

2011

Designing a knowledge distribution simulator

Martin Hill
Cranfield University, UK

Graham Fletcher
Cranfield University, UK

DOI: [10.4225/75/57a83e50befad](https://doi.org/10.4225/75/57a83e50befad)

Originally published in the proceedings of the 12th Australian Information Warfare and Security Conference, Edith Cowan University, Perth Western Australia, 5th -7th December, 2011

This Conference Proceeding is posted at Research Online.

<http://ro.ecu.edu.au/isw/44>

DESIGNING A KNOWLEDGE DISTRIBUTION SIMULATOR

Martin Hill and Graham Fletcher
Cranfield University, Cranfield Defence & Security, United Kingdom
m.hill@cranfield.ac.uk, g.p.fletcher@cranfield.ac.uk

Abstract

To make good decisions, we need to be suitably informed. 'Good' and 'Suitably' in this case depend on the informational needs of the decision and the mechanisms of getting the information to the decision maker in time. The trade-offs in qualities, quantities, timeliness, impacts on other activities, and so on are infamously wickedly complex, and usually buried in a clutter of special circumstances, personality characteristics, environments unsuitable for study, and so on.

Decision-making systems can be explored using case studies and exercises, but these are limited by the expense and time of using real people. A virtual simulator for large scale networks of communities can provide systems to examine that are not otherwise possible, while bearing in mind that simulators only partially reflect real systems.

This paper describes a design for such a simulator framework that can be implemented on an ordinary desktop computer. We intend to use it to exercise and explore various 'knowledge distribution strategies' in order to understand and suggest information communication mechanisms for investigation in the real world, without expecting it to be complete enough to be prescriptive.

We focus on military collaborations as suitably 'eXtreme' environments to exercise these communication mechanisms. Topics for further investigation include isolation, turnover and resilience.

Keywords

Knowledge Distribution, Distributed Collaborations, Distributed Situation Awareness, Decision Support

INTRODUCTION

Where large groups of people are distributed over poor communications networks, different strategies for informing can have very different impacts on the suitability of information received by decision makers, especially as the group and the decision needs within it change over time. Military collaborations are usefully extreme versions of such environments: the organisations can be large, communications are slow, intermittent and fragile, time constraints are hard and tight, outcome stakes are high, stress and emotions are strong, consequences are ethically fraught, evidence is sparse, unreliable [Hill 2010] and drowned in clutter, knowledge sets are complex, and supporting systems are stripped-down yet mission- and safety-critical. Also, the enemy are developing their own strategies to make better, faster decisions than the security force.

Ideally, informing such decision makers *suitably* well means supplying them with relevant, representative and useful data, in time, without overwhelming them ('the right information to the right people in the right time in the right form'). In most real world situations we have to settle for various 'good enough' trade-offs between these ideals, and these trade-offs are 'wickedly complex' and a general theory intractable; Rittel et al (1973) were early adopters of the term 'wicked' to describe systems so complex that they cannot be examined using stereotypical exact-science experiments.

Real world experiments such as case studies, exercises and 'microworld' experiments such as C3Fire are time consuming, expensive to run, and are limited by the case or exercise under study, as Klein (1999) describes when studying fire-fighters, and Andrade (2009) describes more generally. Simulations can be useful to explore large distributed networks of decision makers, however it is extremely difficult to make them realistic enough to be prescriptive predictors; Oreskes et al (1994) go so far as to claim it cannot be formally done for complex systems. Instead the intent behind this simulator is threefold:

1. Provide a mechanism for exploring, comparing and contrasting knowledge distribution strategies
2. Capture and express knowledge about them
3. Distribute that knowledge to practitioners as well as other researchers in an easily understood and interactive graphical form.

The simulator is not intended to answer particular research questions directly, but to provide a framework for asking them. Furthermore, the process of design and implementation will also be a process for finding out what those questions could be; Armour (2000) suggests in his article 'Five Orders of Ignorance' that asking questions we know we want answers for is only a First Order Ignorance; that even in engineering we expect to develop methods to find out what questions we *should* be asking.

This paper focuses on describing an outline design for this simulator, and how significant real world factors will be modelled. It is not intended to describe the 'right' way to design a simulator, but to discuss the things we should consider including to make it workable, and to suggest things that other people will find useful when writing social nets that include information transfer.

We start with some definitions, go on to outline the overall approach, look in more detail at the nature of knowledge and communication, define what we expect to measure, and finish with a brief look at a pilot project that checked the feasibility of the approach.

