Noname manuscript No.
(will be inserted by the editor)

Reasoning about the Impacts of Information Sharing

Yuqing Tang · Federico Cerutti · Nir Oren · Chatschik Bisdikian

Received: date / Accepted: date

Abstract Shared information can benefit an agent, allowing others to aid it in its goals. However, such information can also harm, for example when malicious agents are aware of these goals, and can then thereby subvert the goal-maker's plans. In this paper we describe a decision process framework allowing an agent to decide what information it should reveal to its neighbours within a communication graph in order to maximise its utility. We assume that these neighbours can pass information onto others within the graph. The inferences made by agents receiving the messages can have a positive or negative impact on the information providing agent, and our decision process seeks to assess how a message should be modified in order to be most beneficial to the information producer. Our decision process is based on the provider's subjective beliefs about others in the system, and therefore makes extensive use of the notion of trust with regards to the likelihood that a message will be passed on by the receiver, and the likelihood that an agent will use the information against the provider. Our core contributions are therefore the construction of a model of information propagation; the description of the agent's decision procedure; and an analysis of some of its properties.

Keywords Information sharing · Impacts · Trust · Risk

1 Introduction

Appropriate decision making by an agent operating within a multi-agent system often requires information from other agents. However, unless the system is fully cooperative, there are typically both costs and benefits to divulging information — while the agent may be able to achieve some goals, others might be able to use this information to their own advantage. An agent must therefore weigh up the risks and benefits that information divulgence will bring when deciding how to act. We use "risk" as an abstract notion of concepts such as damage, cost, negative effects, harms, penalty and so on; and "benefit" as an abstract notion of concepts such as utility, gain, rewards and so on. One of the most critical factors in this calculation is the trust placed in the entity to which one is providing the information — an untrusted individual might pass private information onto others, or may

This paper is dedicated to the memory of Chatschik Bisdikian who recently passed away.

We would like to thank Leon van der Torre for his insightful comments on an early draft of this work.

This research was sponsored by US Army Research laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defence, or the UK Government. The US and UK Governments are authorised to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Yuqing Tang
Carnegie Mellon University, Robotics Institute, 5000 Forbes Ave, Pittsburgh, PA 15213, USA
E-mail: yuqing.tang@cs.cmu.edu

Federico Cerutti
University of Aberdeen, School of Natural and Computing Science, King's College,
AB24 3UE, Aberdeen, UK
E-mail: f.cerutti@abdn.ac.uk

Nir Oren
University of Aberdeen, School of Natural and Computing Science, King's College,
AB24 3UE, Aberdeen, UK
E-mail: n.oren@abdn.ac.uk

Chatschik Bisdikian
IBM Research Division,Thomas J. Watson Research Center,P.O. Box 704,Yorktown Heights, NY 10598, USA
E-mail: bisdik@us.ibm.com

act upon the information in a manner harmful to the information provider.

In this paper we seek to provide a decision mechanism for assessing the positive and negative effects of information release to an agent. Using our mechanism, first discussed in Bisdikian et al (2013) and expanded here, the agent can decide how much information to provide in order to maximise its own utility. We situate our work within the context of a multi-agent system. Here, an agent must assess the risk of divulging information to a set of other agents who may, in turn, further propagate the information. The problem the agent faces is to identify the set of information that should be revealed to its neighbours (who will potentially propagate the information further) in order to maximise its own utility.

Within a multi-agent system, the ability of an agent to assess the risk of information sharing is critical when agents have to reach agreement, for example when coordinating, negotiating or delegating activities. In many contexts, agents have conflicting goals, and inter-agent interactions must take the risk of a hidden agenda into account. Thus, a theory of risk assessment for determining the right level of disclosure to apply to shared information is vital in order to avoid undesirable impacts on an information producer.

As a concrete example, consider the work described in Chakraborty et al (2012), where information from accelerometer data attached to a person can be used to make either *white-listed* inferences — that the person desires others to infer, or *black-listed* inferences — which the person would rather not reveal. For example, the person may wish a doctor to be able to determine how many calories they burn in a day, but might not want others to be able to infer their state (e.g. sitting, running or asleep). The person must thus identify which parts of the accelerometer data should be shared in order to enable or prevent their white- or black-listed inferences. While Chakraborty et al (2012) examined how inferences can be made (e.g. that the sharing of the entropy of FFT coefficients provides a high probability of detecting activity level and low probability of detecting activity type), this work does not consider the *impacts* of sharing such information when it is passed on to others.

In this paper we focus on the case where such black- and white-listed inferences can be made by other agents within a system, and seek to identify what information to provide in order to obtain the best possible outcome for the information provider.

To illustrate such a scenario, let us consider a governmental espionage agency which has successfully placed spies within some hostile country. It must communicate with these spies through a series of handlers, some of which may turn out to be double-agents. It must therefore choose what information to reveal to these handlers in order to maximise the benefits that spying can bring to it, while minimising the damage they can do. It is clear that the choices made by the agency depend on several factors. First, it must consider the amount of trust it places in the individual spies and handlers. Second, it must take into account the amount of harm these can do with any information it provides to them. Finally, it must consider the benefits that can accrue from providing its spies with information. The combination of the first and second factors together provide a measure of the negative effects of information sharing. Now when considering the second factor, an additional detail must be taken into account, namely that the information recipients (i.e. the spies) may already have some knowledge which, when combined with the information provided by the agency, will result in additional unexpected information being inferred. Therefore, the final level of harm which the agency may face depends not on the information it provides, but instead on the undesired inferences which hostile spies can make.

The remainder of this paper is structured as follows. In Section 2 we describe our model, outlining the process of decision making that an agent performs in the presence of white- and black-listed inferences. We concentrate on a special case of communication in multi-agent systems, and show how such a case can be reduced to communication between an information provider and consumer (Section 3). We describe the decision procedure in Section 4. We then contrast our approach with existing work in Section 5, and identify several avenues of future work. Section 6 concludes the paper.

2 The Impacts of Information Sharing

We consider a situation where an information producer shares information with one or more information consumers. These consumers can, in turn, forward the information to others, who may also forward it on, repeating the cycle. Furthermore, since a consumer may or may not use the information provided as expected by the provider, the producer must assess the risk it will incur if the provided information is misused. The decision problem faced by the producer is to therefore identify an appropriate message to send to a consumer which will achieve an appropriate balance between desired and undesired impacts. We assume that once information is shared, the producer is unable to control its spread or use further.

We begin by describing a model of such a system. As part of our notation, we use upper-case letters, e.g. X, to represent random variables (r.v.'s); lower-case letters, e.g. x, to represent realisation instances of them, and $F_X(\cdot)$ and $f_X(\cdot)$ to represent the probability distribution and density of the r.v. X respectively; and write $\Pr(\cdot)$ and $\Pr(\cdot \mid \cdot)$ to represent the probability and conditional probability of discrete random variables respectively.

We consider a set of agents able to interact with their neighbours through a set of communication links, as embodied by a communication graph or network. We introduce the concept of a *Framework for Communication Assessment* FCA_0 with respect to a producer ag_0 — which is willing to share some information — that considers the set of agents, the messages that can be exchanged, the communication links of each agent, and the recipients of the information.

These recipients are either *desired* or *undesired*: according to our assumption above, once a producer shared the information, it can not longer control the sharing process. Intuitively, a *desired* recipient is an agent with which the producer is willing — i.e. the risk–benefit trade-off is in favour of benefit — to share information. On the other hand, an *undesired* recipient is an agent to which the producer is not willing to provide information — the risk–benefit trade-off is in favour of risk.

Moreover, we assume that each agent is aware of its neighbours (i.e., those it can directly communicate with), and that the producer agent ag_0 has probabilistic knowledge regarding the network topology. We do not consider how such probabilistic knowledge is obtained in the current paper, noting that in future work we will exploit the Subjective Logic approach (Jøsang 2001) for modeling such knowledge and appropriate machine learning techniques to estimate the parameters of such models.

Definition 1 A *Framework for Communication Assessment* with respect to the producer ag_0 (FCA_0) is a 4-ple:

$\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m \rangle$

where:

- $\mathcal{A}$ is a set of agents ($ag_0 \in \mathcal{A}$);
- $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$ is the set of communication links among agents;
- $\mathcal{M}$ is the set of all the messages that can be exchanged;
- $m \in \mathcal{M}$ is a *goal message* that the producer ag_0 intends to share with other agents; and the producer will need to decide how to share such a message based on the assessment of its impacts.
- $\mathcal{A} \setminus \{ag_0\}$ is the set of *consumers*.

Example 1 To illustrate our proposal, let refine our example, supposing that British Intelligence (BI) has two spies, James and Alec, in place in France. James is very loyal to Britain but not smart. Alec is smart, but his trustworthiness is highly questionable. BI doest not have direct communication links to James and Alec. Messages from BI will need to be delivered through Bob and Alice via a communication network as in Fig. 1. At some point, BI informs the spies that in three weeks France will be invaded by a European country: it hopes that James and Alec can recruit new agents in France thanks to this information. BI does not want to share the information that the invasion will be started by Germany, because they are the only ones aware of these plans, and a leak would result in endangering the high-value information sources who provided the plans to the British Intelligence. Therefore, British Intelligence has to assess the benefit and risk trade-off in order to determine how to inform its spies that France will be invaded by a European country. Formally, we can represent the above example $FCA_{BI} = \langle \mathcal{A}_{BI}, \mathcal{C}_{BI}, \mathcal{M}_{BI}, ag_{BI}, m_{BI} \rangle$, where:

- $\mathcal{A}_{BI} = \{BI, James, Alec, Bob, Alice\}$;
- $\mathcal{C}_{BI} = \{\langle BI, Bob \rangle, \langle BI, Alice \rangle, \langle Bob, Alec \rangle, \langle Bob, James \rangle, \langle Alice, Alec \rangle\}$ (see Figure 1);
- $\mathcal{M} = \{m_1, m_2\}$ with:
  - m_1: France will be invaded by Germany;
  - m_2: France will be invaded by a European country;
- ag_{BI} is the producer;
- $m_{BI} = m_1$ is the message that BI wants to disseminate;
- $\{James, Alec\}$ are the consumers.

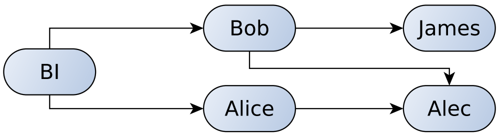

Fig. 1 The communication of the British Intelligence

Given a framework FCA_0, ag_0 will make use of the procedure described in this paper to determine how to share information. This information sharing decision seeks to identify a *policy of disclosures* for the ag_0 with respect to the original messages — the goal of the policy is to reduce the information conveyed in the *delivered messages* to other agents so as to optimize certain criteria, such as maximizing benefits, minimizing risks or optimizing certain trade-offs.

