EVIDENCE MARSHALLING WITH INFERENCE NETWORKS: AN APPLICATION TO HOMELAND SECURITY

Ken R. McNaught[(a)] and Peter Sutovsky[(b)]

Operational & Decision Analysis Group, Dept of Informatics & Systems Engineering,
Cranfield University, Defence Academy of the UK, Shrivenham

[(a)]K.R.McNaught@cranfield.ac.uk, [(b)]P.Sutovsky@cranfield.ac.uk

ABSTRACT
When trying to reason about some adversary's likely intentions, an intelligence analyst frequently needs to combine multiple pieces of evidence observed at different times and having different degrees of relevance, coming from sources with varying degrees of credibility. The evidence marshalling process concerns the structuring of evidence to help analysts and investigators organize their thinking and make better sense of a situation. Here we show how the qualitative structure of a Bayesian network offers a useful approach to evidence marshalling. We propose a framework consisting of four types of nodes, arranged in layers – hypothesis nodes, ground truth nodes, evidence nodes and credibility nodes. An example is presented in the context of homeland security.

Keywords: intelligence analysis, Bayesian network, decision support

1. INTRODUCTION

Not surprisingly, much has been written in recent years about intelligence failures and the need to improve the intelligence analysis process. A good proportion of this is uninformed and unhelpful speculation, grounded in perfect hindsight. However, such a turbulent period affords an opportunity for reflection by the intelligence agencies and it would be surprising if they did not feel that something useful could be learned from recent history.

In addition to the various organizational concerns which have tended to dominate the discussion, another potential avenue for improvement which has been suggested relates to the analyst reasoning process. While Heuer (1999) raised awareness of the potential deleterious effects of various cognitive biases on intelligence analyst reasoning some years ago, recent events, along with a generally wider acceptance of how important such cognitive effects can be (as witnessed by Daniel Kahneman's Nobel Prize for Economics in 2002, for example) have shone the spotlight back in this direction.

Following from Kahneman's observations of two modes of human reasoning, Wastell (2010) provides an excellent discussion of how these relate to the intelligence analyst. The natural reasoning mode can be viewed as a mixture of instinct and experience. When making decisions in this mode, we are making use of our gut feel for a situation and any experience we might have acquired of similar situations in the past, what Klein et al (1986) refer to as recognition primed decision-making. However, it can be shown that over-reliance on this natural reasoning mode quickly leads us astray. For example, many relatively simple problems involving probability or everyday calculations often defy our intuition. Furthermore, it is well known that most people's ability to process more than a few pieces of information at a time is severely limited (Miller 1956). When required to combine several pieces of information, this leads us to take shortcuts. In some repetitive, regular situations this may be sensible and lead to acceptable results but in others, particularly new situations where we may have little experience and where there is much uncertainty, it will lead to sizeable errors and misjudgements. It is in these situations where we need to employ our second 'systematic' mode of reasoning based on logic and rationality.

Furthermore, Wastell (2010) argues that it is here where there is a lack of formal methods to complement an analyst's in-built natural reasoning capability. Without such methods and sufficient training to accompany them so that they become second nature when required, analysts can end up over-relying on natural reasoning.

Evidence marshalling is the name usually given to methods which attempt to organize evidence in some systematic fashion, typically to aid sense-making and to support decision-making by analysts or investigators relating to the case in question. In this paper we outline an approach to evidence marshalling based on inferential networks. This builds on the work of Schum (2001) who has been instrumental in developing a science of evidence. In the next section, we consider this and related work. The following sections discuss a generic framework and present a particular example in the context of homeland security.

2. EVIDENCE MARSHALLING

An item of evidence has several characteristics. These include relevance to the issues being addressed, source credibility, particularly where human intelligence is involved, and timeliness or latency, since typically we might expect a newer observation to carry greater weight than an older one of the same type. In conflict situations when trying to reason about some adversary's likely intentions, it is frequently necessary to combine multiple pieces of evidence observed at different times and having different degrees of relevance, coming from sources with varying degrees of credibility. Identifying unreliable sources can be particularly important in reducing vulnerability to deception, as is identifying common or highly dependent sources. In such high-stakes situations, the pressure on investigators can be very high.

