

Employing Argumentation to Support Human Decision Making: A User Study

Jordan Salvit¹, Zimi Li¹, Senni Perumal², Holly Wall¹,
Jennifer Mangels³, Simon Parsons⁴ and Elizabeth I. Sklar⁴

¹ Department of Computer Science, Graduate Center, City University of New York,
365 Fifth Avenue, New York, NY, USA

² Raytheon BBN Technologies, 10 Moulton Street Cambridge, MA, USA

³ Department of Psychology, Baruch College, City University of New York, 55
Lexington Avenue, New York, NY, USA

⁴ Department of Computer Science, University of Liverpool, Ashton Street,
Liverpool, UK

jordan@jordansalvit.com, zimili.sjtu@gmail.com, holly.e.wall@gmail.com,
senni.peri@gmail.com, jennifer.mangels@baruch.cuny.edu,
s.d.parsons@liverpool.ac.uk, e.i.sklar@liverpool.ac.uk

Abstract. *ArgTrust* is an implementation of a formal system of argumentation in which the evidence that influences a conclusion is modulated according to values of *trust*. This paper reports on a user study, conducted with twenty-two human subjects, which was designed to explore how effective *ArgTrust* is as a decision-support tool. Results of the study are presented, including analysis of participants’ decision choices and opinions about the tool. Preliminary lessons respecting the use of argumentation for interactive decision support tools are extrapolated.

1 Introduction

In systems of argumentation in which arguments are constructed from logical statements [5, 31], an important feature is the way in which elements of the arguments—the premises and rules from which they are constructed—have a bearing on the quality of the arguments. Premises may be undermined and hence defeated. Conclusions may be rebutted, and rules themselves may be undercut. This relationship between the parts and the whole, combined with the relationship between the trust individuals place in information and the provenance of that information [13], has led us to suggest the use of argumentation in situations where trust in information is critical [36, 37]. The key idea is that since argumentation tracks the data used in deriving conclusions, if that data could be related to the sources from which it comes, information about those sources could be used in reasoning about the conclusions.

We subsequently developed a formal argumentation system [42] that allows information about sources—represented in the form of the “trust networks” that are standard in the literature of reasoning about trust—to be combined with arguments. This formal system was implemented in a software system called

ArgTrust [38]. In the work presented here, we report on a user study designed to explore how effective *ArgTrust* is as a decision-support tool for humans. In particular, the aim of the user study was to gather information about how people reason, how they make decisions in uncertain situations, and how they explain their decisions. Participants (i.e., human subjects) used the *ArgTrust* software tool to help visualise a scenario and make sense of information presented that describes elements of the scenario in different ways.

The remainder of this paper is structured as follows. We start in Section 2 with a brief description of the *ArgTrust* system. Then, Section 3 describes the design of the user study, and Section 4 gives the results of the study. Related work is highlighted in Section 5, and then Section 6 concludes.

2 ArgTrust software tool

This section briefly describes the *ArgTrust* software tool and the underlying formal model.

2.1 Theoretical basis

Our formal argumentation system [42] starts with the idea that we want to represent the beliefs of a set of individuals, *Ags*, where each $Ag_i \in Ags$ has access to a knowledge base, Δ_i , containing formulae in some language \mathcal{L} . An *argument* is then:

Definition 1 (Argument). *An argument A from a knowledge base $\Delta_i \subseteq \mathcal{L}$ is a pair (G, p) where p is a formula of \mathcal{L} and $G \subseteq \Delta_i$ such that:*

1. G is consistent;
2. $G \vdash p$; and
3. G is minimal, so there is no proper subset of G that satisfies the previous conditions.

G is called the *grounds* of A , written $G = Grounds(A)$ and p is the *conclusion* of A , written $p = Conclusion(A)$. Any $g \in G$ is called a *premise* of A . The key aspect of argumentation is the association of the grounds with the conclusion, in particular the fact that we can trace conclusions to the source of the grounds.

The particular language \mathcal{L} that we use is \mathcal{L}^{DHC} the language of defeasible Horn clauses—that is, a language in which formulae are either atomic propositions p_i or formulae of the form $p_i \wedge \dots \wedge p_n \Rightarrow c$, where \Rightarrow is a defeasible rule rather than material implication. Inference in this system is by a defeasible form of generalised *modus ponens* (DGMP):

$$\frac{p_1, \dots, p_n \quad p_i \wedge \dots \wedge p_n \Rightarrow c}{c} \quad (1)$$

and if p follows from a set of formulae G using this inference rule alone, we denote this by $G \vdash^{DHC} p$.

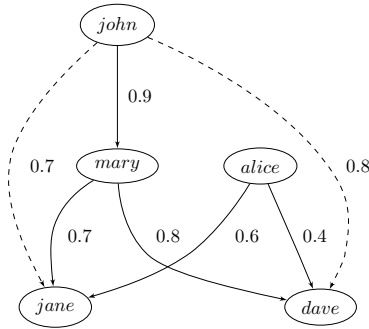


Fig. 1. Social network. Trust is propagated using *tidal trust* (see text).

The set of individuals, Ag_s , are related to each other by a social network that includes estimates of how much individual agents trust their acquaintances, as illustrated in Figure 1. Nodes represent individuals and links between them are annotated with the degree to which one individual trusts another, represented as values between 0 and 1. The input to the network (i.e., information known *a priori*) consists of the nodes and the solid edges. The output of the network (dashed edges) is the degree of trust inferred between any two nodes in the network. We could, for example, apply *TidalTrust* [12] to propagate trust values through the network and relate agents that are not directly connected in the social network.

