McGuinness, B. & Leggatt, A. (2006) Information trust and distrust in a sensemaking task. *Command and Control Research and Technology Symposium*, June 2006, San Diego (DoD Command & Control Research Program).

Information Trust and Distrust in a Sensemaking Task

Barry McGuinness and Andrew Leggatt
BAE Systems, UK

ABSTRACT

This paper reports on a DoD-funded experiment into human trust and distrust of information in the context of a military sensemaking task. Twenty-two British Army majors undertook an intelligence picture compilation task during a simulated coalition engagement with enemy forces. The task required them to evaluate and compile incoming intelligence reports from a variety of sources. One in four of the reports were in fact deficient with respect to some parameter of information quality (correctness, completeness, timeliness, etc.), and the subjects were advised to eliminate any untrustworthy items from the ongoing picture. The appropriateness of their information trust/distrust responses under different conditions, and their ratings of the perceived trustworthiness of information items, were assessed. Two key factors were manipulated: (1) the subjects were given either a high or low prior understanding of the situation at the start of the task; (2) at one point the subjects were presented an information network alert informing them that a breach of the network had occurred and that information quality may have been compromised. In fact, these interventions had little effect on the information trust data. This was found to be due to an overriding effect of the subjects' awareness of the information sources, and the biases and assumptions associated with that.

INTRODUCTION

Enormous advantages are anticipated in the ability of networked information technology to reduce the fog of war. At the same time, however, military organizations are becoming increasingly dependent upon information, and upon the confidence users have in that information. This dependence is widely regarded as a potential Achilles' heel in the network-centric warfighting organization.

For one thing, the sheer richness and availability of electronic information can sometimes be problematic. In rapidly changing circumstances, for example, battlespace information can quickly become outdated and the situational picture may not truly reflect ground truth. Having information that is no longer true yet is still available as part of the current picture can even be worse than having none at all, for any information that is untimely, irrelevant or unreliable can consume valuable human attention – perhaps the greatest bottleneck in any organization (Neus, 2001).

The uncertainties inherent in electronically produced information are also an important human issue (Alberts et al, 2001). Clearly we want our information to be accurate, not vague; yet the apparently high precision of electronically displayed information can sometimes obscure the actual uncertainty or ambiguous nature of the underlying data or data filtering/fusion processes (Waller, 1995). The information as it is displayed might not provide

any indication of such imprecision, creating the impression that one piece of data is as definite as any other.

Another set of problem derives from the various vulnerabilities of information systems and networks. For example, as battlespace information superiority becomes more central to military operations, information systems themselves become critical targets. At the same time, a network is only as secure as its most vulnerable component, and the sheer complexity of deployed information systems makes them still more vulnerable to attack. As a network grows and becomes more fully interconnected, moreover, the mere task of noticing a penetration or penetration attempt becomes extremely difficult (Cox et al, 1999).

It is therefore essential for effective and efficient command and control to maintain high information quality and integrity. The consequences of even a single *failed* attempt or merely a *rumoured* attempt at attack or deception could be devastating, as the mere awareness of possible interference with friendly databases can be expected to greatly inhibit decision-making. As Alberts et al (2001, p.86) explain in *Understanding Information Age Warfare*:

> [R]esearch has shown that even a small amount of wrong information can have a major impact on the quality of situational understanding and lower the chances of high-quality military decisions. Perhaps equally important, users must be able to trust the data and information in the systems supporting them and have confidence in the system's ability to provide them with needed information. Users who do not trust the quality of information available to them or do not have confidence in their information systems are believed (hypothesized) to both act more cautiously (create and select action sets that are risk averse in that they will work even if the available information is incorrect, late, inconsistent, etc.) and more slowly (waiting for confirmatory evidence before they act on emerging patterns, deliberating longer, etc.).

