A descriptive, practical, hybrid argumentation model to assist with the formulation of defensible assessments in uncertain sense-making environments: an initial evaluation

Celeste Groenewald1, Simon Attfield1, Peter Passmore1, B. L. William Wong1, Nadeem Qazi1, Neesha Kodagoda1

Received: 1 January 2018 / Accepted: 2 June 2018 / Published online: 15 June 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

Abstract
This paper presents the preliminary results of our initial, descriptive, practical, hybrid argumentation model, designed for the use by criminal intelligence analysts (from now on referred to as analysts) working with sophisticated visual analytical software in uncertain sense-making environments. Analysts are required to create exhibits (as evidence) for a court of law or as input for decision-making in intelligence-led policing. These exhibits are required to be accurate, relevant and unbiased. Eight experienced criminal intelligence analysts from West Midlands police and the Belgium police evaluated a low-fidelity prototype resembling the first-order argumentation concepts of our initial argumentation model. The evaluation was to assess the applicability and practicality of the first-order argumentation concepts within our model. The preliminary results presented in this paper indicate that most of the first-order argumentation concepts are both applicable and practical and that the participants would use such concepts to construct their rationale from the onset of an analytical activity, if it were included as part of a software application.

Keywords Argumentation · Defensible assessment · Uncertainty · Sense-making · Criminal intelligence analysis

1 Introduction
We have created our initial, descriptive, practical, hybrid argumentation model by combining the research results from numerous previous studies. These previous studies are outlined in Sect. 2. Our model is considered to be descriptive as it describes the argumentation concepts that we think should be considered, to be of practical use in uncertain sense-making environments. We have researched criminal intelligence analysis as an example of an uncertain sense-making environment. Our model is also considered to be a hybrid model as it includes cognitive, schema and task/question-oriented considerations, which do not normally form part of an argumentation model. Hybrid in this context also differs in meaning of Bex et al.’s (2006, 2013) notion of hybrid. Their hybrid argumentation model consists of narrative (Pennington and Hastie 1992), which is used to explain the causality of events that are anchored (justified) in evidence. Bex et al.’s (2006) argumentation model is known and referred to as anchored narratives and this is a term that originated from the work of Waggenaar et al. (1993).

Argumentation maps are used to visually represent the logic of an argument and usually take the form of a hierarchical diagram. Wigmore (1931) used this visual layout to represent evidence and forces. Evidence is the statements or assertions made and the forces are the degree to which the evidence supports (affirmatory) or opposes (negatory) other statements. Both the affirmative and negatory forces play a role in determining the relevance of evidence, as the affirmative supports the argument and the negatory supports the counter-argument (rebuttal). For this reason, lawyers are
still focused on determining and representing evidence and the relevance of evidence when constructing their arguments (Allen et al. 2015).

There are, however, drawbacks with Wigmore’s (1931) proposed layout, as it is considered to be difficult to learn and construct. Toulmin (2003) addressed some of these difficulties and proposed a much simpler and general layout. His layout represents distinct argumentation concepts such as the claim, datum, warrant, backing, qualifier and the rebuttal. The warrant and the backing are considered to illustrate the reasons for why the datum supports the claim, therefore, illustrating the relevance for choosing that datum as evidence. The qualifier represents the degree to which the datum, warrant and backing support the claim, therefore, illustrating the degree of certainty.

Within the evidential reasoning environment, the argumentation maps that are being constructed benefits from having a pre-defined set of information sources (such as the trial case files) and the conclusion that needs to be supported (such as a guilty or an innocent verdict). We have found that in highly uncertain sense-making environments, the criminal intelligence analysts seldom begin their analysis with having sufficient information from which to base their conclusions. This causes criminal intelligence analysts to postpone the process of constructing an argument until the end of the analysis and the sense-making phase. This highlighted to us the need for extra support in constructing arguments in environments where information is absent, ambiguous, incomplete and where it is unclear what the conclusions may turn out to be.

We reckon that if we could tailor the construction of an argument, within software applications, to the practical needs of our analysts, then we could assist them with creating an argument from the onset of their analytical and sense-making activities. The practical needs that we identified as important to support, are the analysts’ activities of observation, conceptualisation and tracing of significant information. This could aid the process of determining relevance as an argumentation concept. We could also possibly assist analysts with their process of determining certainty as an argumentation concept, which could be regarded as the equivalent to Toulmin’s qualifier concept. This could be achieved by supporting analysts’ activities of observation, conceptualisation and tracing of the various fluctuations that could occur within the certainty levels of analysts’ tacit thinking and inferential processes. Certainty in this instance refers to the analysts’ certainty of the outcomes they have achieved during any point of their sense-making process, rather than the certainty of the documents or intelligence they have received. The certainty of intelligence and documents are supported through the use of the 5 × 5 × 5 model (Centrex 2007). Fluid schematisations and fluid process flows could assist the analyst with taking up argumentation from the onset of sense-making activities, as it may be seen as an aid to the sense-making process, rather than a possible hindrance to the flow of thinking and exploration activities.

The next section outlines the relevant literature and research considerations that we have used to develop our initial, descriptive, practical, hybrid argumentation model.

2 Literature

We started our research on how to design argumentation software in highly uncertain sense-making environments, by investigating the cognitive aspect of our analysts. The results are depicted as a sense-making triangle (see Fig. 1)
and incorporate the research of Wong (2014), Wong and Kodagoda (2016) and Gerber et al. (2016).