In Example 1, if the agency knows m_1, that France will be invaded by Germany, it may nevertheless share m_2 — France will be invaded by a European country — based on its policy of disclosure in order to reduce the possible harm it may accrue from sharing this message.

Definition 2 Given a set of agents $\mathcal{A}$, a message $m \in \mathcal{M}$, two agents ag_i, $ag_j \in \mathcal{A}$, we define a *disclosure policy* from ag_i to ag_j as follows:

$\boldsymbol{\pi}_{i,j} : \mathcal{M} \times \mathcal{M} \mapsto [0,1]$.

This disclosure policy prescribes that the original message m will be transformed into a new message m' with probability $\boldsymbol{\pi}_{i,j}(m, m') = \Pr_{i,j}(m' \mid m)$. $\boldsymbol{\pi}_{i,j}$ is a conditional probability matrix indexed by $\mathcal{M} \times \mathcal{M}$ (the set of all possible pairs of messages):

$$\boldsymbol{\pi}_{i,j} = \begin{pmatrix} \Pr(m_1 \mid m_1) & \Pr(m_1 \mid m_2) & \cdots & \Pr(m_1 \mid m_M) \\ \Pr(m_2 \mid m_1) & \Pr(m_2 \mid m_2) & \cdots & \Pr(m_2 \mid m_M) \\ \vdots & \vdots & \ddots & \vdots \\ \Pr(m_M \mid m_1) & \Pr(m_M \mid m_2) & \cdots & \Pr(m_M \mid m_M) \end{pmatrix}$$

where $M = |\mathcal{M}|$ is the size of the message space. The kth ($k = 1...M$) column $\boldsymbol{\pi}_{i,j}$, denoted by $\boldsymbol{\pi}_{i,j}(k)$, specifies a distribution over messages to be forwarded when ag_i receives the kth message m_k[1]. Therefore, we require that each column vector $\boldsymbol{\pi}_{i,j}(k)$ satisfies:

$$\sum_{m_l \in \mathcal{M}} \Pr(m_l \mid m_k) = 1.$$

To clarify the concept of disclosure policy, we elaborate two special forms for disclosure policy. A policy is called a *deterministic policy* if there is a single 1 entry in each column of the $\boldsymbol{\pi}$ matrix. A *direct forward policy*, written $\boldsymbol{\pi}^F$ is the identity matrix. $\boldsymbol{\pi}^F$ specifies that the probability of directly forwarding the original message is 1, namely $\Pr(m \mid m) = 1$, and that the probability of modifying the original message is 0, as $\Pr(m' \mid m) = 0$ for any $m \neq m'$. Another special deterministic disclosure policy, denoted by $\boldsymbol{\pi}^k$, is the one by which the agent always sends the same message m_k out no matter what message it receives. All entries in the kth row of such a matrix are set at 1, with the remainder of the matrix elements being 0.

Example 2 Continuing with Example 1, assume that the disclosure policies $\boldsymbol{\pi}_{\text{Bob, James}}$, $\boldsymbol{\pi}_{\text{Bob, Alec}}$, and $\boldsymbol{\pi}_{\text{Alice, Alec}}$ in the communication networks have been estimated as follows.

Bob always directly forwards the message to James with no modification, and therefore has a direct forward policy:

$$\boldsymbol{\pi}_{\text{Bob,James}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

On the other hand, Bob might reduce the information in the messages probabilistically when forwarding messages to Alec:

$$\boldsymbol{\pi}_{\text{Bob,Alec}} = \begin{pmatrix} 0.2 & 0 \\ 0.8 & 1 \end{pmatrix}$$

which specifies that when receiving message m_1, with probability 0.2 the message m_1 will be forwarded but with a higher probability 0.8 the message m_2 of reduced information will be forwarded. When message m_2 is received, m_2 will be forwarded with probability 1.

Finally, Alice also reduces the information in the messages probabilistically when sending messages to Alec:

$$\boldsymbol{\pi}_{\text{Alice,Alec}} = \begin{pmatrix} 0.9 & 0 \\ 0.1 & 1 \end{pmatrix}$$

which specifies that when receiving message m_1, with probability 0.9 the message m_1 will be forwarded while with probability 0.1 message m_2 is forwarded. When message m_2 is received, m_2 will be forwarded with probability 1.

With these preliminaries in place, we now consider the concepts of *degree of disclosure* and *impacts* of information sharing.

2.1 Degree of disclosure

Definition 3 For two agents ag_i, $ag_j \in \mathcal{A}$ and a message $m \in \mathcal{M}$, $x_{i,j} \in [0,1]$ is the *degree of disclosure* by which agent ag_i will send the message m to agent ag_j, where $x_{i,j} = 0$ implies no sharing and $x_{i,j} = 1$ implies full disclosure between the two agents. We define the *disclosure function* as follows:

$$d : \mathcal{M} \times [0,1] \mapsto \mathcal{M}$$

$d(\cdot,\cdot)$ accepts a message and a degree of disclosure (for that message) as its inputs, and returns a modified message (referred to as the disclosed portion of the original message).

Following the scenario in Example 1, m_1: France will be invaded by Germany, and m_2: France will be invaded by a European country. BI decides to disclose messages to Bob with disclosure degree $x_{BI,Bob} = 0.6$ and BI's the disclosure function will output $d(m_1, 0.6) = m_2$. Here m_2 is a derived message from m_1 whose degree of disclosure is 0.6 relative to m_1 — only the portion m_2 of m_1 is disclosed.

In this paper, we consider a special form of disclosure function d which maps a degree of disclosure $x_{i,j}$ (from ag_i to ag_j) to a deterministic disclosure policy $\boldsymbol{\pi}_{i,j}$ as follows:

- $\boldsymbol{\pi}_{i,j}(m' \mid m) = 1$ iff $d(m, x_{i,j}) = m'$, and
- $\boldsymbol{\pi}_{i,j}(m' \mid m) = 0$ otherwise.

Instead of being deterministic, the degree of disclosure $x_{i,j}$ is usually described by a random variable $X_{i,j}$ with probability density $f_{X_{i,j}}$, then

$$\boldsymbol{\pi}_{i,j}(m' \mid m) = \int_0^1 \delta\left(d(m, x_{i,j}) - m'\right) f_{X_{i,j}}(x_{i,j}) dx_{i,j} \tag{1}$$

where $\delta(\cdot)$ is the Dirac delta function (i.e. a distribution is at an infinitely high, infinitely thin spike at the origin while being 0 everywhere). Equation 1 maps a random variable of the degree of disclosure to a probabilistic disclosure policy. We do not further examine the exact mapping and the mathematical properties between the degree of disclosure and the derived message in this work, leaving this as an avenue of future research.

[1] We can treat the possibility that an agent does not forward a message by assuming a special “empty message” within $\mathcal{M}$.

2.2 Impacts

Given a FCA_0, the decision on how to share information with the recipients must consider the impact that the information recipients can have on the producer. For example, many information providers are *selfish*, i.e., will only share information if doing so provides it with some benefit while having minimal negative impact. However, such benefit and damage may be subjective and uncertain. Therefore, when sharing information, the producer not only considers the benefit it obtains, but must also consider potential risks based on the following:

1. the degrees of disclosure (the disclosure policies) of messages exchanged between two agents;
2. the ability of each agent to infer knowledge from the received (disclosed) message;
3. the *impacts* (i.e. risks and benefits) that the inferred knowledge has on the information producer.

Let us notice here that the messages are not fully disclosed in a sense that the disclosure policy transforms a message into another with possibly less information; thereafter only the transformed message is shared. Also the risk and benefit of sharing information is subjective to each agent. However, since we focus on a single producer agent, from a "cautious" perspective we want to investigate the case where each agent in the (probabilistically) known topology will fully disclose the message it receives to its neighbours. This is a worst case scenario, which can thus be ameliorated by refining the model of other agents, for instance following (Burnett et al 2011a).

Definition 4 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m \rangle$, let $ag_i \in \mathcal{A} \setminus \{ag_0\}$, $Z \in \mathbf{Z}$ be a r.v. which represents the *impact* agent ag_0 receives when sharing the message m with a disclosure policy $\boldsymbol{\pi}_{0,i}$ (with a corresponding disclosure degree random variable $X_{0,i}$) for every neighbor ag_i of the producer ag_0. Let $\boldsymbol{\pi}_0 = \langle \boldsymbol{\pi}_{0,i_1}, ..., \boldsymbol{\pi}_{0,i_N} \rangle$ be a list of disclosure policies where each agent ag_{i_l} ($l = 1...N$) is an immediate neighbor of the producer ag_0; and correspondingly $\mathbf{X}_0 = \langle X_{0,i_1}, ..., X_{0,i_N} \rangle$ be a vector of random variables of the disclosure degrees from the producer to his immediate neighbors. $\mathbf{Z}$ is called the *space of impact* which can be described either by

- a continuous random variable Z whose distribution is described by $F_Z(\cdot; \mathbf{x}_0)$ and $f_Z(\cdot; \mathbf{x}_0)$ (where x_0 is an instance of the disclosure degree random variable vector X_0), or
- a discrete random variable Z whose probability distribution is described by $Pr(z \mid m)$ where $z \in \mathbf{Z}$ is an impact.

The central theme of this paper is centered around Definition 4. More specifically, we focus on 1) how to derive the distribution of impacts from disclosure policies; 2) how to evaluate the benefits, risks and trade-offs of the impacts; and 3) how to determine the disclosure policies accordingly.