As forces become increasingly dependent on a complex and vulnerable information domain, knowing what information can be trusted, and knowing when to distrust the available information, are set to become key human issues. As yet, however, there has been little research of direct relevance to these issues. The causes and effects of errors of information trust are largely unexplored. The research reported here seeks to address the issue in the context of network centric operations (NCO).

SENSEMAKING EXPERIMENT

The research was carried out in 2005-06 by the Advanced Technology Centre (ATC) of BAE Systems in Bristol, UK, as part of a 12-month contract funded by the DoD.[1] Two experiments were conducted. This paper reports the first of these, which was designed to address the possible influence of prior situational awareness/understanding (SA/SU) on the appropriateness of information trust responses (Figure 1). It was hypothesized that better understanding of either (a) *the operational situation* or (b) *the network status* would facilitate better judgements of information trustworthiness.

[1] The UK Defence Academy at Shrivenham (run by Cranfield University) was subcontracted to provide research facilities and experimental support.

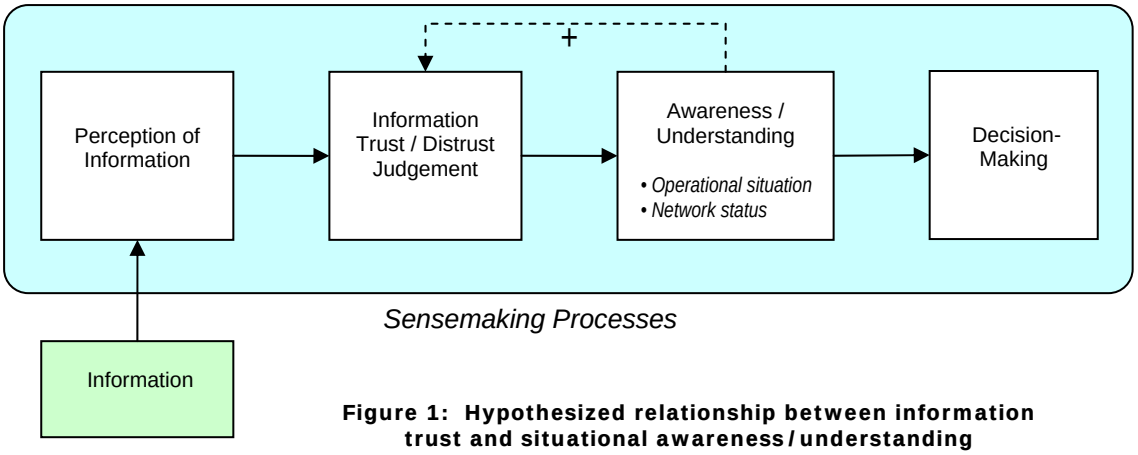

Figure 1: Hypothesized relationship between information trust and situational awareness/understanding

Twenty-two British Army majors performed a sensemaking task designed to elicit multiple information trust/distrust discriminations. The task was set within a fictitious future coalition operation based in Africa, where rebel forces were preparing to attack the capital city of a former French colony, 'Kumbiba'. A French-led multinational HQ was supported by a combination of French, British, American and host-nation armed forces.

The subjects were theoretically located within the intelligence cell of the British element's HQ. In the course of the operation, they received a flow of electronic intelligence reports/messages from multiple sources on the status, position and movement of enemy forces. The content of these messages was automatically translated into appropriate visual icons on an electronic map (Figure 2).

