Therefore, $\int_0^A (f_{r,s}(x) - 1)dx + \int_B^1 (f_{r,s}(x) - 1)dx + \int_A^B (f_{r,s}(x) - 1)dx = 0$
and $\int_0^A (1 - f_{r,s}(x))dx + \int_B^1 (1 - f_{r,s}(x))dx$
$= \int_A^B (f_{r,s}(x) - 1)dx$.
Thus $\int_0^1 |f_{r,s}(x) - 1|dx = \int_0^A 1 - (f_{r,s}(x))dx + \int_B^1 (1 - f_{r,s}(x))dx + \int_A^B (f_{r,s}(x) - 1)dx$
and $\frac{1}{2}\int_0^1 |f_{r,s}(x) - 1|dx = \int_A^B (f_{r,s}(x) - 1)dx$. □

LEMMA 7.

$$\int_0^1 x^r(1-x)^s dx = \frac{1}{r+s+1}\prod_{i=1}^{r} \frac{i}{r+s+1-i}$$

Proof: We use integration by parts.
$\int_0^1 x^r(1-x)^s dx = \int_0^1 x^r d(\frac{-1}{s+1}(1-x)^{s+1})$
$= -\frac{x^r(1-x)^{s+1}}{s+1}|_0^1 + \frac{r}{s+1}\int_0^1 x^{r-1}(1-x)^{s+1}dx$
$= \frac{r}{s+1}\int_0^1 x^{r-1}(1-x)^{s+1}dx$
$= \cdots$
$= \frac{r\cdot(r-1)\cdots 1}{(r+s)\cdot(r+s-1)\cdots(s+1)}\int_0^1 (1-x)^{r+s}dx$
$= \frac{1}{r+s+1}\prod_{i=1}^{r}\frac{i}{r+s+1-i}$. □

LEMMA 8.

$$\lim_{r\to\infty}\sqrt[r]{\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}} = \frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}} \quad (1)$$

Where r is a positive integer.

Proof: This lemma is used in the next lemma, to show that the right side of an equation approaches a constant, where the equation has duplicated roots, and then the two roots of the equation approach that duplicated root.
$\lim_{r\to\infty}\frac{1}{r}\ln\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}$
$= \lim_{r\to\infty}\frac{1}{r}\ln(\prod_{i=1}^{r} i\prod_{i=1}^{r}\frac{1}{\alpha r+r+1-i})$
$= \lim_{r\to\infty}\frac{1}{r}\ln(\prod_{i=1}^{r} i\prod_{i=1}^{r}\frac{1}{\alpha r+i})$
$= \lim_{r\to\infty}\frac{1}{r}\sum_{i=1}^{i=r}\ln\frac{i}{\alpha r+i}$
$= \lim_{r\to\infty}\frac{1}{r}\sum_{i=1}^{i=r}\ln\frac{\frac{i}{r}}{\alpha+\frac{i}{r}}$
$= \int_0^1 \ln\frac{x}{\alpha+x}dx$
$= \ln\frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}}$
Therefore,
$\lim_{r\to\infty}\sqrt[r]{\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}} = \frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}}$. □

LEMMA 9.

$$\lim_{r\to\infty}A(r) = \lim_{r\to\infty}B(r) = \frac{1}{1+\alpha} \quad (2)$$

Where r is a positive integer.

Proof: The idea is to show that $A(r)$ and $B(r)$ are two roots of an equation $g(x)=\beta(r)$. If $\lim_{r\to\infty}\beta(r)=\beta$ and the equation $g(x)=\beta$ has duplicated roots of α, then we have $\lim_{r\to\infty}A(r) = \lim_{r\to\infty}B(r) = \alpha$
By Lemma 6, $A(r)$ and $B(r)$ are two roots for the equation
$x(1-x)^\alpha = \sqrt[r]{\int_0^1 x^r(1-x)^{\alpha r}dx}$
since

$\lim_{r\to\infty}\sqrt[r]{\int_0^1 x^r(1-x)^{\alpha r}dx}$
$=\lim_{r\to\infty}\sqrt[r]{\frac{1}{\alpha r+r+1}\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}}$ (by Lemma 7)
$=\frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}}$ (by Lemma 8)
$=\frac{1}{1+\alpha}(1-\frac{1}{1+\alpha})^\alpha$
since $x(1-x)^\alpha$ achieves its maximum at $x=\frac{1}{1+\alpha}$, and $x=\frac{1}{1+\alpha}$ is the only root for the equation
$x(1-x)^\alpha=\frac{1}{1+\alpha}(1-\frac{1}{1+\alpha})^\alpha$
Therefore,
$\lim_{r\to\infty}A(r)=\lim_{r\to\infty}B(r)=\frac{1}{1+\alpha}$. $\square$