Example 3 Continuing with Example 2, BI will need to compute two disclosure policies $\boldsymbol{\pi}_{\text{BI, Bob}}$ and $\boldsymbol{\pi}_{\text{BI, Alice}}$ to determine how to forward messages to Bob and Alice respectively. As the framework already specifies the message m_1 to be the goal message which BI wants to disseminate, we just need to compute the first columns $\boldsymbol{\pi}_{\text{BI, Bob}}(1)$ and $\boldsymbol{\pi}_{\text{BI, Alice}}(1)$ of the two policy matrices $\boldsymbol{\pi}_{\text{BI, Bob}}$ and $\boldsymbol{\pi}_{\text{BI, Alice}}$ — BI need to choose between the following among two vectors: $\begin{pmatrix}1\\0\end{pmatrix}$ or $\begin{pmatrix}0\\1\end{pmatrix}$ for column $\boldsymbol{\pi}_{\text{BI, Bob}}(1)$ and column $\boldsymbol{\pi}_{\text{BI, Alice}}(1)$ regarding Bob and Alice respectively. As we don't need to compute the second column[2], without loss of generality, we assume that BI needs to choose between two policy matrices: $\boldsymbol{\pi}^{m_1} = \begin{pmatrix}1 & 1\\0 & 0\end{pmatrix}$ or $\boldsymbol{\pi}^{m_2} = \begin{pmatrix}0 & 0\\1 & 1\end{pmatrix}$. We call such choices *candidate disclosure policies*. Now let's denote the candidate disclosure policies of BI by: $\boldsymbol{\pi}^{m_1}_{\text{BI, Bob}}$, $\boldsymbol{\pi}^{m_1}_{\text{BI, Alice}}$, $\boldsymbol{\pi}^{m_2}_{\text{BI, Bob}}$, and $\boldsymbol{\pi}^{m_2}_{\text{BI, Alice}}$ respectively,

3 Communication Networks

We now turn our attention to communication between agents who must send information via intermediaries. We show that under some special conditions, such communication can be abstracted as direct communication in terms of *equivalent disclosure policies*. In order to show this result, we introduce two operators combining disclosure policies when information is shared in this way. The first operator *discounts* the disclosure policies based on agents within the message path, while the second operator *fuses* information which may have traveled along multiple paths.

3.1 Discount operator

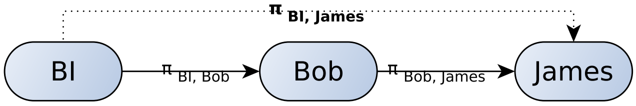

Fig. 2 Combining disclosure policies: single communication path (Definition 5).

Figure 2 depicts the simplest case where the first operator can be applied. Let us suppose that there are three agents

[2] The second column of BI's disclosure policy deals with the case that m_2 is the goal message to be shared. Having the second column in the producer's disclosure policy as a placeholder helps to unify the notion of disclosure policy for both producer and the other agents.

BI, Bob, and James, and a piece of information m is shared by BI with Bob, as per BI's disclosure policy $\boldsymbol{\pi}_{BI,Bob}$ — the message sent to Bob is m' with probability $\boldsymbol{\pi}_{BI,Bob}(m,m')$. Bob then shares m' with James, sending James a message m'' with probability $\boldsymbol{\pi}_{Bob,James}(m',m'')$. To model this message propagation scenario, we define the following discount operator:

Definition 5 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m\rangle$, let $\langle ag_i, ag_k\rangle \in \mathcal{C}$ and $\langle ag_k, ag_j\rangle \in \mathcal{C}$ be a communication path from agent ag_i to ag_j with corresponding disclosure policies $\boldsymbol{\pi}_{i,k}$ and $\boldsymbol{\pi}_{k,j}$ respectively. We define a discount operator $\odot$ which computes the *equivalent disclosure policy* from ag_i to ag_j:

$$\boldsymbol{\pi}_{i,j} = \boldsymbol{\pi}_{i,k} \odot \boldsymbol{\pi}_{k,j}.$$

Due to the mapping between disclosure policies and disclosure degrees (see Equation 1), the discount operator is used to compute an *equivalent degree of disclosure* (a random variable) from ag_i to ag_j which is defined as:

$$X_{i,j} = X_{i,k} \odot X_{k,j}$$

where $X_{i,k}$ and $X_{k,j}$ are the random variables of the degrees of disclosure from ag_i to ag_k and from ag_k to ag_j respectively.

Note that we make the natural assumption that the operator $\odot$ satisfies the associativity:

$$(\mathbf{a}_{i,j} \odot \mathbf{a}_{j,k}) \odot \mathbf{a}_{k,l} = \mathbf{a}_{i,j} \odot (\mathbf{a}_{j,k} \odot \mathbf{a}_{k,l})$$

In this paper, we instantiate the discount operator for our matrix form of disclosure policies through matrix multiplication:

$$\begin{aligned}\boldsymbol{\pi}_{i,j} &= \boldsymbol{\pi}_{i,k} \odot \boldsymbol{\pi}_{k,j} \\ &= \boldsymbol{\pi}_{i,k} \times \boldsymbol{\pi}_{k,j}. \end{aligned} \tag{2}$$

This instantiation of the Discount operator essentially assumes a form of local Markov property — that how the agents apply their disclosure policies only depends on the message they received, and is independent of the messages' transportation history. Effectively, this discount operator models a Markov Chain over message passing.

Example 4 Continuing with Example 2, although there is no direct communication link between BI and James, we can compute the equivalent disclosure policy from BI to James as:

$$\boldsymbol{\pi}_{BI,James} = \boldsymbol{\pi}_{Bob,James} \times \boldsymbol{\pi}_{Bi,Bob}.$$

Recall that $\boldsymbol{\pi}_{\text{Bob,James}} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$. Assume that BI chooses to use a candidate disclosure policy: $\boldsymbol{\pi}^{m_1}_{BI,Bob} = \begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$ (the other candidate disclosure policy is $\boldsymbol{\pi}^{m_2}_{BI,Bob}$), then the equivalent disclosure policy from BI to James can be computed as:

$$\begin{aligned}\boldsymbol{\pi}_{BI,James} &= \boldsymbol{\pi}_{Bob,James} \times \boldsymbol{\pi}_{BI,Bob} \\ &= \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \times \begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix} \\ &= \begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}\end{aligned}$$

Note that this example demonstrates an extreme case of the disclosure policies where all the policies are deterministic.

3.2 Fusion operator

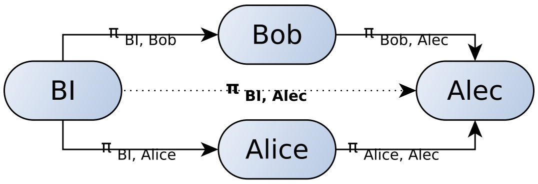

Fig. 3 Combining disclosure policies: multiple communication paths (Definition 6).

Our second operator deals with the case where there are multiple paths that a message can traverse before reaching an information consumer. This is the case depicted in Fig. 3, where BI can send a message along several communication paths in order to increase the likelihood that it reach its destination (Alec). BI shares information with Bob and Alice, but discloses these messages following different policies $\boldsymbol{\pi}_{BI,Bob}$ and $\boldsymbol{\pi}_{BI,Alice}$. Afterwards, both Bob and Alice will disclose the received messages to Alec following their own disclosure policies $\boldsymbol{\pi}_{Bob,Alec}$ and $\boldsymbol{\pi}_{Alice,Alec}$. To model this, we define a fusion operator (in Definition 6 below) which when used with the discount operator defined in Definition 5 above, provides us with an equivalent disclosure policy as if the message was sent directly from BI to Alec ($\boldsymbol{\pi}_{BI,Alec}$).

Definition 6 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m\rangle$, let ag_{k_l} $(l = 1, ...N)$ be an intermediate agent which has direct links between ag_i and ag_j (where N is the number of such intermediate agents): i.e. $\langle ag_i, ag_{k_l}\rangle \in \mathcal{C}$ and $\langle ag_{k_l}, ag_j\rangle \in \mathcal{C}$ is a communication path from agent ag_i to ag_j for every intermediate agent ag_{k_l}. The corresponding policies are denoted by $\boldsymbol{\pi}_{i,k_l}$ and $\boldsymbol{\pi}_{k_l,j}$. We define a *fusion operator* $\oplus$ which computes the *equivalent disclosure policy* from ag_i to ag_j over all alternative communication paths:

$$\boldsymbol{\pi}_{i,j} = \oplus_{l=1}^{N} \left(\boldsymbol{\pi}_{i,k_l} \odot \boldsymbol{\pi}_{k_l,j}\right).$$

By the mapping between disclosure policies and disclosure degrees (see Equation 1), the fusion operator to compute an *equivalent degree of disclosure* from ag_i to ag_j can be defined as:

$$X_{i,j} = \oplus_{l=1}^{N} \left(X_{i,k_l} \odot X_{k_l,j} \right)$$

where X_{i,k_l} and $X_{k_l,j}$ are the random variables of the degrees of disclosure from ag_i to ag_{k_l} and from ag_{k_l} to ag_j respectively.

Again in this paper we only consider the fusion operator for the matrix forms of disclosure policies:

$$\begin{aligned} \boldsymbol{\pi}_{i,j} &= \oplus_{l=1}^{N} \left(\boldsymbol{\pi}_{i,k_l} \odot \boldsymbol{\pi}_{k_l,j} \right) \\ &= \left(\sum_{l=1}^{N} \left(\boldsymbol{\pi}_{i,k_l} \times \boldsymbol{\pi}_{k_l,j} \right) \right) \times C \end{aligned} \qquad (3)$$

where C is $|\mathcal{M}|\times 1$ matrix. Each entry c_u of C is a renormalized constant for the uth column in the equivalent policy matrix $\boldsymbol{\pi}_{i,j}$ to guarantee that every column can still be summed up to 1 to ensure the resulting policy matrix is a column-wise conditional probability matrix. This fusion operator is a special case of the mixing or average operator over Dempster-Shafer probabilities adapted from Sentz and Ferson (2002).

Example 5 Continuing with Example 2, there are two paths from BI to Alec: 1) BI — Bob — Alec, and 2) BI — Alice — Alec. First we discount the disclosure policies along the path BI — Bob — Alec:

$$\boldsymbol{\pi}_{BI,Alec}^{Bob} = \boldsymbol{\pi}_{Bob,Alec} \times \boldsymbol{\pi}_{Bi,Bob}.$$

Second we discount the disclosure policies along the path) BI — Alice — Alec:

$$\boldsymbol{\pi}_{BI,Alec}^{Alice} = \boldsymbol{\pi}_{Alice,Alec} \times \boldsymbol{\pi}_{Bi,Alice}.$$

Then we fuse the two paths:

$$\boldsymbol{\pi}_{BI,Alec} = (\mathbf{x}_{BI,Alec}^{Bob} + \mathbf{x}_{BI,Alec}^{Alice}) \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

where the constant $\frac{1}{2}$ is used to re-normalize the merged probabilities. Note that this choice of $\frac{1}{2}$ as the re-normalizing constant assumes that the two paths have equal influence over the distributions of the delivered message so that the two distributions are averaged together. Other operators (for examples, those from the Dempster-Shafer theory literature Sentz and Ferson (2002)) can be applied with respect to additional background knowledge about the nature of different paths and how they interact.