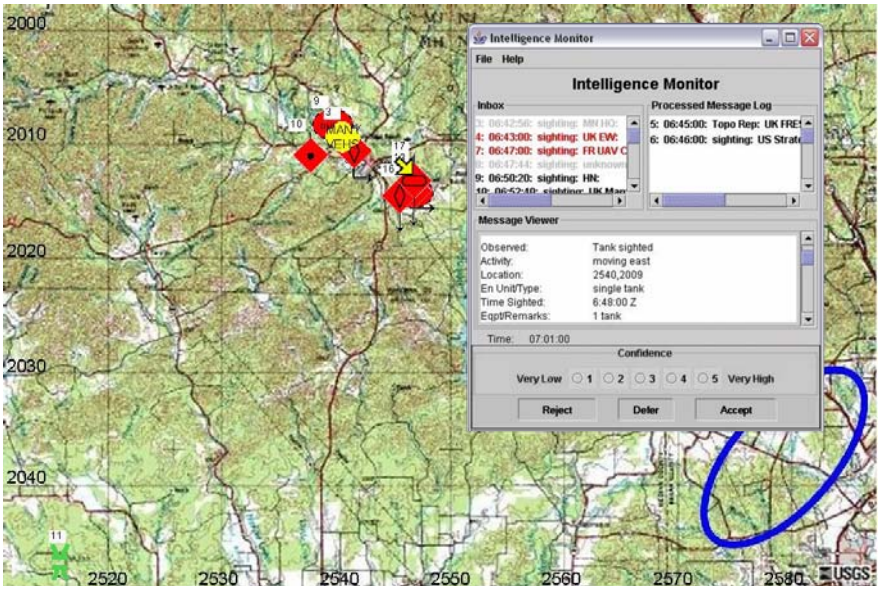

Figure 2: Screenshot of the electronic map and message software used by the subjects

The subjects' task was to monitor the incoming information in order to assess the enemy's course of action. However, they were also advised that some of the available information might be unreliable or unusable in some way, and if so should be filtered out from the ongoing picture compilation process. In fact, 25% of all messages had some kind of information quality 'defect', namely one of the following:

Incorrectness	•	the information does not actually fit reality
Inaccuracy	•	the information is not sufficiently precise to be of use
Incompleteness	•	some key information is missing from the report
Untimeliness	•	the information is received too late to be of use
Irrelevance	•	the information has no bearing on the operational situation
Inconsistency	•	the information does not fit the emerging pattern of events

Independent variables

The causal factor of interest was the subjects' degree of situational awareness/understanding while performing the task. This was manipulated through two independent variables.

- *Situation Briefing*
- *Network Alert*

In a repeated-measures design, all subjects performed the task twice, using two similar but non-identical versions of the scenario. In one, the subject was given a comprehensive situation briefing, describing in detail both the *current operational situation* (including probable disposition and locations of red forces) and the *security status of the intelligence information network*. This was designed to provide a relatively high degree of initial SA/SU. In the other run, a minimal briefing was given whereby initial SA/SU would be relatively low.

During one of their runs the subject was also handed an Alert message informing them that "a breach of the intelligence network" had been detected and that "information quality may be temporarily compromised." (In fact, the ratio of good information to flawed information remained the same at 3:1.) In the other run, no such alert was given.

The intent behind these interventions was to test the hypothesis that the subject's awareness/understanding of either the current operational situation or the network's status had an influence on how much information or what kinds of information they would choose to trust or distrust. The combinations of conditions were equally distributed over the two runs and across the subjects so as to balance out any order effects.

Task

For each run of the task, the subjects were presented 48 messages, 4 from each of 12 sources. The variety of sources, one of which was simply designated 'unknown', was chosen to be a realistic fit for the scenario (Figure 3). The pre-written messages were delivered electronically using a system akin to email (Figure 4).