Proof of Theorem 1 $c(r)$ is increasing where $r>0$ $c'(r)=\frac{d}{dr}\int_{A(r)}^{B(r)}(\frac{(x^r(1-x)^{\alpha r}}{\int_0^1 y^r(1-y)^{\alpha r}dy}-1)dx$

$$\begin{aligned}
&= B'(r)(\tfrac{B^r(r)(1-B(r))^{\alpha r}}{\int_0^1(y^r(1-y)^{\alpha r}dy}-1)\\
&\quad -A'(r)(\tfrac{A^r(r)(1-A(r))^{\alpha r}}{\int_0^1(y^r(1-y)^{\alpha r}dy}-1)\\
&\quad +\int_{A(r)}^{B(r)}\tfrac{d}{dr}(\tfrac{(x^r(1-x)^{\alpha r}}{\int_0^1 y^r(1-y)^{\alpha r}dy}-1)dx\\
&= \int_{A(r)}^{B(r)}\tfrac{d}{dr}\tfrac{(x^r(1-x)^{\alpha r}}{\int_0^1 y^r(1-y)^{\alpha r}dy}dx\\
&= \tfrac{1}{d^2}(\int_{A(r)}^{B(r)}\tfrac{d}{dr}(x^r(1-x)^{\alpha r})\int_0^1 y^r(1-y)^{\alpha r}dy\\
&\quad -\int_{A(r)}^{B(r)}(x^r(1-x)^{\alpha r})\tfrac{d}{dr}\int_0^1 y^r(1-y)^{\alpha r}dy)\\
&= \tfrac{1}{d^2}(\int_{A(r)}^{B(r)}(x^r(1-x)^{\alpha r})\ln(x(1-x)^\alpha)\int_0^1 y^r(1-y)^{\alpha r}dy\\
&\quad -\int_{A(r)}^{B(r)}(x^r(1-x)^{\alpha r})\int_0^1 y^r(1-y)^{\alpha r}\ln(y(1-y)^\alpha)dy)\\
&= \tfrac{1}{d^2}\int_0^1\int_{A(r)}^{B(r)}x^r(1-x)^{\alpha r}y^r(1-y)^{\alpha r}\ln\tfrac{x(1-x)^\alpha}{y(1-y)^\alpha}dxdy
\end{aligned}$$

where $d=\int_0^1 y^r(1-y)^{\alpha r}dy$ According to Lemma 5 $x^r(1-x)^{\alpha r}>y^r(1-y)^{\alpha r}$ when $x\in[A(r),B(r)]$ and $y\in(0,A(r)]\cup[B(r),1)$ so we have
$\int_0^{A(r)}\int_{A(r)}^{B(r)}x^r(1-x)^{\alpha r}y^r(1-y)^{\alpha r}\ln\frac{x(1-x)^\alpha}{y(1-y)^\alpha}dxdy>0$
and
$\int_{B(r)}^1\int_{A(r)}^{B(r)}x^r(1-x)^{\alpha r}y^r(1-y)^{\alpha r}\ln\frac{x(1-x)^\alpha}{y(1-y)^\alpha}dxdy>0$
since
$\int_{A(r)}^{B(r)}\int_{A(r)}^{B(r)}x^r(1-x)^{\alpha r}y^r(1-y)^{\alpha r}\ln\frac{x(1-x)^\alpha}{y(1-y)^\alpha}dxdy=0$
we have $c'(r)>0$, so $c(r)$ is increasing when $r>0$. $\square$

LEMMA 10. *Define* $L(r)=\frac{1}{\int_0^1 f(x,r)dx}\int_0^{A(r)}f(x,r)dx$ *and* $R(r)=\frac{1}{\int_0^1 f(x,r)dx}\int_{B(r)}^1 f(x,r)dx$. *Where*
$f(x,r)=x^r(1-x)^{\alpha r}$ *Then*
$\lim_{r\to\infty}L(r)=0$ *and* $\lim_{r\to\infty}R(r)=0$

Proof: We only need to show that $\lim_{r\to\infty}L(r)=0$. Since $\lim_{r\to\infty}R(r)=0$ can be proved similarly. The idea is to show that $L(r)$ is the remainder of the Taylor expansion of $(A+1-A)^{\alpha r+r}$

$$\int_0^A x^r(1-x)^{\alpha r}dx$$
$$= \int_0^A x^r d(\tfrac{-1}{\alpha r+1}(1-x)^{\alpha r+1})$$
$$= \tfrac{-1}{\alpha r+1}x^r(1-x)^{\alpha r+1}|_0^A + \tfrac{r}{\alpha r+1}\int_0^A x^{r-1}(1-x)^{\alpha r+1}dx$$
$$= \tfrac{r}{\alpha r+1}\int_0^A x^{r-1}(1-x)^{\alpha r+1}dx - \tfrac{1}{\alpha r+1}A^r(1-A)^{\alpha r+1}$$
$$= \cdots$$
$$= \tfrac{1}{\alpha r+r+1}\prod_{i=1}^{r}\tfrac{i}{\alpha r+i}(1-(1-A)^{\alpha r+r+1})$$
$$- \sum_{i=1}^{r}\prod_{j=i}^{r}\tfrac{j}{\alpha r+r+1-j}\tfrac{A^i}{i}(1-A)^{\alpha r+r+1-i}$$