In decision-making situations, argumentation can help in two ways. First, it is typical that from the data a given individual Ag_i has about a situation, we can construct a set of arguments that may conflict with each other. We might have an argument (G, p) in favour of some decision option, and another argument $(G', \neg p)$ against it (in this case, we say that the arguments *rebut* each other). We might also have a third argument $(G'', \neg g)$ where $g \in G$ is one of the grounds of the first argument (in this case we say that $(G'', \neg g)$ *undermines* (G, p)). Finally, we might have a fourth argument $(G''', \neg i)$ where i is one of the conclusions to one of the defeasible rules in (G, p) . (This is another form of rebut, rebuttal of a sub-argument.) Argumentation provides a principled way—or rather a number of alternative ways—for Ag_i to establish which of a conflicting set of arguments it is most reasonable to *accept* [4].

Second, the grounds of an argument G , can be related back to the sources of the information that constitutes the grounds. If that information comes from some individual Ag_j that Ag_i knows, then Ag_i can *weight* the information according to how much they trust Ag_j (an extension of Liau’s [29] principle that you believe information from individuals that you trust). The same principle can be applied to other sources of information⁵. This weight can be used to re-

⁵ For example, military intelligence traditionally separates information into that which comes from human sources, that which comes from signals intercepts, and that which

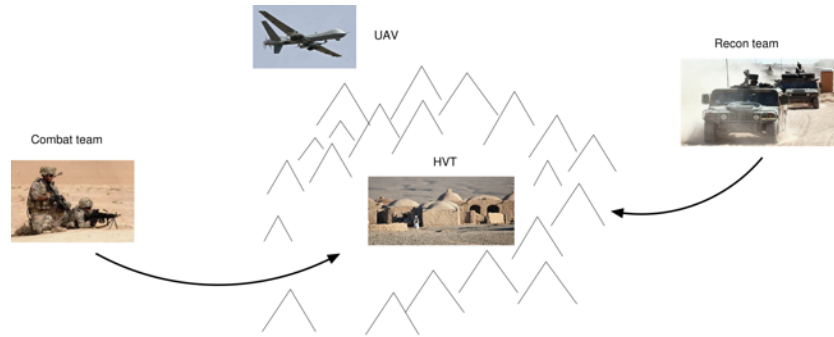


Fig. 2. The example scenario

solve conflicts between arguments. It is possible to provide the decision maker with links between information that feeds into a decision and the source of that information, allowing them to explore the effect of trusting particular sources.

To see more concretely how this can be useful, let's look at a simple decision-making example, loosely based on Operation Anaconda [33] and depicted in Figure 2. In this example, a decision is being made about whether to carry out an operation in which a combat team will move into a mountainous region to try to apprehend a high value target (HVT) believed to be in a village in the mountains.

We have the following information:

1. If there are enemy fighters in the area, then an HVT is likely to be in the area.
2. If there is an HVT in the area, and the mission will be safe, then the mission should go ahead.
3. If the number of enemy fighters in the area is too large, the mission will not be safe.
4. UAVs that have flown over the area have provided images that appear to show the presence of a significant number of camp fires, indicating the presence of enemy fighters.
5. The quality of the images from the UAVs is not very good, so they are not very trusted.
6. A reconnaissance ("recon") team that infiltrated the area saw a large number of vehicles in the village that the HVT is thought to be inhabiting.
7. Since enemy fighters invariably use vehicles to move around this is evidence for the presence of many enemy fighters.
8. Informants near the combat team base claim that they have been to the area in question and that a large number of fighters are present.

comes from imagery. All of these sources can be rated with some measure of trustworthiness.

9. In addition, we have the default assumption that missions will be safe, because in the absence of information to the contrary we believe that the combat team will be safe.

Thus there is evidence from UAV imaging that sufficient enemy are in the right location to suggest the presence of an HVT. There is also some evidence from informants that there are too many enemy fighters in the area for the mission to be safe. Since informants are paid, their incentive is often to make up what they think will be interesting information and so they are not greatly trusted. However, this conclusion is supported by the findings of the reconnaissance team who are highly trusted.

We might represent this information as follows (the numbers in parentheses indicate the correspondence between the logic representations, below, and the relevant piece(s) of information, above)⁶:

- (1) $InArea(enemy) \Rightarrow HVT$
- (2) $HVT \wedge Safe(mission) \Rightarrow Proceed(mission)$
- (3) $InArea(enemy) \wedge Many(enemy) \Rightarrow \neg Safe(mission)$
- (4, 5) $InArea(campfires)$
- (4) $InArea(campfires) \Rightarrow InArea(enemy)$
- (6) $InArea(vehicles)$
- (7) $InArea(vehicles) \Rightarrow Many(enemy)$
- (7, 8) $Many(enemy)$
- (9) $Safe(mission)$

From this information, we can construct arguments such as:

$$\left(\left(\begin{array}{l} InArea(campfires), \\ InArea(campfires) \Rightarrow InArea(enemy), \\ InArea(enemy) \wedge Safe(mission) \Rightarrow HVT, \\ HVT \Rightarrow Proceed(mission) \end{array} \right), Proceed(mission) \right)$$

which is an argument for the mission proceeding, based on the fact that there are campfires in the area, which suggest enemy fighters, that enemy fighters suggest the presence of an HVT, and that the presence of an HVT (along with the default assumption that the mission will be safe) suggests that the mission should go ahead.

