

# Partial Observable Update for Subjective Logic and Its Application for Trust Estimation

Lance Kaplan<sup>a</sup>, Murat Şensoy<sup>b,e,\*</sup>, Supriyo Chakraborty<sup>c</sup>, Geeth de Mel<sup>d</sup>

<sup>a</sup>*US Army Research Lab, Adelphi, MD 20783, USA*

<sup>b</sup>*Department of Computer Science, Ozyegin University, TR*

<sup>c</sup>*Department of Electrical Engineering, UCLA, Los Angeles, CA 90095, USA*

<sup>d</sup>*IBM T. J. Watson Research Center, Hawthorne, NY, USA*

<sup>e</sup>*Department of Computing Science, University of Aberdeen, UK*

---

## Abstract

Subjective Logic (SL) is a type of probabilistic logic, which is suitable for reasoning about situations with uncertainty and incomplete knowledge. In recent years, SL has drawn a significant amount of attention from the multi-agent systems community as it connects beliefs and uncertainty in propositions to a rigorous statistical characterization via Dirichlet distributions. However, one serious limitation of SL is that the belief updates are done only based on completely observable evidence. This work extends SL to incorporate belief updates from partially observable evidence. Normally, the belief updates in SL presume that the current evidence for a proposition points to only one of its mutually exclusive attribute states. Instead, this work considers that the current attribute state may not be completely observable, and instead, one is only able to obtain a measurement that is statistically related to this state. In other words, the SL belief is updated based upon the likelihood that one of the attributes was observed. The paper then illustrates properties of the partial observable updates as a function of the state likelihood and illustrates the use of these likelihoods for a trust estimation application. Finally, the utility of the partial observable updates is demonstrated via various simulations including the trust estimation case.

*Keywords:* subjective logic, partial observations, Dirichlet distributions, uncertain information, trust management

---

## 1. Introduction

Decision support systems have demonstrated great utility in many different applications in the commercial and military sectors [1, 2, 3]. The performance

---

\*Corresponding author

*Email addresses:* [lkaplan@ieee.org](mailto:lkaplan@ieee.org) (Lance Kaplan), [murat.sensoy@ozyegin.edu.tr](mailto:murat.sensoy@ozyegin.edu.tr) (Murat Şensoy), [supriyo@ucla.edu](mailto:supriyo@ucla.edu) (Supriyo Chakraborty), [grdemel@us.ibm.com](mailto:grdemel@us.ibm.com) (Geeth de Mel)

of such systems has been steadily improving over the years as they incorporate big data [4] and collect data about users' preferences [5]. However, in all of these applications, such systems simply retrieve existing information. There is a desire for decision support systems to augment human intelligence beyond retrieving information by predicting the evolution of current events to help recommend potential actions. This ability is termed situational awareness [6] and is currently a human cognitive function. There have been some efforts to achieve situation awareness through multi-agent architectures [7, 8]. Furthermore, situation awareness concepts and principles can in turn enhance multiagent systems in a wide-range of applications [9, 10]. The problem is that the world is uncertain. Quantum issues aside, one might argue that complete knowledge of the state of the world at any instance (including the cognitive state of all human actors in this world) enables one to predict how the world state evolves over time. However, the complete knowledge at the necessary precision is unattainable, and at the fidelity at which one can model the world in a computer (or in ones head), it tends to be highly probabilistic. For instance, when one drives to work in the morning, there is a probability that they will encounter traffic. Of course, one can listen to the radio and nowadays they can also use the GPS in their smartphones to determine if this probability is near one or zero. However, there is always a chance that a very recent accident will cause one to get stuck in unforeseen traffic. The determination of this exact probability is obtained by collecting evidence similar to rolling a weighted die. There is always uncertainty associated to finite evidence, and the uncertainty is significant when the evidence is limited. In short, decision support systems must be able to reason and predict over such uncertainty.

Because of the prevalence of uncertainty, reasoning under uncertainty has a very rich history [11]. Certainly, probability distributions provide a representation of uncertainty, and Bayesian approaches provide sound methods to manipulate these distributions. However, many argue that Bayes' rule is unable to properly handle conflicting evidence, and many alternatives to probability theory such as belief theory [11, 12, 13] and fuzzy set theory [14] have emerged to enable fusion of information under these conflicts. At this point, it is unclear that the issue is with Bayesian reasoning per se, or with exploitation of a limited version of Bayes (first order versus higher order [15]).

Recently, *Subjective Logic* (SL) has emerged as a probabilistic logic that exploits second order Bayes to incorporate uncertainty by connecting the belief mass assignment (BMA) to a Dirichlet distribution [16, 17]. It provides a rigorous and computationally efficient method to represent and reason over human generated or automated beliefs in face of uncertainty. Applications of SL include trust management [18], Bayesian networks [19], and fusion [20, 17]. In short, SL provides effective tools to manage and combine beliefs over a set of mutually exclusive attribute states from multiple human or software agents. At a given point in time, an agent's belief, known as a subjective opinion, is the result of a prior belief and a set of observations. The uncertainty of the belief represents the reliance on the prior, and the uncertainty decreases as the agent incorporates more observations to form the beliefs over the set of attribute

states.

SL includes many operations to fuse and infer over subjective opinions from many different agents. These opinions are built through a set of observations. Each observation is the occurrence of one of the mutually exclusive attribute states as observed at a given time. The collection of observations forms the evidence, which is the number of times each of the mutually exclusive attribute states manifest in the past observations. This is akin to rolling a weighted die multiple times and tabulating the results of the rolls. For instance, the probability of encountering traffic is formed by counting the number of past instances of encountering traffic or not over a stretch of roadway at a similar time of day.

To make these notions even more concrete, let us consider another motivating example where one wants to understand the criminal activity within a city. Specifically, one wants to understand if a crime happens, what is the probability that the crime occurs in any one of the city districts. Without any initial data, one might look at socio-economic factors to develop an initial set of probabilities. Over time, one can log where a crime occurs and start to use these observations to update the probabilities. Clearly, as more observations are logged, the certainty associated with the generated probabilities increases. SL is well suited to infer the probabilities of a crime occurring in the districts and the uncertainty associated to these probabilities.

Now, let us assume that one is interested in where criminals live. The question is now when a crime occurs, what is the probability that the perpetrator lives in a particular district. Like before, one can start with a prior set of probabilities based upon the socio-economics factors. Furthermore, when a crime occurs, the location of the crime is readily available in the police report. However, the identity of the perpetrator may or may never be discovered. Therefore, it is generally not possible to log where the perpetrator lives. Sometimes, this information can be determined with great certainty when the criminal is caught. Most likely, one can only incorporate the occurrence of the crime as a *partial observation* through statistical models that link the probability of where the perpetrator lives conditioned on where a crime occurs. For instance, a criminal may not operate in his/her immediate neighborhood where he/she can easily be identified, and a criminal may not want to venture too far away either. This contextual information can help answer the questions of the distribution of criminals over the various districts within a city. This scenario is an example of a geospatial abduction problem (GAP) [21]. SL is suited to tackle such applications, but the notions of how to incorporate statistical (and not just hard) evidence of the appearance of an attribute (the home district of a perpetrator) need to be developed within the SL framework.

Another example of statistical evidence occurs in reputation systems. SL has been incorporated in reputation systems to form opinions about the trustworthiness of other agents to act honestly in a financial transaction [22]. It assumes, however, that one can clearly label a prior transaction as being honest or not. Many times, the trustor can only determine the likelihood that the trustee actually had cut corners and provided a product inferior than advertised.

This work extends the theory of SL to enable the update of the BMA under such partial observations where one piece of evidence is not a result of the die roll, but a measurement of the die roll indicating the likelihood of the result. Our preliminary work considered sequential updating of opinions from single partial observations [23] by approximating the updated posterior distribution by a Dirichlet distribution through moment matching. This paper extends these results in two respects. First, it considers an update that incorporates multiple observations. Second, the preliminary work in [23] assumes that the likelihoods are known precisely without demonstrating how they are obtained. This paper shows an example of how to derive the likelihoods for a trust estimation application where trustworthy behavior after each transaction is not directly observed. Specifically, this paper demonstrates the likelihood update for trust estimation where agents share opinions about various propositions. To this end, the likelihood that an agent’s opinion is an honest report given the trustees opinion about the same proposition is formulated. The ability to rigorously update the subjective opinions about agents from these partial observations provides a novel and important contribution in the area of trust estimation.

The rest of this paper is organized as follows. We present the related work in Section 2. Some basic concepts about Dirichlet distributions and the mathematical notations used in this paper are given in Section 3. Section 4 reviews SL and demonstrates belief updates in SL for fully visible observations. Then, Section 5 expands SL to incorporate partial updates. Examples of the behavior of partial updates are presented in Section 6, and Section 7 derives the information consistency likelihood for trust estimation. Simulations to evaluate the performance of partial observation updates in general and for the trust estimation application are provided in Section 8. Finally, Section 9 provides concluding remarks.

## 2. Related Work

We group prior work into two broad categories. The first category includes work that focus on the development of SL as a logic technique. The second category outlines the effort in the direction of trust management and reputation based systems.

### 2.1. Development of SL

This work is most directly influenced by Jøsang’s efforts to develop SL [16, 17]. SL is a type of probabilistic logic that explicitly takes uncertainty and belief ownership into account. A formal introduction to SL is presented in Section 4. In general, SL is suitable for modeling and analyzing situations involving uncertainty and incomplete knowledge. It builds upon the extensive work of evidential reasoning under uncertainty with extensive influence from the seminal theory of Dempster and Shafer [11]. Arguments in SL are subjective opinions about propositions. The key aspect of SL is that it expresses an opinion as a Dirichlet distribution and maps the parameters of this distribution to a basic

belief assignment. The interpretation of an opinion as a Dirichlet distribution leads to a rigorous calculus for many logical operations within SL [16, 17, 19, 20, 24, 25].

Within the construct of belief theory, much effort has been devoted to understand how conflicts during fusion of opinions should be spread over various belief assignments [12, 13]. Unfortunately, the notion of the best fusion rule is still a matter of debate because it ultimately is situation dependent. In essence, the best fusion rule depends on how the beliefs were formed in the first place. By relating opinions to Dirichlet distributions, SL provides a rigorous means to reason over opinions about propositions whose attributes manifest probabilistically over independent observations. For independent observers of the instantiations of such propositions, cumulative fusion is optimal [25]. While much of the effort in belief theory is devoted to fusion of beliefs, little effort has been made to consider how these beliefs were formed in the first place.

In many applications in SL, beliefs within opinions emerge from direct observations. In these cases, SL has proven very powerful. For instance in multi-agent systems, SL is used in a large range of research problems from trust modeling [26, 27, 24, 28] and argumentation [29] to knowledge representation [30] and information fusion [31].

## 2.2. Trust Management Systems

The Internet has enabled people from all over the world to interact in social and/or business contexts without necessarily any physical, e.g., face to face, connection. This is all possible due to trust and reputation systems such as online reviews that enable people to evaluate the reliability and veracity of potential associates. Numerous trust and reputation systems have been proposed in the literature, and we refer the reader to some excellent surveys for a more comprehensive view of the technology [32, 33, 22]. Because this paper applies the partial SL update to trust estimation, this subsection focuses on reviewing related trust management systems that represent trust as a beta distribution.

Jøsang and Ismail proposed the Beta Reputation System (BRS) [34]. In the BRS, modeling trustworthiness of an agent  $y$  is reduced to computing the subjective opinion for the binary proposition “ $y$  is trustworthy”. The corresponding opinion is represented as a beta distribution with parameters  $(r+1, s+1)$ , where  $r$  and  $s$  are the numbers of positive and negative outcomes of interactions with the agent  $y$ , respectively. In BRS, trust opinions about the agent are collected from a number of reputation providers and these opinions are fused. However, some malicious agents may disseminate misleading opinions.

Whitby *et al.* extended BRS to filter out misleading opinions provided by the malicious agents. This approach filters out those opinions that do not comply with the significant majority by using an *iterated filtering approach* [26]. Hence, this approach assumes that the majority of reputation providers honestly share their opinions; liars are in the minority.

Teacy *et al.* proposed TRAVOS [27], a trust modeling approach similar to BRS. The main difference between BRS and TRAVOS is the way they filter misleading opinions. While BRS uses the majority of ratings to filter out

unfair ratings from information sources, TRAVOS uses the personal observations about these sources to derive some evidence about their trustworthiness. That is, TRAVOS keeps a history of information sources and their opinions about propositions. To measure the trustworthiness of a source, an information consumer counts how many times their opinions agree and disagree for the same propositions. The number of agreements and disagreements are taken as amounts of positive and negative evidence and used to model trust in the source using beta distributions. Hence, unlike BRS, TRAVOS does not assume that the majority of the provided opinions are trustworthy; however, it requires an opinion history of a source to estimate its trustworthiness.

Zhang and Cohen proposed the personalized approach to calculate trust based on SL. This approach measures the trustworthiness of agents using two metrics [35, 36]: 1) private reputation calculated by comparing the opinion of the source with the personal observations and 2) public reputation estimated by comparing the opinion of the source and the opinions from other sources about the same propositions.

In BRS, TRAVOS, and the personalized approach, subjective opinions about the trustworthiness of agents are calculated for a fixed base rate (i.e., 0.5). Burnett *et. al* proposed to learn trustworthiness of agents using machine learning methods based on their observable features. Then, the estimated trust values are stereotypical and used as base rates in SL-based trust models [28]. That is, the effect of the stereotypical trust values are reduced as more evidence about the trustworthiness is observed. When there is no evidence observed about the trustworthiness of agents, the stereotypical trust values determine their trustworthiness.

Jøsang *et. al* explored different types of trust propagation using SL [18, 37]. Trust propagation models how new trust relationships can be derived from pre-existing trust relationship. Trust transitivity is a relatively well-studied type of trust propagation. Fusion of derived trust values is an important element of trust propagation. Jøsang *et. al* investigated a set of models that can be implemented using SL operators. Later, Jøsang and Bhuiyan proposed a method for optimal trust network analysis using SL [38]. Their approach is shown to be optimal since it does not require simplifications over trust graphs (e.g., by removing relationships).