Now if BI chooses $\boldsymbol{\pi}_{Bi,Bob}^{m_1}$ (to send the message m_1 to Bob) and $\boldsymbol{\pi}_{Bi,Alice}^{m_2}$ (to send m_2 to Alice), we can fuse (estimate) an equivalent disclosure policy from BI to Alec:

$$\begin{aligned} &\boldsymbol{\pi}_{BI,Alec}^{m_1,m_2} \\ &\quad = \left(\boldsymbol{\pi}_{Bob,Alec} \times \boldsymbol{\pi}_{Bi,Bob}^{m_1} + \boldsymbol{\pi}_{Alice,Alec} \times \boldsymbol{\pi}_{Bi,Alice}^{m_2} \right) \\ &\quad\quad \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \\ &\quad = \begin{pmatrix} 0.1 & 0.1 \\ 0.9 & 0.9 \end{pmatrix} \end{aligned}$$

3.3 Equivalent indirect disclosure policy

Having described the discount and fusion operators $\odot$ and $\oplus$, we are ready to compute the equivalent indirect disclosure policy between the producer and any information consumer in the communication network:

Definition 7 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m \rangle$, let $N^I(ag_i) = \{ag_j \mid \langle ag_i, ag_j \rangle \in \mathcal{C}\}$ be the set of incoming agents neighbouring ag_i. We can define an *equivalent indirect disclosure policy* between the producer ag_0 and the information consumer $ag_q \neq ag_0$ as:

$$\boldsymbol{\pi}_{0,q} = \oplus_{ag_k \in N^I(ag_q)} \left(\boldsymbol{\pi}_{0,k} \odot \boldsymbol{\pi}_{k,q} \right) \qquad (4)$$

where $\boldsymbol{\pi}_{0,k}$ is recursively computed using the Equation 4 until a direct communication link is reached (at this point, the predetermined disclosure policy is used in the computation).

Note that the disclosure policies of intermediate agents might be determined following a complex decision making process, such as the trade-offs of benefits and risks we will introduce in Section 4. The disclosure policies of intermediate agents are usually not known to the producer. However, we assume that such policies can be estimated using appropriate methods by considering knowledge or beliefs regarding the intermediate agents.

In this paper, we do not further detail the operators over the communication network. More discussions regarding the appropriateness of these operators over communication networks (or trust networks) can be found in Tang et al (2011).

3.4 Distribution of delivered messages

We now characterize the distribution of messages that can be received by an information consumer. Let $m_k \in \mathcal{M}$ be the kth message the producer ag_0 shares in the first place. In this case, we can represent the distribution of messages from the producer with the kth basis unit vector: $\mathbf{e}^{m_k} = \left[0 \ldots 1 \ldots 0\right]^T$ which is a column vector where only the entry corresponding to message m_k is set to 1 and all other entries are set to 0. Now let agent ag_q be the information consumer. The distribution of messages that ag_q will receive can be computed as following:

- First we compute the equivalent disclosure policy from ag_0 to agent ag_q:

  $$\boldsymbol{\pi}_{0,q} = \oplus_{ag_k \in N^I(ag_q)} \left(\boldsymbol{\pi}_{0,k} \odot \boldsymbol{\pi}_{k,q}\right).$$

- Then we compute the *distribution of delivered messages* for agent ag_q:

  $$\mathbf{x}_{0,q} = \begin{pmatrix} Pr_q(m_1) \\ Pr_q(m_2) \\ \vdots \\ Pr_q(m_M) \end{pmatrix} = \boldsymbol{\pi}_{0,q} \times \mathbf{e}^m$$

As the initial goal message m is specified in a framework $FCA_0\ \langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m\rangle$, the producer's choices in disclosure policies (resp. the degrees of disclosure in the continuous case) will determine the a distribution of delivered messages to a consumer ag_q. Therefore, when $FCA_0\ \langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m\rangle$ is fixed, the following three notations can derive from each other:

- the equivalent disclosure policy from the producer ag_0 to a consumer ag_q, denoted by $\boldsymbol{\pi}_{ag_0,q}$,
- the equivalent random variable of the degree of disclosure from the producer ag_0 to a consumer ag_q, denoted by $X_{ag_0,q}$, and
- the distribution of delivered messages from the producer ag_0 to a consumer ag_q, denoted by $\mathbf{x}_{ag_0,q}$.

Therefore in this paper we use the $\boldsymbol{\pi}_{ag_0,q}$, $x_{ag_0,q}$, and $\mathbf{x}_{ag_0,q}$ interchangeably (especially in the continuous version of our model) when the $FCA_0\ \langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m\rangle$ is fixed.

Example 6 Continuing with Example 5, recall that the framework specifies that the goal message that BI intends to share is m_1. Assume that BI selects in its disclosure policies to send message m_1 to Bob while sending m_2 to Alice. Given that the goal message distribution is a unit basis vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (i.e., intent to share message m_1), we can now compute a distribution over all possible messages delivered to Alec governed by the disclosure policies:

$$\begin{aligned} &\mathbf{x}^{m_1,m_2}_{BI,Alec} \\ &\quad = (\boldsymbol{\pi}_{Bob,Alec} \times \boldsymbol{\pi}^{m_1}_{Bi,Bob} + \boldsymbol{\pi}_{Alice,Alec} \times \boldsymbol{\pi}^{m_2}_{Bi,Alice}) \\ &\qquad \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &\quad = \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix} \end{aligned}$$

Please note that if Alec wants to recover the original message that the producer intends to share from this distribution, he will need to know all the agents' disclosure policies (probabilistically) in the network and he can only recover a distribution of the original messages instead of a deterministic one. How much he can recover the original message can be measured by information theoretical concepts, such as mutual-information and conditional entropy, based on the distribution over all pairs of messages — the message it receives and the original message that the producer intends to share. This will be in the line of our future study.

4 The Decision Process

Having computed the distribution of messages that the consumer agent will receive when a message m is shared through a communication network, we can now turn our attention to the core of the decision process for assessing impact, which is an integration of the estimation of the *inferred knowledge* and the estimation of impacts of the inferred knowledge that the message has on the producer ag_0.

Definition 8 Given a $FCA_0\ \langle \mathcal{A}, \mathcal{C}, \mathcal{M}, ag_0, m_0\rangle$, for a distribution of messages (represented as a random variable x_q over the messages $\mathcal{M}$) that the consumer ag_q can receive, we describe a distribution $\mathbf{y}$ of the amount of knowledge that ag_q can infer from $\mathbf{x}$ as a random variable $y_q = I(x_q) \in \mathbf{Y}$ which is either

- a continuous random variable whose cumulative distribution and density function are $F_{I_q}(\cdot; x_q)$ and $f_{I_q}(\cdot; x_q)$ respectively; or
- a discrete random variable whose distribution is $Pr(y_q \mid m_i)$.

$\mathbf{Y}$ is called the space of *inference*.

Working with discrete random variables, the inference of agent ag_q corresponds to the following *inference matrix*:

$$I_q = \begin{pmatrix} Pr(y_1 \mid m_1) & Pr(y_1 \mid m_2) & \cdots & Pr(y_1 \mid m_M) \\ Pr(y_2 \mid m_1) & Pr(y_2 \mid m_2) & \cdots & Pr(y_2 \mid m_M) \\ \vdots & \vdots & \ddots & \vdots \\ Pr(y_N \mid m_1) & Pr(y_N \mid m_2) & \cdots & Pr(y_N \mid m_M) \end{pmatrix}$$

Each entry $Pr(y_i \mid m_j)$ represents the probability that agent ag_q makes inference $y_i \in \mathbf{Y}$ when receiving message $m_i \in \mathcal{M}$. The ith column of I_q corresponds to the inference distribution that can be made from receiving a message $m_i \in \mathcal{M}$. Note that every column will sum up to 1 as we require that $\Sigma_{j=1}^{N} Pr(b_j \mid m_i) = 1$ for a valid conditional probability. The size of matrix I_q is $N \times M$ where $N = |\mathbf{Y}|$ and $M = |\mathcal{M}|$. As I_q is a valid conditional probability table, there are $(N-1) \times M$ number of independent parameters in I_q.

Example 7 Continuing with Example 1, let the inference space $\mathbf{Y} = \{y_0, y_1\}$ where

- y_0: "France will be invaded by a European country" (equivalent to m_2), and
- y_1: "France will be invaded by Germany" (equivalent to m_1).

BI believes that an information consumer will make such an inference with probability $u(m_1)_q$ and $1 - u(m_1)_q$ respectively if receiving message m_1, and with probability $u(m_2)_q$ and $1 - u(m_2)_q$ respectively if receiving message m_2. Clearly, we have to distinguish between James and Alec's ability to make inferences:

$$I_{James} = \begin{pmatrix} u_{James}(m_1) & u_{James}(m_2) \\ 1 - u_{James}(m_1) & 1 - u_{James}(m_2) \end{pmatrix} = \begin{pmatrix} 0 & 0.8 \\ 1 & 0.2 \end{pmatrix}$$

while

$$I_{Alec} = \begin{pmatrix} u_{Alec}(m_1) & u_{Alec}(m_2) \\ 1 - u_{Alec}(m_1) & 1 - u_{Alec}(m_2) \end{pmatrix} = \begin{pmatrix} 0 & 0.4 \\ 1 & 0.6 \end{pmatrix}.$$

When receiving message m_1: "France will be invaded by Germany", both James and Alec will trivially infer y_1: "France will be invaded by Germany" as they are identical. When receiving message m_2: "France will be invaded by a European country", as James is not clever he will only be able to infer y_1: "France will be invaded by Germany" with a probability 0.2 while with probability 0.8 James will stay with the original information of m_2. On the other hand, as Alec is smart he will be able to make additional inferences to reach y_1 with a higher probability of 0.6 while given the information y_0 (i.e. m_2) with a relatively low probability of 0.4.

4.1 Reasoning about impacts

As we have previously discussed, the provision of information enables a recipient to make inferences which have an impact on the information producer. We capture this impact as a random variable over an impact space $\mathbf{Z}$ (see Definition 4). In Definition 4, our target is to establish a probabilistic link between the disclosure policy (resp. the degree of disclosure and the derived probabilistic distribution of messages) and the impact. However, we have not yet specified how such a probabilistic link can be established. In this section, we establish this link through inference.