The inner triangle is the inference triangle and describes the process of inference-making as a combination of deductive, inductive and abductive inferential process types that are interlinked. The active inferential process type at any particular point in time is influenced by a combination of factors, such as the analyst's experience, domain and situation knowledge as well as the availability of information. Each factor determines which inferential process type will be at the forefront of the analyst's thinking and reasoning process. The second triangle is the anchoring triangle and describes the sense-making process in terms of anchoring, laddering and associative questioning (Wong and Kodagoda 2016). Gerber et al. (2016) added the third insight triangle, by describing the role that intuition and leap-of-faith play in highly uncertain sense-making environments to gain insight. All three triangles may work together in a complex combination of processes and sequences and may form an integral part of the tacit processes of human thinking and reasoning. The work of Wong (2014) and Wong and Kodagoda (2016) relies greatly on the foundational research of Klein et al. (2007) and Khaneman (2011) on the human thinking processes.

Passmore et al. (2015) described the importance and functionality of different types of evidential structuring and reasoning approaches, as found within a wide set of literature dedicated to the research of uncertain sense-making environments. The evidential structuring and reasoning approaches described by Passmore et al. (2015) encompassed argumentation schemas (Wigmore 1931; Wage-naar et al. 1993; Toulmin 2003; Bex et al. 2006; Allen et al. 2015), narrative (Pennington and Hastie 1992; Rao 2003; Bruner 2004; Segel and Heer 2010; Attfield and Blandford 2011; Chapin et al. 2013) and thematic sorting (Pirolli and Card 2005; Attfield and Blandford 2011; Rooney et al. 2014) and the role that each served during sense-making and analytical activities. Passmore et al. (2015) urged that any software application design that is aimed at supporting thinking and reasoning throughout analytical and problem-solving tasks, should incorporate a hybrid of structuring and reasoning approaches.

Selvaraj et al. (2016) conducted research in understanding crime schematisation within criminal intelligence analysis and proposed the concept of “think steps”. “Think steps” has been defined by Selvaraj et al. (2016) as “providing a template that allows the analyst to approach the case, decompose it into separate elements and classify associated data accordingly”. Although Klein et al. (2007) defined the use of frames (as schematisation) in their data frame model, Selvaraj et al. (2016) have tailored their work specifically for the criminal intelligence analysis environment, thus allowing for greater understanding on how analysts think and work.

Klein and Klinger (1991) researched how people make decisions in natural settings where time constraints inhibited them from using deliberate analytical methods, such as Multi-Attribute Utility Analysis (MAUA) and Decision Analysis. Decisions in these situations where based on prior experiences to “meet the needs of the situation” and to “recognise and classify a situation” (Klein and Klinger 1991). This hinted towards the use of System 1 thinking where there was no time for the utilisation of System 2 thinking (Khaneman 2011). Klein and Klinger (1991) produced a macrocognition model, which explained that there are a multitude of factors which influenced decision making within a natural setting as appose to a controlled laboratory setting. The macro-cognitive model of Klein et al. (2003) outlines the supporting macro-cognitive processes as: maintaining common ground; developing mental models; turning leverage points into courses of management; uncertainty management; attention management; mental stimulation and storytelling.

‘Uncertainty management’ within criminal intelligence analysis has been researched by Groenewald et al. (2017a) and they outlined various problems (or issues) that an analyst could encounter as part of a crime schema and the properties of uncertainty that accompanied each. Each issue adds to the level of uncertainty in relation to analysts’ thinking and reasoning efforts. Issues can crop up during any phase of the analysis process, thus casting a long-lasting shadow of uncertainty onto the analyst’s mind. Analysts have thus developed expert strategies to work their way through each type of uncertainty (skepticism, suspiciousness, complexity, obscurity, disparity, gaps, misconceptions, exhausted options; errors (data quality) and mental blocks) as and when it surfaces during analytical activities.

Moore and Dunham (1995) refers to ‘attention management’ within coordinated activities as, “when team members help each other direct their attention to signals, activities and changes that are important” (in Klein et al. 2005). As our analysts work mostly independently, the system could take on the role as an assisting team member and assist the analyst with attention management. Groenewald et al. (2017b) concentrated on how analysts observe significant information and categorised it under attention management, because the analyst is required to decide which information is important and worthy of their attention at different stages of the analysis process.

Other research related to this area is based on Rasmussen and Goodstein's (1985) decision tree, which attempts to explain the process from the point of becoming alert of new information to the point of making the decision of either acting or ignoring the information. This research is popular in research areas related to control management systems, such as found within the railway industry. Dadashi et al. (2017) asked the question whether too much information is
hindering railway operators from making effective decisions. Software applications in these environments often have alarms that signal when automatically collected field data reaches pre-defined thresholds. The users in these communities are required to decide if the alarm is valid for concern and if it requires action, thus contributing to the concept of ‘attention management’.

Our user base is faced with a different challenge where they often lack information. They often have to dig for information and try to identify what may be important, which may be a difficult task for a software programme to assist with through the use of alarms. It is, therefore, a different problem, but is still related to the same concept of ‘attention management’ and the question on how to design software applications that could assist users in these scenarios.

Klein et al. (2005) explained that common ground refers to, “the pertinent mutual knowledge, mutual beliefs and mutual assumptions that support interdependent actions in some joint activity.” The analysts we interviewed were mostly single operators working individually on sense-making tasks. We did, however, consider that they might have a need to manage significant information during the sense-making task. This could loosely be considered as a type of common ground between the analyst, their thoughts and the outputs from the various systems they use, as analysts need to keep track of their many different findings and attempts to solve the sense-making problem they are presented with. Two separate studies have been conducted. The first covered the management of significant information and is described by Groenewald et al. (2017b) as a lifecycle consisting out of cataloguing, comparing and tracking activities. The second study was by means of an interview with an experience criminal intelligence analyst from West Midlands police regarding the purpose of the Day Book (Groenewald et al. 2018).