DEFINITIONS

We define, for the purposes of this design, the following terms, which are used in the model as described in the following section:

- **FACTS** describe the real world, allowing for some uncertainty. For this simulator, these are name/value pairs such as 'X=5'.
- **Information** informs; it is a fact or group of facts that alter the decision makers perception ('situation awareness') of the real world. For this simulator, information consists of facts relayed from point to point; there is no aggregation or interpretation (yet)
- **Knowledge** is the accumulated set of facts known.
- **Knowdes** are entities in the simulated world that are capable of knowing things; they are nodes in the communication meshes that have knowledge.
- **Contexts** are a set of categories attached to facts and the area of the world that each Knowde is located in.
- **Communication** is the act of a knowde sending facts down communication links that may be received as sent by other knowdes on the communication link.
- **Decisions** are modelled as problems to solve, as equations such as $X+Y+Z=?$, which specify the elements that the decision maker needs to know and gives a (poor) measure of how close we are to solving it.
- **Knowledge Distribution Strategy** defines the set of rules that describe how and what a Knowde communicates. This may be global, or assigned to groups of Knowdes.

We include some basic notions of uncertainty, but we are not yet concerned with misinformation; that is, things that the decision maker thinks are facts but do not describe the real world.

OUTLINE DESIGN

Structure

The core design is a reasonably common form of network graph consisting of nodes and communication links. The nodes know things ('Knowdes') and the communication links are used to pass facts between them, and can be represented visually as symbols connected by lines (see Diagram 1).

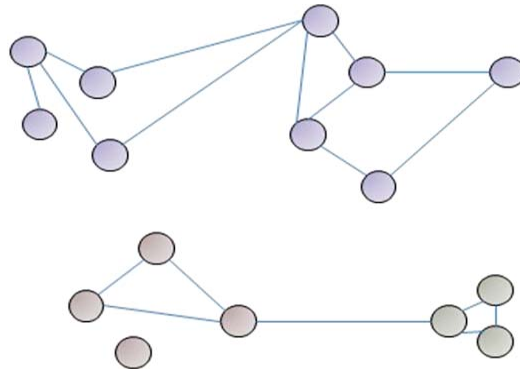


Diagram 1 – An arbitrary example graph network of nodes and communication links
Colour, distance, shape and line detail represent features of the model

An example of how we represent the detail of an individual Knowde symbol and its communication links is given in Diagram 2. These are described in more detail in the following sections.

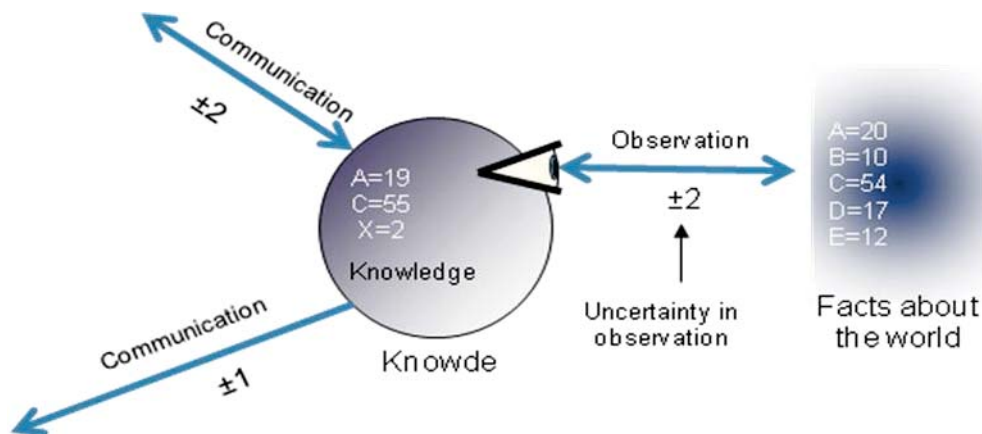


Diagram 2: Representing a Knowde

Process

A simulated ‘run’ consists of ‘turns’ repeated until some end state is reached, such as an elapsed time, or a proportion of decisions can be made, or the system reaches some kind of equilibrium and there is little change between steps.