While traditionally, these investigators were largely expected to rely on their experience and intuition to make sense of a situation, there is now increasing interest in providing them with some form of decision support. This in no way diminishes the value of experience or the need for intuition. It simply recognises the scale of the task faced by the investigator and attempts to supplement their capabilities by utilising modern technology.

One relatively well-known method of systematic analysis, proposed by Heuer (1999) to overcome confirmation bias in particular, is Analysis of Competing Hypotheses (ACH). In uncertain situations, there are usually several plausible explanations for an action or an observation. From these, we can select a number of alternative competing hypotheses. However, human nature tends to make us look for evidence which confirms our favourite hypothesis rather than that which would disconfirm it or support other hypotheses. We might also place more weight on evidence which confirms our favourite hypothesis and less weight on evidence which casts doubt on it. The result is that we often stick with our early favourite hypothesis for too long, even when considerable disconfirming evidence is building up.

In relating items of evidence to multiple competing hypotheses, the intention is to keep minds open, and avoid getting stuck in a favoured hypothesis too early. Of note is the emphasis of the approach on the importance of negative evidence (i.e. the definite absence of some indicator) and the varying diagnosticity of different pieces of evidence, i.e. how well a given piece of evidence can discriminate between the hypotheses under consideration.

While it could certainly be claimed that ACH is a form of evidence marshalling, this latter term usually signifies a more comprehensive approach to relating evidence and hypotheses. For example, in the experimental, visual analytic 'Jigsaw' system (Stasko et al. 2008) developed to help intelligence analysts navigate a vast array of potentially relevant documents, provision is made for a 'shoebox' which is essentially an evidence marshalling tool. Such a tool can help an analyst to organize the available evidence, so aiding the construction of a coherent case.

The evidence marshalling process, described by Schum (2001), concerns the structuring of evidence to help investigators organize their thinking and make better sense of a case. It can include creative elements related to the construction of narratives or explanations, the identification and analysis of evidence gaps, and notions of evidence thresholds to take different actions. In the case of homeland security, these could include more intrusive surveillance or making an arrest. The potential for deception must also be considered explicitly (Elsaesser and Stech 2007).

Schum (2001) outlines a number of methods in tiers of varying complexity concerned with organizing evidence. One of the more complex is the Wigmore chart (Wigmore 1937). This was devised by the legal scholar John Wigmore in order to map the structure of a legal argument. As Schum rightly observes, such a construct bears considerable similarity to a particular type of modern probabilistic graphical model, usually known as a Bayesian network (Pearl 1988 and Jensen 2001). Although there are significant differences, both can be described as inferential reasoning networks.

In this paper, we explore the potential for Bayesian networks to be used as a tool for evidence marshalling in the context of homeland security. As well as organizing existing evidence, such a tool can help to encourage thinking about new avenues of enquiry, and highlight gaps in the evidential support for a hypothesis. With often very limited resources, support is required to identify the most promising lines of enquiry. In our view, decision aids such as the inferential reasoning networks presented here can help in such situations. Furthermore, they can help to address some of the problems and limitations encountered in the communication of uncertain information in the intelligence field as described, for example, by Weiss (2008).

3. PROPOSED FRAMEWORK

Figure 1 displays a proposed, generic inferential reasoning network. This consists of four layers of nodes: high-level hypotheses, expected ground truth activities related to these hypotheses, evidence (or perhaps its absence) relating to the expected activities, and finally evidence credibility nodes which help in distinguishing evidence collected from different sources.

The kind of propositions contained in the first three layers bear some similarity to the three levels of propositions discussed in forensic science – namely, crime level, activity level and source level (Taroni et al, 2006). In that domain, hypotheses at the crime level tend to revolve around the guilt or innocence of one or more suspects.

In the intelligence domain, the first layer of nodes contains the key, high-level hypotheses of interest to the analyst, e.g. the target of a planned attack, the intentions of a suspected terrorist cell or the role of an individual

in a terrorist organization. In each of these situations, there will be several possibilities, corresponding to the multiple competing hypotheses in the ACH method. While alternatives which are truly mutually exclusive would normally be accommodated within a Bayesian network as different states within a single hypothesis node, for this kind of application we propose that each competing hypothesis is explicitly represented as a separate node in the network. This makes the analyst keep the full set of possibilities in mind at all times and makes it easier to follow which items of evidence and sources relate to which hypotheses.