We can build other arguments from the available information, and, since these will conflict, then compute a subset that are *acceptable*. (Approaches to this computation are discussed in [4].) We can build other arguments from the full information that is available. For example, from the informants' information

⁶ While stressing that this is purely illustrative — a real model of this example would be considerably more detailed.

we can conclude that there are many enemies in the area and hence the mission will not be safe:

$$\left(\left(\begin{array}{l} InArea(vehicles), \\ InArea(enemy), \\ InArea(vehicles) \Rightarrow Many(enemy) \\ InArea(enemy) \\ \wedge Many(enemy) \Rightarrow \neg Safe(mission) \end{array} \right) , \neg Safe(mission) \right)$$

This conflicts with the previous argument by undermining the assumption about the mission being safe. Since in our scenario the informants are not highly trusted, the first argument is not *defeated* and so is then *acceptable*. The relation between trust in the source of an argument and defeat between arguments is explored in [37]. Given all the information from the scenario, we can also construct an argument against the safety of the mission based on information from the recon team. Since the recon team is highly trusted, this argument would defeat the argument for the mission to proceed, rendering it not acceptable.

2.2 Implementation

An initial version (*v1.0*) of *ArgTrust* was described in [43]. Here we present some aspects of the current version, *v2.0*. Like *v1.0*, this current version takes as input an XML file in a format which we sketch here. First, we have a specification of how much sources of information are trusted, for example:

```
<trustnet>
  <agent> recon </agent>
  ...
  <trust>
    <truster> me </truster>
    <trustee> recon </trustee>
    <level> 0.95 </level>
  </trust>
  ...
</trustnet>
```

which specifies the individuals involved (including “me”, the decision maker) and the trust relationships between them, including the level of trust (specified as a number between 0 (no trust) and 1 (completely trustworthy)). The current implementation uses these values to compute the trust that one agent places on another using a choice of TidalTrust [12] or the mechanism described in [49].

The XML file also contains the specification of each individual’s knowledge, for example:

```
<beliefbase>
  <belief>
    <agent> recon </agent>
    <fact> enemy_in_area </fact>
```

```

    <level> 0.9 </level>
  </belief>
  ...
<belief>
  <agent> me </agent>
  <rule>
    <premise> many_enemy </premise>
    <conclusion> not safe </conclusion>
  </rule>
  <level> 1.0 </level>
</belief>
  ...
</beliefbase>

```

Here the numbers reflect the belief each individual has in its information about the world.

From this data, and a query about a particular proposition, *ArgTrust* constructs arguments for that proposition by backward chaining. Once these arguments have been constructed, *ArgTrust* examines each formula used in the derivation of these arguments to identify if there are arguments with conclusions that attack these formulae. Each formula in those attacking arguments are then examined in turn. (And so on.) Once the full set of arguments is constructed, the grounded semantics [7] are applied and the conclusions labelled IN, OUT or UNDEC [4]

ArgTrust v2.0 extends the previous version [44, 42] by implementing a more robust and flexible data model. *ArgTrust v2.0* uses a SQL database and the Python programming language (for reasons outlined below), in place of Java (which was employed for *ArgTrust v1.0*), or instead of using a more traditional, logic programming language. The language choice was largely made in order to simplify the recursive methods for storing the data and traversing it in different ways. In a mySQL⁷ database, we maintain arguments as a relationship to a set of trees that represent the logical steps needed to arrive at the argument's conclusion. Thus, to return to our Operation Anaconda example, the combination of premise

$$InArea(campfires)$$

with rule

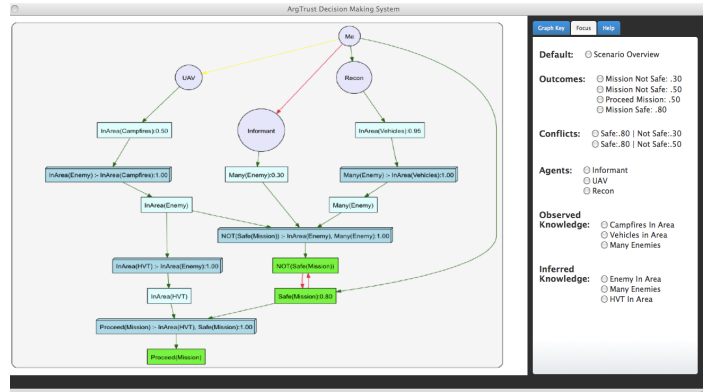
$$InArea(campfires) \Rightarrow InArea(enemy)$$

to infer conclusion

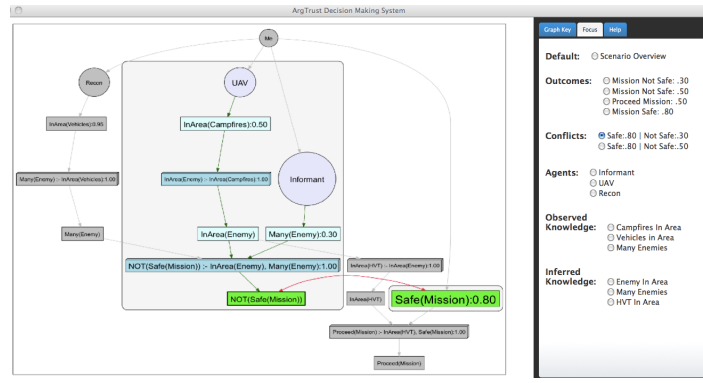
$$InArea(enemy)$$

would be represented as a tree in which each of the above formulae was a node, and arcs led from premise to rule to conclusion. The representation allows us to easily overlap arguments that share predicates or rules. Thus, if we had another argument with conclusion *InArea(enemy)*, we would represent the two arguments together as a tree with a single conclusion node.