For each ‘turn’, each Knowde can choose to:

- Communicate facts depending on the available communication links
- Learn facts as they are received from communication links
- Learn facts by observing the world
- Do work (a ‘work done’ counter is incremented)

These are all mutually exclusive activities, and declaring and resolving the priorities between them for each Knowde are the key elements of the **Knowledge Distribution Strategies**. Each Knowde is allocated some ‘attention’ for each turn, and this is consumed as it communicates, listens, observes and does work.

KNOWLEDGE

Facts are observed by Knowdes as elements of the world around them, becoming part of their Knowledge – their list of facts known.

The Knowledge Distribution Strategy may allow Knowdes to remember which facts they have told the other Knowdes that they are aware of on the communication links, and which facts they have been told by which other Knowdes. This leads to a ‘shared’ set of facts: you told me a fact, so I know that you know it, and you know that I now know it, and you know that I know that you know. This can be extended to 3rd parties; you told me a fact given to you by a 3rd party, so I know that you know it, I know the 3rd party knows it, and you know that I know that.

This is represented as what a Knowde knows of another Knowde, and is recursive; within the internal box of “what I think you know” is the box of “what I think you think I know”, and so on (see Diagram 3).

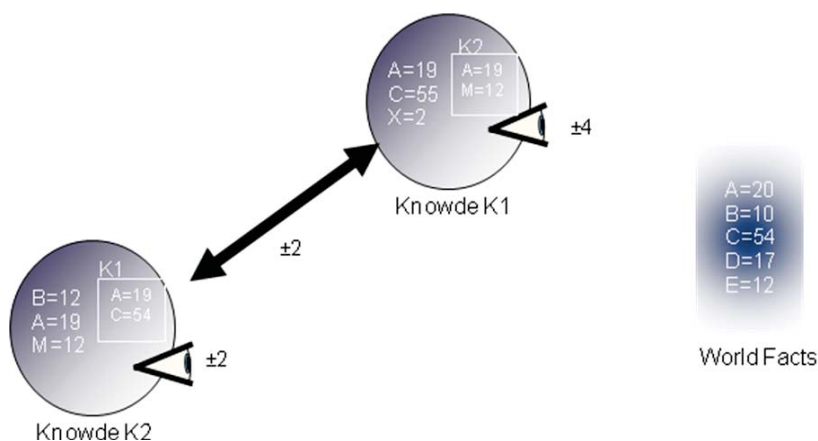


Diagram 3: Knowing what other Knowdes know

Context

Real world communication depends heavily on background context; the more background you share, the less you need to communicate to pass the relevant facts along.

These are represented as background 'context shading' and are coded as simple categories (such as ‘Military’ or ‘Computer Software’) that need not be exclusive. These can be applied to groups of Knowdes by assigned job titles, or by their location in some background arbitrary space. The longer a Knowde is immersed in a context, the more it 'absorbs' and while it is away it 'forgets'; a Knowde’s familiarity with a context is given as a simple proportion (0-100%).

Facts may be rated or coloured by context, and communicating such facts is faster and more reliable the more the Knowdes are coloured by the same Context as the fact.

Fade

People forget, and this is simply modelled as a rate of loss of known facts and context that can be varied depending on the experiment.

COMMUNICATION

Communication links describe the transfer of facts and the constraints of doing so, such as availability, bandwidth and time taken. They are represented by lines with arrows showing the direction facts are sent, as dotted or dashed to show availability, and thicker lines for higher bandwidth. 'Broadcasts' are represented as hexagonal nodes which show who is connected on so-called all-informed nets, as availability may still mean individuals are not informed.

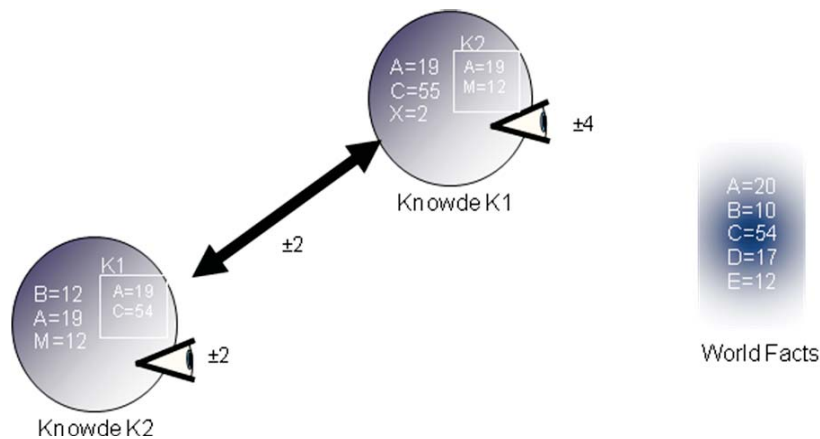


Diagram 4: A mesh of Knowdes

There may be several communication links between the same individuals. For example, military briefings provide very intermittent high bandwidth communications between soldiers and commanders, and radios provide intermittent but more frequent low bandwidth communications.