Definition 9 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, M, ag_0, m \rangle$, let y_q be a random variable regarding the inference that a consumer ag_q can make when the producer ag_0 disseminates the message m through the communication network. We define the impact of ag_q's inferences on the producer ag_0 as a real random variable $z_q = Z_q(y_q) \in \mathbf{Z}$ (as defined in Definition 4) which is either

- a continuous random variable whose cumulative distribution and density function are $F_{Z_q}(\cdot; y_q)$ and $f_{Z_q}(\cdot; y_q)$ respectively, or
- a discrete random variable whose distribution is $Pr(z_q \mid y_q)$.

The discrete case can be treated in a similar manner to the inference matrix, that is, via the use of an an *impact matrix* which encodes the conditional probabilities of an impact:

$$Z_q = \begin{pmatrix} Pr(z_1 \mid y_1) & Pr(z_1 \mid y_2) & \cdots & Pr(z_1 \mid y_N) \\ Pr(z_2 \mid y_1) & Pr(z_2 \mid y_2) & \cdots & Pr(z_2 \mid y_N) \\ \vdots & \vdots & \ddots & \vdots \\ Pr(z_K \mid y_1) & Pr(z_K \mid y_2) & \cdots & Pr(z_K \mid y_N) \end{pmatrix}$$

Each entry $Pr(z_k \mid y_j)$ represents the probability that agent ag_q can cause impact $z_k \in \mathbf{Z}$ to the producer ag_0 when ag_q can make inference $y_j \in \mathbf{Y}$. The jth column of Z_q corresponds to the impact distribution that can occur if the jth inference y_j is made by the consumer agent. Again, every column will sum up to 1 as $\Sigma_{k=1}^{K} Pr(z_k \mid y_j) = 1$. $|Z_q| = K \times N$ where $N = |\mathbf{Y}|$ and $K = |\mathbf{Z}|$. Z_q has $N \times (K-1)$ independent parameters. With an impact matrix Z_q, we layout the corresponding impact evaluation into a vector $\mathbf{z}_q$, defined as follows:

$$\mathbf{z}_q = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{pmatrix}$$

where $K = |\mathbf{Z}|$. Entry z_k in $\mathbf{z}_q$ is the k-th impact that agent ag_q can make to the producer. $\mathbf{z}_q$ is called the *impact vector* (an impact distribution) on the producer ag_0 by agent ag_q.

We concentrate on two types of impact, namely the benefits and risks of the inferences made by the consumer on the producer. We respectively refer to these as the *benefits* and *risks* to the producer.

- Benefit $\mathbf{B}$: Let $b_q \in \mathbf{B}$ be the producer ag_0's evaluation of the benefit of inferences a consumer ag_q can make following the receipt of a message. Following Definition 9, we model benefit via either a continuous random variable $B_q(y_q)$ with cumulative distribution and density function $F_{B_q}(\cdot; y_q)$ and $f_{B_q}(\cdot; y_q)$ respectively; or a discrete random variable whose distribution is $Pr(b_q \mid y_q)$.
- Risk $\mathbf{R}$: Let $r_q \in \mathbf{R}$ be the producer ag_0's evaluation of the risk of the harm of inferences a consumer ag_q can make following the receipt of a message. Following Definition 9, we model risk via either a continuous random variable $R_q(y_q)$ with cumulative distribution and density function $F_{R_q}(\cdot; y_q)$ and $f_{R_q}(\cdot; y_q)$ respectively; or a discrete random variable whose distribution is $Pr(r_q \mid y_q)$.

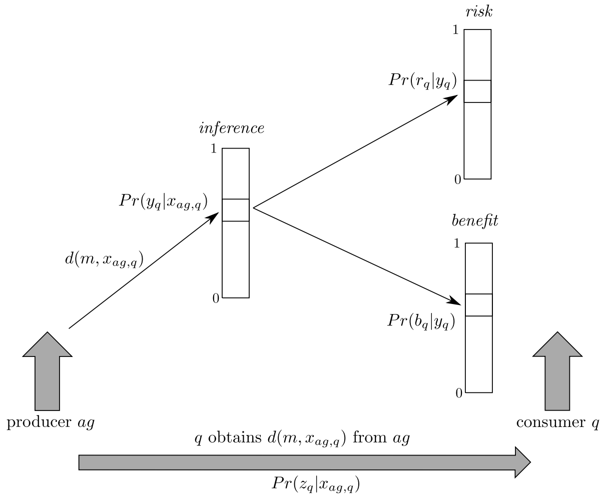

Fig. 4 The probabilities of inference and trust following sharing.

For notational clarity, we explicitly list benefit vector $\mathbf{b}_q$ and risk vector $\mathbf{r}_q$ respectively as follows.

$$\mathbf{b}_q = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{K^B} \end{pmatrix}$$

where $K^B = |\mathbf{B}|$. Entry b_k in $\mathbf{b}_q$ is the k-th benefit the producer can obtain regarding agent ag_q.

$$\mathbf{r}_q = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{K^R} \end{pmatrix}$$

where $K^R = |\mathbf{R}|$. Entry r_k is the k-th risk the producer is concerned with regarding agent ag_q.

Figure 4 provides a graphical interpretation of *inference* (Definition 8) and *impact* (Definition 9), distinguishing between risk and benefit, when a producer ag_0 communicates a message m according to his disclosure policies $\boldsymbol{\pi}_{0,j}$ to a neighbor ag_j of his. Given the derived disclosure policy $\boldsymbol{\pi}_{0,q}$ towards the final consumer in the communication network, the consumer might infer information y_q drawn from all possible inferences. This results in an impact z_q drawn from the space of possible impacts, which is conditioned on the inference y_q made by the consumer. This impact z_q can be either a risk (r_q) or a benefit (b_q).

Example 8 Continuing with Example 7, for each of the possible inferences there are two impacts: the risk and the benefit. We consider two levels of risk, represented in a vector as $\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$. We associate a utility cost with these two outcomes: $r_1 = 20K$ (20000) and $r_2 = 100K$ (100000) respectively, as depicted in the Figure 4. In other words, that the risk r_1 is $20K$ represents the situation that James or Alec gets captured when they are trying to recruit new agents. On the other hand, that the impact r_2 is $100K$ represents that James or Alec leaks that the country being invaded is Germany leading to the loss of a high value information sources in Germany for the UK government. Let us characterize the risk impact probability based on the inference obtained with the following parameters: $w(y_0)_q = Pr_q(r_1 \mid y_0)$ and $w(y_1)_q = Pr_q(r_1 \mid y_1)$. Applying the matrix formalism, the risk impact distribution matrices of James and Alec are as follows.

$$\begin{aligned} Z^R_{James} &= \begin{pmatrix} w(y_0)_{James} & w(y_1)_{James} \\ 1 - w(y_0)_{James} & 1 - w(y_1)_{James} \end{pmatrix} \\ &= \begin{pmatrix} 0.9 & 0.8 \\ 0.1 & 0.2 \end{pmatrix}. \end{aligned}$$

Z^R_{James} means that when the risk materializes, if y_0: “France will be invaded by a European country” is inferred, with a high probability 0.9 it will result in a low cost $20K$ outcome; while with a low probability 0.1 it will result in the high cost $100K$ outcome. This is the case because the information in the inference y_0 rarely causes the high cost. On the other hand, if y_1: “France will be invaded by Germany”is inferred, because James is loyal, he won’t provide the enemy with full information even he is captured. Therefore, in this case with a probability of 0.8, the impact will have a cost of $20K$, while the likelihood of a $100K$ cost outcome is 0.2.

Similarly for Alec, we have

$$\begin{aligned} Z^R_{Alec} &= \begin{pmatrix} w(y_0)_{Alec} & w(y_1)_{Alec} \\ 1 - w(y_0)_{Alec} & 1 - w(y_1)_{Alec} \end{pmatrix} \\ &= \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \end{aligned}$$

Z^R_{Alec} describes the case where when y_0 — “France will be invaded by a European country” is inferred, it will result in the $20K$ outcome with a likelihood of 0.9, and in a $100K$ cost outcome with a likelihood of 0.1. On the other hand, if y_1 (“France will be invaded by Germany”) is inferred, Alec’s loyalty results in a $100K$ cost with a likelihood of 0.9, and the low $20K$ cost with a likelihood of only 0.1.

Benefit is treated in a similar manner. We consider two levels, represented in a vector as $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$. We associate two benefit levels, $b_1 = 10K$, and $b_2 = 120K$, as depicted in the Figure 4. In our running example the benefit impact b_1 of $10K$ could represent the situation that either James or Alec recruit some spies but these spies are not as valuable as expected. On the other hand, that the benefit impact b_2 of $120K$ represents the case where James or Alec recruits a number of spies that are highly valuable. Returning to our matrix representation, the benefit impact distribution matrices of James and Alec are as follows.

$$Z^B_{James} = \begin{pmatrix} 0.6 & 0.5 \\ 0.4 & 0.5 \end{pmatrix}.$$

Z^B_{James} means that when the benefit materializes, if y_0: "France will be invaded by a European country" is inferred, then we will obtain a benefit of $10K$ with a likelihood of 0.6, while the benefit of $120K$ will occur with a likelihood of 0.4. Alternatively, if James infers y_1, the likelihood of obtaining benefit $10K$ or $120K$ both become 0.5.

Similarly for Alec, we have

$$Z^B_{Alec} = \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.8 \end{pmatrix}$$

Meaning that inference y_0 results in a equal chance of the two benefits materialising, while inference y_1 makes the $120K$ outcome more likely (0.8) than the $10K$ outcome (0.2).

In Bisdikian et al (2013) we show that several interesting properties hold in the case of continuous r.v.'s. These properties can be found in Appendix A. However, in this paper we concentrate on discrete random variables and the properties obtained in such a model.

The random variable B_q implicitly captures an aspect of the producer's trust in the consumer, as it reflects the former's belief that the consumer will utilise the information in the manner it desires. Similarly, the random variable R_q captures the notion of *distrust* in the consumer, describing the belief that the consumer will utilise the information in a harmful manner. Note that when considering repeated interactions, these random variables will evolve as the producer gathers experience with various consumers. The evolution of these random variables can be captured as an incremental process of estimating the parameters of the probabilistic models behind these random variables. The probabilistic estimation of the disclosure policy matrix, the inference matrix, the benefit and risk impact matrices can be viewed as special case of such kinds of random variable evolution. In the current work we assume that a steady state has been reached with regards to the parameter estimates, allowing us to ignore the problem of updating the distribution.