⁷ <http://www.mysql.com>



(a) A high-level view of the argument graph



(b) A more focussed view of one argument for the mission being unsafe.

Fig. 3. Two argument graphs from the *ArgTrust* interface

Another important piece of the data model is its flexibility to receive new attributes and easily facilitate reconstructing arguments for the conclusions at hand. For example, our experience is that users each have different senses of what “very trustworthy” means. Therefore, we built the system in such a way that changing values to belief levels or trust levels does not require completely reloading the scenario, instead entails just changing a parameter value.

The underlying *ArgTrust* inference engine can be invoked in four different modes: (1) as a command-line tool; (2) as a visualisation tool; (3) as an interactive decision support tool; and (4) as a back-end reasoning engine. In command-line tool mode, a user can load an XML file, modify its contents on the *ArgTrust* command line, and pose queries to the inference engine. The system will respond by outputting text that reports the status of arguments supporting the query. In visualisation tool mode, the system produces output in a graphical display of

the resulting arguments—here the result of inference is an *argument graph* (see below) like that in Figure 3(a). In interactive decision support mode, users step through a *decision scenario* and analyse it interactively. In back-end reasoning engine mode, *ArgTrust* is called by another program—which might itself have an interactive front-end. Input is in the form of an XML file, as with the previous three modes; and output is also presented in the form of an XML file, where the burden of communicating the content of the output to a human user becomes the responsibility of the calling program. An example of this mode has been implemented and tested in related work involving a human-robot environment [3].

In visualisation and interactive decision support modes, *ArgTrust* makes use of *argument graphs* to visualise complex scenarios and assign probabilities to all the possible outcomes. These graphs, which are distinct from the attack graphs common in the literature and which are also often called “argument graphs”⁸, represent the relationship between the facts and rules that make up the arguments, and the relationships between the arguments themselves. A full explanation of the graphs can be found in [42], along with the translation into graph-theoretic terms of the usual ideas of *extension* and the *acceptability* of arguments.

The next section, below, describes the user interface developed for the interactive decision support mode. Then, Sections 3 and 4 describe a user study designed to evaluate the effectiveness of this mode.

2.3 User interface

The interactive decision support mode of *ArgTrust v2.0* includes an interface which allows users to manipulate the argument graphs at different levels of detail, and to focus on individual components of an argument, in order to better understand a scenario in its entirety and reach an informed conclusion.

When used in this mode, there are four main components to the *ArgTrust* interface, each accessed by clicking on *tabs* in the user interface:

- **Review Scenario** tab: This component shows the text-based narrative describing a scenario and allows users to read through it before progressing. This tab is persistent, allowing users to revisit the scenario text at any time.
- **Review Trust/Beliefs/Rules** tab: These three tabs allow a user to change the belief level for each individually trust/belief/rule from a specified agent. The corresponding sentence in the scenario is displayed as well, facilitating the user’s ability to set the level more accurately. After setting a belief level, the user can navigate to one of the last two tabs to see how it impacts the argument. This process can be iterative, as the user understands and learns their own process for defining belief and trust levels.
- **Trust+Beliefs** tab: In this part of the system, the combination of an agent’s trust values and belief values are displayed to illustrate the belief value for the decision maker in a particular proposition. The goal is to fill the logical gap

⁸ See, for example, Figure 1 in [47].

between setting trust values and belief values by displaying the combination of both.

- **Argument Graph** tab: This component displays the argument graph that corresponds to the scenario. Scenarios are broken down into arguments which are built from bits of knowledge (e.g., facts or evidence), rules (or beliefs), and the resulting conclusions. Facts and rules are linked to individuals (sources of information or agents). Arrows connect facts, rules and conclusions together to form a chain. A chain is referred to as an argument. (A chain can have only one arrow linking two nodes, or multiple arrows linking more than two nodes.) Each argument ends in a conclusion, and every conclusion is assigned a probability. The user can control the amount of information displayed in the graph by selecting “zoom level” and “detail” options and “focus”.
 - **Zoom-level and Detail-level** controls: Located in the upper right-hand corner of the argument graph panel are the zoom and detail controls. The zoom buttons allow users to visually zoom in and out of the graph (i.e., magnifying the visual display, but not changing the content). The detail slider allows users to adjust the level of detail displayed in the graph (i.e., changing the content to be more refined or more abstract). At the highest level of detail (most abstract), only the conclusions and their corresponding probabilities are displayed. Alternatively, at the lowest level of detail (most refined), all sources, i.e., agents, beliefs, facts, rules are shown.
 - **Focus** controls: The focus feature, located in the sidebar, enables users to focus on individual arguments of a graph. The graph updates to highlight the chosen conclusion or piece of knowledge, allowing users to focus in on that particular piece of the scenario.

3 User study design

We conducted a user study designed to evaluate the effectiveness of the *ArgTrust* interactive decision support mode. Two scenarios were developed for the user study, including narratives and logical representations of information contained in each narrative, such as in the example outlined in Section 2.1. One scenario is short, relatively simple and was designed as a training exercise; the other scenario is longer, more complex and was designed as an evaluation exercise.