Oren *et. al* proposed to use SL in an argumentation framework [29] to support evidential reasoning. Sensors are used within the proposed framework as sources of evidence. The collected evidence is used to form subjective opinions over arguments and do evidential reasoning. They introduced a dialogue game and a decision procedure allowing agents to decide which arguments to advance. Oren *et. al* showed that an SL-based approach for argumentation can cope with a number of problems that existing frameworks could not deal with.

Semantic Web is an important vision for the next generation of the Web. Unfortunately, knowledge representation standards for the Semantic Web cannot accommodate probabilistic and uncertain information. On the other hand, SL is a powerful probabilistic logic that can represent and reason with uncertainty. Colin *et. al* argued that SL and the Semantic Web can mutually benefit from

each other. They proposed three different extensions to enhance Semantic Web with SL to accommodate uncertainty and probabilistic reasoning [30].

SL has a set of fusion operators that are very useful to combine evidence (i.e., subjective opinions) from different sources. Trust approaches such as TRAVOS and BRS are based on *cumulative fusion* operator of SL [25]. However, more complicated fusion operators have also been developed. Sensoy *et. al* proposed to detect and resolve conflicts between subjective opinions about the same propositions before fusion [31]. They showed that the conflict resolution before fusion may significantly improve the performance of fusion in multi-agent settings.

In all the approaches described above, subjective opinions are calculated based on the totally observable evidence. This is a significant limitation, since the evidence may not be totally observable in many settings. In this paper, we propose a technique that incorporates partially observable evidence while calculating subjective opinions. Our approach has a potential to enhance a wide range of existing applications of SL in multiagent systems and other domains.

### 3. Mathematical Notation and Properties of the Dirichlet Distribution

All vectors in this paper are column vectors and are represented in boldface such as  $\mathbf{x}$  where the  $k$ -th element is given by  $x_k$ . The transpose of  $\mathbf{x}$  is denoted as  $\mathbf{x}^T$ .<sup>1</sup> The elementary vector  $\mathbf{e}_j$  is a vector of all zeros except that the  $j$ -th element is one, i.e.,  $e_{j,k} = \delta_{j,k}$  where  $\delta_{j,k}$  is the Kronecker delta that is zero for  $k \neq j$  and one for  $k = j$ .

The Dirichlet distribution is a probability distribution function (pdf) for possible values of the probability mass function (pmf)  $\mathbf{p}$  that describes the probability for the manifestation of the particular state from the  $K$  attribute states. It is characterized by  $K$  parameters  $\boldsymbol{\alpha}$  and is given by

$$f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K p_i^{\alpha_i-1} & \text{for } \mathbf{p} \in \mathcal{S}_K, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where  $\mathcal{S}_K$  is the  $K$ -dimensional unit simplex,

$$\mathcal{S}_K = \left\{ \mathbf{p} \mid \sum_{i=1}^K p_i = 1 \text{ and } 0 \leq p_1, \dots, p_K \leq 1 \right\},$$

and

$$B(\boldsymbol{\alpha}) = \int_{\mathcal{S}_K} \left( \prod_{i=1}^K p_i^{\alpha_i-1} \right) d\mathbf{p} \quad (2)$$

$$= \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^K \alpha_i\right)} \quad (3)$$

---

<sup>1</sup>In this paper, the transpose operator is used only for vectors, and the use of  $T$  in  $x^T$  for a scalar  $x$  refers to the variable  $T$ .

is the  $K$ -dimensional multinomial beta function [39]. The  $\beta$  in the subscript of  $f_\beta(\cdot)$  is used to signify that the pdf is Dirichlet. A general pdf is represented without the subscript as  $f(\cdot)$ . When  $K = 2$ , the Dirichlet distribution simplifies to the beta distribution.

The values for the  $\alpha_k$ 's relative to each other are equivalent to the expected value for the Dirichlet distribution, i.e.,

$$m_k = \frac{\alpha_k}{S}, \quad (4)$$

and the second order (noncentral) moment is given by

$$v_k = \frac{\alpha_k(\alpha_k + 1)}{S(S + 1)}, \quad (5)$$

for  $k = 1, \dots, K$  where  $S$  is the Dirichlet strength given by the sum of the  $\alpha$ 's, i.e.,

$$S = \sum_{i=1}^K \alpha_i. \quad (6)$$

It is easy to see that the second order moment can be expressed by the mean and Dirichlet strength as

$$v_k = m_k \frac{m_k S + 1}{S + 1}, \quad (7)$$

and the Dirichlet strength can be computed from the moments of any of the  $K$  marginals as

$$S = \frac{m_k - v_k}{v_k - m_k^2}. \quad (8)$$

Note that for a general distribution, the ratio given in (8) varies over  $k$ .

The Dirichlet distribution peaks near its mean value (4).<sup>2</sup> The Dirichlet strength given by (8) represents the “spread” or variance of the Dirichlet distribution around its peak. Equivalently, it represents the *strength* in the confidence of the mean to characterize the actual ground truth for  $\mathbf{p}$ . As  $S$  increases, the peak becomes higher and narrower. In the limit, as  $S \rightarrow \infty$ , the Dirichlet converges to a Dirac delta function.

The partial observation update formulation will exploit various properties for the expectation of Dirichlet distributions. Let  $E_\alpha[g(\mathbf{p})]$  represent the the expected value of  $g(\mathbf{p})$  when the distribution of  $\mathbf{p}$  is Dirichlet with parameter  $\alpha$ . When  $g(\mathbf{p})$  is a polynomial function of the elements of  $\mathbf{p}$ , the order of the expectation can be simplified via two properties. The first property alters the Dirichlet parameters for the expectation.

**Property 1**

$$E_\alpha[p_k g(\mathbf{p})] = \frac{\alpha_k}{S} E_{\alpha + \mathbf{e}_k}[g(\mathbf{p})]$$

---

<sup>2</sup>As the Dirichlet strength increases to infinity, the peak and mean become arbitrarily close to each other.



**Proof** See Appendices.

The second property preserves the Dirichlet parameters for the expectation for reducing the order of a  $T$ -th order (noncentral) moment.

**Property 2**

$$E_{\alpha}[p_{i_1}p_{i_2}\cdots p_{i_T}] = \frac{\alpha_{i_T} + \delta_{i_1,i_T} + \cdots + \delta_{i_{T-1},i_T}}{S + T - 1} E_{\alpha}[p_{i_1}p_{i_2}\cdots p_{i_{T-1}}]$$

**Proof** See Appendices.

#### 4. Subjective Logic

SL is a probabilistic logic where propositions such as the location of a crime in a city can take on one of  $K$  mutually exclusive attributes, e.g, city districts, at any observation time [16, 17]. The subjective opinion characterizes the belief in the probabilities that any of the  $K$  attributes will appear at a given observation time. The subjective opinion also characterizes the uncertainty related to these beliefs. Formally, SL considers a frame of  $K$  mutually exclusive singletons by providing a belief mass  $b_k$  for each singleton  $k = 1, \dots, K$  and providing an overall uncertainty mass of  $u$ . These  $K + 1$  mass values are all non-negative and sum up to one, i.e.,

$$u + \sum_{k=1}^K b_k = 1, \quad (9)$$

where  $u \geq 0$  and  $b_k \geq 0$  for  $k = 1, \dots, K$ .

SL also includes a base rate probability  $a_k$  for each singleton and a non-informative prior weight  $W$ . The collection of all the parameters for agent  $s$  about proposition  $x$  forms agent  $a$ 's subjective multinomial opinion  $\omega_s^x = [ (\mathbf{b}_s^x)^T \quad u_s^x \quad (\mathbf{a}^x)^T \quad W ]^T$ , where the superscripts and subscripts indexes the proposition and agent, respectively. In most cases in this paper, when the proposition and agent are implicit, the superscripts and subscripts are not used. The base rate values represent initial (or prior) information about the probability of a singleton emerging for any given observation. The inclusion of the belief and uncertainty values along with the base rates and non-informative prior weight represent the accrued evidence regarding the probability of any singleton appearing in an observation. Specifically, these values map to a Dirichlet distribution for the possible pmf that is controlling how singletons appear in observations. The parameters of the Dirichlet distribution are related to the multinomial opinion values via

$$\alpha_k = \frac{Wb_k}{u} + Wa_k. \quad (10)$$

Likewise, using (9), solving for  $b_k$  and  $u$  in (10) for  $k = 1, \dots, K$ , leads to the mapping of  $\alpha$  to the multinomial opinions

$$u = \frac{W}{S}, \quad (11a)$$

$$b_k = \frac{u}{W}(\alpha_k - Wa_k). \quad (11b)$$

The uncertainty is inversely proportional to the Dirichlet strength  $S$ . Note that the disbelief for any particular attribute is the sum of the other beliefs, i.e.,

$$d_k = \sum_{i \neq k} b_i \quad (12)$$

When the size of the frame is  $K = 2$ , the opinion simplifies to the special case of a binary opinion for a binary logic.

The expectation probabilities  $\hat{p}_k$  of a subjective opinion is defined as the mean of its corresponding Dirichlet distribution representation. Given (4) and (10), it is clear that

$$\hat{p}_k = b_k + ua_k. \quad (13)$$

When the Dirichlet distribution represents the posterior after incorporation of a number of observations,  $\hat{\mathbf{p}}$  represents the minimum mean square error (mmse) estimate of the ground truth appearance probabilities given these observations.

The cumulative fusion of two subjective opinions consists of mapping opinions into the Dirichlet parameters, summing up the parameters while taking into account not to double-count the baseline rates, and then mapping back into the multinomial opinion space [20]. This method for fusing implies that subjective opinions are formed by observations that increment the Dirichlet parameters so that fused opinions account for the aggregate of these increments. Given that the current multinomial opinion corresponds to Dirichlet parameters  $\boldsymbol{\alpha}$ , then the prior distribution for  $\mathbf{p}$  is  $f_\beta(\mathbf{p}|\boldsymbol{\alpha})$ . The updating of the opinion is accomplished by tabulating which attributes manifest for each observations. These manifestations are controlled by the ground truth probabilities  $\mathbf{p}$ , and the probability of the observations given  $\mathbf{p}$  is given by the multinomial distribution. Since the Dirichlet is the conjugate prior of the multinomial distribution, the posterior is Dirichlet where the Dirichlet parameters are updated by incrementing the parameter corresponding to which attribute appears. Specifically, for a given observation, the probability of observing the attribute as the  $i$ -th singleton, i.e.,  $z = i$ , given the pmf is  $\mathbf{p}$  is simply  $\text{prob}(z = i|\mathbf{p}) = p_i$ . Therefore, it is easy to show that the posterior for  $\mathbf{p}$  given the observation  $z$  is

$$f(\mathbf{p}|z = i) = \frac{p_i f_\beta(\mathbf{p}|\boldsymbol{\alpha})}{\int_{\mathcal{P}} p_i f_\beta(\mathbf{p}|\boldsymbol{\alpha}) d\mathbf{p}}, \quad (14)$$

$$= f_\beta(\mathbf{p}|\boldsymbol{\alpha} + \mathbf{e}_i). \quad (15)$$

In short, when a single observation reports the occurrence of the  $i$ -th singleton, the updated Dirichlet parameters are simply

$$\alpha_k^+ = \alpha_k + \delta_{k,i} \quad (16)$$

for  $k = 1, \dots, K$ . The updated belief values and uncertainties are:

$$b_k^+ = \frac{Wb_k + u\delta_{k,i}}{W + u}, \quad (17)$$

$$u^+ = \frac{Wu}{W + u}.$$

This updated multinomial opinion represents the expected pmf for the singletons via (13), and the uncertainty is related to the variance of the underlying assumed Dirichlet distribution of the singleton pmf. Overall, a multinomial opinion is formed by simply counting the occurrences of singletons to update the Dirichlet parameters, and equivalently, the multinomial opinion values. Typically, the prior weight  $W = 2$ . It represents the strength of the prior in influencing updated beliefs relative to the observation as seen in (17).

When removing the prior from the Dirichlet parameters, the resulting values represent the equivalent number of times a particular attribute is observed when rolling the dice. It is well known that for  $N$  dice rolls with ground truth probabilities  $\mathbf{p}$ , the Cramer-Rao lower bound (CRLB) for each marginal is

$$\text{CRLB}_k = \frac{p_k(1-p_k)}{N}. \quad (18)$$

Given that the estimate  $\hat{\mathbf{p}}$  as a maximum *a posteriori* estimate is asymptotically Gaussian, unbiased and efficient, we can approximate the variance of the estimate via

$$\text{VAR}_k \approx \frac{\hat{p}_k(1-\hat{p}_k)u}{W}, \quad (19)$$

where the Dirichlet strength  $W/u$  is the equivalent number of dice rolls when including the prior as reflected in the expectation probability  $\hat{p}_k$ . Furthermore, one can define bounds

$$\begin{aligned} U_p &= \hat{p}_k + \text{erfinv}(\gamma)\sqrt{2 \cdot \text{VAR}_k}, \\ L_p &= \hat{p}_k - \text{erfinv}(\gamma)\sqrt{2 \cdot \text{VAR}_k}, \end{aligned} \quad (20)$$

such the ground truth should be bounded by  $L_p \leq p_k \leq U_p$  with probability  $\gamma$ . Note that  $\text{erfinv}(\cdot)$  is the inverse of the standard error function.

The fully observable SL update is completely amenable to the Dirichlet distribution because the posterior is Dirichlet. This is not true for partial observable updates as shown in the next section. In fact, many operations that exist in SL for multinomial or just for binary opinions are not completely amenable to a mapping to the Dirichlet distribution in the sense of fusion and updates from observations. One example is the ‘‘and’’ or multiplication operation for binary opinions [40]. SL is a tractable framework, but it approximates belief propagation via parameters of a Dirichlet distribution. For any operation in SL, the operands are assumed to follow the Dirichlet distribution. A Dirichlet distribution is fitted to the output of the operation in a manner that preserves the mean. However, to maintain the properties of SL, the variance is approximated. Usually, the uncertainty is maximized under logical constraints. We assert that the accuracy of the uncertainty is how close  $\gamma$  is to the fraction of times the ground truth is actually bounded by (20), and we believe that the quality of any SL operator should also be evaluated based upon this uncertainty criterion.