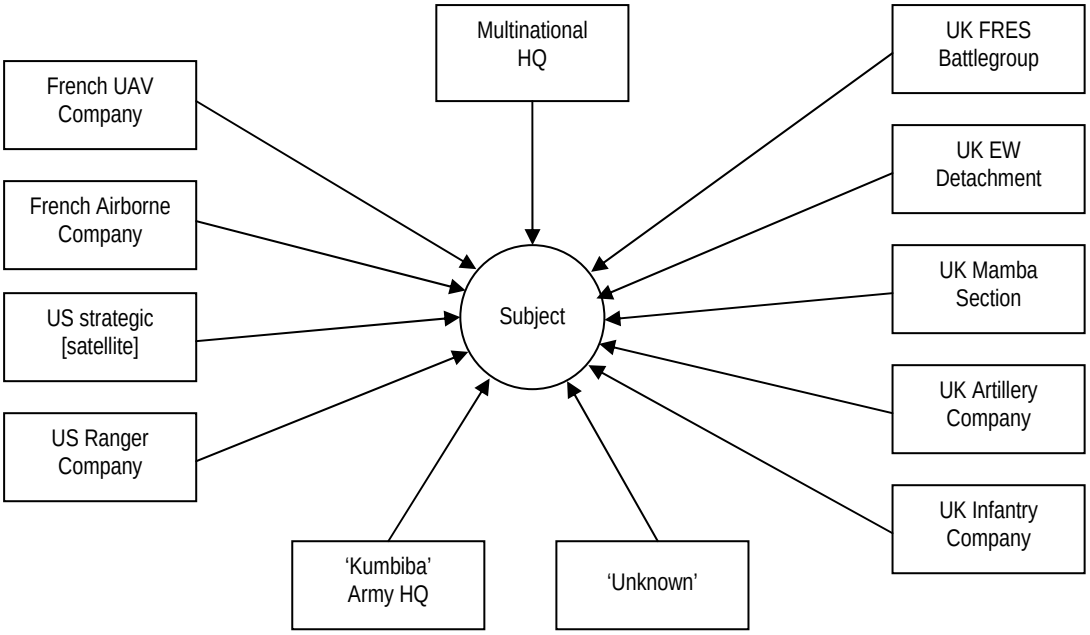

Figure 3: The information received by the subject was attributed to 12 sources

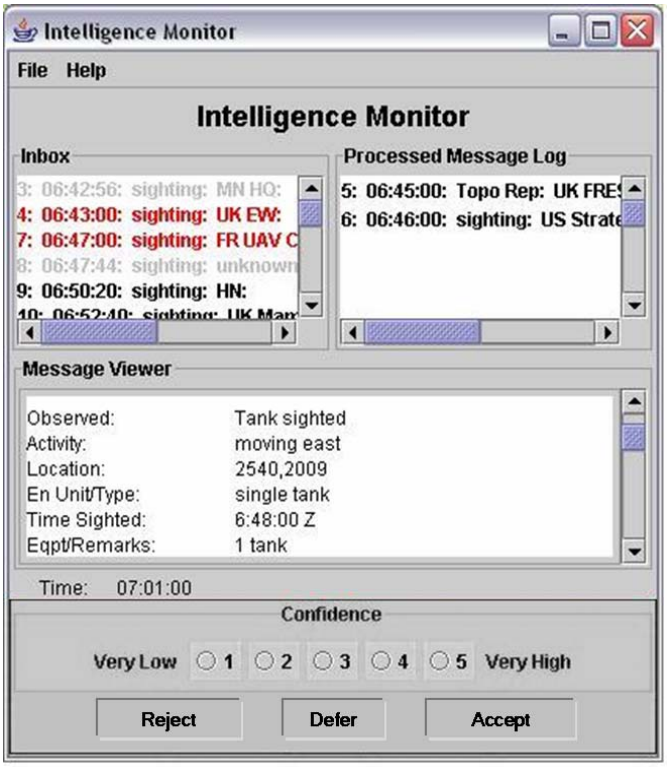

Figure 4: Screenshot of the subjects' electronic message reader

Messages arrived in the subjects' Inbox at pseudo-random intervals of approximately two messages per minute. After 16 messages had been delivered, the flow of messages was automatically paused and the subject was asked to review the current set of messages in their Inbox using the following steps:

1. Select a header in the Inbox (the message text then appears in the Message Viewer).
2. Give that message a confidence rating of 1–5 (where 1 = very low, 5 = very high).
3. Select either **ACCEPT** to retain that message as part of the current intelligence picture (this transfers it to the 'Processed Message Log'), or **REJECT** to remove it (this leaves it in the Inbox but the header is shown with a red typeface).