Not all communication is deliberate or explicit. ‘Water cooler moments’, for example, describe those accidental but useful conversations between people without explicit communication links. These are modelled as ‘free’ facts transferred between Knowdes independent of the explicit communication links.

Uncertainty must be included if we are to represent how confirmation and duplication improves the resilience of knowledge, which is not clear if we assume all communication is perfect (and so need only be made once). This is represented as errors introduced during communication, and one of the characteristics of each communication link is the range of the uncertainty that will be randomly applied to all facts sent over it.

MEASUREMENTS AND RESULTS

The simulator is intended to provide a visual tool for exploring and demonstrating the effects of knowledge distribution strategies for further investigation, rather than hard measurements to model and prescribe real world behaviour in detail. Metrics therefore should be treated carefully; response curves and comparisons between strategies are more useful than standalone values.

Measurements will include differences between ground truth and knowledge, completeness of community knowledge, and decision outcomes, and are logged for every turn for each Knowde and summarised for each Context. These can then be plotted and compared graphically across different strategies.

PILOT SIMULATOR

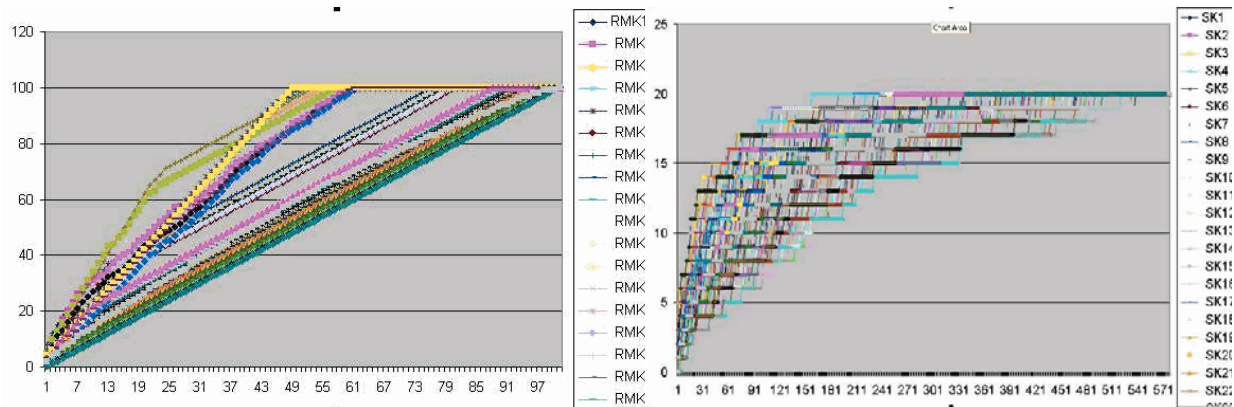
A partial implementation has been developed in java to understand what would be required of the development environments, toolsets, result formats, execution hardware, and so on, and to check the feasibility of the design.

Each run generated up to 100 nodes, with random communication links adjusted to ensure all nodes were connected to at least one other node. Around 20 runs were made for each Knowledge Distribution Strategy.



Diagram 5: Pilot simulator screenshot, showing 2nd run, several steps in Blue Knowdes indicate complete knowledge, and lines are blue when between blue Knowdes.

Results were produced in the form of tables containing a row per step, with each row containing the number of facts known by each Knowde at the end of the step, and some indicators of 'knowledge completeness' across the complete system. These tables were read into a spreadsheet to be converted to graphs and compared visually.



(a) Tell Everyone Something Only Once

(b) Tell Everyone Something Randomly

Diagram 6: Plot for each Knowde; X= step, Y = facts known

Some simple strategies – “Tell Everyone Something Randomly”, “Tell Everyone Something Only Once”, “Tell Everyone Something in a Round Robin” – were applied. The difference in efficiency shown in diagram 6 is expected (random communication took hundreds of steps to inform all Knowdes of 20 facts, whereas more careful direction took tens of steps to inform all Knowdes of 100 facts), as is the shape in diagram 6b, with the slowing approach to complete knowledge as the last few facts are propagated. The near-linear learning in diagram 6a is more surprising and interesting, but may be a result of the pilot Knowdes being able to communicate without loss of attention.