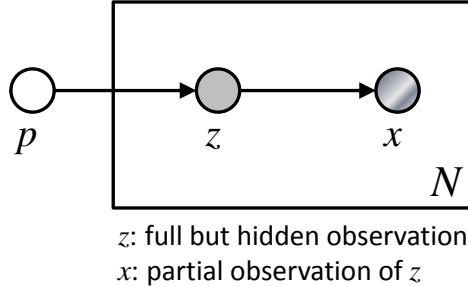


Figure 1: Plate notation describing  $N$  full and partial observations of the hidden attribute states used to estimate the state probabilities  $\mathbf{p}$ . Traditional subjective logic presumes that  $\mathbf{x}$  is the full state observation  $z$ .

## 5. Measurement Updates for Subjective Logic

Usually, it is not possible to update beliefs in attributes by directly observing the attribute at a given sampling time. Rather, a measurement is made that is statistically related to the occurrence of the attribute. The measurement  $\mathbf{x}$  can be viewed as a partial observation, i.e., a random vector drawn from a distribution that depends on the full (but hidden) observation  $z$  of the attribute state (see Figure 1). The likelihood of each of the attribute states is given by the probability of the measurement conditioned on the singletons, i.e.,

$$l_k^{(t)} = f(\mathbf{x}|z = k). \quad (21)$$

The superscript  $t$  simply indexes the observation, which happened to lead to the measurement  $\mathbf{x}$ . The superscript is only used in cases where multiple observations are used to update the subjective opinion at once.

This paper refers to (21) as the partial observation analogous to a partial observation in partial observable Markov decision processes [41]. In essence, the true observations are hidden similar to states in a hidden Markov model [42]. The likelihood function is very natural in state estimation/track filtering [43] and in pattern classification [44]. It represents abductive reasoning where the partial observation that is measured  $\mathbf{x}$  is the consequent, the full observation  $z$  is the antecedent, and  $z$  is inferred through the likelihoods  $f(\mathbf{x}|z = i)$  that serve as the conditionals. This abduction process is employed when deductive reasoning [19, 45] is not possible because the number of possible states for  $\mathbf{x}$  is too large to learn the necessary probabilities of  $z$  conditioned on all manifestations of  $\mathbf{x}$ . On the other hand, the machine learning literature provides many tools to discover the likelihood functions through neural networks, kernel methods, etc. [46, 47]. Nevertheless, learning the likelihood functions is in general still a difficult and not completely solved task as it requires a large number of

ground truthed data samples, and the likelihoods can be sensitive to operating/environmental conditions of how the data is collected. In many cases, the likelihood function emerges naturally from a physical or generative model, and Section 7 provides such an example for trust estimation. This section expands SL so that it can update beliefs given these partial observations by assuming the likelihood functions are known perfectly. Future work will investigate the implications of incorporating uncertainty about the likelihood values that can arise when a machine learning technique is employed to estimate them.

### 5.1. Naïve Belief Update

The naïve approach for the partial observation update is to spread the mass of the Dirichlet update in (16) via the normalized likelihood

$$\alpha_k^+ = \alpha_k + \frac{l_k}{\sum_{i=1}^K l_i}. \quad (22)$$

For the case of a visible update where the value of  $z$  is known, i.e.,  $l_k = \delta_{i,k}$ , then (22) simplifies to (16). While this naïve approach can be viewed as a generalization of the visible observation update, it does not yield a posterior Dirichlet distribution that is a good fit to the actual posterior distribution of the observation probabilities  $\mathbf{p}$ , i.e., the pmf describing the occurrence of the attribute states. For multiple observation  $\mathbf{l}^{(t)}$  for  $t = 1, \dots, T$ , then (22) can be applied in sequence. It is easy to see that the final parameter values  $\alpha^+$  do not depend on the ordering of how the observations are updated.

### 5.2. Actual Posterior after Measurement Update

The likelihood update determines the posterior observation probabilities  $\mathbf{p}$  given the current subjective opinion and measurement. Then one fits a Dirichlet density to the posterior in order to approximate the updated subjective opinion. This derivation starts with the joint pdf of the measurement, the hidden observation, and the observation probabilities conditioned on the current multinomial opinion, which is

$$\begin{aligned} f(\mathbf{x}, z = i, \mathbf{p}|\alpha) &= f(\mathbf{x}|z = i)\text{prob}(z = i|\mathbf{p})f_\beta(\mathbf{p}|\alpha), \\ &= l_i p_i f_\beta(\mathbf{p}|\alpha). \end{aligned} \quad (23)$$

Then marginalization to remove the hidden variable  $z$  leads to

$$f(\mathbf{x}, \mathbf{p}|\alpha) = \left( \sum_{i=1}^K l_i p_i \right) f_\beta(\mathbf{p}|\alpha), \quad (24)$$

so that the posterior for the observation probabilities after the measurement update is

$$f(\mathbf{p}|\alpha, \mathbf{x}) = \frac{1}{C} \left( \sum_{i=1}^K l_i p_i \right) f_\beta(\mathbf{p}|\alpha) \quad (25)$$

where  $C$  is the normalizing constant so that the posterior integrates to one. Note that (25) is invariant to the scaling of the likelihood. When the likelihood is zero for all classes except one, then (25) simplifies to (14), which means that the observation of the target class is completely observable. On the other hand, when all classes have equal likelihoods, (25) simplifies to  $f_{\beta}(\mathbf{p}|\boldsymbol{\alpha})$ , which means the updated beliefs are equivalent to the previous belief. In other words, the measurement is vacuous for the case of equal likelihoods. Clearly, the naïve approach given by (22) is not properly updating beliefs for the vacuous case.

In [23], the method of moments is used to approximate (25) by a Dirichlet to determine the updated  $\boldsymbol{\alpha}^+$ . This updating process can be applied in sequence for multiple observations. However, unlike the naïve method, the ordering of the observations does lead to different results. Instead, it is possible to update the belief after collecting the likelihoods for  $T$  partial observations. This does provide a more accurate subjective opinion because the Dirichlet approximation is evoked less often. This section provides various methods to perform an update from multiple partial observations at various tradeoffs of complexity. All of the methods start with the posterior that is the result of applying (25) in sequence so that after  $T$  observations

$$f(\mathbf{p}|\boldsymbol{\alpha}, \mathcal{X}^T) = \frac{1}{C} \prod_{t=1}^T \left( \sum_{i=1}^K l_i^{(t)} p_i \right) f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}), \quad (26)$$

where  $\mathcal{X}^T = \{\mathbf{1}^{(1)}, \dots, \mathbf{1}^{(T)}\}$  is the set of  $T$  partial observations. The Dirichlet approximation needs to compute the moments of the posterior. Algorithm 1 provides the basic method of moments calculation based upon the primitive calculation

$$\text{calcint}(\boldsymbol{\alpha}; T, \mathbf{1}^{(1)}, \dots, \mathbf{1}^{(T)}) = \int_{\mathcal{S}_K} \prod_{t=1}^T \left( \sum_{i=1}^K l_i^{(t)} p_i \right) f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}) d\mathbf{p}. \quad (27)$$

It is easy to see that (27) is the normalizing constant  $C$  in (26). The Dirichlet distribution is characterized by  $K$  parameters. The first  $K - 1$  degrees of freedom are used to exactly match the means of the marginal of (26) so that the relative values of  $\boldsymbol{\alpha}^+$  are obtained within a scaling factor. The final degree of freedom adjusts the Dirichlet strength to minimize the squared error between the marginal variances of (26) and that of the Dirichlet approximation, i.e.,

$$S^+ = \arg \min_{S^*} \sum_{k=1}^K \left( v_k - m_k \frac{m_k S^* + 1}{S^* + 1} \right)^2. \quad (28)$$

Note that only the noncentral moments need to be considered since the means match exactly. Solving for the point where the derivative with respect to  $S^*$  goes to zero leads to line 10 of Algorithm 1. It is easy to show that for  $K = 2$ , the variance fit is also exact. For  $K > 2$ , the number of Dirichlet parameters does not offer enough degrees of freedom to fit both the means and variances of the  $K$  marginals. As will be shown later, it is possible for some of the Dirichlet

---

**Algorithm 1** Partial Observable Update
 

---

 Input: Subjective opinion  $\omega = (\mathbf{b}, u)$  and  $T$  partial observations  $\mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}$ 

 Implicit: Base rate probability  $\mathbf{a}$  and the non-informative prior weight  $W$ 

 Output: Updated subjective opinion  $\omega^+ = (\mathbf{b}^+, u^+)$ 

- 1:  $\alpha = \frac{W\mathbf{b}}{u} + W\mathbf{a}$
- 2:  $S = \sum_{i=1}^K \alpha_i$
- 3:  $d = 0$
- 4: **for**  $i = 1, \dots, K$  **do**
- 5:    $m_i = \text{calcint}(\alpha + \mathbf{e}_i; T, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}) \frac{\alpha_i}{S}$
- 6:    $v_i = \text{calcint}(\alpha + 2\mathbf{e}_i; T, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}) \frac{\alpha_i}{S} \frac{\alpha_i + 1}{S + 1}$
- 7:    $d = d + m_i$
- 8: **end for**
- 9:  $\mathbf{m} = \frac{\mathbf{m}}{d}, \mathbf{v} = \frac{\mathbf{v}}{d}$
- 10:

$$S^+ = \frac{\sum_{i=1}^K (m_i - v_i) m_i (1 - m_i)}{\sum_{i=1}^K (v_i - m_i^2) m_i (1 - m_i)} \quad (29)$$

$$11: S^+ = \max \left\{ S^+, \frac{W a_1}{m_1}, \dots, \frac{W a_K}{m_K} \right\}$$

$$12: \alpha^+ = \mathbf{m} S^+$$

$$13: \mathbf{b}^+ = \frac{1}{S^+} (\alpha^+ - W\mathbf{a}), u^+ = \frac{W}{S^+}$$


---

parameters to decrease in value. The max operation in line 11 of Algorithm 1 ensures that the subjective beliefs never become zero.

The three methods for updating  $T$  partial observations simply differ in how `calcint` is actually computed. The first method recurses over the attributes. Specifically, `calcint` can be written as a sum of expected values

$$\text{calcint}(\alpha; T, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}) = \sum_{i_1, \dots, i_T=1}^K \left( \prod_{t=1}^T l_{i_t}^{(t)} \right) E_{\alpha} \left[ \prod_{t=1}^T p_{i_t} \right]. \quad (30)$$

Then, by Property 1,

$$\begin{aligned} \text{calcint}(\alpha; T, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}) &= \sum_{i_1, \dots, i_T=1}^K \left( \prod_{t=1}^T l_{i_t}^{(t)} \right) \frac{\alpha_{i_T}}{S} E_{\alpha + \mathbf{e}_{i_T}} \left[ \prod_{t=1}^{T-1} p_{i_t} \right], \\ &= \sum_{i_T=1}^K l_{i_T}^{(T)} \frac{\alpha_{i_T}}{S} E_{\alpha + \mathbf{e}_{i_T}} \left[ \sum_{i_1, \dots, i_{T-1}=1}^K \prod_{t=1}^{T-1} l_{i_t}^{(t)} p_{i_t} \right], \\ &= \sum_{i_T=1}^K l_{i_T}^{(T)} \frac{\alpha_{i_T}}{S} \text{calcint}(\alpha + \mathbf{e}_{i_T}; T-1, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T-1)}). \end{aligned}$$

Algorithm 2 uses this attribute-based recursion to compute `calcint`.

---

**Algorithm 2** Attribute-based Recursion
 

---

```

calcint ( $\alpha; T, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}$ )
1:  $S = \sum_{i=1}^K \alpha_i$ 
2: if  $T = 1$  then
3:    $y = \frac{1}{S} \sum_{i=1}^K l_i^{(1)} \alpha_i$ 
4: else
5:    $y = 0$ 
6:   for  $i = 1, \dots, K$  do
7:      $y = y + \frac{l_i^{(T)} \alpha_i}{S} \text{calcint}(\alpha + \mathbf{e}_i; T - 1, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T-1)})$ 
8:   end for
9: end if
10: return  $y$ 

```

---

The second method recurses over the partial observations. Using Property 2, (30) can be rewritten as

$$\begin{aligned}
\text{calcint}(\alpha; T, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}) &= \sum_{i_1, \dots, i_T=1}^K \left( \prod_{t=1}^T l_{i_t}^{(t)} \right) \frac{\alpha_{i_T} + \delta_{i_1, i_T} + \dots + \delta_{i_{T-1}, i_T}}{T - 1 + \sum_{i=1}^K \alpha_i} E_{\alpha} \left[ \prod_{i=1}^{T-1} p_{i_t} \right], \\
&= \frac{\sum_{i_T=1}^K l_{i_T}^{(T)} \alpha_{i_T}}{S + T - 1} E_{\alpha} \left[ \sum_{i_1, \dots, i_{T-1}=1}^K \prod_{t=1}^{T-1} l_{i_t}^{(t)} p_{i_t} \right] \\
&\quad + \frac{1}{S + T - 1} \left( \sum_{t'=1}^{T-1} E_{\alpha} \left[ \sum_{i_1, \dots, i_{T-1}=1}^K \tilde{l}_{i_t}^{(t, t', T)} p_{i_t} \right] \right), \\
&= \frac{\sum_{i_T=1}^K l_{i_T}^{(T)} \alpha_{i_T}}{S + T - 1} \text{cacint}(\alpha; T - 1, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T-1)}) \\
&\quad + \frac{1}{S + T - 1} \left( \sum_{t'=1}^{T-1} \text{calcint}(\alpha, T - 1, \tilde{\mathbf{l}}^{(1, t', T)}, \dots, \tilde{\mathbf{l}}^{(T-1, t', T)}) \right),
\end{aligned}$$

where

$$\tilde{\mathbf{l}}^{(t, t', T)} = \delta_{t, t'} \left( \mathbf{l}^{(t)} \circ \mathbf{l}^{(T)} \right) + (1 - \delta_{t, t'}) \mathbf{l}^{(t')} \quad (31)$$

and  $\circ$  is the Hadamard (elementwise) product. Algorithm 3 uses this observation-based recursion to compute **calcint**.