The user study procedure involved multiple steps. First, participants were asked to provide demographic (e.g., gender and age) and background (e.g., level of experience working with computers and decision-making tools) information by filling out a “Demographic and Background” survey, for statistical purposes in order to describe the population of human subjects and to satisfy reporting requests by funding agencies. Then participants completed a short training exercise, using the short and simple scenario mentioned above (shown in Figure 4), to familiarise them with the notion of decision making under uncertainty and to give them a preliminary experience using the ArgTrust system and user interface.

Your grandparents are coming to visit you in New York City, and they are arriving at the airport shortly. They get anxious when visiting big cities, so you promised to meet them at the airport and escort them to their hotel. You had planned to take the train to the airport straight from work. Right before you planned to leave, your co-worker tells you there was an earlier incident at a station and that train line is experiencing delays. You text a friend, who you know lives near that train line, to confirm. Your friend tells you that she left her house at the usual time and arrived to work on time, without experiencing any problems with the train. **Do you risk taking the train, which may be delayed, or do you take a taxi instead (more expensive, but quicker)?**

Fig. 4. Narrative of short, simple training scenario: “Grandparents Scenario”

Next, participants were presented with a text-based narrative describing a more complex scenario (the longer scenario, mentioned above and shown in Figure 5). They were asked to analyse the details of the scenario and come to a decision about an action to take with respect to the scenario. Once they made their decision, they were then asked to report on why and how they made that decision, via an on-line “pre-survey”. Participants were asked to provide as much detail as possible regarding their thought processes.

Finally, participants were given the same scenario and asked to reconsider their decision, this time with the aid of the ArgTrust tool. Participants were asked to use the ArgTrust user interface to explore the data describing the scenario and then report on how they utilised the software in their decision-making process, by completing an on-line “post-survey”. The study took approximately 60 minutes to complete.

4 User study results

The user study was conducted in three sessions, where each session was conducted in a different location and involved a different set of participants. The first two sessions were conducted in a university setting and participants were undergraduate and graduate students: the first group were psychology students (Group I) and the second group was computer science students (Group II). The third session was conducted at an engineering company and participants were technical employees of the company (Group III). Each session was conducted following the same procedure (outlined Section 3). Twenty-two participants completed the user study; demographics are shown in Table 1. All participants were well educated: 12 participants had a Masters Degree or above, and 7 had a PhD. Nobody reported previous experience with computer-based decision making tools, but almost everyone (21 of 22) claimed prior experience with data management tools, such as Microsoft Office. Only 2 of the 22 participants had previously encountered the concepts of “logical argumentation” or “argumentation graphs”.

A week ago, a powerful earthquake struck Brax causing widespread devastation to the countrys infrastructure and leaving over 10,000 dead and over 50,000 injured. The two cities in Brax that were hit the hardest are Waga and Tapel. The Braxian Government and the UN have requested global assistance to launch the largest humanitarian relief operation ever executed. The Braxian Military, with its extensive and modern military force and airlift capability, is leading the effort and coordinating the international response. You are an Intelligence Analyst at your desk in the Operations Center of the main Forward Operating Base (FOB) in Tapel, monitoring the flow of data and reports coming in related to conditions, casualties and relief requirements. You have direct communications with the other FOB location in Waga, which was likewise affected by the earthquake. There are two rebel insurgent cells operating in the region: Reds and Lions. Each one is vying for power with the population, local and national politicians. Each one is seeking to take advantage of the situation to consolidate their political positions and establish local control with their rebel militia forces. The rebel militia forces have access to only small arms weapons and limited explosives. The rebel militias are stirring up the local population to protest the incompetence of the Braxian government. Braxian military forces are now stretched thin, trying to defend against the rebel militia forces while, at the same time, leading humanitarian rescue efforts in the wake of the earthquake. It has been 6 days since the earthquake hit Brax. Your Army Commander has asked you to answer the following Priority Intelligence Requirement (PIR): Which rebel militia cell is encouraging the most violence against the Braxian government? You have the following information (the order of the items listed is arbitrary):

- The Braxian Military reports that they have encountered many attacks/incidents of violence involving Red rebels and only some incidents of violence involving Lion rebels.
- Many incidents of violence by a rebel group imply that it is creating/encouraging much violence whereas some incidents of violence by a rebel group imply that it is not creating much violence.
- Sources of information include: Braxian Department of State, Braxian First Responders, Braxian Officials, Braxian Civilians, International Civilians and Open Media (like newspapers). Collectively, these sources of information reports only few incidents each, which makes information from them incomplete.
- Twitter feeds are inundated with reports of violence which are often contradictory. Twitter feeds are not considered very reliable.
- The Braxian Military reconnaissance reports that they have seen lots of vehicles outside the Lion Headquarters both in Tapel and in Waga.
- The presence of large number of vehicles outside a rebel militia headquarters can indicate that the rebel militia is planning many attacks/incidents of violence on relief personnel.
- Members of the rebel Lion militia who are paid by the Braxian government to inform on their comrades indicate that they have been directed to increase violence and use small arms against the Braxian military.
- A rebel group that may be planning many attacks as well as directing its members to increase violence could be a group that will create much violence.

You have to decide which rebel militia the Braxian Military efforts should focus on defending against.

Fig. 5. Narrative of longer, more complex evaluation scenario: “Humanitarian Relief Scenario”

Table 1. User study participants: 22 human subjects in total.

	female	male	age 18 – 24	age 25 – 39
Group I: psychology students	6	0	2	4
Group II: computer science students	1	6	4	3
Group III: technical professionals	2	7	0	9
total	9(41%)	13(59%)	6(28%)	16(72%)

The detailed results of the study are available from the authors on request, and in-depth statistical analysis of study data is underway. Here we present an overview of the most notable findings.