The second layer of nodes contains variables which represent a number of activities typically associated with the hypotheses under consideration. These should reflect the modus operandum of the terrorist organization and provide a bridge between the hypotheses and the observable evidence. Nodes in the second layer are usually considered to be not directly observable themselves, i.e. there might always be an element of doubt about their truth or falsity.

Here we have referred to nodes in the second layer in Figure 1 as 'ground truth' nodes. This is intended to convey the notion that such variables effectively describe the true nature of the situation. However, that, of course, is typically hidden from us for a long period of time and some ground truth variables may always remain a matter of dispute and never be revealed. As in the forensic science domain, we expect nodes at this level to most often be associated with activities of various types, e.g. training recruits, making explosives, and preparing for an attack.

In contrast, evidence nodes in the third layer correspond to observations that have been made or can be realistically expected to be made in a useful timeframe. Their observation, or for that matter lack of observation, will depend in a probabilistic sense on one or more ground truth variables. Nodes in this layer are directly observable and cover the gamut of evidence types including human intelligence, signals intelligence, imagery, etc. It is when we become aware of such evidence that we revise our beliefs in the second layer nodes and then, in turn, revise our beliefs in the top-level hypotheses. Naturally, some items of evidence will lead to greater revisions than others. Furthermore, an item of negative evidence, i.e. the definite lack of some expected indicator, should also lead to appropriate revisions in our beliefs.

The final set of nodes we have labelled 'credibility' nodes. Each evidence node which has been observed is associated with a corresponding credibility node. These recognise that some pieces of evidence are more trustworthy than others. This might be because of the source of the evidence, e.g. an experienced field agent vs an unknown informant vs an informant with a long track record. It can also reflect the circumstances in which the evidence was collected. A credibility node has not been attached to all evidence nodes. This is because some of them are simply included as potential evidence nodes which have not yet been resolved. The lack of some evidence item might be due to that item not having been searched for or because after searching it has not been found. Only in this latter case have we associated a credibility node with the absence of the item, which is then an observation in its own right. In these cases, the credibility rating would reflect both the difficulty and the thoroughness of the search. It is much more likely that an observed absence of an evidence item reflects its genuine absence when an exhaustive search has taken place rather than a superficial one.

In terms of its structure, this type of network is qualitatively the same as a Bayesian network (Pearl, 1988). As such it is possible to quantify the nodes with probability distributions to facilitate true probabilistic inference. However, this would require the elicitation of many uncertain probabilities and so is not recommended for routine application, although there may be occasions when it is desirable. Nonetheless, the BN's qualitative structure provides a logical framework for qualitative reasoning.

Within our proposed framework, both evidence nodes and credibility nodes can be opened within the software tool we are developing to store and retrieve information deemed relevant to them. For example, an evidence node may contain a detailed description of the evidence itself, the identity or other information about the source of the evidence, and links to relevant documents or images. A credibility node may contain information about the source's track record and the chain of custody which the item of evidence has experienced. Based on these, some overall assessment of the evidence's credibility may be recorded. Essentially this should moderate the extent to which our beliefs in higher level ground truth and hypotheses nodes are updated in the light of this item of evidence.

The workspace shown below the network represents the type of information that an analyst is prompted for by our prototype tool. This is a free text area which allows the investigator to record their beliefs as time progresses and evidence unfolds. There are three main categories – a summary of the analyst's current understanding of the situation, an analysis of evidence gaps and key uncertainties, and finally a list of actions required. This is intended to encourage thoughtful reflection as the evidential picture unfolds, as well as the explicit recognition and analysis of evidence gaps, contradictions and uncertainties, including possible deception activities by the adversary in question. Finally, the investigator is invited to make a list of required actions such as requests for additional information, requests for resources, suggested new leads to investigate, current leads to drop, etc, based on the foregoing analysis. This helps to create an audit trail, which can be time-stamped, of what was done, when and why during the course of an investigation, clearly linking these decisions with the beliefs and possibilities being considered at the time and providing a logical justification for them. Such an approach also supports collaborative working, making it easier for co-workers

and shift workers to understand each other and pick up where the other one has left off.