The studies outlined above by Wong (2014), Passmore et al. (2015), Wong and Kodagoda (2016), Gerber et al. (2016), Selvaraj et al. (2016) and Groenewald et al. (2017a, b, 2018) all contributed to the construction of our initial, descriptive, practical, hybrid argumentation model. Figure 2 illustrates our model and each colour coding refers to the research contribution of each of the mentioned research teams.

![Fig. 2 Depiction of our initial, descriptive, practical and hybrid argumentation model with research colour codings (Groenewald et al. 2018)](image-url)
• Study A (purple coding): interview with an experienced operational intelligence analyst from West Midlands Police, UK (Groenewald et al. 2018)
• Study B (green coding): think-steps: a field study with four criminal intelligence analysts (Selvaraj et al. 2016).
• Study C (yellow coding): sense-making triangle: multiple qualitative studies to determine how criminal intelligence analysts think. (Wong 2014; Wong and Kodagoda 2016; Gerber et al. 2016)
• Study D and E (blue and pink coding): sense-making issues and managing significance: qualitative studies of eleven cognitive task analysis interviews with five experienced operational intelligence analysts (Groenewald et al. 2017a, b)
• The white areas are covered by the literature.

Figure 3 depicts each of the first-order argumentation concepts that we translated into ‘widgets’ to simulate a web-based application as part of our low-fidelity prototype.

We wanted to evaluate the applicability and practicality of the first-order argumentation concepts in our initial argumentation model and the next section describes the methodology we used for the evaluation study.

3 Methodology

The evaluation study spanned over 2 days with eight experienced criminal intelligence analysts from West Midlands police (UK) and the Belgium police. Each evaluation was 19 min in duration. The aim of the evaluation study was to determine the applicability and practicality of the argumentation concepts that we have previously constructed through the course of multiple preceding research studies (see Sect. 2). The first-order argumentation concepts were presented to the participants by means of a low-fidelity prototype, which consisted of grouped text boxes in a Microsoft Word document. The sections below describe the setup of the evaluation study and the evaluation proceedings.

3.1 Evaluation study setup

3.1.1 Implementation and representation of our initial, practical, hybrid, argumentation concepts

The first-order argumentation concepts were presented in a Microsoft Office (MS) word document. The first-order argumentation concepts are: Terms of Reference (TOR); Request for Information (RFI); Question, task/sub-task, visualisation/source, assertion, conclusion, schema/think-steps/theme, product and entity list. Each first-order argumentation concept was created by means of grouping text boxes together to give the visual effect of widgets in a web-based application (see Fig. 3). Each first-order argumentation concept was displayed at the top of the MS word document to simulate icons on a horizontal toolbar. In addition to the first-order argumentation concepts, we have included a range of connector icons such as arrows of various colours (black, red and green) and styles (solid and dashed lines). The MS word document acted as a low-fidelity prototype.

3.1.2 Physical setup

A large high-definition touch monitor was attached to a laptop with extended display. The participant had his/her own ‘qwerty’ keyboard and mouse. The MS word document was

![Fig. 3](image-url) Representation of each first-order argumentation concept as a ‘widget’ in our low-fidelity prototype
maximised on the external display monitor. The screen capturing software was visible on the laptop’s display.

3.1.3 Recording media

Each participant’s interactions with the prototype were recorded using BB Flashback Recording Software (Blueberry Software 2017). It captured the participant’s screen interactions. In addition to this, the participant’s voice was recorded using a separate voice recorder. The entire session was also recorded using a high-definition video recorder on a tri-pod, which recorded the participant’s screen over his/her right shoulder. This was a backup in case the computer recording software or the voice recorder failed.

3.2 Evaluation study proceedings

3.2.1 Study process

At the beginning of the study, each participant was given a consent form to sign and was made aware that the screen and the voice of the participant would be recorded. Due to the nature of the work that the participants are involved in, we attempted not to include their faces in the video recordings. The researcher gave each participant approximately 15 min overview of the different first-order argumentation concepts and the data that were available to the participant during the scenario. Only the column headers of the data were explained to the participants and not the actual content of the data. The participants were instructed to create a copy of each argumentation concept (as widgets) in the low-fidelity prototype that they wanted to work with. As we were interested in the applicability and practicality of the argumentation concepts, we allowed the participants to alter the ‘widgets’ as they deemed necessary. This allowed us to reduce the amount of possible participant frustration with the limited functionality and simplicity of the prototype we used. We also allowed the participants to add the functionality that they deemed necessary and to alter the widgets to suit their needs. By doing so allowed us (the researchers) the possibility for discovering concepts that our model did not include (or missed) as a result of our previous research activities. Each participant was given a further 60 min to work through the given scenario. The participants were allowed to stop at any given time. Participants were asked to use the think-aloud protocol (Kussmaul and Tirkkonen-Condit 1995), so that we could attempt to capture their thinking and reasoning processes as they worked with the low-fidelity prototype. By using the think-aloud protocol, we also hoped to be able to determine if their words match their actions on the screen. After the study, the participant was asked to complete a questionnaire consisting of numerous Likert scales (Likert 1932) to assess the participant’s thoughts on the applicability and practicality of the argumentation concepts. After the questionnaire, each participant was asked a few open-ended questions, based on the observations from the researcher during the study.