1. When we look at the conclusions that participants drew about the scenario, we find that 5 people, nearly a quarter of the test subjects, *changed their decision as a result of using the software*. This suggests that presenting the information in the argument graphs revealed things about the scenario that were initially missed by the participants.
2. The above conclusion is supported by the result that we see in the answer to the question “How well would you say you understood the scenario?”. Responding on a 10–point scale, the score for Group II participants rose from 7.9 to 8.6 and the score for Group III participants rose from 6.9 to 7.2. However, Group I reported a decrease in understanding from 7.8 to 7. If we look at the numbers of participants reporting an increase/decrease in understanding rather than the average score, we find that 3 people in Group I reported better understanding and 2 reported worse understanding, 3 people in Group II reported better understanding and none reported worse understanding, and 3 people in Group III reported better understanding and 2 reported worse understanding. Overall, that makes 9 reporting better understanding and 7 reporting worse understanding.
3. We also asked participants how hard they found it to analyse the scenario. Here the results were more mixed. Group II and Group III both showed a decrease in difficulty from pre-software to post-software (4.7 to 4.6 and 5.4 to 5, respectively) on the same scale as before. Group I, however, showed a large rise in difficulty (5.1 to 7). If we again look at numbers of participants reporting, we find that 1 person from Group I reports the task is less demanding with the software and 4 report it is more demanding, and the figures for Group II and Group III are 4 (less demanding) and 2 (more) and 2 (less) and 2 (more) respectively. Thus, overall, we have 7 reporting it is less demanding and 8 reporting it is more demanding; amongst participants with technical backgrounds, more (6) reported that the task was less demanding when using the ArgTrust software than those who reported the task was more demanding (4).

Overall, then, the data provides evidence that participants who are presented with argumentation-based support for making decisions found that using the software tool helped their understanding (both directly reported and inferred

- The tool is successful in representing the factors that go into making a decision and displaying the relationships between those facts and how they influence a decision.
- It somewhat helped filter out the information that is not necessarily true.
- It helped me break the component information down a little bit but I had already created an outline of my own that helped me just as much if not more. Seeing the beliefs and the trust broken down was helpful, though.
- It helped to tell me what I'm supposed to think. . . how much I'm supposed to trust people, and how I was supposed to interpret the statements in the given scenario. However, it bothered me that a tool was telling me how to simplify a complex problem, since I don't believe the tool can possibly take all the details and subtleties into account. but if I accept that a complex situation can and must be simplified, then yes, the tool is helpful as a place to plug in parameters and let it do the math.
- By breaking down the situation into smaller bits and displaying how much you believe each situation to be true, it was much easier to make decisions because I was considering the situation asked only, not the entire situation.
- When I put how I felt into numbers, it organized and simplified my concerns and weighed all of the factors into the equation for me. It made it easier to see the results.

Fig. 6. Comments from the test subjects.

from the changes made in their decisions), albeit at the cost of increased difficulty in reaching a decision.

In addition, we can derive some support for argumentation-based decision making from the freeform comments made by participants are given in Figure 6. Though the comments also contain some negative sentiments, we believe that these indicate that some users understand the benefits of structuring decisions in the way that we do in *ArgTrust*.

5 Related work

There are three main areas of work that are related to the results reported here: modelling trust; reasoning about trust using argumentation; and argumentation-based decision making. We briefly cover each of these below.

5.1 Modelling trust

As computer systems have become increasingly distributed, and control of those systems has become more decentralised, computational approaches to trust have become steadily more important [15]. Some of this work has directly been driven by changes in technology, for example considering the trustworthiness of nodes in peer-to-peer networks [1, 9, 22], or dealing with wireless networks [14, 23, 41]. Other research has been driven by changes in the way that technology is used,

especially involving the Internet. One early challenge is related to the establishment of trust in e-commerce [32, 39, 50], and the use of reputation systems to enable this trust [25, 26]. Another issue is the problem of deciding which of several competing sources of conflicting information one should trust [2, 6].

Additional issues have arisen with the development of the social web, for example, the questions of how social media can be manipulated [27, 28] and how one should revise one's notions of trust based on the past actions of individuals [17]. In this area is some of the work that is most relevant for that we describe here, work that investigates how trust should be propagated through networks of individuals [16, 19, 24, 49], and we have drawn on this in our implementation of ArgTrust.

5.2 Reasoning about trust using argumentation

The second area of work to consider is that which looks at the use of argumentation to handle trust. While the literature on trust is considerable, prior work on argumentation and trust is much more sparse. Existing work includes Harwood's application of argumentation techniques to networks of trust and distrust [18], Stranders' coupling of argumentation with fuzzy trust measures [40], Matt's [30] combination of arguments with statistical data to augment existing trust models, Villata's use of metalevel argumentation to describe trust [46], and Oren's [34] coupling of argumentation and subjective logic (used in [19] to handle trust measures). However, none of this covers the same ground as our work.

5.3 Argumentation-based decision making

The third area of work to consider is that on argumentation-based decision making. Here the work by Fox and colleagues [8, 10] showed that constructing arguments for and against a decision option, and then simply combining these arguments⁹ could provide a decision mechanism that rivalled the accuracy of probabilistic models. This basic method was extended in [11, 35] to create a symbolic mechanism that, like classical decision theory, distinguished between belief in propositions and the values of decision outcomes, while [20] showed the usefulness of arguments in communicating evidence for decision options to human users. More recent work on argumentation and decision making is described in [21, 45].