Corollary 1 *Assume that the impact z is independent of the message m given the inferred information y. Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, ag_0, m\rangle$, $\mathbf{x}_{0,q}$ (the delivered message distribution), I_q (the inference matrix) and Z_q (the impact matrix) of agent ag_q, the distribution of the impact $\widetilde{\mathbf{z}}_q$ that agent ag_q can have on the producer ag_0 can be computed as follows.*

$$\widetilde{\mathbf{z}}_{0,q} = Z_q \times I_q \times \mathbf{x}_{0,q}$$

where

$$\widetilde{\mathbf{z}}_{0,q} = \begin{pmatrix} Pr(z_1) \\ Pr(z_2) \\ \vdots \\ Pr(z_K) \end{pmatrix}$$

Here, entry $Pr(z_k)$ is the probability with which agent ag_q causes the kth impact to ag_0. $Pr(z_k)$ is the marginal probability over all possible messages and inferences. $\widetilde{\mathbf{z}}_{0,q}$ is called the the impact distribution vector *of a consumer ag_q incurred by the producer ag_0.*

Example 9 Continuing with our running example, we can now combine the estimation of delivered message distribution, the inference distribution and impact distribution together to establish a distribution over all possible impacts. Assume that BI releases message m_1 to Bob and message m_2 to Alice, then the distributions of delivered messages to James and Alec are:

$$\mathbf{x}^{m_1,m_2}_{0,James} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathbf{x}^{m_1,m_2}_{0,Alec} = \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix}$$

We can compute the risk impact distributions for James and Alec as:

$$\begin{aligned} \widetilde{\mathbf{z}}^R_{James} &= Z^R_{James} \times I_{James} \times \mathbf{x}^{m_1,m_2}_{0,James} \\ &= \begin{pmatrix} 0.9 & 0.8 \\ 0.1 & 0.2 \end{pmatrix} \times \begin{pmatrix} 0 & 0.8 \\ 1 & 0.2 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} \\ \widetilde{\mathbf{z}}^R_{Alec} &= Z^R_{Alec} \times I_{Alec} \times \mathbf{x}^{m_1,m_2}_{0,Alec} \\ &= \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \times \begin{pmatrix} 0 & 0.4 \\ 1 & 0.6 \end{pmatrix} \times \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix} = \begin{pmatrix} 0.39 \\ 0.61 \end{pmatrix} \end{aligned}$$

We can compute the benefit impact distributions for James and Alec as:

$$\begin{aligned} \widetilde{\mathbf{z}}^B_{James} &= Z^B_{James} \times I_{James} \times \mathbf{x}^{m_1,m_2}_{0,James} \\ &= \begin{pmatrix} 0.6 & 0.5 \\ 0.4 & 0.5 \end{pmatrix} \times \begin{pmatrix} 0 & 0.8 \\ 1 & 0.2 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \\ \widetilde{\mathbf{z}}^B_{Alec} &= Z^B_{Alec} \times I_{Alec} \times \mathbf{x}^{m_1,m_2}_{0,Alec} \\ &= \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.8 \end{pmatrix} \times \begin{pmatrix} 0 & 0.4 \\ 1 & 0.6 \end{pmatrix} \times \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix} = \begin{pmatrix} 0.31 \\ 0.69 \end{pmatrix} \end{aligned}$$

Definition 10 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, ag_0, m\rangle$, $\widetilde{\mathbf{z}}_q$ (the impact probability distribution vector) and $\mathbf{z}_q$ (the impact vector) regarding agent ag_q. The *expected impact* regarding agent ag_q is

$$\mathbb{E}\{Z_q\} = \mathbf{z}^T_q \times \widetilde{\mathbf{z}}_q.$$

Since the impact can be either a benefit or a risk, Z_q can be specialised in Z^B_q (*benefits probability matrix*) or Z^R_q (*risk probability matrix*). Correspondingly, the distribution of impact can be either a benefit distribution or a risk distribution, $\widetilde{\mathbf{b}}_q$ (*benefits distribution vector*) or $\widetilde{\mathbf{r}}_q$ (*risk distribution vector*); the impact vector can be either a benefit vector $\mathbf{b}_q$ or a risk vector $\mathbf{r}_q$; the expected impact $\mathbb{E}\{z_q\}$ can either be the expected benefit $\mathbb{E}\{B_q\}$ or the expected risk $\mathbb{E}\{R_q\}$.

Corollary 2 *Assume that the impact z is independent of the message m given the inferred information y. For agent ag_q, given a delivered message distribution $\mathbf{x}_{0,q}$; the inference matrix I_q; the benefit impact matrix Z_q^B; the risk impact matrix Z_q^R; the benefit vector $\mathbf{b}_q$; and the risk vector $\mathbf{r}_q$. The expected benefit $\mathbb{E}\{B_q\}$ and the expected risk $\mathbb{E}\{R_q\}$ that agent ag_q has towards ag_0 is be computed as follows.*

$$\mathbb{E}\{B_q\} = \mathbf{b}_q^T \times Z_q^B \times I_q \times \mathbf{x}_{0,q} \tag{5}$$

$$\mathbb{E}\{R_q\} = \mathbf{r}_q^T \times Z_q^R \times I_q \times \mathbf{x}_{0,q} \tag{6}$$

If we assume that benefit and risk are comparable (as done in decision theory when expressed as utilities), we can now define the net benefit of sharing information as follows.

Definition 11 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, ag_0, m\rangle$, the *net benefit* for the producer to share information through the communication network $\mathcal{C}$ with $ag_q \in \mathcal{A} \setminus \{ag_0\}$ is described by $C_q = B_q - R_q$. The *expected net benefit* is defined as $\mathbb{E}\{C_q\} = \mathbb{E}\{B_q - R_q\}$.

Corollary 3 *Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, ag_0, m\rangle$. Assume that the impact z is independent of the message m given the inferred information y. For an agent ag_q, given the delivered message distribution $\mathbf{x}_{0,q}$; the inference matrix I_q; the benefit impact matrix Z_q^B; the risk impact matrix Z_q^R; the benefit vector $\mathbf{b}_q$; and the risk vector $\mathbf{r}_q$. The expected net benefit $\mathbb{E}\{C_q\}$ that agent ag_q can provide to the producer ag_0 can be computed as follows:*

$$\mathbb{E}\{C_q\} = \left(\mathbf{b}_q^T \times Z_q^B - \mathbf{r}_q^T \times Z_q^R\right) \times I_q \times \mathbf{x}_{0,q} \tag{7}$$

Assume that there is a bijection between the spaces of benefit impact and risk impact, namely $B_q(y_q) = f(R_q(y_q))$ where f is a bijection. Further assume that $B_q(y_q)$ and $f(R_q(y_q))$ have the same distribution after the mapping f represented by the matrix Z_q. Then the expected net benefit can be simplified to be as follows.

$$\mathbb{E}\{C_q\} = \left(\mathbf{b}_q^T - \mathbf{r}_q^T\right) \times Z_q \times I_q \times \mathbf{x}_{0,q}.$$

Definition 12 Given a FCA_0 $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, ag_0, m\rangle$, the *global net benefit* is

$$C = \sum_{ag_j \in \mathcal{A}\setminus\{ag_0\}} C_j.$$

By the linearity of expectation, the *expected global net benefit* $\mathbf{E}\{C\}$ is

$$\mathbf{E}\{C\} = \sum_{ag_j \in \mathcal{A}\setminus\{ag_0\}} \mathbf{E}\{C_j\}.$$

Example 10 Continuing with Example 9, we again assume that BI delivers message m_1 to Bob and m_2 to Alice, we can then compute the expected risk, expected benefit, and the expected net benefit obtained by James and Alec receiving the messages as follows.

$$\begin{aligned}
\mathbf{E}\{R_{James}\} &= \mathbf{r}^T \times Z^R_{James} \times I_{James} \times x_{0,James} \\
&= 36K \\
\mathbf{E}\{R_{Alec}\} &= \mathbf{r}^T \times Z^R_{Alec} \times I_{Alec} \times x_{0,Alec} \\
&= 68.96K \\
\mathbf{E}\{B_{James}\} &= \mathbf{b}^T \times Z^R_{James} \times I_{James} \times x_{0,James} \\
&= 65K \\
\mathbf{E}\{B_{Alec}\} &= \mathbf{b}^T \times Z^R_{Alec} \times I_{Alec} \times x_{0,Alec} \\
&= 86.12K \\
\mathbf{E}\{C_{James}\} &= (\mathbf{b}^T \times Z^R_{James} - \mathbf{r}^T \times Z^R_{James}) \\
&\quad \times I_{James} \times x_{0,James} \\
&= 29K \\
\mathbf{E}\{C_{Alec}\} &= (\mathbf{b}^T \times Z^R_{Alec} - \mathbf{r}^T \times Z^R_{Alec}) \\
&\quad \times I_{Alec} \times x_{0,Alec} \\
&= 17.16K
\end{aligned}$$

To simplify the scenario, we ignore the benefit and risk of disclosing messages to Bob and Alice. The expected global net benefit is then as follows.

$$\mathbf{E}\{C\} = 29K + 17.16K = 46.16K.$$

4.2 Decision making

Given a FCA $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m\rangle$, we now have an integrated model which relates message sharing, disclosure policies, communication network, inference estimation, impact estimation to compute the expected global net benefits. The decision making problem can now be posed as requiring the computation, for the producer ag_0, a list of disclosure policies, one for each of its neighbors such that they can lead to the maximum global net benefit.

To formalize the decision making problem, let us start with the disclosure policies. Let

$$N^O(ag_0) = \{ag_k \mid \langle ag_0, ag_k\rangle \in \mathcal{C}\}.$$

be the set of *outgoing neighbors* of the producer ag_0. Let

$$\Pi_{0,k} = \{\boldsymbol{\pi}_{0,k}\}$$

be the set of all possible disclosure policies that the producer can use to deliver message to its neighbor $ag_k \in N^O(ag_0)$. The space of all possible policies of the producer is then defined as

$$\Pi_0 = \prod_{ag_k \in N^o(ag_0)} \Pi_{0,k}.$$

Therefore a choice $\boldsymbol{\pi}_0 \in \Pi_0$ of the producer's disclosure policies to its neighbors can be captured by the vector

$$\boldsymbol{\pi}_0 = \langle \boldsymbol{\pi}_{0,j_1}, \boldsymbol{\pi}_{0,j_2}, ..., \boldsymbol{\pi}_{0,j_N} \rangle$$

where $ag_{j_i} \in N^O(ag_0)$ is a neighbor of producer in the communication network. As discussed in Section 3, every choice of a disclosure policy vector $\boldsymbol{\pi}_0^k \in \Pi_0$ will result in an equivalent indirect disclosure policy for each agent $ag_q \in \mathcal{A} \setminus \{ag_0\}$, we denote such an equivalent indirect policy corresponding to the policy vector $\boldsymbol{\pi}_0^k$ as $\boldsymbol{\pi}_{0,q}^k$.