3.2.2 Scenario

The fictional scenario made use of anonymised crime data from West Midlands police. Due to the limited functionality of the prototype, the researchers only used a small subset of the available data. The scenario described an assault case where the victim identified the main offender. The main offender denied the attack. The second offender could not be identified by the victim, as the second offender wore a ski-mask and did not say anything. The participant was given a list of possible offenders based on their offending location (they committed crimes in the same location as the main offender). The information was displayed in a MS Excel spreadsheet where each offender’s information was on a different worksheet. The available data for each offender was: name, surname, gender, age, sex, town, street, postcode, beat number, victim profile, type of assaults, location coordinates, distance from offence and crime record numbers. There were a total of eight offenders of which seven were male and one was a female. The participant was instructed to use the available data to recommend who the likely second offender was and to use the prototype (MS Word document) to record their rationale.

3.2.3 Limitations

The low-fidelity prototype was supposed to have run alongside the main Visual Analytics for Sense-Making in Criminal Intelligence Analysis (VALCRI) application, to allow participants to make screen captures of the available visualisations (maps, timelines, charts, etc.) and to use those screen captures as part of their rationale. Unfortunately, not all of the equipments made it to the experiment location. Due to the limited functionality of the prototype, the researchers only used a small subset of the available data. The selection of crime records for the study was not available on the day that the study was suppose to have taken place. This resulted in abandoning the VALCRI application altogether and to load the crime data into an MS Excel spreadsheet. Although this was unfortunate, the participants were already very familiar with the MS Word and the MS Excel applications, which increased the amount of time the participant spent with the prototype as opposed to learning new functionality. By only having the crime data available in a spreadsheet, inhibited the participants from using the first-order visualisation argumentation concept as part of their rationale. The modus operandi (MO) details were also not available as part of the anonymised crime record selection and the...
participants were informed of this limitation. The limitations turned out to be a positive effect, as it increased the uncertainty and complexity of the data, thus forcing the participants to think of alternative possibilities and solutions. The ‘qwerty’ keyboard hindered the Belgium participants from typing frustration-free and as researchers, we should make attempts in the future to assist our participants with having region-friendly equipment. The next section outlines the preliminary results of our study.

4 Results

The main research question was: Did our initial, practical, hybrid argumentation model capture the first-order concepts correctly? To assess this, we made use of a Likert scale (Likert 1932) questionnaire which captured the participant’s views on the following categories: about each participant, general details, concept applicability, sense-making orientation, sense-making exploration, sense-making verification, practical use and other feedback.

4.1 Score calculation

As we had too few participants to perform statistical analysis on the data, we used a simplistic score calculation to visualise the degree of consensus from the participants. This is purely used for visualisation purposes and not to replace any statistical analysis that one would normally provide with larger population sizes. Each Likert scale was assigned a value ranging from −2 to 2, where −2 represents ‘strongly disagree’, 0 represents ‘neither agree or disagree’ and 2 represents strongly agree. The score for a particular question is calculated as the sum of the values each participant provided.

An example to illustrate the scores would be as follow: For a given question, participants 1, 2 and 3 answered ‘disagree’ which sums up to −3; participants 4, 5 and 6 answered ‘strongly disagree’ which sums up to −6 and participants 7 and 8 answered ‘agree’ which sums up to 2. The total score would equate to −7. Figure 4 illustrates how the scores could be visually presented on a xy-axis. −7 is plotted in red on the negative side of the x-axis. The y-axis represents each question on the questionnaire. This should indicate to the reader that most participants are in consensus in disagreeing with question 1. Although question 3 is also on the negative side of the x-axis, it is closer to zero and should be indicative that the participants had a less stronger consensus on the question. We hope that these visualisations assist our readers with the interpretation of the results. Having said that, our participants are experienced criminal intelligence analysts with 74 years of experience between them, so we highly regard and appreciate their feedback.

Below follows a brief set of results for each section of the questionnaire.

4.2 About each participant

This section was used to capture the experience and job function of the participants. The eight participants had a combined on-the-force experience of 97.5 years and a combined experience of 74 years as analysts. Of the eight participants, two were in managerial roles with no experience as analysts. By having two participants in managerial roles, we were able to consider the different perceptions on perceived value of our prototype. Table 1 below provides a summary of the participant’s roles and expertise.

![Fig. 4 Example of visualising questionnaire scores](image)

| Table 1 Questionnaire results—Sect. 4.2 |
|----------------------------------------|
| Question | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 |
|-----------|----|----|----|----|----|----|----|----|
| Main job function | Operational crime analyst | Chief of analyst service | Higher analyst | Strategic/ performance analyst | Deputy chief Of department | Strategic analyst | Head of department | Operational crime analyst |
| Experience in main job function (in years) | 24 | 15 | 15 | 4 | 0.5 | 1 | 38 | 24 |
| Experience as an analyst (in years) | 12 | 15 | 23 | 11 | 0 | 1 | 0 | 24 |
4.3 General details

This section was used to determine if analysts normally record their rationale and if so, do they do it from the onset or after the analysis? We used the term ‘rationale’ as opposed to ‘argument’ to fit in with the vocabulary that our participants understand and use. Our main focus was to understand if our participants normally record their rationale from the onset of their analytical activities, as we strive to create an argumentation model that is practical enough to be used from the onset of sense-making tasks. The reason for this being that argumentation maps from the evidential reasoning literature, requires evidence/facts to be available before the logic of an argument can be externalised in a visual format. In uncertain sense-making environments, the analysts are first required to evaluate many different data sources, before they can offer propositions, assertions, facts or evidence.

The questions presented to the participants were:

1. I normally record most of my task procedures during my analysis task.
2. I normally record most of my questions during an analysis task.
3. I normally construct an argument (rationale) from the onset (beginning) of the analysis task.
4. I normally construct an argument (rationale) at the end of the analysis task.
5. I normally find it easy to know what I’ve rationalised and what the justification for the rationale was.
6. I can normally return to an analysis task after period of absence (e.g. weekend, weeks) and easily resume the activities without much effort in understanding what I’ve rationalised previously.