The relationship that we consider between argumentation and decision-making is different from all of this work. All the above work tries to build argumentation systems that identify the best decision to take. Even [20], which is closest to what we are doing, tries to identify the best decision and explain it to a human user in terms the human can understand. In work that evaluates the effectiveness of the systems, the aim is to show that the system gets the decisions right [8, 48]. Our focus, in contrast, is just to present information and test

⁹ Even the very straightforward mechanism of counting arguments for and against.

whether the users find the information to be useful—whether it helps them to feel more comfortable with their decisions, and whether they alter their opinion as a result of being able to use our software tool to visualise and manipulate the information on which they base their decisions.

6 Summary

We have described the *ArgTrust v2.0* software system designed to help users balance information from multiple sources and draw conclusions from that information. The system can be invoked in any of four modes, one of which is as an interactive decision support tool. A user study examining the efficacy of ArgTrust running in this mode has been conducted and was described here. Twenty-two human subjects participated, from three distinct groups of users. The non-technical users had a mixed reaction to the software system, and a not insignificant percentage of those users felt more confused about making a decision with uncertain, conflicting information after using the software. We surmise that this is because the non-technical users were unfamiliar with symbolic representations and graph-based visuals than technical users. In contrast, technical users felt more comfortable with their decisions after using the software.

The conclusions we draw from these results is that perhaps non-technical users should receive more training than is apparently necessary for technical users. In future work, we will extend *ArgTrust* to operate in dynamic environments, where information changes as users are trying to make decisions. Further user testing is also planned.

Acknowledgements

This work was supported by NSF grant #1117761, by the Army Research Office under the Science of Security Label grant (SoSL), by the Army Research Laboratory under the Network Science Collaborative Technology Agreement and by a University of Liverpool Research Fellowship. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of the funders.

References

1. Z. Abrams, R. McGrew, and S. Plotkin. Keeping peers honest in EigenTrust. In *Proceedings of the 2nd Workshop on the Economics of Peer-to-Peer Systems*, 2004.
2. B. T. Adler and L. de Alfaro. A content-driven reputation system for the Wikipedia. In *Proceedings of the 16th International World Wide Web Conference*, Banff, Alberta, May 2007.
3. M. Q. Azhar, Eric Schneider, Jordan Salvit, Holly Wall, and Elizabeth I. Sklar. Evaluation of an argumentation-based dialogue system for human-robot collaboration. In *Proceedings of the Workshop on Autonomous Robots and Multirobot Systems (ARMS) at Autonomous Agents and MultiAgent Systems (AAMAS)*, St Paul, MN, USA, May 2013.

4. P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 2011.
5. P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235, 2001.
6. X. L. Dong, L. Berti-Equille, and D. Srivastava. Integrating conflicting data: The role of source dependence. In *Proceedings of the 35th International Conference on Very Large Databases*, Lyon, France, August 2009.
7. P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n -person games. *Artificial Intelligence*, 77:321–357, 1995.
8. J. Emery, R. Walton, A. Coulson, D. Glasspool, S. Ziebland, and J. Fox. Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (RAGs): Qualitative evaluation with simulated patients. *British Medical Journal*, 319(7201):32–36, 1999.
9. M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica. Free-riding and white-washing in Peer-toPeer systems. In *Proceedings of the 3rd Annual Workshop on Economics and Information Security*, 2004.
10. J. Fox, A. Glowinski, C. Gordon, S. Hajnal, and M. O’Neil. Logic engineering for knowledge engineering: design and implementation of the oxford system of medicine. *Artificial Intelligence in Medicine*, 2(6):323–339, 1990.
11. J. Fox and S. Parsons. Arguing about beliefs and actions. In A. Hunter and S. Parsons, editors, *Applications of uncertainty formalisms*. Springer-Verlag, 1998.
12. J. Golbeck. *Computing and Applying Trust in Web-based Social Networks*. PhD thesis, University of Maryland, College Park, 2005.
13. J. Golbeck. Combining provenance with trust in social networks for semantic web content filtering. In *Proceedings of the International Provenance and Annotation Workshop*, Chicago, Illinois, May 2006.
14. K. Govindan, P. Mohapatra, and T. F. Abdelzaher. Trustworthy wireless networks: Issues and applications. In *Proceedings of the International Symposium on Electronic System Design*, Bhubaneswar, India, December 2010.
15. T. Grandison and M. Sloman. A survey of trust in internet applications. *IEEE Communications Surveys and Tutorials*, 4(4):2–16, 2000.
16. R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on the World Wide Web*, 2004.
17. C-W Hang, Y. Wang, and M. P. Singh. An adaptive probabilistic trust model and its evaluation. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, Estoril, Portugal, 2008.
18. W. T. Harwood, J. A. Clark, and J. L. Jacob. Networks of trust and distrust: Towards logical reputation systems. In D. M. Gabbay and L. van der Torre, editors, *Logics in Security*, Copenhagen, Denmark, 2010.
19. A. Jøsang, R. Hayward, and S. Pope. Trust network analysis with subjective logic. In *Proceedings of the 29th Australasian Computer Society Conference*, Hobart, January 2006.
20. P. N. Judson, J. Fox, and P. J. Krause. Using new reasoning technology in chemical information systems. *Journal of Chemical Information and Computer Sciences*, 36:621–624, 1996.
21. A. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *2nd International Conference on Autonomous Agents and Multi-Agent Systems*, New York, NY, 2003. ACM Press.