Finally, given a *FCA* $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m \rangle$, we can formalize the decision problem as an optimization problem on the expected global net benefit:

$$\begin{aligned}\boldsymbol{\pi}_0^* &= \arg\max_{\boldsymbol{\pi}_0^k \in \Pi_0} \mathbf{E}\{C\} \\ &= \arg\max_{\boldsymbol{\pi}_0^k \in \Pi_0} \sum_{ag_q \in \mathcal{A} \setminus \{ag_0\}} \mathbf{E}\{C_q\} \\ &= \arg\max_{\boldsymbol{\pi}_0^k \in \Pi_0} \sum_{ag_q \in \mathcal{A} \setminus \{ag_0\}} \Big(\big(\mathbf{b}_q^T \times Z_q^B - \mathbf{r}_q^T \times Z_q^R \big) \\ &\qquad \times I_q \times \boldsymbol{\pi}_{ag_0,q}^k \times \mathbf{e}^m \Big) \qquad (8)\end{aligned}$$

where $\mathbf{e}^m$ is a unit basis vector which represents a message distribution where the probability of the message m is 1 and the probabilities of all other messages are 0s (see Section 3.4). m is the message that the producer intends to share as specified in the framework *FCA* $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m \rangle$.

Additional trade-offs can be introduced to further tune the decision making. For example, we can require that there are no agents in the network that can cause the producer a negative net-benefit. However, these trade-offs must be introduced carefully so that the resultant optimization problem can still be solved or approximated efficiently. Regarding computational efficiency, the decision making problem as described above can involve the search of an exponential growth joint disclosure policy space of all the neighbors of the producer if we only allow the producer to employ deterministic disclosure policies. If we allow probabilistic disclosure policies for the producer, then numerical techniques (such as gradient ascent and other appropriate numerical optimization techniques) can applied to search the joint policy space more efficiently.

Example 11 Continuing with Example 10, now we can consider all possible messages that BI can deliver to Bob and Alice. Assume that all BI's policies are deterministic, and BI only needs to worry about how to deliver message m_1: France will be invaded by Germany. As there are only two choices in our example: either the original message m_1 or the message m_2 with reduced information, we can, without loss of generality for this problem constrain the space of all possible policies for BI on Bob or on Alice to be:

$$\Pi_{BI,Bob} = \Pi_{BI,Alice} = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \right\}$$

where the first policy always forward message m_1, denoted by $\boldsymbol{\pi}^{m_1}$, while the second policy always forward message m_2, denoted by $\boldsymbol{\pi}^{m_2}$. The space of all disclosure policies of BI is then

$$\Pi_{BI} = \Pi_{BI,Bob} \times \Pi_{BI,Alice}$$

which contains 4 elements corresponding to the 4 possible choices of the message combinations to forward to Bob and Alice. We list the corresponding expected global net benefits as follows:

$$\begin{aligned}&\mathbf{E}_{\boldsymbol{\pi}^{m_1}_{BI,Bob}, \boldsymbol{\pi}^{m_1}_{BI,Alice}}\{C\} = 40.58K \\ &\mathbf{E}_{\boldsymbol{\pi}^{m_1}_{BI,Bob}, \boldsymbol{\pi}^{m_2}_{BI,Alice}}\{C\} = 46.16K \\ &\mathbf{E}_{\boldsymbol{\pi}^{m_2}_{BI,Bob}, \boldsymbol{\pi}^{m_1}_{BI,Alice}}\{C\} = 39.42K \\ &\mathbf{E}_{\boldsymbol{\pi}^{m_2}_{BI,Bob}, \boldsymbol{\pi}^{m_2}_{BI,Alice}}\{C\} = 45K.\end{aligned}$$

As a result,

$$\begin{aligned}C^* &= \max_{\boldsymbol{\pi}_0^k \in \Pi_0} \mathbf{E}\{C\} = 46.16K \\ \boldsymbol{\pi}_0^* &= \arg\max_{\boldsymbol{\pi}_0^k \in \Pi_0} \mathbf{E}\{C\} = \langle \boldsymbol{\pi}^{m_1}_{BI,Bob}, \boldsymbol{\pi}^{m_2}_{BI,Alice} \rangle\end{aligned}$$

meaning that sending message m_1 to Bob and sending message m_2 to Alice is the disclosure policy which achieve the highest expected global net benefit $46.16K$ — a best trade-off criteria on benefits and risks in an average sense.

5 Discussion and Future Work

The work described in this paper makes use of a trust model implicitly captured via disclosure policies as the core input to the decision making process. Our probabilistic underpinnings are intended to be sufficiently general to enable it to be instantiated with arbitrary models, such as Jøsang and Ismail (2002); Teacy et al (2006). Unlike these models, our work is not intended to compute a specific trust value based on some set of interactions, but rather to decide how to use the trust value output by the models.

The use of trust within a decision making system is now a prominent research topic; the interested reader is referred to Castelfranchi and Falcone (2010); Urbano et al (2013) for an overview. However, most work in this area assumes that agents will interact with some most trusted party, as determined by the trust model. This assumption reflects the basis of trust models on action and task delegation rather than information sharing. Burnett et al (2011b) is an exception to this trend; while still considering tasks, Burnett explicitly takes into account the fact that dealing with a trusted party may be more expensive, and thus leads to a lower utility when a task has relatively low potential harmful effects. Burnett's model therefore considers both risk and reward when selecting agents for interaction. However, Burnett situated his work using utility theory, while the present work allows for a more complex impact space to be used.

Another body of work relevant to this paper revolves around information leakage. Work such as Mardziel et al (2011) considers what information should be revealed to an agent given that this agent should not be able to make specific inferences. Unlike our work, Mardziel et al (2011) does not consider the potential benefits associated with revealing information.

Finally, there is a broad field of research devoted to assessing risk in different contexts. As summarised in Wang and Williams (2011), which compares seven definitions of trust[3], the notion of risk is the result of some combination of uncertainty about some outcome, and a (negative) payoff for an intelligent agent and his goals. While this definition is widely accepted (with minor distinctions), different authors have different point of view when it comes to formally define what is meant by *uncertainty*. In Kaplan and Garrick (1981), instead of providing a formal definition of risk, the authors introduce a scenario-based risk analysis method, considering (i) the *scenario*, (ii) its *likelihood*, and (iii) the *consequences* of that scenario. They also introduce the notion of *uncertainty* in the definition of likelihood and of consequences. Doing so allows them to address the core problem of such models, viz. that complete information of all possible scenarios is required. The connection between risk and trust has been the subject of several studies, e.g. Tan and Thoen (2002) shows a formal model based on epistemic logic for dealing with trust in electronic commerce where the risk evaluation is one of the components that contribute to the overall trust evaluation, Das and Teng (2004) proposes a conceptual framework showing the strict correspondence between risk and some definition of trust, Castelfranchi and Falcone (2010) discusses the connection between risk and trust in delegation. However, to our knowledge our work is the first attempt to consider risk assessment in trust-based decision making about information sharing.

There are several potential avenues for future work. First, we have assumed that trust acts as an input to our decision process, and have therefore not considered the interplay between risk and trust. We therefore seek to investigate how both these quantities evolve over time. To this end, we will investigate the connections between the our approach and those based on game theory, such as Goffman (1970). Another aspect of work we intend to examine is how the trust process affects disclosure decisions by intermediate agents with regards to the information they receive. We note that agents might not propagate information from an untrusted source onwards, as they might not believe it. Such work, together with a more fine grained representation of the agents' internal beliefs could lead to interesting behaviours such as agents lying to each other Caminada (2009). Other scenarios of interest can be easily envisaged, and they will be investigated in future work. For instance, a slightly modified version of the framework proposed in this paper can be used for determining the disclosure policies in order to be reasonably sure that a desired part of the message will actually reach a specific agent with which we do not know how to communicate. This is the situation when an organisation tries to reach an undercover agent by sharing some information with the enemy, hoping that somehow the relevant pieces of information will eventually reach the agent. Our long term goal is to utilise our approach to identify which message to introduce so as to maximise agent utility, given a knowledge rich (but potentially incomplete or uncertain) representation of a multi-agent system.

6 Conclusions

In this paper we described a framework enabling an agent to determine how much information it should disclose to others in order to maximise its utility. This framework assumes that any disclosure could be propagated onwards by the receiving agents, and that certain agents should not be allowed to infer some information, while it is desirable that others do make inferences from the propagated information. We showed that our framework respects certain intuitions with regards to the disclosure policies used by an agent, and also identified how an information provider should disclose information in order to achieve some form of equilibrium with regards to its benefits and risks. Potential applications can be envisaged in strategic contexts, where pieces of information are shared across several partners which can result in the achievement of a hidden agenda.

It is clear that from a computational complexity point of view, using algorithms derived directly from our formalism without optimization, our approach lies in $O(M^3 \times N^2)$ (where M is the upper bound of the size of the message space, inference space, and impact space and $O(M^3)$ comes from the complexity applying the naive matrix multiplication algorithm; N is the number of agents and $O(N^2)$ comes from the complexity of computing the equivalent disclosure

[3] Although not considered in Wang and Williams (2011), the definition provided in Castelfranchi and Falcone (2010) follows the others.

policies using Equation 4 by evaluating all equivalent disclosure policy $\boldsymbol{\pi}_{0,k}$ for every agent ag_k through dynamic programming) in the worst case in evaluating the net benefit when the disclosure policies are determined. However, we believe that in real world domains, our matrices will often be sparse, and that our approach will therefore scale better in the average case. As for the search over the space of disclosure policies for the producer, if we only allows deterministic policies, the worst-case complexity will be in $O(M^N)$ (exponential in the number of neighboring agents of the producer) using algorithms directly derived from our formalism; when probabilistic disclosure policies are allowed, we can apply numerical techniques to approximate the solutions. Also if we generalize the fusion operator over arbitrary paths (not requiring that the paths have the same starting agent and the same ending agent), and we require that the fusion operator distributes over the discount operator. With these constraints in place, we can further derive the property that the maximization operator of net benefit can distribute over communication paths. With these additional requirement, we will be able to search the disclosure policy space for an optimal net-benefit with complexity linear in the number of neighboring agents. In this paper we focus on the fundamental formalisms; we intend, as future work, to further pursue more efficient algorithms and evaluate our system on a real world coalition information sharing task in order to determine both its performance and accuracy. However, obtaining a sufficiently large data set for this evaluation is proving challenging.