Table 2 below summarises the results of Sect. 4.3 of the questionnaire and Fig. 5 is a graphical representation of the degree of participant consensus, based on the calculated score (see score calculation at the beginning of Sect. 4 for more details on how the score was calculated).

The results indicate that overall, the participants did not normally record their tasks and questions during analysis (see Q1). Participant 2 explained that the degree of rationale depends on the nature and difficulty of the analytical task they faced in the workplace. This is outlined in the following comment, “P2: The answers from above depend on the sort of analysis task we have. Going from more strategic to more tactical and the degree of difficulty of the question”.

The strategic analysts (P4 & P6) did not construct their rationale from the onset, whereas the other participants did (see Q3). Participant 4 outlined that the externalisation of their rationale at each stage of the analysis would be very time consuming as outlined in the following comment, “P4: Too time consuming to record rationale at each stage of analysis”. Most participants indicated that they construct their rationale at the end of their analysis (see Q4), which is expected, as they would require to first find the evidence/facts to construct their argument/rationale/ feedback to communicate it to their relevant audiences.

Some participants found it normally easy to know what their rationale was and the justifications (see Q5), others were on the fence, as indicated by the ‘neither agree or disagree’ scale. Similar feedback was given for resuming analytical tasks after a period of absence (see Q6).

Participant 7, who has a managerial position, emphasised that this section was not relevant to him/her as he/she does not carry out analytical tasks. This is illustrated by the following comment, “P7: For me really difficult to answer, because I never do an analyst’s job!”

![Table 2 Questionnaire results—Sect. 4.3](image_url)
4.4 Concept applicability

This section was to determine if the first-order argumentation concepts that we created were applicable to the participants. This assists with the justification of the validity of the argumentation concepts that we have created and published previously (Groenewald et al. 2018).

The entity list was not initially specified as a first-order argumentation concept, as we were unsure of the degree of practicality and applicability of the concept. As we have previously published work on the lifecycles of entities (Groenewald et al. 2017b), we made the decision to upgrade the entity list to a first-order concept for this study, to evaluate its stance as part of our argumentation concept map. This decision proved to be fruitful as the results in subsequent sections will show.

The questions presented to the participants were:

1. The ‘Terms of Reference’ concept was applicable to me.
2. The ‘Request for Information’ concept was applicable to me.
3. The ‘Question’ concept was applicable to me.
4. The ‘Task/Sub-Task’ concept was applicable to me.
5. The ‘Visualisation/Source’ concept was applicable to me.
6. The ‘Assertion’ concept was applicable to me.
7. The ‘Conclusion’ concept was applicable to me.
8. The ‘Schema/Think-Steps/Theme’ concept was applicable to me.
9. The ‘Product’ concept was applicable to me.
10. The ‘Entity List’ concept was applicable to me.

Table 3 summarises the results of Sect. 4.4 and Fig. 6 is a graphical representation of the degree of participant consensus, based on the calculated score (see score calculation at the beginning of Sect. 4 for more details on how the score was calculated).

| Q# | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | Score |
|----|----|----|----|----|----|----|----|----|-------|
| 1  | –  | A  | SA | A  | D  | A  | SA | –  | 6     |
| 2  | A  | A  | A  | –  | A  | A  | SA | A  | 6     |
| 3  | A  | A  | A  | SA | A  | D  | SA | A  | 8     |
| 4  | A  | –  | SA | SA | A  | A  | A  | –  | 8     |
| 5  | SA | D  | –  | –  | D  | –  | A  | –  | 1     |
| 6  | SA | A  | –  | A  | A  | A  | A  | –  | 1     |
| 7  | SA | A  | –  | –  | A  | A  | –  | –  | 5     |
| 8  | SA | A  | SA | –  | A  | SA | A  | –  | 9     |
| 9  | SA | –  | –  | D  | D  | –  | A  | –  | 1     |
| 10 | SA | A  | SA | D  | A  | –  | A  | –  | 6     |

The results indicate that the participants found all of the first-order concepts applicable. The exceptions were the visualisation and the product concepts. As outlined in Sect. 2, we expected that the visualisation concept would score low due to the absence of maps, graphs and other visual analytical aids. The task that we have given to the participants did not indicate that they had to create a report, so the participants did not have a need to group different analytical results together to create products.

We were pleased to see that the entity list was well received and that most participants used it as a central point for constructing their rationale. Participant 7 did not use the entity list at all, but afterwards explained that he/she missed it, because the MS Word document was zoomed in to such a degree that it concealed some icons from view. This is outlined in the following comment, “P7: Entity list and Schema/ThinkSteps I missed, but would use them probably the most.” Participant 4 indicated that the entity list required more options, which means that participant 4 wanted to easily modify the list with the required information. This is outlined by the following comment, “P4: Entity List needs more options”. Participant 4 also indicated that the look and feel of some of the first-order concepts were similar and could
possibly be represented as a single icon, with the functionality to toggle between them. This is outlined by the following comment, “P4: Terms of Reference (TOR) and Request for Information (RFI) one similar. Assertion and Conclusion is similar”.

### 4.5 Sense-making orientation

This section was to determine if using the concepts would assist the participants with orienting themselves in the analysis process. This refers to observation, conceptualisation and tracing activities during their analysis.

The questions presented to the participants were:

1. The ability of creating concepts assisted me with getting started.
2. The ability of creating concepts assisted with dividing the analytical problem into manageable pieces.
3. The ability of creating concepts assisted me with providing an overview of my analytical process.
4. The ability of creating concepts assisted me with keeping track of where I was in my analytical process.