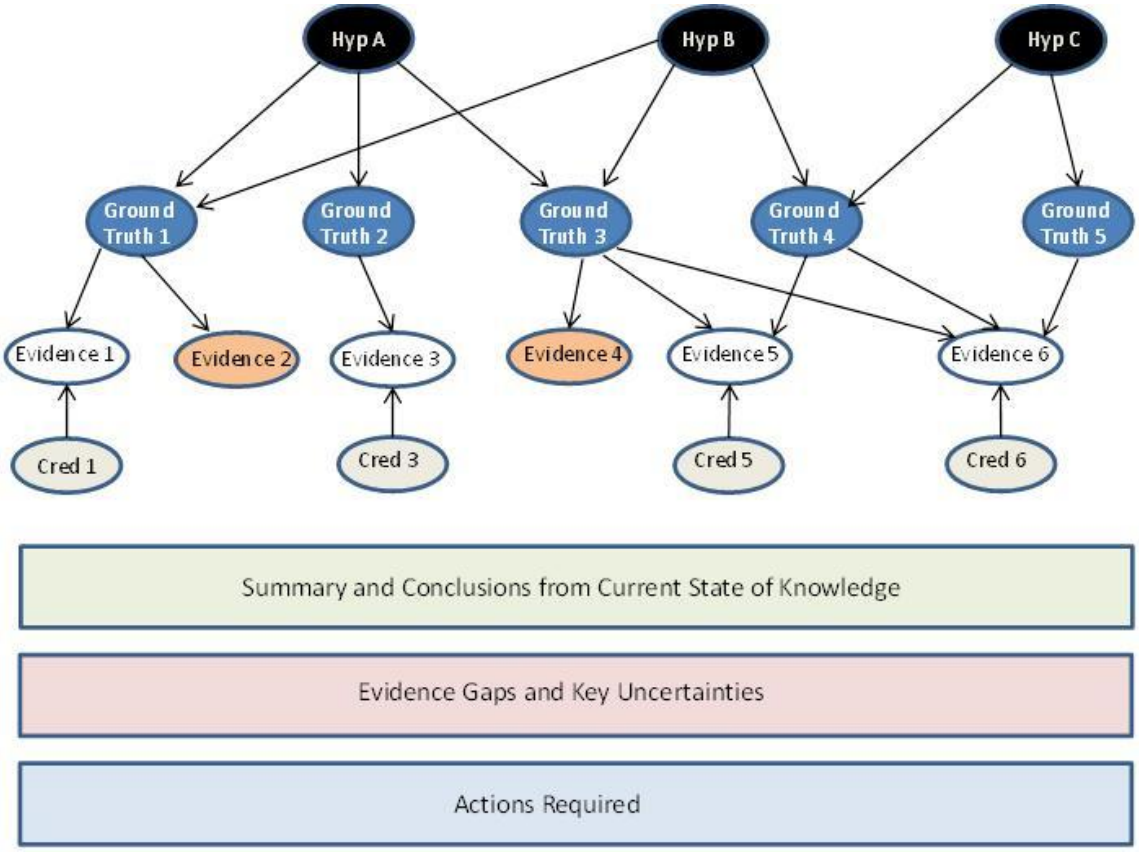

Figure 1: Generic inferential reasoning network and associated workspace.

4. EXAMPLE APPLICATION

In this example, an intelligence agency is concerned with a potential threat from a known terrorist organization. They have become concerned with a high level of chatter, the nature of which is often symptomatic of an impending attack by this organization. Furthermore, they have recent human intelligence that an attack is to be mounted by this organization on a particular target (here denoted as Target A). The source of this intelligence (here denoted as Source 1) is considered credible by the agency and has provided correct intelligence about this organization in the past. A separate source (Source 2) has also provided intelligence of an impending attack by the same organization but has not identified a specific target. In response to the received intelligence, the agency has examined CCTV footage relating to target A. This has identified some suspicious activity some time ago consistent with possible reconnaissance on Target A. However, there is no evidence of a recent dry run on this target. Rudner (2009) describes elements of the modus operandum for Al Qaeda attacks on infrastructure targets. This kind of information is useful in identifying activities to include in the second layer of nodes.