The final method uses numerical integration (the rectangular method) to calculate **calcint**. Specifically, the argument of the integral is calculated over  $\Delta$  points for each of the  $K$  dimensions for a total of  $\Delta^K$  points. This grid based integration becomes more accurate as the number of points  $\Delta$  increases.

Overall, the partial observable update method can be calculated by one of three methods. The computational complexity for these updates methods is summarized in Table 1. The attributed-based recursion branches out by a factor of  $K$  (the number of attributes) at each recursion through  $T$  levels with a



---

**Algorithm 3** Observation-based Recursion
 

---

```

calcint ( $\alpha; T, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T)}$ )
1:  $S = \sum_{i=1}^K \alpha_i$ 
2:  $y = \frac{\sum_{i=1}^K l_i^{(T)} \alpha_i}{S+T-1}$ 
3: if  $T > 1$  then
4:    $y = y$ calcint( $\alpha; T-1, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(T-1)}$ )
5:   for  $t = 1, \dots, T-1$  do
6:     for  $t' = 1, \dots, T-1$  do
7:        $\tilde{\mathbf{l}}^{(t')} = \delta_{t',t} \left( \mathbf{l}^{(t')} \circ \mathbf{l}^{(T)} \right) + (1 - \delta_{t',t}) \mathbf{l}^{(t')}$ 
8:     end for
9:      $y = y + \frac{1}{S+T-1} \text{calcint}(\alpha; T-1, \tilde{\mathbf{l}}^{(1)}, \dots, \tilde{\mathbf{l}}^{(T-1)})$ 
10:  end for
11: end if
12: return  $y$ 

```

---

Method	Complexity	Analytically Exact	Approximation Frequency
Attribute-based	$O(K^{T+1})$	Yes	Every $T$ observations
Observation-based	$O(K^2 T!)$	Yes	Every $T$ observations
Grid-based Integration	$O(TK^2 \Delta^K)$	No	Every $T$ observations
Sequential $T = 1$	$O(TK^2)$	Yes	Every observation

Table 1: Computational complexity of partial observation updates for  $T$  observations over  $K$  attributes via different methods.

leaf complexity of  $K$ . Given the  $2K$  calls to **calcint** in Algorithm 1, the overall complexity is  $O(K^{T+1})$ . Likewise, the observation-based recursion branches out by a factor of  $t$  (the number of remaining observations) at each recursion through  $T$  (the total number of observations) levels. The computational complexity at the leaf is  $O(K)$ . Overall, the complexity of the partial observable update via observation-based recursion is  $O(K^2 T!)$ . Both the attribute-based and observation-based recursions provide exact solutions within numerical precision. Finally, grid-based integration requires the calculation and summation of  $\Delta^K$  likelihood values where each likelihood value requires  $O(TK)$  to calculate. Thus, the total complexity of partial observable updates by grid based integration is  $O(TK^2 \Delta^K)$ . Clearly the accuracy of grid based integration increase proportionally with  $\Delta$  at the cost of increased complexity.

As a means to control computational complexity, one can sequentially run the partially observable update one observation at a time. The sequential update was initially proposed in [23]. Using either the attribute or observation-based recursion for  $T = 1$ , it is easy to see that in Algorithm 1, the means and noncentral second order moments are given by

$$m_k = \frac{\alpha_k + \frac{l_k^{(1)} \alpha_k}{\sum_{i=1}^K l_i^{(1)} \alpha_i}}{S+1}, \quad (32)$$

$$v_k = \frac{(\alpha_k + 1) \left( \alpha_k + 2 \frac{l_k^{(1)} \alpha_k}{\sum_{i=1}^K l_i^{(1)} \alpha_i} \right)}{(S + 1)(S + 2)}. \quad (33)$$

The equations match those given in [23]. The only difference is that the Dirichlet strength in [23] was erroneously set so that the Dirichlet parameters never decrease. While this ensured that the subjective beliefs never become negative, it also forced the uncertainty lower than it should be, which in turn lowered the influence of later observations. As a result, the performances presented in [23] are worse than what Algorithm 1 would exhibit.

For a single observation, the complexity is  $O(K^2)$  so that the overall complexity for sequentially updating  $T$  observations is  $O(TK^2)$ . Sequential updating is attractive as its complexity is linear with respect to the number of observations  $T$ , which is significantly less than the exponential complexity of the attribute and observation-based recursions. However, the sequential update makes more Dirichlet approximations. The grid-based integration is also linear with respect to  $T$ , but it requires significantly more computations than sequential updating to achieve good accuracy due to the  $\Delta^K$  term. One could also sequentially update the measurements in steps of  $T > 1$ . Section 8 investigates both the execution time and the accuracy of these various update methods.

## 6. Partial Observation Update Properties and Examples

For  $T$  fully observable updates, the Dirichlet parameters can only increase, and the overall Dirichlet strength increases from  $S$  to  $S + T$ . For partially observable observations, the behavior of the updated parameters is more interesting. For instance, it is possible for partial observations to cause a decrease in some of the Dirichlet parameters. Therefore, Algorithm 1 has to include the maximum operator to possibly adjust the Dirichlet strength so that the beliefs never become negative. This section provides some theory and examples to demonstrate how partial observations update the Dirichlet parameters, and in turn, the subjective opinion.

First, it is shown that  $T$  partially observable updates can never increase the Dirichlet strength from  $S$  to over  $S + T$ .

**Property 3** *Given a subjective opinion with an uncertainty of  $u$ , the updated uncertainty  $u^+$  after incorporation of  $T$  partial observations is lower bounded such that  $u^+ \geq \frac{W}{u+T}$ . Furthermore, the lower bound is achieved if and only if all observations are fully visible.*

**Proof** *See Appendices.*

Property 3 demonstrates that partial observations can never decrease the uncertainty lower than what visible (full) observations can accomplish. Conversely, as discussed in Section 5.2, vacuous observations do not change the opinion and do not decrease uncertainty. While an ordinary partial observations is somewhere between vacuous and fully visible, the vacuous case does not

represent an upper bound for the updated uncertainty. Usually, an observation will decrease the uncertainty. However, this is not always the case as will be illustrated below. To understand better how partial observations alter opinions, we consider a single partial observation  $T = 1$  of a binary opinion with a uniform prior, i.e.,  $a_k = 0.5$  for  $k = 1, 2$  and  $W = 2$ . The certainty, which is the complement of the uncertainty, (or equivalently the Dirichlet strength), depends on either how decisive the measurement is in terms of rejecting an attribute, i.e., the minimum likelihood is near zero, or how the observation likelihood correlates with the opinion (or equivalently the Dirichlet parameters). The updated opinions are invariant to the overall scaling of the likelihood values as the scaling is absorbed in the constant  $C$  in the prior as seen in (26). Thus for binomial opinions, the likelihood can be parameterized by a scalar likelihood ratio  $\Lambda = l_1/l_2$  so that  $\mathbf{l} = [\Lambda \ 1]^T$ . When  $\Lambda = \infty$  or  $\Lambda = 0$ , the observation is fully visible where attribute  $k = 1$  or  $k = 2$  is observed, respectively. Likewise, when  $\Lambda = 1$ , the likelihoods for the binary attributes are equal, and the observations is vacuous. Note that the relevance between the partial observation  $\mathbf{x}$  and the corresponding full observation  $z$  in the sense of [51] can be derived from  $\Lambda$ . When *Lambda* approaches zero or one, the full and partial observations become completely relevant, but when  $\Lambda$  is equal to one, they are irrelevant.

For the first set of examples, the initial uncertainty is set to  $u = 0.1$ . Figure 2 plots the updated Dirichlet and opinion scores as a function of  $\Lambda$ . In these plots  $\Lambda \geq 1$ , because the  $\Lambda < 1$  case is equivalently explained by transposing the Dirichlet parameters  $\alpha_1$  and  $\alpha_2$ , or equivalently  $b_1$  and  $b_2$ . Figure 2(a) shows that the incremental update in  $\alpha_1$ , i.e.,  $\Delta\alpha_1 = \alpha_1^+ - \alpha_1$ , is bounded by zero and one, but  $\alpha_2$  is never positive as shown in Figure 2(b). In fact, as shown later in this section, the increment update of at least one Dirichlet parameter will always be non-positive after a  $T = 1$  update. When  $\Lambda = 1$ , the measurement is vacuous, and the incremental update is always zero. This means that updated SL opinion scores are equivalent to the initial scores. As  $\Lambda \rightarrow \infty$ , the observation becomes visible and the incremental update is one for  $\alpha_1$  and zero for  $\alpha_2$ . For a given value of  $\Lambda$ , the incremental update becomes larger as the initial value of  $\alpha_1$  grows. Note that  $\alpha_1 + \alpha_2 = 20$  since  $u = 0.1$  and  $W = 2$  (see (11a)). In other words, the measurement strength increases as the Dirichlet parameters demonstrate more bias towards the singleton being espoused by the observation likelihood.

Figures 2(c)-(e) show the updated opinion scores as a function of  $\Lambda$  for the first set of examples. Note that the curves in these plots correspond to the curves in Figure 2(a) as the opinion scores and Dirichlet parameters are related via the bijective transformations given in (10) and (11). For the vacuous observations, the uncertainty remains at  $u = 0.1$ . When the initial opinion favors the second attribute, i.e.,  $\alpha_2 > \alpha_1$ , the uncertainty initially grows as  $\Lambda$  increases. In these cases, the initial opinion conflicts with the indecisive partial observations, which means these observations move the attribute beliefs closer together in the updated opinion. Since the observations are indecisive, the updated uncertainty increases so that future observations can repair errors if these observations are incorrect. As the partial observations become more

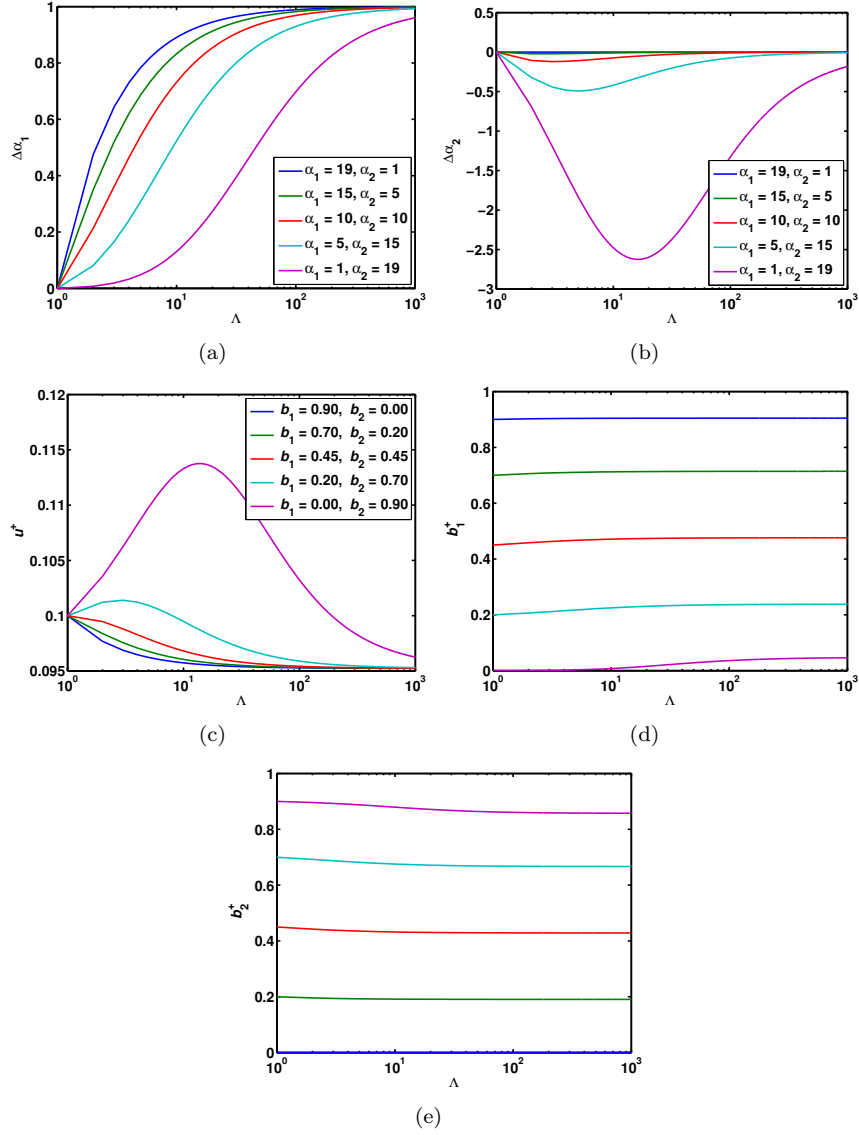


Figure 2: Updated parameters for various skew in Dirichlet parameters when the likelihood of the observation is  $\mathbf{l} = [\Lambda \ 1]^T$ ,  $u = 0.1$ ,  $\mathbf{a} = [0.5 \ 0.5]^T$ , and  $W = 2$ : (a) Incremental update on  $\alpha_1$ , (b) incremental update on  $\alpha_2$ , (c) updated  $u$ , (d) updated  $b_1$ , and (e) updated  $b_2$ .

decisive, i.e., as  $\Lambda \rightarrow \infty$ , the observation is taken to be correct and the updated uncertainty decreases to the lower limit as stipulated in Property 3. Given that  $u = 0.1$  initially and  $W = 2$ , then by (17), the uncertainty decreases to a value of  $u \approx .0952$ . Eventually, all updated opinions reach this same lower limit whether or not the observation conflicts with the initial opinion. The conflicting observations simply converge to this limit slower.

Figures 2(d) and (e) demonstrate that when the observation likelihood is in agreement with beliefs, the beliefs do not change very much, as the decrease in the uncertainty can only change the beliefs by at most a value of 0.048. On the other hand, when the observation likelihood is in disagreement with the initial beliefs, mass is transferred between the two beliefs, and the transferring of the mass becomes more pronounced as either the observation becomes more confident, i.e.,  $\Lambda$  increases, or as the initial opinion scores provides more mass for  $b_2$  than  $b_1$ .