Table 4 summarises the results of Sect. 4.5 and Fig. 7 is a graphical representation of the degree of participant consensus, based on the calculated score (see score calculation at the beginning of Sect. 4 for more details on how the score was calculated).

The results indicate that the participants perceived the first-order argumentation concepts as helpful in assisting them with getting started. The participants also found the concepts useful as an aid to divide the analytical problem into manageable pieces, which assisted them with maintaining an overview of their findings and their analytical progress.

Some of the participants were also part of a second independent evaluation on the day, so they were already mentally exhausted by the time they sat down to evaluate our prototype. Although not desirable, it was interesting to see how the participants perceived the activity using the first-order argumentation concepts whilst being mentally exhausted. Participant 3 was one such participant who was already mentally taxed by the time he/she took part in our study. Although at a disadvantage, participant 3 was the only participant who, after reviewing his/her results, started to question the validity of the victim’s statement and provided an alternative scenario of a possible feud between the victim and the main offender. Participant 3 made the recommendation that the relationship between the victim (male), the main offender (male) and a third offender (female) from the data, be looked at further, as it may be a case of jealousy. Participant 3 commented that although he/she was tired, the concept of mapping out thoughts in the manner we provided was an interesting methodology to use. This is outlined in the following comment, “P3: I was in a circular process and getting tired, so I started to shortcut. However, mapping thoughts is an interesting methodology”.

Participant 4 indicated that the first-order argumentation concepts would assist with the scoping phase of a complex analytical problem and referred to it as a project plan. As participant 4 is a strategic analyst, it makes sense that a person is such a position would want to keep track of the intelligence they have and the options that allow them to consider to move forward. This is outlined in the following comment, “P4: I didn’t use this to ‘project plan’ the task, but with something more complex, I think this would be really useful in the scoping phase.”

Participant 7 indicated that he/she struggled to maintain an overview of their analytical task, but put it against his/her lack of experience. This is outlined in the following comment, “P7: Because my lack of experience it was difficult to maintain the overview.” This is a helpful piece of information, as it allows us with the opportunity to delve deeper into this participant’s results, to try and determine the areas where a less experienced analyst could struggle.

Table 4 Questionnaire results—Sect. 4.5

| Q # | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | Score |
|-----|----|----|----|----|----|----|----|----|-------|
| 1   | A  | A  | A  | A  | A  | A  | A  | –  | 7     |
| 2   | A  | A  | A  | SA | A  | A  | A  | –  | 8     |
| 3   | SA | A  | A  | –  | A  | A  | D  | A  | 6     |
| 4   | SA | A  | D  | SA | A  | A  | –  | A  | 7     |
4.6 Sense-making exploration

This section was to determine if the first-order argumentation concepts assisted the participants with exploring the analytical problem, such as in identifying next steps, important areas and maintaining an overall understanding of what was going on.

The questions presented to the participants were:

1. Having the concepts visible assisted me with identifying what the next analytical step should be (e.g. delve deeper or to start a new search).
2. Having the concepts visible assisted me with identifying important areas
3. Having the concepts visible assisted me with understanding what was going on

Table 5 summarises the results of Sect. 4.6 and Fig. 8 is a graphical representation of the degree of participant consensus, based on the calculated score (see score calculation at the beginning of Sect. 4 for more detail on how the score was calculated).

The results indicate that most participants found that the first-order argumentation concepts assisted them with determining what the next analytical step in their thinking process should be (see Q1). Most participants were also able to determine which areas of the analysis were important to them (see Q2) and made use of the available certainty indicators.

Having the concepts visible assisted all of the participants with understanding what was going on in their analysis (see Q3). Participant 2 found that the detail the concepts provided to be useful as indicated by the following comment, “P2: I’m not used to work in this kind of detail in describing think-steps, tasks and subtasks, but it is helpful to do it in this kind of detail”.

Participant 3 indicated that the certainty indicators were helpful in re-evaluating existing concepts as new information became available. This particular participant used both types of certainty indicators (sequential colour scale and check boxes). This is indicated by the following comment, “P3: I think having the chart/graph to colour as well as the certain/plausible/believable checkboxes as really helpful. I could change it as more information came in.”

Participant 4 indicated that by having the concepts visible assisted with identifying gaps in his/her analysis process and that it served as a prompt for new lines of enquiry. This is indicated by the following comment, “P4: Helped me identify which parts of the task I had missed. It provides useful prompts for lines of enquiry / methods of dissemination”.

4.7 Sense-making verification

This section was to determine if the concepts assisted the participants with verifying their rationale.

The questions presented to the participants were:

1. Having the concepts visible assisted me with differentiating between certain and uncertain areas.
2. Having the concepts visible assisted me with identifying where the gaps in my understanding was.
3. Having the concepts visible assisted me with identifying where my focus should be.
4. Having the concepts visible assisted me with identifying the weak areas in my rationale.
5. Having the concepts visible assisted me with identifying the strong areas in my rationale.

Table 6 below summarises the results of Sect. 4.7 and Fig. 9 is a graphical representation of the degree of participant consensus, based on the calculated score (see score calculation at the beginning of Sect. 4 for more details on how the score was calculated).