The specific items of evidence which need to be considered by the intelligence analyst in this example are as follows:

- High level of Chatter observed, consistent with impending attack somewhere
- HUMINT received from Source 1 that an imminent attack is planned on Target A
- Source 1 has previously provided HUMINT that person X belongs to organization Z
- Search through CCTV in vicinity of Tgt A reveals possible reconnaissance some time ago.
- No recent evidence of further reconnaissance or dry run observed.

This evidence set would lead to an inferential reasoning network along the lines of that presented in Figure 2. An imminent threat to Target A is the most obvious hypothesis to consider. However, an alternative explanation might be that Target A is part of a deception, a decoy for an attack on a different target. Yet another hypothesis is that no imminent attack is planned and the supporting evidence for an attack is mistaken, possibly due to errors and misunderstandings or possibly the result of a deliberate hoax. Multiple explanations or hypotheses such as these should be considered for the evidence available so far, and expanded or contracted as necessary. As new hypotheses are considered, these generate ideas for

possible new items of evidence and new avenues of enquiry.

The workspace below the network in this example might point to the lack of evidence suggesting pre-attack planning on any target other than A. This may lead to an action to request a search of relevant CCTV footage in the vicinity of other likely alternative targets. The human intelligence from Source 1 is key and needs to be scrutinized. In addition, the accuracy of previous intelligence provided by this source needs to be checked. In this example that relates to the information that individual X belonging to organization Z. Obviously, the more correct information that X has previously provided, the higher is the credibility of any new intelligence that they provide. Also relevant, however, is the ease of attainability and usefulness of the previously correct intelligence. A source who regularly contributes correct and useful intelligence which is difficult to obtain should expect a higher credibility rating than one who provides intelligence which is less useful and more widely available.

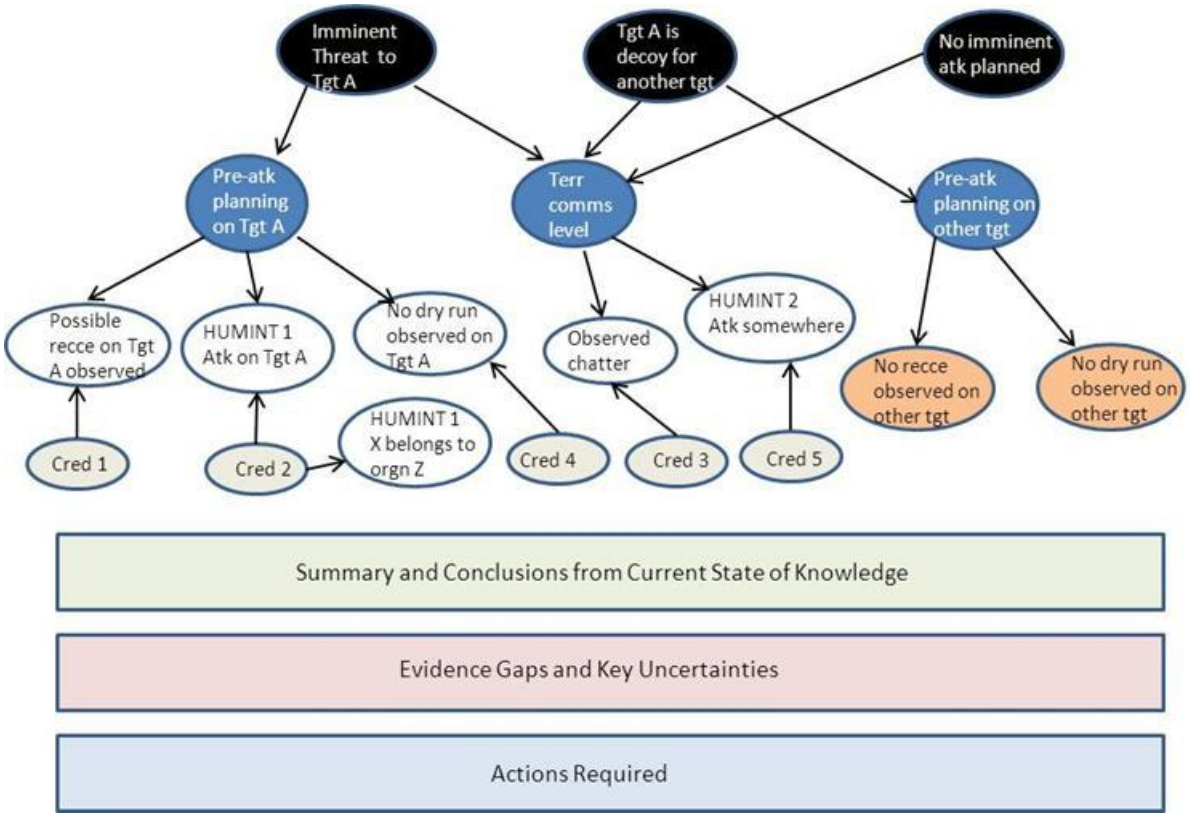

Figure 2: Specific example inferential reasoning network and associated workspace.