The next set of examples demonstrate how Dirichlet and opinion scores change for various values of uncertainty. In these examples, the belief masses are equal, i.e.,  $b_1 = b_2 = (1 - u)/2$ . Again, these plots demonstrate the transition from a vacuous observation, i.e.,  $\Lambda = 1$ , to a completely visible observation, i.e.,  $\Lambda = \infty$ . Figure 3(a) demonstrates that for a fixed  $\Lambda$ , the incremental update on  $\alpha_1$  becomes larger as uncertainty decreases except when  $u = 1$ . Figure 3(b) shows that this trend reverses for the incremental update on  $\alpha_2$ . This increment update is never positive. Furthermore, it is zero for  $u = 1$ , which is an artifact of the maximum operator in Algorithm 1 that regulates beliefs from becoming negative. The use of this regulation is also why the  $u = 1$  case demonstrates the largest (instead of the smallest) increment in  $\alpha_1$ . As the initial uncertainty goes to zero, the increase in the incremental  $\alpha_1$  saturates. Figures 3(c)-(e) show that the observations have greater impact when the initial uncertainty is higher. This is simply a fact that the differences between the initial and updated values for visible observations are larger (see (17)).

In all these examples, the observation likelihood is positively correlated to the current opinion when  $\alpha_2 > \alpha_1$  or  $b_2 > b_1$ , and it is negatively correlated when  $\alpha_2 < \alpha_1$  or  $b_2 < b_1$ . As can be seen from these examples, the decrease in uncertainty is highly dependent on the how the observation likelihood correlates to the current opinion. For instance, positive correlation leads to a larger decrease. For stronger opinions, i.e., larger spread between  $\alpha_1$  and  $\alpha_2$ , or for stronger observations, i.e., larger  $\Lambda$ , the decrease in uncertainty is more pronounced. Furthermore, the degree of the decrease in uncertainty also depends on the value of uncertainty for the current belief. For multiple measurement updates using sequential updating, the ordering of the observation likelihoods in updating the beliefs matters, as some observation likelihoods better correlate with the current opinion than others. However, as well be demonstrated in the next section, sequential updating can achieve almost as good performance as incorporating all  $T$  observations at once. Therefore, the ordering of the sequential updates should have limited impact on performance.

The negative increments for  $\alpha_2$  in Figs. 2(b) and 3(b) is a usual consequence for the  $T = 1$  update.

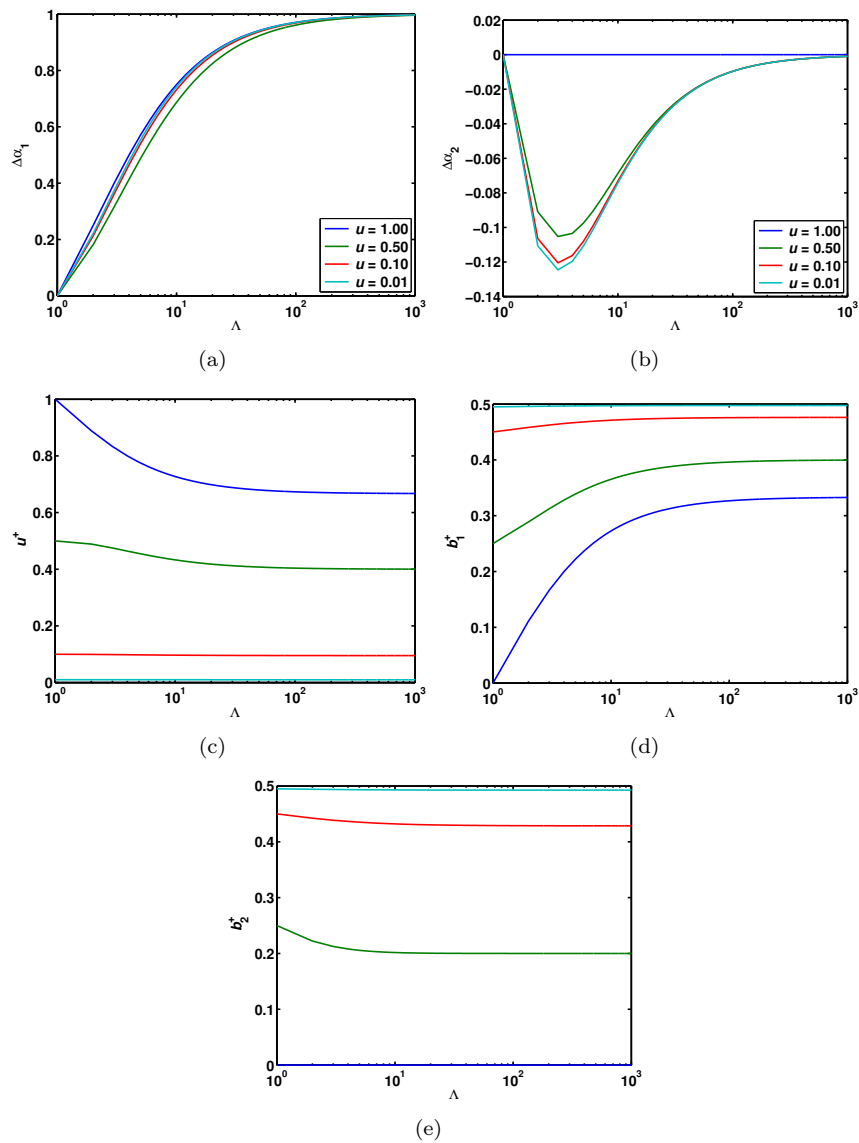


Figure 3: Updated parameters for various levels of uncertainty when the likelihood of the observation is  $\mathbf{l} = [\Lambda \ 1]^T$ ,  $b_1 = b_2$ ,  $\mathbf{a} = [0.5 \ 0.5]^T$ , and  $W = 2$ : (a) Incremental update on  $\alpha_1$ , (b) incremental update on  $\alpha_2$ , (c) updated  $u$ , (d) updated  $b_1$ , and (e) updated  $b_2$ .

**Property 4** *The Dirichlet parameter for the attribute associated to the smallest likelihood value will never increase in value after a single  $T = 1$  partial observable update. The only cases when this Dirichlet parameter does not decrease in value is when the observation is vacuous, fully observable, or the belief for the attribute is zero.*

**Proof** *See Appendices.*

Note that in Figures 2(b) and 3(b),  $\alpha_2$  corresponds to the attribute with the smallest likelihood, and its increment only achieves zero when  $\Lambda = 1$ ,  $\Lambda \rightarrow \infty$ , or when  $u = 0$  so that  $b_2 = 0$ .

## 7. Consistency Likelihood and Trust Estimation

This section derives the likelihood for a trust estimation application where a trustor agent is receiving subjective opinions from various sources, i.e., other agents) about various binary propositions. For instance, the binary proposition refers to the probability that the agent will be *satisfied* by a particular vendor based up previous transactions. The trustor agent is collecting these opinions about various vendors from various agents. At the same time, the trustor is able to form its own subjective opinions about the vendors based upon its own transactional history. The trustor wants to form an opinion about each agent to determine the probability that an agent does not manipulate its subjective opinion before reporting it. Then, the trustor can understand how to discount and fuse all the agent reports to form more certain opinions about the various vendors.

At a given instance, the trustor asks all the source agents to return their binary opinions about a particular proposition (e.g., a vendor). Given that agent  $i$  reports a binary opinion  $\omega_i$  such that the opinion corresponds to Dirichlet parameters  $\alpha_i$ , then it means that the agent has collected  $r_i = \alpha_{i,1} - 1$  and  $s_i = \alpha_{i,2} - 1$  positive and negative evidences (e.g., transactional experiences with a vendor).<sup>3</sup> If the opinion reported by agent  $i$  is truthful (i.e.,  $z_i = 1$ ), the distribution for the pmf  $\mathbf{p}$  of the proposition (about a vendor) due to this report is given by the beta distribution<sup>4</sup> with parameters  $[ r_i + 1 \quad s_i + 1 ]^T$ , i.e.,

$$f(p_1|z_i = 1, \omega_i) = \frac{p_1^{r_i}(1-p_1)^{s_i}}{B([ r_i + 1 \quad s_i + 1 ]^T)}. \quad (34)$$

for  $0 \leq p_1 \leq 1$ .

On the other hand, when source  $i$  is not truthful (i.e.,  $z_i = 2$ ), there may or may not be any relationship between the reported opinion (about the vendor) and the evidence that source  $i$  actually collected. For now, we assume that an untruthful agent reports a random opinion that does not favor any value for  $\mathbf{p}$ .

<sup>3</sup>Here it assumed that  $W = 2$  and the prior is  $\mathbf{a} = [ 0.5 \quad 0.5 ]^T$ .

<sup>4</sup>The beta distribution is equivalent to the Dirichlet distribution when  $K = 2$ .

In other words, the distribution for  $\mathbf{p}$  given that the source's report is untruthful is uniform, i.e.,

$$f(p_1|z_i = 2, \boldsymbol{\omega}_i) = 1. \quad (35)$$

If the ground truth values are known to be  $\mathbf{p}$  for the current proposition (about a vendor), the probability that the trustor agent (source 0) collected evidence associated to Dirichlet parameters  $\boldsymbol{\alpha}_0 = [r_0 + 1 \quad s_0 + 1]$  about the same proposition is given by the binomial distribution

$$f(\boldsymbol{\omega}_0|p_1) = \frac{(r_0 + s_0)!}{r_0!s_0!} p_1^{r_0} (1 - p_1)^{s_0}. \quad (36)$$

This binomial distribution assumes that the collected evidences  $r_0, s_0$  are integers. To be more general, we extend the probability for any subjective opinion by replacing the factorials with gamma functions using the fact that  $\Gamma(r + 1) = r!$  for non-negative integer  $r$ . Then, the probability that the trustor has direct opinion  $\boldsymbol{\omega}_0$  is given by

$$f(\boldsymbol{\omega}_0|p_1) = \frac{1}{(r_0 + s_0 + 1)B([r_0 + 1 \quad s_0 + 1]^T)} p_1^{r_0} (1 - p_1)^{s_0}. \quad (37)$$

The likelihood that source  $i$  is truthful is the probability that the trustor formed an opinion  $\boldsymbol{\omega}_0$  conditioned on a truthful source  $i$  reporting  $\boldsymbol{\omega}_i$ , i.e.,

$$l_1 = p(\boldsymbol{\omega}_0|z_i = 1, \boldsymbol{\omega}_i). \quad (38)$$

This likelihood is easily calculated using the conditional probability of the observed opinion and the distribution of  $\mathbf{p}$  given the reported evidence

$$l_1 = \int_0^1 f(\boldsymbol{\omega}_0|p_1) f(p_1|z_i = 1, \boldsymbol{\omega}_i) dp_1. \quad (39)$$

Using (34) and (37), this likelihood simplifies to

$$l_1 = \frac{1}{r_0 + s_0 + 1} \frac{B([r_0 + r_i + 1 \quad s_0 + s_i + 1]^T)}{B([r_0 + 1 \quad s_0 + 1]^T)B([r_i + 1 \quad s_i + 1]^T)}. \quad (40)$$

Similarly, the likelihood that the source  $i$  is untruthful is computed using (35) and (37) in a similar manner to be

$$l_2 = \frac{1}{r_0 + s_0 + 1}. \quad (41)$$

The trustor can compute trust for source  $i$  over multiple rounds by comparing its observed opinion to the opinion reported by the source. Each round can consist of opinions for a different proposition as long as the observed and reported opinions are about the same proposition. The above models provide a rigorous method to compute either the likelihood that source  $i$  is truthful or the likelihood that it is not reporting the actual observed opinions to the trustor



agent. These likelihoods can be used to update a beta reputation system for source  $i$  since the reported and collected opinions form a partial observation about the veracity of source  $i$  with the likelihoods of trustworthy and non-trustworthy behavior given by (40) and (41), respectively. For this case, the likelihood ratio

$$\Lambda_i = \frac{l_1}{l_2} = \frac{B([r_0 + r_i + 1 \quad s_0 + s_i + 1]^T)}{B([r_0 + 1 \quad s_0 + 1]^T)B([r_i + 1 \quad s_i + 1]^T)} \quad (42)$$

uniquely determines whether or not the current evidence points towards a truthful source  $i$  or not. When  $\Lambda_i \gg 1$  or  $\Lambda_i \ll 1$ , the observation about agent  $i$  becomes fully visible, and the corresponding Dirichlet parameter is incremented by one. When  $\Lambda_i = 1$ , there is no evidence favoring either hypothesis. In other words, the evidence is vacuous, and the Dirichlet parameters are not updated.

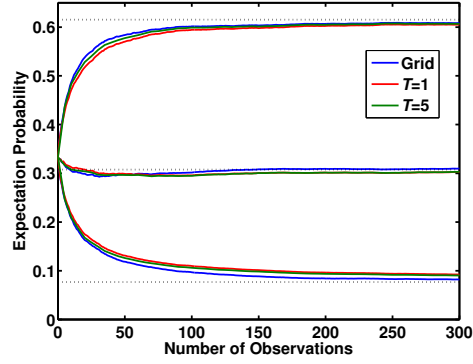
## 8. Simulations

This section incorporates simulations to evaluate the utility of the various extensions of SL for partial observable measurements. The first simulated example considers a toy problem where a proposition can manifest into one of  $K = 3$  attributes where the appearance probabilities are  $\mathbf{p} = \begin{bmatrix} \frac{8}{13} & \frac{4}{13} & \frac{1}{13} \end{bmatrix}$ . For each observation, the measurement conditioned on the appearance of the  $j$ -th attribute is a 3D vector  $\mathbf{x} = \mathbf{e}_j + \mathbf{n}$ , where  $\mathbf{n}$  is Gaussian measurement noise of mean zero and a covariance matrix of  $\Sigma = \sigma^2 I$ , i.e.,

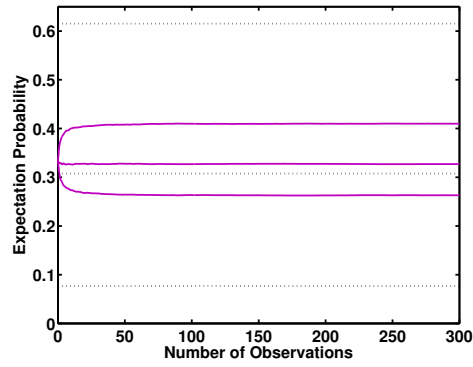
$$f(\mathbf{x}|z = j) = \frac{1}{(\sqrt{2\pi\sigma^2})^K} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{e}_j\|^2 \right\}. \quad (43)$$

Note that in the data space, the distances between the clusters representing the  $K$  attributes, e.g., distance between attribute centroids, are all equal. The spread  $\sigma^2$  controls the confidence of the resulting likelihoods. Here  $\sigma^2 = 1$ , which represents fairly poor class separation. Using this measurement model, we simulated 300 partial observations of the attributes of the proposition. The process was repeated over 1000 Monte Carlo simulations. Given the cluster separation when  $\sigma^2 = 1$ , the largest likelihood in  $\mathbf{l}$  actually corresponds to the currently appearing attribute about 63% of the time.