Networks of several hundred nodes ran to 'complete knowledge' in a few hundred steps, and took a few seconds on an ordinary desktop computer, which suggests that the overall design is feasible, and demonstrates the performance characteristics to inform simulator populations.

INITIAL TOPICS FOR INVESTIGATION

This paper, for reasons of space, focuses mainly on the design for the framework, and this section can only briefly touch on the research topics that we intend to pursue.

Our intent is to build and extend rather than attempt to complete a complex simulator in a single step. The designed framework given here should be sufficient to suggest answers for some simple research questions, such as:

- **Isolation/moving people:** what are the trade-offs in moving people between communities of various

isolations (with few communication links between the communities) that brings 'far off' facts yet require 'education' in the local environment?

- **Population ageing:** as experienced people retire and ignorant new people enter a community, the community must spend time educating the new people. The bigger the community and knowledge size the more time this takes.
- **Resilience:** If communication links or nodes or groups of nodes are removed, what strategies ensure there is still suitable information 'near' the decision maker?
- **Moving decision points:** Where are the optimal places that receive the right facts to make decisions? What needs to be communicated to move decision making roles?

Further development could provide mechanisms for examining:

- **Updating decisions:** Decisions usually change over time; by adding such changes, we introduce the need for timeliness.
- **Changing world:** values of facts in the world change. Are there rates, populations and/or network types that swamp at certain rates of changing world?
- **Not knowing what you need to know:** In this design, we assume that a decision maker knows the names of all the elements required to satisfy the decision equation. Unknown elements – that is, elements that some the decision maker is not aware of, but are available – can be added. Mechanisms for discovering what needs to be known will be needed.
- **Aggregation & Interpretation:** Information is rarely passed unchanged around collaborative networks; it is aggregated and interpreted. Mechanisms for bundling facts as, perhaps, results of equations could model this.
- **Priorities:** timeliness is key to informing decision makers, and ways of labelling and communicating priorities
- **Errors, uncertainty, clutter and lies** will be subject of another paper

CONCLUSION

We have described the design for a simulator that will provide a means to investigate knowledge distribution across large scale distributed collaborations that is not feasible using real people. Characteristics of distributed collaborations – such as the consumption of attention by communication – have been briefly described and captured in the model. We are and should be wary of attempting or assuming completeness.

A pilot prototype has been developed and demonstrates that this design is feasible, is within the technical capability of the authors, and can be run on an ordinary desktop computer. Comparing results of some simple knowledge distribution strategies indicate that even the pilot implementation produces interesting results, but we must be careful to investigate and allow for effects of the model that are not reflected in the real world. Such investigations are drivers for this project: to develop better understanding of real world collaborations.

The visual representations provide a means to show practitioners some of the effects of various strategies, which can help to inform non-expert practitioners (such as military commanders and commercial managers) , and so transfer abstract concepts from study to the field. The simulator should be a suitable mechanism for distributing knowledge across scattered organisations about how we could and should not distribute knowledge across scattered organisations.

REFERENCES

- Armour, Phillip (2000), article “Orders of Ignorance” *Communications of the ACM* (October 2000).
<http://cacm.acm.org/magazines/2000/10/7556-the-five-orders-of-ignorance/abstract>
- C3Fire is a ‘microworld’ simulated forest-fire fighting exercise used to study small group command and communications at the Defence Academy, UK, and elsewhere. <http://www.c3fire.org>
- Klein, Gary (1999) Book “Sources of Power: How People Make Decisions,” ISBN 0262611465
- Hill, Martin & Salt, John (2010) Conference Proceedings “Life and Death Decisions using Sparse, Unreliable Evidence” *ECIME 2010 p15*, <http://www.academic-conferences.org/pdfs/ECIME10-booklet.pdf>

Oreskes, Naomi; Shrader-Frechette, Kristin & Belitz, Kenneth (1994) Article “Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences” *Science*, vol 263, 4th February 1994.
<http://www.sciencemag.org/content/263/5147/641>

Rittel, Horst and Webber, Melvin (1973) Article “Dilemmas in General Theory of Planning” *Policy Sciences* 4, 1973, <http://www.cc.gatech.edu/~ellendo/rittel/rittel-dilemma.pdf>