22. S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The EigenTrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th World Wide Web Conference*, May 2004.
23. C. Karlof and D. Wagner. Secure routing in wireless sensor networks: attacks and countermeasures. *Ad Hoc Networks*, 1:293–315, 2003.
24. Y. Katz and J. Golbeck. Social network-based trust in prioritized default logic. In *Proceedings of the 21st National Conference on Artificial Intelligence*, 2006.
25. T. Khopkar, X. Li, and P. Resnick. Self-selection, slipping, salvaging, slacking and stoning: The impacts of. In *Proceedings of the 6th ACM Conference on Electronic Commerce*, Vancouver, Canada, June 2005. ACM.
26. B. Khosravifar, J. Bentahar, A. Moazin, and P. Thiran. On the reputation of agent-based web services. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 1352–1357, Atlanta, July 2010.
27. J. Lang, M. Spear, and S. F. Wu. Social manipulation of online recommender systems. In *Proceedings of the 2nd International Conference on Social Informatics*, Laxenburg, Austria, 2010.
28. K. Lerman and A. Galstyan. Analysis of social voting patterns on Digg. In *Proceedings of the 1st Workshop on Online Social Networks*, Seattle, August 2008.
29. C-J. Liau. Belief, information acquisition, and trust in multi-agent systems — a modal logic formulation. *Artificial Intelligence*, 149:31–60, 2003.
30. P-A. Matt, M. Morge, and F. Toni. Combining statistics and arguments to compute trust. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagents Systems*, Toronto, Canada, May 2010.
31. S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.
32. L. Mui, M. Moteashemi, and A. Halberstadt. A computational model of trust and reputation. In *Proceedings of the 35th Hawai'i International Conference on System Sciences*, 2002.
33. S. Naylor. *Not a Good Day Day to Die: The Untold Story of Operation Anaconda*. Berkley Caliber Books, New York, 2005.
34. N. Oren, T. Norman, and A. Preece. Subjective logic and arguing with evidence. *Artificial Intelligence*, 171(10–15):838–854, 2007.
35. S. Parsons and S. Green. Argumentation and qualitative decision making. In *Proceedings of the 5th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 1999.
36. S. Parsons, P. McBurney, and E. Sklar. Reasoning about trust using argumentation: A position paper. In *Proceedings of the Workshop on Argumentation in Multiagent Systems*, Toronto, Canada, May 2010.
37. S. Parsons, Y. Tang, E. Sklar, P. McBurney, and K. Cai. Argumentation-based reasoning in agents with varying degrees of trust. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems*, Taipei, Taiwan, 2011.
38. Simon Parsons, Elizabeth I. Sklar, Jordan Salvit, Holly Wall, and Zimi Li. ArgTrust: Decision making with information from sources of varying trustworthiness (Demonstration). In *Proceedings of Autonomous Agents and Multiagent Systems (AAMAS)*, St Paul, MN, USA, May 2013.
39. P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In M. R. Baye, editor, *The Economics of the Internet and E-Commerce*, pages 127–157. Elsevier Science, Amsterdam, 2002.

40. R. Stranderson, M. de Weerd, and C. Witteveen. Fuzzy argumentation for trust. In F. Sadri and K. Satoh, editors, *Proceedings of the Eighth Workshop on Computational Logic in Multi-Agent Systems*, volume 5056 of *Lecture Notes in Computer Science*, pages 214–230. Springer Verlag, 2008.
41. Y. Sun, W. Yu, Z. Han, and K. J. R. Liu. Trust modeling and evaluation in ad hoc networks. In *Proceedings of the 11th Annual IEEE Global Communications Conference*, pages 1862–1867, 2005.
42. Y. Tang, K. Cai, P. McBurney, E. Sklar, and S. Parsons. Using argumentation to reason about trust and belief. *Journal of Logic and Computation*, 22(5):979–1018, 2012.
43. Y. Tang, K. Cai, E. Sklar, and S. Parsons. A prototype system for argumentation-based reasoning about trust. In *Proceedings of the 9th European Workshop on Multiagent Systems*, Maastricht, Netherlands, November 2011.
44. Yuqing Tang, Elizabeth I. Sklar, and Simon Parsons. An Argumentation Engine: ArgTrust. In *Proceedings of the Workshop on Argumentation in Multiagent Systems (ArgMAS) at Autonomous Agents and MultiAgent Systems (AAMAS)*, Valencia, Spain, June 2012.
45. P. Tolchinsky, S. Modgil, U. Cortes, and M. Sanchez-Marre. Cbr and argument schemes for collaborative decision making. In *Proceedings of the First International Conference on Computational Models of Argument*, pages 71–82, Liverpool, UK, 2006.
46. S. Villata, G. Boella, D. M. Gabbay, and L. van der Torre. Arguing about the trustworthiness of the information sources. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, Belfast, UK, 2011.
47. G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logics in Artificial Intelligence*, 2000.
48. R. Walton, C. Gierl, H. Mistry, M. P. Vessey, and J. Fox. Evaluation of computer support for prescribing (CAPSULE) using simulated cases. *British Medical Journal*, 315:791–795, 1997.
49. Y. Wang and M. P. Singh. Trust representation and aggregation in a distributed agent system. In *Proceedings of the 21st National Conference on Artificial Intelligence*, Boston, MA, 2006.
50. B. Yu and M. Singh. Distributed reputation management for electronic commerce. *Computational Intelligence*, 18(4):535–349, 2002.