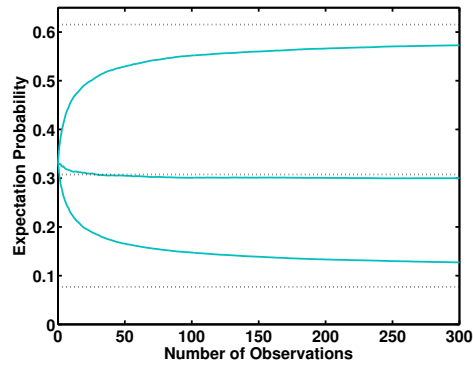
Figure 4 compares sequentially updating  $T = 1$  or  $T = 5$  partial observations using exact calculations against updating all available observations at once via the grid-based approximation with  $\Delta = 0.001$ . All updates are consistent in that as more observations are incorporated, the expectation probabilities are converging to the ground truth. Tables 2 and 3 summarize the plots by providing the average bias and root mean squared (rms) error for each of the appearance probabilities. The figure and tables do show that increasing  $T$  or using all available observations does improve the bias, but only very slightly. The rms error seems to be insensitive to the choice of how many partial observations to incorporate at once. In short, performing a sequence of single partial updates



(a)



(b)



(c)

Figure 4: Expectation probability of subjective opinion versus the number of accumulated partial observations: (a) Proposed likelihood updates, (b) naïve updates, and (c) monotonic likelihood updates from [23].

Estimate	$T = 1$	$T = 2$	$T = 3$	$T = 4$	$T = 5$	All $T$
$p_1$	0.0184	0.0174	0.0165	0.0158	0.0152	0.0127
$p_2$	0.0072	0.0074	0.0076	0.0076	0.0076	0.0029
$p_3$	0.0256	0.0248	0.0240	0.0234	0.0228	0.0141
All	0.0171	0.0165	0.0160	0.0156	0.0152	0.0099

Table 2: The absolute bias for the three estimates of  $\mathbf{p}$  averaged over the partial observable measurements.

Estimate	$T = 1$	$T = 2$	$T = 3$	$T = 4$	$T = 5$	All $T$
$p_1$	0.0749	0.0750	0.0750	0.0751	0.0752	0.0746
$p_2$	0.0682	0.0685	0.0689	0.0691	0.0694	0.0693
$p_3$	0.0531	0.0528	0.0524	0.0521	0.0518	0.0507
All	0.0660	0.0661	0.0661	0.0662	0.0662	0.0656

Table 3: The rmse for the three estimates of  $\mathbf{p}$  averaged over the partial observable measurements.

is just about as accurate as incorporating all  $T$  observations at once despite evoking the Dirichlet approximation many more times.

Figure 4(b) shows the performance of the naïve update. The naïve update is clearly biased. It actually converges to the expected likelihood vector for the given measurement and appearance models, which usually does not correspond to the ground truth appearance probability. As the class separation become infinitely good, i.e.,  $\sigma^2 \rightarrow 0$ , the expected likelihood vector does correspond to the appearance probability. On the other hand, as the class separation becomes infinitely poor, i.e.,  $\sigma^2 \rightarrow \infty$ , the expected likelihood become uniformly  $1/3$ . In that case, the partial SL updates are vacuous, and the expectation probability is also uniformly  $1/3$ . The SL update correctly returns a completely uncertain  $u = 1$  opinion, but the uncertainty from the naïve update erroneously goes to zero.

Figure 4(d) shows the performance of the partial observation update that we initially proposed in [23]. In that method, the Dirichlet strength was erroneously increased so that the Dirichlet parameters  $\alpha$  never decrease after each update. Therefore, we refer to this update as the *monotonic* method. The resulting expectation probabilities are still unbiased, but the faster decrease in uncertainty causes the convergence to slow down tremendously as compared to Figure 4(a).

Finally, Figure 5 demonstrates the accuracy of the uncertainty in the updated opinions to actually represent the spread of the expectation probability. For each update method, there are 300,000 estimates of  $\mathbf{p}$  and we tabulated the fraction of times the corresponding ground truth fall within the bounds given by (20) for a desired probability. All the proper partial observation update methods demonstrate a fraction that equals the desired probability. The naïve and monotonic methods do not straddle the slope one line. This could be attributed to the bias. However, the bias is virtually zero for the monotonic method cal-

Method	$t_2/t_1$	$t_3/t_2$	$t_4/t_3$	$t_5/t_4$
Attribute	1.9778	3.1771	3.2863	2.9742
Observation	2.0382	2.8867	3.7419	4.9128

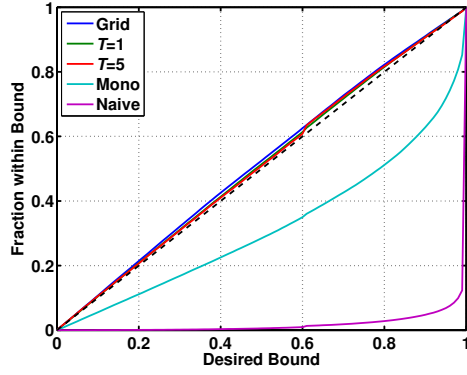
Table 4: Ratio of average execution times.

culating  $p_2$ . Even in that case, the unnecessarily larger Dirichlet strength (or smaller uncertainty) leads to tighter bounds so that a smaller fraction of ground truth values fall within the bounds.

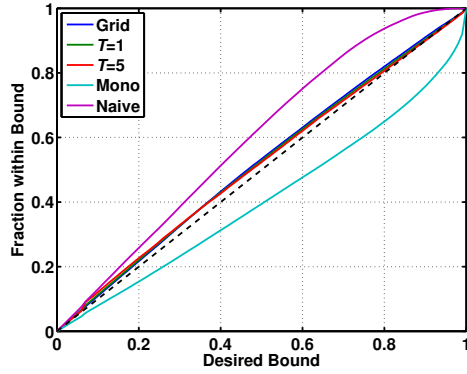
Figure 6 plots the average execution time to implement each update. Table 4 provides the ratio of times for incorporating  $T + 1$  observations over the time to incorporate  $T$  observations. These execution times are consistent with the algorithm complexity as attributed-based recursion increases with  $T$  as  $K^T$  and observation-based recursion increases as  $T!$ . Figure 6(b) does show that the grid based approximation increases linearly with respect to  $T$ , but it is more than three orders of magnitude slower than a sequential  $T = 1$  updates. It should be noted that by  $T = 8$ , both exact recursion methods should exceed the grid based approximation in execution time due to their exponential increase with respect to  $T$ .

The next simulation involves evaluation of one agent’s opinions about the ability of other agents to provide accurate opinions for various propositions. Ten agents are simulated where the each agent is able to formulate an opinion about a particular binary proposition by collecting between zero to ten fully visible observations. The ground truth appearance probability for the proposition and the number of collected observations are randomly drawn from the uniform distribution. Nine of the agents reports their opinions to the trustor agent. Each of the reporting agents manipulates their opinion about the proposition with a given probability. If the opinion is manipulated, the agent reports a uniformly distributed random opinion. Otherwise, the agent reports its true opinion. The trustor agent uses its own opinion about the proposition to formulate an updated opinion of the nine agents via the  $T = 1$  partial observable update using the consistency calculation from Section 7 for the likelihoods. The trustor opinions about the other nine source agents are formed sequentially over 500 propositions.

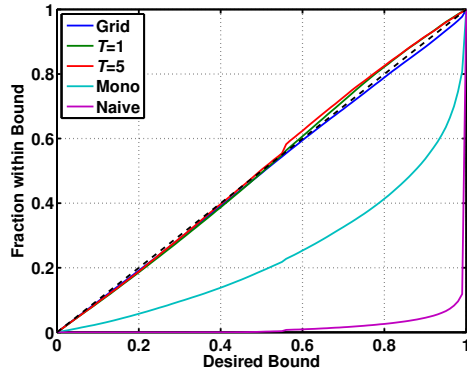
Usually trust reputation systems treat trust updates as fully observable. As in [48], the positive evidence is incremented by one when the absolute difference between the expectation probabilities of the trustor and reporting agent is smaller than a threshold. Likewise, negative evidence is incremented by one when this difference is larger than another larger threshold. This threshold techniques is used as one baseline method in the simulations. TRAVOS is a reputation system that estimates the trustworthiness of inaccurate sources through a binning method that measures the probability of accuracy of a source [27]. Specifically, it tabulates the trustor agent’s opinion whenever a reporting agent’s opinion fall under a particular bin, and then it computes the probability that the trustor agent’s opinion cover this bin. This estimated probability of



(a)

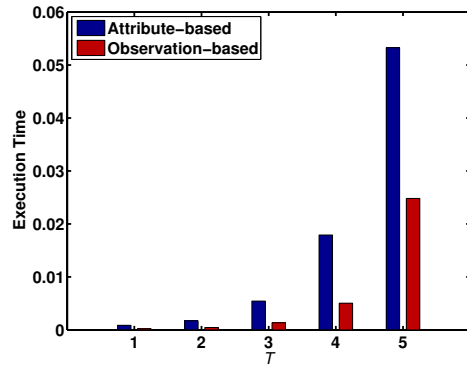


(b)

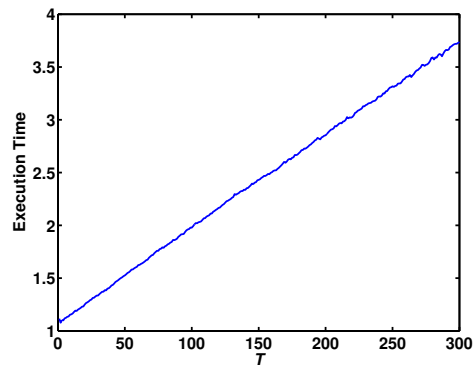


(c)

Figure 5: Comparison of fraction of estimates that fall within uncertainty bounds versus the desired probability  $\gamma$  for the various update methods: (a)  $p_1$ , (b)  $p_2$ , and (c)  $p_3$ .



(a)



(b)

Figure 6: Average executions time to update  $T$  partial observations: (a) Exact updates and (b) grid based approximation with  $\Delta = 0.001$ .

accuracy by TRAVOS serves as the second baseline method. Unfortunately, TRAVOS does not report an uncertainty associated with this estimate so that the estimate cannot be converted into a subjective opinion.<sup>5</sup>

Figure 7 plots the expectation probabilities for the trustor agent’s opinion about the other agents as it collects reported propositions as partial observations. The dotted lines in Figures 7 represent the ground truth probabilities that a particular agent manipulates its reported opinion. Clearly, the likelihood update provides a consistent estimate that converges to the ground truth. The thresholding and binning methods are not converging to the ground truth, but they are consistent in that they preserve the ordering of the ground truth. The binning method of TRAVOS appears to exhibit much less bias than thresholding. However, TRAVOS appears to converge slower than the other two methods. Figure 7(d) demonstrates the accuracy of the uncertainty obtained from the likelihood updates. The plot does not show the TRAVOS result as the confidence in the probability of accuracy is not provided as part of the TRAVOS computations. Overall, this simulation demonstrates the a well formulated likelihood can be determined from modeling for updating SL opinions from partially observable measurements.

TRAVOS then uses the probability of source accuracy to discount the sources opinion for a proposition followed by consensus fusion of the discounted opinions. Specifically, the distribution of  $\mathbf{p}$  for the proposition is modeled as a mixture of a beta distribution representative of the reported opinion and a uniform distribution where the weights are given by the probability of source accuracy. TRAVOS then approximates this mixture distribution as a beta distribution to formulate a discounted opinion by the method of moments methods similar to its use in this paper. Finally, the discounted methods are fused via the SL consensus operator. We implemented this discounting and fusion method to compare the quality of the resulting opinions when using the source accuracy opinions as shown in Figure 7. We also simulated the *whitewashing* attacks in Peer-to-Peer systems, where malicious agents leave the society when their reputation decrease and join back with new identities to whitewash their bad reputation and abuse the system. To this end, after each round of sharing propositional opinions, each reporting agent has a probability of  $P_l$  to leave, and it is replaced by a new reporting agent with equivalent trustworthiness. The trustor agent resets its trust opinion whenever an agent leaves. The trust estimates presented in Figure 7 represent the case when  $P_l = 0$ .

Figure 8 reports the rms error of the expectation probabilities for the fused opinions over the 500 propositions for different values of  $P_l$ . Clearly, discounting demonstrates significantly better accuracy than simple consensus fusion. The simple consensus fusion does not incorporate the trust estimates of the reporting agents; thus, its performance does not depend of  $P_l$ . All three discounting methods exhibit larger error as  $P_l$  increases. The thresholding method exhibits

---

<sup>5</sup>Many trust and reputation systems do provide reliability measures for the reported trust/reputation values as discussed in [33].