The results indicate that most participants felt that the first-order concepts assisted them with the verification process of their rationale. More participants were able to identify the strong areas (see Q5) of their rationale than the weak areas (see Q4), but they also indicated that they were able to identify gaps in their understanding (see Q2) and where to place their focus (see Q3). It appears that the certainty indicators that we have added to some of the concepts assisted most participants with differentiating between certain and uncertain areas of their rationale (see Q1). This is best

Table 5 Questionnaire results—Sect. 4.6

| Q # | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | Score |
|-----|----|----|----|----|----|----|----|----|-------|
| 1   | A  | A  | D  | A  | A  | A  | A  | D  | 3     |
| 2   | A  | A  | D  | A  | A  | A  | A  | D  | 4     |
| 3   | A  | A  | A  | A  | A  | A  | A  | A  | 8     |
illustrated with participant 4’s comment, “P4: Particularly the idea of how certain I was about my hypothesis”.

Participant 3 did indicate that the process of constructing a rationale, as we provided, was not playful enough. This is indicated in with the following comment, “P3: I think once I’d exhausted the data, I could then think freely at end. I didn’t feel like I did this in the given task - putting steps down felt too logical / not playful.”

4.8 Practical use

As our argumentation model is intended to be a practical model for use in the workplace, this section tried to assess how practical the concepts were and if the participants would like to have it as part of an analytical and sense-making application such as VALCRI.

The questions presented to the participants were:

1. I would use all of the concepts if they were available as part of an application.
2. I would only ever use some of the concepts, if they were available as part of an application.
3. I would not use any of the concepts, if they were available as part of an application.
4. I would prefer to use the concepts from the onset (beginning) of the analytical process, if they were available as part of an application.
5. I would prefer to use the concepts at the end of the analytical process, if they were available as part of an application.
6. Using the concepts would assist me with remembering what my rationale was and the justifications for each rationale, if they were part of an application.
7. Using the concepts would assist me with easily resuming my activities after a period of absence (e.g weekend/weeks) without much effort in understanding what my rationale was.

Table 7 summarises the results of Sect. 4.8 and Fig. 10 is a graphical representation of the degree of participant consensus, based on the calculated score (see score calculation at the beginning of Sect. 4 for more details on how the score was calculated).

The results indicate that most participants would use the first-order argumentation concepts as part of an application (see Q1 and Q2), but their use would depend on the type of analytical task they are presented with. This is outlined by participant 4’s comment, “P4: I would probably use in different ways depending on the task”. It was very clear that the participants would not ignore the concepts and that they would definitely make use of them as indicated by the feedback from question 3.
Although some participants indicated that they currently construct their rationale from the onset of an analytical problem (see question 3 in Sect. 4.3 where score = −1), the results from question 4 in this section, indicate that more participants started to favour the idea of creating a rationale from the onset of an analytical problem (score = 5).

There is also a favourable increase in the views of participants that the concepts could assist them with remembering what their rationale and the justification for their rationale was with a score of 8 (see contrast to question 5 in Sect. 4.3 where score = 3). There is also a favourable increase in the participants’ view that they could easily resume an analytical activity after a period of absence from the task with a score of 11 (see contrast to question 6 in Sect. 4.3 where score = 1).

This concludes the preliminary results of the evaluation of the first-order argumentation concepts in our initial, practical, hybrid, argumentation model. The next section outlines the discussion and the conclusion.

### 5 Discussion and conclusion

We have constructed our initial, practical, hybrid, argumentation model by combining the results of multiple studies that we have conducted over the past 3 years as part of the VALCRI project. This paper outlined the preliminary results of an evaluation study of the first-order argumentation concepts of our argumentation model. We have translated the first-order argumentation concepts into a visual representations (as widgets) as part of a low-fidelity prototype, which simulated a web-based application. Eight experienced criminal intelligence analysts from West Midlands police and the Belgium police evaluated the applicability and practicality of our first-order argumentation concepts.

The results indicate that most of the first-order argumentation concepts are both applicable and practical and that the participants would use such concepts to construct their rationale from the onset of an analytical activity, if it were included as part of an analytical and sense-making application such as VALCRI. The participants did stress that our prototype is in need of fine-tuning and that it needs to be more fluid in its functionality, such as with selecting and editing the argumentation concepts.

We still have a wealth of information to analyse before we can determine the actual value and success of our attempts to construct a practical, hybrid argumentation model for use in uncertain sense-making environments, such as found in criminal intelligence analysis. There are areas of our model that we need to revisit, such as the concept of ‘entities’ and if it should be upgraded to a first-order concept. We also need to re-evaluate our terminology for the concepts of proposition, assertion and hypothesis, as these appeared to have been used interchangeably or at least differently to what we anticipated. We also have a number of additional hypotheses that we can attempt to answer by exploring the results of our evaluation in more depth. One of our hypotheses is that when our argumentation model is used from the onset of an analytical activity, it should assist people with navigating their way through the sense-making problems and the uncertainties that they have (and come across) at different stages of their sense-making process. Although the preliminary results indicate that our argumentation concepts assisted the participants with their sense-making orientation, exploration and verification activities, we still need to delve deeper into the data before we can make any claims.

There are also opportunities to investigate the layout and the physical construction of each participant’s rationale using our first-order argumentation concepts, which could provide insights for software developers on how and what to include in their software. There is also room to investigate the certainty indicators that we provided and how the participants interacted with them.

We conclude this paper with a few final comments from our participants: “P2: Interesting exercise where I learned something I will use in the future”, “P3: I like this and it is similar to business modelling, so it is an interesting exercise. I would like to reuse this in the office”. “P4: New for me to provide such rationale. I think it would be very useful for investigation/evidential analysis for disclosure at court”. “P7: The concepts are a very interesting ‘concept’, which could be a winner when fine-tuned.”