5. CONCLUSION

Experience and intuition will always be invaluable core ingredients for reasoning and decision-making in the domain of intelligence analysis. However, that does not obviate the requirement for more formal tools to support an analyst's systematic mode of reasoning. Inexperienced analysts, in particular, may benefit from such support. In this paper, we have highlighted the potential of Bayesian networks to provide a logical and intuitive approach to support an analyst's reasoning process. In particular, we have suggested a framework consisting of four types of nodes, emphasizing the distinction between observable evidence nodes and typically unobservable ground truth nodes. However, in the approach outlined here we have made no use of the quantitative aspect of BNs - that will be explored in a separate paper. Our purpose in this paper is to demonstrate that even without explicit probability distributions, the qualitative support to logical reasoning provided by BNs can still be substantial. Such inferential reasoning networks show how hypotheses, propositions and observations or evidence are related and offer a useful framework for collaborative working and reasoning under uncertainty. Such a framework also supports explicit consideration of an adversary's deception activities, an aspect which will also be developed further in a future paper.

ACKNOWLEDGMENT

The authors are grateful for funding provided by the UK's EPSRC under Grant Number EP/H023135/1.

REFERENCES

Elsaesser, C. and Stech, F., 2007. Detecting deception. In: Kott, A. and McEneaney, W.M., eds. *Adversarial Reasoning – Computational Approaches to Reading the Opponent's Mind*. Boca Raton, FL: Chapman & Hall, 101-124.

Heuer, R., 1999. *The Psychology of Intelligence Analysis*. Washington DC: Center for the Study of Intelligence, CIA.
Klein, G.A., Calderwood, R. and Clinton-Cirocco, A. 1986. Rapid decision making on the fire ground. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 30(6), 576-580.
Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
Rudner, M.,2009. Protecting critical energy infrastructure through intelligence. *International Journal of Intelligence and Counter-Intelligence*, 21, 635-660.
Schum, D.A., 2001. Evidence marshaling for imaginative fact investigation. *Artificial Intelligence and Law*, 9, 165-188.
Stasko, J., Gorg, C. and Liu, Z., 2008. Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, 7, 118-132.
Taroni, F., Aitken, C., Garbolino, P. and Biedermann, A., 2006. *Bayesian Networks and Probabilistic Inference in Forensic Science*. Chichester: Wiley.
Wastell, C.A., 2010. Cognitive predispositions and intelligence analyst reasoning. *International Journal of Intelligence and Counter-Intelligence*, 23, 449-460.
Weiss, C., 2008. Communicating uncertainty in intelligence and other professions. *International Journal of Intelligence and Counter-Intelligence*, 21, 57-85.
Wigmore, J.H., 1937. *The Science of Judicial Proof: As Given by Logic, Psychology and General Experience and Illustrated in Judicial Trials*, 3rd ed. Boston, MA: Little, Brown.

AUTHORS' BIOGRAPHIES

Ken R. McNaught is a senior lecturer in Operational Research (O.R.) at Cranfield University's School of Defence and Security situated at the UK's Defence Academy in Shrivenham. He leads the Operational and Decision Analysis Group where his research interests include simulation, combat modelling and decision support, particularly making use of probabilistic graphical approaches such as Bayesian networks and influence diagrams. He also teaches on a number of specialized MSc courses, including Military Operational Research and Defence Simulation and Modelling.

Peter Sutovsky is a research fellow in the Operational and Decision Analysis Group at Cranfield University's School of Defence and Security in Shrivenham. He is currently finishing a PhD in the area of probabilistic graphical modelling for disease outbreak detection in the University of Pittsburgh's Department of Biomedical Informatics.