To our knowledge, this work is the first to take trust and risk into account when reasoning about information sharing, and we are pursuing several exciting avenues of future work in order to make the framework more applicable to a larger class of situations.

References

Bisdikian C, Tang Y, Cerutti F, Oren N (2013) A framework for using trust to assess risk in information sharing. In: ChesÃśevar C, Onaindia E, Ossowski S, Vouros G (eds) Agreement Technologies, Lecture Notes in Computer Science, vol 8068, Springer Berlin Heidelberg, pp 135–149, DOI 10.1007/978-3-642-39860-5_11

Burnett C, Norman TJ, Sycara K (2011a) Trust decision-making in multi-agent systems. In: Proceeding IJCAI'11 Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume One, AAAI Press, pp 115–120, DOI 10.5591/978-1-57735-516-8/IJCAI11-031

Burnett C, Norman TJ, Sycara K (2011b) Trust decision-making in multi-agent systems. In: Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume One, AAAI Press, IJCAI'11, pp 115–120

Caminada MW (2009) Truth, Lies and Bullshit; distinguishing classes of dishonesty. In: Social Simulation workshop (SS@IJCAI), pp 39–50

Castelfranchi C, Falcone R (2010) Trust theory: A socio-cognitive and computational model. Wiley Series in Agent Technology

Chakraborty S, Raghavan KR, Srivastava MB, Bisdikian C, Kaplan LM (2012) Balancing value and risk in information sharing through obfuscation. In: Proceedings of the 15th Int'l Conf. on Information Fusion (FUSION '12)

Das TK, Teng BS (2004) The Risk-Based View of Trust: A Conceptual Framework. Journal of Business and Psychology 19(1):85–116, DOI 10.1023/B:JOBU.0000040274.23551.1b

Goffman E (1970) Strategic Interaction. Basil Blackwell Oxford

Jøsang A (2001) A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9(3):279–311

Jøsang A, Ismail R (2002) The beta reputation system. In: Proceedings of the 15th Bled Electronic Commerce Conference

Kaplan S, Garrick BJ (1981) On the quantitative definition of risk. Risk Analysis 1(1):11–27

Mardziel P, Magill S, Hicks M, Srivatsa M (2011) Dynamic enforcement of knowledge-based security policies. In: Proceedings of the 24th IEEE Computer Security Foundations Symposium, pp 114–128

Sentz K, Ferson S (2002) Combination of evidence in Dempster-Shafer theory, vol 4015. Citeseer

Tan YH, Thoen W (2002) Formal aspects of a generic model of trust for electronic commerce. Decision Support Systems 33(3):233–246, DOI 10.1016/S0167-9236(02)00014-3

Tang Y, Cai K, McBurney P, Sklar E, Parsons S (2011) Using argumentation to reason about trust and belief. Journal of Logic and Computation DOI 10.1093/logcom/exr038

Teacy WTL, Patel J, Jennings NR, Luck M (2006) Travos: Trust and reputation in the context of inaccurate information sources. Autonomous Agents and Multi-Agent Systems 12(2):183–198

Urbano J, Rocha A, Oliveira E (2013) A socio-cognitive perspective of trust. In: Ossowski S (ed) Agreement Technologies, Law, Governance and Technology Series, vol 8, Springer Netherlands, pp 419–429, DOI 10.1007/978-94-007-5583-3_23

Wang X, Williams MA (2011) Risk, uncertainty and possible worlds. In: Privacy, security, risk and trust (passat), IEEE Third International Conference on Social Computing (SOCIALCOM), pp 1278–1283, DOI 10.1109/PASSAT/SocialCom.2011.130

A The Case of Continuous Random Variables

By utilising Definitions 8 and 9 we can describe the impact of disclosing a message to the consumers on the producer ag_0.

Proposition 1 *Given a FCA $\langle \mathcal{A}, \mathcal{C}, \mathcal{M}, m\rangle$; a consumer $ag_q \in \mathcal{A}$; and the equivalent degree of disclosure $x_{0,q}$ of the producer over ag_q. Let y_q be the information inferred by ag_q according to the r.v. $I_q(x_{0,q})$ (with probability $\approx f_{I_q}(y_q; x_{0,q})\, dy_q$). Then, assuming that the impact z_q is independent of the degree of disclosure distribution $x_{0,q}$ given the inferred information y_q, ag_0 expects an impact z_q described by the r.v. $Z_q(x_{0,q})$ with density:*

$$f_{Z_q}(z_q; x_{0,q}) = \int_0^1 f_{Z_q}(z_q; y_q)\, f_{I_q}(y_q; x_{0,q})\, dy_q.$$

Proof

$$\begin{aligned} F_{Z_q}(z_q; x_{0,q}) &= \Pr\{Z_q \leq z_q | x_{0,q}\} \\ &= \int_0^1 \Pr\{Z_q \leq z_q, I_q = y_q | x_{0,q}\}\, dy_q \\ &= \int_0^1 \Pr\{Z_q \leq z_q | I_q = y_q, x_{0,q}\} f_{I_q}(y_q; x_{0,q})\, dy_q \\ &= \int_0^1 F_{Z_q}(z_q; y_q)\, f_{I_q}(y_q; x_{0,q})\, dy_q, \end{aligned}$$

The density function is easily derived from the distribution $F_{Z_q}(z_q; x_{0,q})$ since $f_{Z_q}(z_q; x_{0,q}) = \frac{d}{dz_q} F_{Z_q}(z_q; x_{0,q})$. □

Moreover, any time we need a single value characterisation of a distribution, we can exploit the same idea of descriptors of a random variable, by introducing descriptors for trust and risk.

Definition 13 Let $h(\cdot)$ be a function defined on $[0, 1]$, and $y \in [0, 1]$ be a level of inference. We define

$$t_h^{Z_q}(x) = \int_0^1 h(w)\, f_{Z_q}(w; y)\, dw, \tag{9}$$

to be the *y-trust descriptor* induced by $h(\cdot)$.

We can do the same to obtain a impact descriptor:

Definition 14 Let $h(\cdot)$ be a function defined on $[0, 1]$, and $x \in [0, 1]$ be a level of disclosure.We define

$$t_h^{Z_q}(x) = \int_0^1 h(w)\, f_{Z_q}(w; x)\, dw, \tag{10}$$

to be the *x-impact descriptor* induced by $h(\cdot)$.

Typical $h(\cdot)$ include the moment generating functions, such as $h(k) = k, k^2$, etc., and entropy $h(k) = -ln(f_K(k))$ for the density of some r.v. K. In the following we use the expectation as the risk descriptor, leaving consideration of other possible functions for future work.

Finally, let us illustrate two notable properties of our model. The first one is with regards to the case where a consumer can derive the full original message, which, unsurprisingly, leads to the worst case impact.

Proposition 2 *When a consumer is capable of gaining maximum knowledge, then $f_I(y; x) = \delta(y-1)$, where $\delta(\cdot)$ is the Dirac delta function, and $F_Z(z; x) = F_Z(z) \triangleq F_Z(z; 1)$, i.e., the risk coincides with the 1-trust (Definition 9).*

Proof By the definition of the inference r.v. $I(x)$, when ag_x is believed to gain maximum knowledge then the density $f_I(y; x)$ carries all its weight at the point $y = 1$ for all x. Hence, $f_I(y; x) = \delta(y-1)$ and it follows from the definition of the Dirac delta function, see also Prop. 1

$$\begin{aligned} F_Z(z; x) &= \int_0^1 F_Z(z; y)\, f_I(y; x)\, dy \\ &= \int_0^1 F_Z(z; y)\, \delta(y-1)\, dy = F_Z(z; 1). \end{aligned} \tag{11}$$

□

The second property pertains to the case where agent ag_0 shares information with more than one consumer. Such situations are typically non-homogeneous as the trust and impact levels with regards to each consumer are different. Clearly, it is beneficial to identify conditions where these impacts balance (and, hence, indicate crossover thresholds) across the multiple agents.

For two agents ag_1, ag_2 having corresponding inference and behavioural trust distributions $F_{I_j}(y; x)$ and $F_{Z_j}(z; y)$, $j \in \{1, 2\}$, for the shared information to have similar impact, x_1 and x_2 should be selected, such that the following holds.

$$\begin{gathered} F_{Z_1}(z; x_1) = F_{Z_2}(z; x_2) \Leftrightarrow \\ \int_0^1 F_{Z_1}(z; y)\, f_{I_1}(y; x_1)\, dy = \int_0^1 F_{Z_2}(z; y)\, f_{I_2}(y; x_2)\, dy. \end{gathered} \tag{12}$$

Note that the above relationship implies the r.v.s Z_1 and Z_2 are drawn from the same distribution. Such a requirement is typically unrealistic. Therefore in general one may want to consider equalities on the average, such as, finding x_1 and x_2 satisfying the following for appropriate $g(\cdot)$ functions.

$$\mathbb{E}\{g(Z_1(x_1))\} = \mathbb{E}\{g(Z_2(x_2))\}, \tag{13}$$

Proposition 3 *Given that $g(z) = z$, in order to attain the same level of impact when ag_0 shares information with ag_1, ag_2, the degrees of disclosure x_1 and x_2 for ag_1, ag_2 respectively must satisfy the following.*

$$\mathbb{E}_{I_1}\{\mathbb{E}\{Z_1(x_1)|I_1\}\} = \mathbb{E}_{I_2}\{\mathbb{E}\{Z_2(x_2)|I_2\}\}. \tag{14}$$

Proof The case where $g(z) = z$ corresponds to the regular averaging operator, and (13) *becomes:*

$$\begin{aligned} &\int_0^1 \int_0^1 z f_{Z_1}(z; y)\, f_{I_1}(y; x_1)\, dy\, dz \\ &\quad = \int_0^1 \int_0^1 z f_{Z_2}(z; y)\, f_{I_2}(y; x_2)\, dy\, dz \\ &\Leftrightarrow \\ &\int_0^1 f_{I_1}(y; x_1) \left[\int_0^1 z f_{Z_1}(z; y)\, dz\right] dy \\ &\quad = \int_0^1 f_{I_2}(y; x_2) \left[\int_0^1 z f_{Z_2}(z; y)\, dz\right] dy \\ &\Leftrightarrow \\ &\mathbb{E}_{I_1}\{\mathbb{E}\{Z_1(x_1)|I_1\}\} \\ &\quad = \mathbb{E}_{I_2}\{\mathbb{E}\{Z_2(x_2)|I_2\}\}. \end{aligned}$$

□