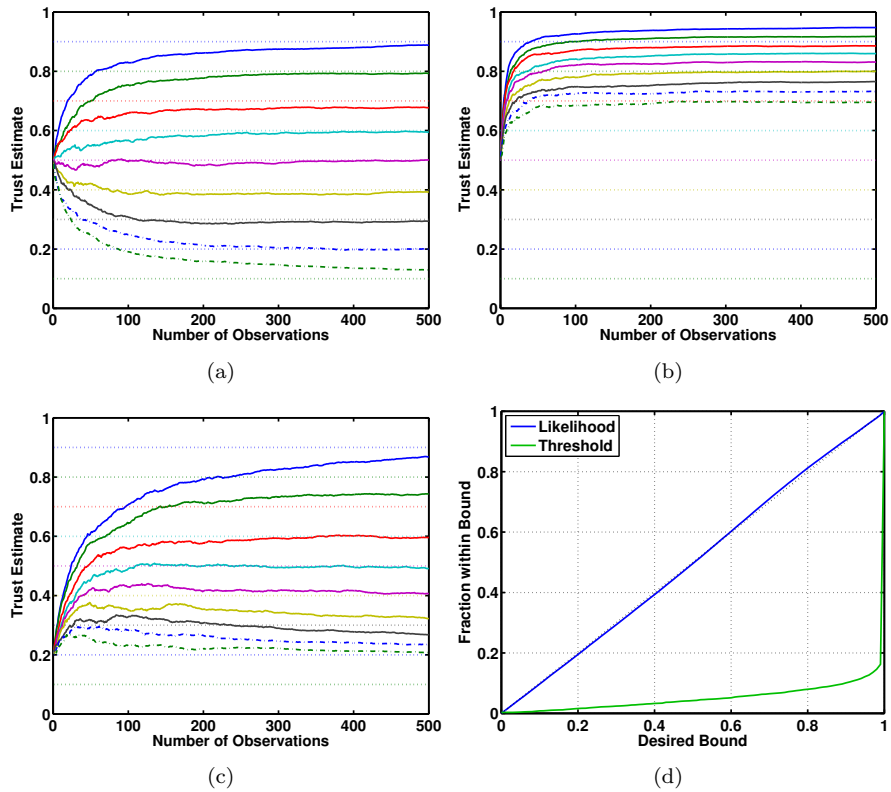


Figure 7: Quality of trust estimates for the unreliable sources: (a) Expectation versus observations accumulated using consistency likelihood, (b) expectation versus observations accumulated by thresholding, (c) expectation versus observations accumulated for TRAVOS, and (d) comparison of fraction of estimates that fall within uncertainty bounds versus the desired probability  $\gamma$ .



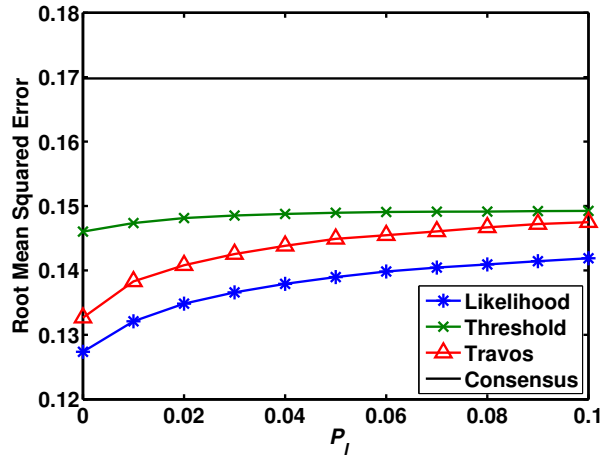


Figure 8: The rmse for expectation probability of the fused reported proposition from unreliable sources versus the probability of agents leaving  $P_l$  after each round of sharing propositional opinions.

the largest error, but it is less sensitive to  $P_l$ . The likelihood trust estimates lead to the best fused results. TRAVOS is much more sensitive to  $P_l$  due to the slower convergence of the trust estimates for the source agents. TRAVOS outperforms thresholding but as  $P_l$  goes to 0.1, the performance of both appear to be converging. As  $P_l$  gets large, all discounting methods are basically discounting all reported opinions for a given proposition by the prior trust estimate and fusing these opinions with the non-discounted directly observed opinion of the trustor agent.

When the propositions in this simulation represent opinions about other agents such as sellers, this simulation represents a classical trust reputation system where the trustor agent fuses direct evidence of the seller along with discounted indirect evidence from the referral agents. The discounting of the referral agents' opinions then represents transitivity of trust whether it is done as in TRAVOS [27] or as in SL [37]. These discounting methods can be justified when manipulated opinions are completely random. However, as shown in [49], the posterior opinion of the trustor's opinion about any seller given his direct evidence and the aggregate of the opinions from all referral agents about all sellers is a sum of Dirichlet distributions. The modular process of discounting opinions followed by consensus fusion in SL only approximates the discounted opinions and subsequently the fused consensus opinions using pure Dirichlet distributions. While the work in [49] provides the form for the (fused) posterior distribution for the trustworthiness of each seller, it does not provide a practical algorithm to compute it as the number of Dirichlet modes in the distribution grows exponentially with the number of opinions. Future work needs to investigate when and when not it is reasonable to use the Dirichlet approximation to

compute an estimate of the seller trustworthiness posteriors.

## 9. Conclusions

SL presumes that direct evidence exists to develop subjective opinions. Most reputations systems uses ad-hoc techniques to determine which type of behavior (good, bad or indeterminant) is being observed so that the proper beta (or Dirichlet) parameters are incremented. Despite the fact that direct evidence is usually not available, it is usually possible to calculate the likelihood that the observation is indicative of a particular attribute. In some applications, a physical model can be employed to determine the likelihood. In other cases, machine learning methods exist to discover the likelihood function, but learning the likelihood can be challenging due to limited availability of training data and variability in how the data can be collected. This paper expands SL to enable such partial observable updates to opinions from these likelihood values. Furthermore, a first order model is used to derive the likelihood functions for a trust estimation application based upon consistency between reported and observed subjective opinions about various propositions.

The evidential strength of the partial observable is upper bounded by that of the fully visible observation. Interestingly, the partial observable updates do not simply increment Dirichlet parameters. In general, some parameters decrement slightly while other increase. A partially confident observation that conflicts with the current belief can actually increase uncertainty, which never occurs for fully visible (or confident) observations.

SL approximates the collected evidence in an opinion so that it corresponds to a Dirichlet distribution. In the partial observation update, a Dirichlet distribution is fitted to the actual posterior by using the method of moments. It is possible to incorporate multiple observations at once before evoking the Dirichlet approximation. However, the multiple observation update exhibits significantly more computational complexity, and the simulations demonstrate negligible performance gains in terms of the expectation probability of the resulting SL opinions reflecting the ground truth. The simulations also demonstrate that the uncertainty of the updated opinion does reflect the known confidence that the resulting expectation probability reflects the ground truth.

When the likelihood functions are discovered via machine learning techniques, there is naturally some uncertainty associated to the likelihoods. We do have some preliminary work that investigates such uncertainty in this likelihood [50]. Uncertain likelihoods is actually a form of SL abduction, where the partial observation  $\mathbf{x}$  is the consequent, the hidden observation  $z$  is the antecedent, and the likelihoods  $f(\mathbf{x}|z = i)$  serve as the conditionals. This form of abduction is different than subjective logic abduction [17, 19, 51] where the antecedent is not a discrete observation, but rather a proposition characterized by the appearance probabilities for its possible mutually exclusive attributes. The abduction in this paper updates the opinion about the appearance probabilities of a proposition by abducting the update from the partial observation  $\mathbf{x}$ . The work presented in this paper is actually a special case of such abduction when the likelihood

(or conditional) uncertainty is zero, and future work will expand this abduction framework to consider the collection of observations to form subjective opinions about the likelihood.

Finally, the uncertainty resulting from many SL operators does not necessarily lead to performance bounds that reflect the proper proximity to the ground truth. This happens either because the uncertainty is not derived to match the variance of the actual resulting distribution for the appearance probabilities, or because the resulting distribution is not well approximated by a Dirichlet distribution. For example, the uncertainty related to the fused propositional opinions in the trust example from Section 8 are not consistent with the ground truth. As shown in [49], the derived opinions about propositions may not be well represented by Dirichlet distribution. Actually, trust and propositional opinions need not be summarized in a discounted opinion. Rather, it can be better to fuse trust and propositional opinions of all agents jointly. Overall, SL is attractive because of its computational simplicity. However, future work needs to investigate under what conditions when SL can accurately evoke the Dirichlet approximation, and when SL should not compress the representation. Such work needs to consider various applications evoking different SL operators.

## Acknowledgments

We are deeply indebted to the late Chatschik Bisdikian whose own efforts served as the inspiration for this work, and we are saddened that he was unable to help us bring this paper to fruition. We also want to thank Gene Whipps for helping to review an early draft of this paper.

Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Numbers W911NF-06-3-0001 and W911NF-13-1-0243. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Dr. Şensoy thanks to the U.S. Army Research Laboratory for its support under grant W911NF-14-1-0199 and The Scientific and Technological Research Council of Turkey (TUBITAK) for its support under grant 113E238.

## References

- [1] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligence Systems*, 9th ed. Prentice Hall, 2011.
- [2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Waltham, MA: Elsevier, 2012.

- [3] E. Blasch, E. Bosse, and D. A. Lambert, Eds., *High-level Information Fusion Management and Systems Design*. Artech House, 2012.
- [4] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, “Big data, analytics and the path from insights to value,” *MIT Sloan Management Review*, vol. 52, no. 2, pp. 21–31, 2011.
- [5] T. Joachims and F. Radlinski, “Search engines that learn from implicit feedback,” *IEEE Computer*, vol. 40, no. 8, pp. 34–40, Aug. 2007.
- [6] M. R. Endsley and D. J. Garland, Eds., *Situation Awareness Analysis and Measurement*. Mahwah, NJ: CRC, 2000.
- [7] T. Nguyen, “A multi-agent architecture for situation awareness,” in *Knowledge-Based Intelligent Electronic Systems*, vol. 2, May 1997, pp. 502–507.
- [8] G. Chen, E. P. Blasch, H. Chen, and K. Pham, “Multi-agent modeling and analysis for space situation awareness,” in *SPIE Newsroom: Defense & Security*, 2009.
- [9] M. Lorenz, J. D. Gehrke, H. Langer, I. J. Timm, and J. Hammer, “Situation-aware risk management in autonomous agents,” in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 2005, pp. 363–364.
- [10] G. Castellano, M. G. Cimino, A. M. Fanelli, B. Lazzerini, F. Marcelloni, and M. A. Torsello, “A multi-agent system for enabling collaborative situation awareness via position-based stigmergy and neuro-fuzzy learning,” *Neurocomputing*, 2014.
- [11] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [12] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, Eds., *Advances in the Dempster-Shafer Theory of Evidence*. New York: John Wiley & Sons, Inc., 1994.
- [13] F. Smarandache and J. Dezert, Eds., *Advances and Applications of DSmT for Information Fusion: Collected Works*. Infinite Study, 2004.
- [14] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice Hall, 1995.
- [15] J. Baron, “Second-order probabilities and belief functions,” *Theory and Decision*, vol. 23, no. 1, pp. 25–36, Jul. 1987.
- [16] A. Jøsang, “A logic for uncertain probabilities,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, no. 3, pp. 279–311, Jun. 2001.
- [17] A. Josang, *Subjective Logic*. online, Jul. 2011, draft book in preparation.

- [18] A. Jøsang, S. Marsh, and S. Pope, “Exploring different types of trust propagation,” in *Proc. of the 4th International Conference on Trust Management (iTrust)*, Pisa, Italy, May 2006.
- [19] A. Jøsang, “Conditional reasoning with subjective logic,” *Journal of Multiple-Valued Logic and Soft Computing*, vol. 15, no. 1, pp. 5–38, 2008.
- [20] A. Josang, “The consensus operator for combining beliefs,” *Artificial Intelligence Journal*, vol. 142, no. 1–2, pp. 157–170, Oct. 2002.
- [21] P. Shakarian, V. S. Subrahmanian, and M. L. Sapino, “GAPs: Geospatial abduction problems,” *ACM Trans. on Intelligent Systems and Technology*, vol. 3, no. 1, Oct. 2011.
- [22] A. Jøsang, R. Ismail, and C. Boyd, “A survey of trust and reputation systems for online service provision,” *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, Mar. 2007.
- [23] L. M. Kaplan, S. Chakraborty, and C. Bisdikian, “Fusion of classifiers: A subjective logic perspective,” in *Proc. of the IEEE Aerospace Conference*, Big Sky, MT, Mar. 2012.
- [24] A. Jøsang, R. Hayward, and S. Pope, “Trust network analysis with subjective logic,” in *Proceedings of the 29th Australasian Computer Science Conference*, 2006, pp. 85–94.
- [25] A. Jøsang, J. Diaz, and M. Rifqi, “Cumulative and averaging fusion of beliefs,” *Information Fusion*, vol. 11, no. 2, pp. 192–200, Apr. 2010.
- [26] A. Whitby, A. Jøsang, and J. Indulska, “Filtering out unfair ratings in bayesian reputation systems,” *The ICFAIN Journal of Management Research*, vol. 4, no. 2, pp. 48–64, 2005.
- [27] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, “TRAVOS: Trust and reputation in the context of inaccurate information sources,” *Autonomous Agents and Multi-Agent Systems*, vol. 12, no. 2, pp. 183–189, Mar. 2006.
- [28] C. Burnett, T. J. Norman, and K. Sycara, “Stereotypical trust and bias in dynamic multiagent systems,” *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 2, pp. 26:1–26:22, 2013.
- [29] N. Oren, T. J. Norman, and A. Preece, “Subjective logic and arguing with evidence,” *Artificial Intelligence*, vol. 171, no. 10-15, pp. 838–854, 2007.
- [30] D. Ceolin, A. Nottamkandath, and W. Fokkink, “Subjective logic extensions for the semantic web,” in *URSW*, 2012, pp. 27–38.
- [31] M. Sensoy, G. de Mel, T. Pham, L. Kaplan, and T. J. Norman, “Tribe: Trust revision for information based on evidence,” in *Proceedings of 16th International Conference on Information Fusion*, 2013.