### Acknowledgements

The research results reported here have received funding from the European Union Seventh Framework Programme (FP7/2007–2013) through Project VALCRI, European Commission Grant Agreement N° FP7-IP-608142, awarded to B.L. William Wong, Middlesex University and partners. We are also very grateful for the enthusiasm of the police analysts in sharing with us their experiences that made this report possible. A special thanks to Nadeem Qazi for his time and effort to get the evaluation data into a usable format.
References

Allen C, Taylor C, Nairns J (2015) Practical guide to evidence. Routledge, Abingdon
Attfield S, Blandford A (2011) Making sense of digital footprints in team-based legal investigations: the acquisition of focus. Hum Comput Interact 26(1–2):38–71
Bex F, Verheij B (2013) Legal stories and the process of proof. Artif Intell Law 21(3):253–278
Bex F, Prakken H, Verheij B (2006) Anchored narratives. Reason Evid Jurix 152:11–20. http://www.florisbex.com/papers/Jurix2006.pdf
Blueberry Software (2017) BB flashback recording software. https://www.flashbackrecorder.com/download/
Bruner J (2004) Making stories: law, literature, life. Harvard University Press, Cambridge (ISBN-13:978-067401099484)
CENTREX (2007) Centrex practice advice: introduction to intelligence-led policing. http://www.fairplayforchildren.org/pdf/1291430265.pdf
Chapin L, Attfield S, Okoro EM (2013) Predictive coding, storytelling and god: narrative understanding in e-discovery. British Journal of Industrial Relations
Dadashi N, Golightly D, Sharples S (2017) Seeing the woods for the trees: the problem of information inefficiency and information overload on operator performance. Cogn Technol Work 19(4):561–570. https://doi.org/10.1007/s10111-017-0451-1
Gerber M, Wong BLW, Kodagoda N (2016) How analysts think: intuition, leap of faith and insight. In: Proceedings of the human factors and ergonomics society annual meeting. 60(1):173–177. Sage CA; Los Angeles, CA: SAGE Publications
Goodstein LP, Rasmussen J (1985) Decision support in supervisory intelligence and security informatics conference (EISIC), 1985 European. pp 16–23. IEEE
Pennington N, Hastie R (1992) Explaining the evidence: tests of the story model for juror decision making. J Personal Soc Psychol 62(2):189
Pirolli P, Card K (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In Proceedings of international conference on intelligence analysis. 5:2–4
Rao S (2003) Making sense of making stories: law, literature, life. Law Libr J 95:455
Rooney C, Attfield S, Wong BLW, Choudhury S (2014) INVISQUE as a tool for intelligence analysis: the construction of explanatory narratives. Int J Hum Comput Interact 30(9):703–717
Segel E, Heer J (2010) Narrative visualization: Telling stories with data. IEEE Trans Vis Comput Gr 16(6):1139–1148
Selvaraj N, Attfield S, Passmore P, Wong BLW (2016) How analysts think: think-steps as a tool for structuring sensemaking in criminal intelligence analysis. In: Intelligence and security informatics conference (EISIC), 2016 European. pp 68–75. IEEE
Toulmin SE (2003) The uses of argument. Cambridge University Press, Cambridge
Wagenaar W, Koppen P, Crombag H (1993). Anchored narratives: the psychology of criminal evidence. https://www.researchgate.net/publication/232456113_Anchored_Narratives_The_Psychology_of_Criminai_Evidence
Wigmore JH (1931) The principles of judicial proof or the process of proof as given by logic, psychology, and general experience, and illustrated in judicial trials, 2nd edn. Little, Brown and Company, Boston
Wong BLW (2014) How analysts think (?): early observations. In: Intelligence and Security Informatics Conference (JISIC). pp. 296–299. IEEE
Wong BLW, Kodagoda N (2016) How analysts think: anchoring, lagging and associations. In: Proceedings of the human factors and ergonomics society annual meeting. 60(1):178–182. Sage CA; Los Angeles, CA: SAGE Publications
Wong BLW, Seidler P, Kodagoda N, Rooney C (2018) Supporting variability in criminal intelligence analysis: From expert intuition to critical and rigorous analysis. In: Leventakis G, Haberfeld MR (eds) Community-oriented policing and technological innovations. Springer, New York pp 64–84. https://doi.org/10.1007/978-3-319-89294-8
Klein G, Feltovich PJ, Bradshaw JM, Woods DD (2005) Common ground and coordination in joint activity. Organ Simul 53:139–184
Klein G, Phillips JK, Rall EL, Peluso DA (2007) A Data-frame theory of sensemaking. In Hoffman RR (Ed.) Expertise out of context proceedings of the sixth international conference on NDM. pp 113–155. Erlbaum
Kussmaul P, Tirkkonen-Condit S (1995) Think-Aloud protocol analysis in translation studies. TTR: Traduction, Terminologie, Réédition 8(1):pp 177–199. https://doi.org/10.7202/037201ar
Likert R (1935) A technique for the measurement of attitudes. Arch Psychology. 22(140):55
Moore C, Dunham PJ (eds) (1995) Joint attention: Its origins and role in development. Lawrence Erlbaum Associates, Hillsdale
Passmore PJ, Attfield S, Kodagoda N, Groenewald C, Wong BLW (2015) Supporting the externalisation of thinking in criminal intelligence analysis. In: Intelligence and security informatics conference (EISIC), 2015 European. pp 16–23. IEEE
Selvaraj N, Attfield S, Passmore P, Wong BLW (2016) How analysts think: think-steps as a tool for structuring sensemaking in criminal intelligence analysis. In: Intelligence and security informatics conference (EISIC), 2016 European. pp 68–75. IEEE
Toulmin SE (2003) The uses of argument. Cambridge University Press, Cambridge
Klein G, Ross KG, Moon BM, Klein DE, Hoffman RR, Hollnagel E (2003) Macrocognition. IEEE Intell Syst 18(3):81–85