So

$$L(r) = \tfrac{1}{\int_0^1 x^r(1-x)^{\alpha r}dx}\int_0^A x^r(1-x)^{\alpha r}dx$$
$$= (\alpha r+r+1)\prod_{i=1}^{r}\tfrac{\alpha r+r+1-i}{i}\int_0^A x^r(1-x)^{\alpha r}dx$$
$$= 1-(1-A)^{\alpha r+r+1}$$
$$-(\alpha r+r+1)\sum_{i=1}^{r}\begin{pmatrix}\alpha r+r\\ i-1\end{pmatrix}\tfrac{A^i}{i}(1-A)^{\alpha r+r+1-i}$$
$$= (\alpha r+r+1)(\int_0^A(x+1-A)^{\alpha r+r}dx$$
$$- \sum_{i=1}^{r}\int_0^A\begin{pmatrix}\alpha r+r\\ i-1\end{pmatrix}x^{i-1}(1-A)^{\alpha r+r+1-i}dx)$$

where $\begin{pmatrix}\alpha r+r\\ k\end{pmatrix} = \prod_{i=1}^{k}\tfrac{\alpha r+r+1-i}{i}$ for any positive integer k. Since

$$(x+1-A)^{\alpha r+r} = \sum_{i=0}^{\infty}\begin{pmatrix}\alpha r+r\\ i\end{pmatrix}x^i(1-A)^{\alpha r+r-i}$$

so we have

$$L(r) = (\alpha r+r+1)\sum_{i=r}^{\infty}\int_0^A\begin{pmatrix}\alpha r+r\\ i\end{pmatrix}x^i(1-A)^{\alpha r+r-i}dx$$
$$= (\alpha r+r+1)\sum_{i=r}^{\infty}\begin{pmatrix}\alpha r+r\\ i\end{pmatrix}\tfrac{A^{i+1}}{i+1}(1-A)^{\alpha r+r-i}$$
$$\le \tfrac{\alpha r+r+1}{r}A\sum_{i=r}^{\infty}\begin{pmatrix}\alpha r+r\\ i\end{pmatrix}A^i(1-A)^{\alpha r+r-i}$$
$$= \tfrac{\alpha r+r+1}{r}A((A+1-A)^{\alpha r+r} - \sum_{i=0}^{r-1}\begin{pmatrix}\alpha r+r\\ i\end{pmatrix}A^i(1-A)^{\alpha r+r-i})$$

since

$\sum_{i=0}^{r-1}\begin{pmatrix}\alpha r+r\\ i\end{pmatrix}A^i(1-A)^{\alpha r+r-i}$ is the Taylor expansion of $(A+1-A)^{\alpha r+r} = 1$, so

$$\lim_{r\to\infty} 1 - \sum_{i=0}^{r-1} \binom{\alpha r + r}{i} A^i (1-A)^{\alpha r + r - i} = 0$$

and by Lemma 9 $\lim\limits_{r\to\infty} \frac{\alpha r + r + 1}{r} A = 1$. Therefore,
$\lim_{r\to\infty} L(r) = 0$ and similarly $\lim_{r\to\infty} R(r) = 0$. $\square$

LEMMA 11. $\lim_{r\to\infty} c(r) = 1$

Proof: Let $f = \frac{x^r(1-x)^{\alpha r}}{\int_0^1 x^r(1-x)^{\alpha r} dx}$. Then we have
$c(x) = \int_0^1 f(x)dx - L(x) - R(x) - (B - A)$
since $\int_0^1 f(x)dx = 1$, $\lim_{r\to\infty} B - A = 0$ (by Lemma 9) and $\lim_{r\to\infty} L(r) = \lim_{r\to\infty} R(r) = 0$ (by Lemma 10). So
$\lim_{r\to\infty} c(r) = 1$

LEMMA 12. $\lim_{r\to 0} c(r) = 0$.

Proof: We only give a sketch of the proof. Let $f(x) \leq M$ when $r < 1$. For $\forall \epsilon > 0$, let $\delta = \frac{\epsilon}{2(M+1)}$, since $\frac{x^r(1-x)^{\alpha r}}{\int_0^1 x^r(1-x)^{\alpha r} dx}$ approaches to 1 uniformly in the interval $[\delta, 1-\delta]$, when $r \to 0$. So $\exists\, r_0 > 0$ such that,
$|f(x) - 1| < \epsilon$ when $r < r_0, x \in [\delta, 1-\delta]$. So when $r < r_0$,
$c(r) = \frac{1}{2}\int_0^1 |f(x) - 1| dx$
$= \frac{1}{2}(\int_0^\delta |f(x) - 1| dx + \int_\delta^{1-\delta} |f(x) - 1| dx \int_{1-\delta}^1 |f(x) - 1| dx)$
$< \frac{1}{2}((M+1)\delta + \epsilon + (M+1)\delta) = \epsilon$. Hence we have $\lim\limits_{r\to 0} c(r) = 0$. $\square$