- [32] S. D. Ramchurn, D. Huynh, and N. R. Jennings, “Trust in multi-agent systems,” *The Knowledge Engineering Review*, vol. 19, no. 1, pp. 1–25, Mar. 2004.
- [33] J. Sabater and C. Sierra, “Review on computational trust and reputation models,” *Artificial Intelligence Review*, vol. 24, no. 1, pp. 33–60, Sep. 2005.
- [34] A. Jøsang and R. Ismail, “The beta reputation system,” in *Proceedings of the Fifteenth Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy*, June 2002, pp. 48–64.
- [35] J. Zhang and R. Cohen, “A comprehensive approach for sharing semantic web trust ratings,” *Computational Intelligence*, vol. 23, no. 3, pp. 302–319, 2007.
- [36] J. Zhang, M. Şensoy, and R. Cohen, “A detailed comparison of probabilistic approaches for coping with unfair ratings in trust and reputation systems,” in *Proceedings of the 6th Annual Conference on Privacy, Security and Trust (PTS)*, 2008, pp. 189–200.
- [37] A. Jøsang, T. Ażderska, and S. Marsh, “Trust transitivity and conditional belief reasoning,” in *6th IFIP WG 11.11 International Conference, IFIPTM 2012*, Surat, India, May 2012, pp. 68–83.
- [38] A. Josang and T. Bhuiyan, “Optimal trust network analysis with subjective logic,” in *The Second International Conference on Emerging Security Information, Systems and Technologies*, Aug 2008, pp. 179–184.
- [39] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*. New York: Wiley, 2000, vol. 1.
- [40] A. Jøsang and D. McAnally, “Multiplication and comultiplication of beliefs,” *International Journal of Approximate Reasoning*, vol. 38, no. 1, pp. 19–51, 2004.
- [41] E. J. Sondik, “The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs,” *Operations Research*, vol. 26, no. 2, pp. 282–304, Mar. 1978.
- [42] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [43] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT: YBS, 2011.
- [44] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: John Wiley & Sons, Inc., 2001.

- [45] L. M. Kaplan, M. Sensoy, Y. Tang, S. Chakraborty, C. Bisdikian, and G. de Mel, “Reasoning under uncertainty: Variations of subjective logic deduction,” in *Proc. of the International Conference on Information Fusion*, Istanbul, Turkey, Jul. 2013.
- [46] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall, 1994.
- [47] T. Hoffman, B. Schölkopf, and A. J. Smola, “Kernel methods in machine learning,” *The Annals of Statistics*, vol. 36, no. 3, pp. 1171–1220, Jun. 2008.
- [48] M. Sensoy, G. de Mel, L. Kaplan, T. Pham, and T. J. Norman, “TRIBE: Trust revision for information based on evidence,” in *Proc. of the International Conference on Information Fusion*, Istanbul, Turkey, Jul. 2013.
- [49] T. Muller and P. Schweitzer, “On beta models with trust chains,” vol. 401, pp. 49–65, 2013.
- [50] L. M. Kaplan, S. Chakraborty, and C. Bisdikian, “Subjective logic with uncertain partial observations,” in *Proc. of the International Conference on Information Fusion*, Singapore, Jul. 2012.
- [51] A. Jøsang and F. Sambo, “Inverting conditional opinions in subjective logic,” in *Proceedings of the 20th International Conference on Soft Computing (Mendel 2014)*, Brno, Czech Republic, Jun. 2014.
- [52] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1972.

## Appendix A. Proof of Property 1

### Proof

$$E_{\alpha}[p_k g(\mathbf{p})] = \int_{S_K} p_k g(\mathbf{p}) \frac{\prod_{i=1}^K p_i^{\alpha_i - 1}}{B(\alpha)} d\mathbf{p} \quad (\text{A.1})$$

$$= \frac{B(\alpha + \mathbf{e}_k)}{B(\alpha)} \int_{S_K} g(\mathbf{p}) \frac{\prod_{i=1}^K p_i^{\alpha_i - 1 + \delta_{i,k}}}{B(\alpha + \mathbf{e}_k)} d\mathbf{p} \quad (\text{A.2})$$

$$= \frac{\alpha_k}{S} E_{\alpha + \mathbf{e}_k}[g(\mathbf{p})] \quad (\text{A.3})$$

The jump from (A.2) to (A.3) is the result of expressing the beta function via (3) and exploiting the well known recursion property for Gamma functions, i.e.,  $\Gamma(n + 1) = n\Gamma(n)$  [52].

## Appendix B. Proof of Property 2

**Proof**

$$E_{\boldsymbol{\alpha}}[p_{i_1} p_{i_2} \cdots p_{i_T}] = \int_{S_K} (p_{i_1} p_{i_2} \cdots p_{i_T}) \frac{\prod_{i=1}^K p_i^{\alpha_i - 1}}{B(\boldsymbol{\alpha})} d\mathbf{p} \quad (\text{B.1})$$

$$= \frac{B(\boldsymbol{\alpha} + \mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_T})}{B(\boldsymbol{\alpha})} \quad (\text{B.2})$$

Likewise,

$$E_{\boldsymbol{\alpha}}[p_{i_1} p_{i_2} \cdots p_{i_{T-1}}] = \frac{B(\boldsymbol{\alpha} + \mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_{T-1}})}{B(\boldsymbol{\alpha})}, \quad (\text{B.3})$$

and the ratio of the two terms is

$$\frac{E_{\boldsymbol{\alpha}}[p_{i_1} p_{i_2} \cdots p_{i_T}]}{E_{\boldsymbol{\alpha}}[p_{i_1} p_{i_2} \cdots p_{i_{T-1}}]} = \frac{B(\boldsymbol{\alpha} + \mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_T})}{B(\boldsymbol{\alpha} + \mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_{T-1}})}, \quad (\text{B.4})$$

$$= \frac{\Gamma(\alpha_{i_T} + \delta_{i_1, i_T} + \cdots + \delta_{i_{T-1}, i_T} + 1) \Gamma(S + T - 1)}{\Gamma(\alpha_{i_T} + \delta_{i_1, i_T} + \cdots + \delta_{i_{T-1}, i_T}) \Gamma(S + T)} \quad (\text{B.5})$$

$$= \frac{\alpha_{i_T} + \delta_{i_1, i_T} + \cdots + \delta_{i_{T-1}, i_T}}{S + T - 1}. \quad (\text{B.6})$$

The jump from (B.4) to (B.5) is due to (3), and (B.6) is the recursion property of the Gamma function, i.e.,  $\Gamma(n + 1) = n\Gamma(n - 1)$ .

## Appendix C. Proof of Property 3

**Proof** First, we show that  $S^+$  is upper bounded by  $S + T$ . To this end, define  $S'_k$  as the Dirichlet strength that matches the  $k$ -th marginal of the posterior distribution, which is

$$S'_k = \frac{m_k - v_k}{v_k - m_k^2}, \quad (\text{C.1})$$

so that the noncentral second order moment relates to the mean as

$$v_k = m_k \frac{m_k S'_k + 1}{S'_k + 1}. \quad (\text{C.2})$$

Substitution of (C.2) into (29) yields

$$S^+ = \frac{\sum_{k=1}^K \frac{S'_k}{S'_k + 1} m_k^2 (1 - m_k)^2}{\sum_{k=1}^K \frac{1}{S'_k + 1} m_k^2 (1 - m_k)^2}. \quad (\text{C.3})$$

Because the functions  $S/(S + 1)$  and  $1/(S + 1)$  are monotonically increasing and decreasing, respectively, it is easy to see that  $S^+$  is bounded by

$$\min_k S'_k \leq S^+ \leq \max_k S'_k. \quad (\text{C.4})$$



Now, we show that  $S'_k \leq S + T$  for  $k = 1, \dots, K$ . The posterior (26) is actually a weighted sum of Dirichlets where the Dirichlet strength of each term is  $S + T$ , i.e.,

$$f(\mathbf{p}|\boldsymbol{\alpha}, \mathcal{X}^T) = \sum_{j=1}^N w_j f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}_j^*), \quad (\text{C.5})$$

where the nonnegative weights sum to one and  $\sum_k \alpha_{j,k}^* = S + T$ . Note that the weights are nonnegative because all likelihood values are nonnegative, and the number of terms  $N = K^T$ . Then,

$$m_k = \sum_{j=1}^N w_j m_{j,k}, \quad (\text{C.6})$$

$$v_k = \sum_{j=1}^N w_j m_{j,k} \frac{m_{j,k}(S+T)+1}{S+T+1}, \quad (\text{C.7})$$

where  $m_{j,k}$  is the  $k$ -th marginal mean of the  $j$ -th Dirichlet term. Now,

$$S'_k = \frac{\sum_{j=1}^N w_j m_{j,k} \left(1 - \frac{m_{j,k}(S+T)+1}{S+T+1}\right)}{\sum_{j=1}^N w_j m_{j,k} \frac{m_{j,k}(S+T)+1}{S+T+1} - \left(\sum_{j=1}^N w_j m_{j,k}\right)^2}. \quad (\text{C.8})$$

Because

$$\left(\sum_{j=1}^N w_j m_{j,k}\right)^2 \leq \sum_{j=1}^N w_j m_{j,k}^2, \quad (\text{C.9})$$

then

$$S'_k \leq \frac{\frac{S+T}{S+T+1} \sum_{j=1}^N w_j m_{j,k} (1 - m_{j,k})}{\sum_{j=1}^N w_j m_{j,k} \left(\frac{m_{j,k}(S+T)+1}{S+T+1} - m_{j,k}\right)}, \quad (\text{C.10})$$

which simplifies to

$$S'_k \leq S + T. \quad (\text{C.11})$$

Since this bound holds for all  $k$  and  $S^+$  is bounded by (C.4), then  $S^+ \leq S + T$ , and this proves the corresponding bound for  $u^+$ . Clearly, the upper bound in (C.11) is achieved if all the observations are fully visible. We now show that if the observations are not all fully visible, the bound is not achievable. The upper bound is achieved by  $S'_k$  only if the bound in (C.9) becomes an equality. This occurs only when  $m_{j,k}$  is a constant with respect to  $j$ . Then, each Dirichlet term in (C.5) share the same means and Dirichlet strength, which means they are the same Dirichlet, and the posterior is a Dirichlet with a strength of  $S + T$ . Now lets say that the  $T$ -th observation is not fully visible so that at least two elements  $m$  and  $n$  of  $\mathbf{1}^{(T)}$  are nonzero where  $m \neq n$ . Then one of the Dirichlet terms in (C.5) will have a parameter  $\boldsymbol{\alpha}_{j_1}^* = \boldsymbol{\alpha} + \mathbf{e}_{i_1} + \dots + \mathbf{e}_{i_{T-1}} + \mathbf{e}_m$  due to the  $m$ -th nonzero element in  $\mathbf{1}^{(T)}$ . In this case, another Dirichlet term must exist with

parameter  $\alpha_{j_2}^* = \alpha + \mathbf{e}_{i_1} + \dots + \mathbf{e}_{i_{T-1}} + \mathbf{e}_n$  due to the  $n$ -th nonzero element in  $\mathbf{l}^{(T)}$ . Clearly, the two Dirichlet terms do not represent the same distribution as  $\alpha_{j_1}^* \neq \alpha_{j_2}^*$ . This violates the equality condition in (C.9) and the upper bound in (C.11) is loose.

#### Appendix D. Proof of Property 4

**Proof** Let  $k^* = \arg \min_k l_k$ . First, we show that the updated unregulated Dirichlet strength  $S^+$  is upper bounded by the Dirichlet strength required for  $\alpha_{k^*}^+ = \alpha_{k^*}$ . To this end, inserting the first and second order moments after a single update given by (32) and (33) into (29) leads to (see [23] for a detailed derivation):

$$S'_k = \frac{S+1}{1 + \frac{(S+2)}{\left(\sum_{i=1}^K \alpha_i l_i^{(1)}\right) \left(\frac{1+\bar{\alpha}_k}{l_k} + \frac{1+\alpha_k}{l_k}\right)}}, \quad (\text{D.1})$$

where

$$\bar{\alpha}_k = \sum_{i \neq k} \alpha_i, \quad (\text{D.2a})$$

$$\bar{l}_k = \frac{1}{\bar{\alpha}_k} \sum_{i \neq k} \alpha_i l_i, \quad (\text{D.2b})$$

represent the Dirichlet parameter and average likelihood, respectively, associated to the complement (or disbelief) of the  $k$ -th attribute. Since  $l_k, \bar{l}_k \geq l_{k^*}$ ,

$$S'_k \leq \frac{S+1}{1 + \frac{(l_{k^*})}{\sum_{i=1}^K \alpha_i l_i}}, \quad (\text{D.3})$$

for any  $1 \leq k \leq K$ . Therefore, by (C.4), the right hand side of (D.3) is also an upper bound for  $S^+$ . The bound in (D.3) is the desired bound for the Dirichlet parameter to keep  $\alpha_{k^*}^+ = \alpha_{k^*}$ . To see this, the updated Dirichlet parameter is obtained by multiplying  $S^+$  by (32) to obtain

$$\alpha_{k^*}^+ \leq \alpha_{k^*}. \quad (\text{D.4})$$

This upper bound is achieved only when all  $S'_k$  achieve the upper bound in (D.3). Inspection of (D.1) lead to the fact that this happens only when all likelihood values are equal ( $l_k = l_{k^*}$ ), i.e., a vacuous observation, or when only one element of  $\hat{\mathbf{l}}^{(1)}$  is nonzero (a fully visible observation) so that either  $l_k$  or  $\bar{l}_k$  is zero. The only other possible case for  $\alpha_{k^*}^+$  not to decrease is if the the maximum operator in Algorithm 1 increases the Dirichlet strength so that the  $b_{k^*}$  does not become negative. In other words, the Dirichlet strength is set so that  $\alpha_{k^*}^+ = W a_{k^*}$ . Unless  $\alpha_{k^*} = W a_{k^*}$  the Dirichlet parameter is reduced. In other words,  $b_{k^*}$  must equal to zero for the parameter value not to decrease.