There were three such reviews per run: (1) after 16 messages, (2) after 32 messages, and (3) after 48 messages.

Dependent variables

The performance attribute of interest was the appropriateness (and any biasing) of subjects' trust and distrust of the messages received. Information trust was operationalized through the following dependent variables:

- *Ratings of confidence in each message*
- *Rates of error in accepting and rejecting messages*

There are, in fact, two distinct and potentially critical errors of judgement that can occur in relation to trusting information:

Type 1 error • Trusting and acting on information that is not in fact trustworthy (i.e. *mistrust*).

Type 2 error • Distrusting and ignoring information that is, in fact, trustworthy (for which we have coined the term "*misdistrust*").

Both types of error were measured. Aside from providing error rates, the ACCEPT/REJECT response data also lend themselves to analysis using the framework of Signal Detection Theory (SDT), as shown in Table 1:

Table 1: Contingency table of stimulus-response outcomes

		Response type	
		ACCEPT	REJECT
Information type	GOOD	**Hit**	**Miss** (type 2 error)
	BAD	**False Alarm** (type 1 error)	**Correct Rejection**

SDT can provide two very useful statistics that are computed from the proportion of hits and false alarms obtained in an experiment. One is *d'*, which represents the degree of sensitivity or ability to differentiate between 'signal' versus 'non-signal' stimuli, as shown by a subject's responses. In the present experiment, *d'* represents the subjects' ability to differentiate good from bad information. A second measure provided by SDT is *ß*, representing the degree of response bias (if any) in stimulus discrimination. In this case, *ß* represents the subjects' tendency to either trust (accept) or distrust (reject) potentially unreliable information.

A variety of other measures were taken to assist the interpretation of the results. These included:

- *Responses to multiple-choice questions addressing the subjects' understanding of the enemy course of action*
- *Self-ratings of confidence in those responses*
- *Ratings of the perceived reliability of each source*
- *Questionnaire responses to assess personal trust/distrust attitudes*
- *Personal background details*

RESULTS

There was no significant effect of SA/SU on the message confidence ratings, and there were only two minor effects on the type 1 and type 2 error rates as follows:

- When the subjects had been given only the minimal briefing (i.e. low initial SA/SU), they committed more type 1 errors (accepting flawed information) during their *first* message review only ($F_{1,21} = 5.404$; $p < 0.05$).

- When the subjects were given a network alert, they committed *fewer* type 2 errors (rejecting good information) during their *third* message review only ($F_{1,21} = 4.360$; $p < 0.05$).

Turning to the SDT statistics, the average value of *d'* for the subjects in this experiment was found to be 0.70. This is in fact quite a low value, suggesting that the overall discrimination of good from bad information was not good. There was no significant effect of any SA/SU manipulation on *d'*. There was, however, a very strong correlation between *d'* and the ratings of the *perceived reliability of each source* ($r = 0.91$, $p < 0.005$). As we can see in Figure 5, the implication of this finding is that the more a source is trusted (for no other reason than its apparent identity), the better a subject is at differentiating good vs. bad information from that source.

On the basis of this finding, further analysis of the data revealed that the main factor affecting the subjects' ACCEPT/REJECT responses was indeed their awareness of the source from which a piece of information had supposedly derived. It was noticed, for example, that there was a strong similarity between the message confidence ratings and the subjects' ratings of source reliability (Figure 6). The correlation between the two was found to be extremely high ($r = 0.968$, $p < 0.005$). This suggested that the subjects may have been basing their assessments of information items almost entirely on their awareness of the information's source, along with

whatever assumptions or biases went with that. Given that all sources produced exactly the same amount of good quality information, this was not an ideal strategy.

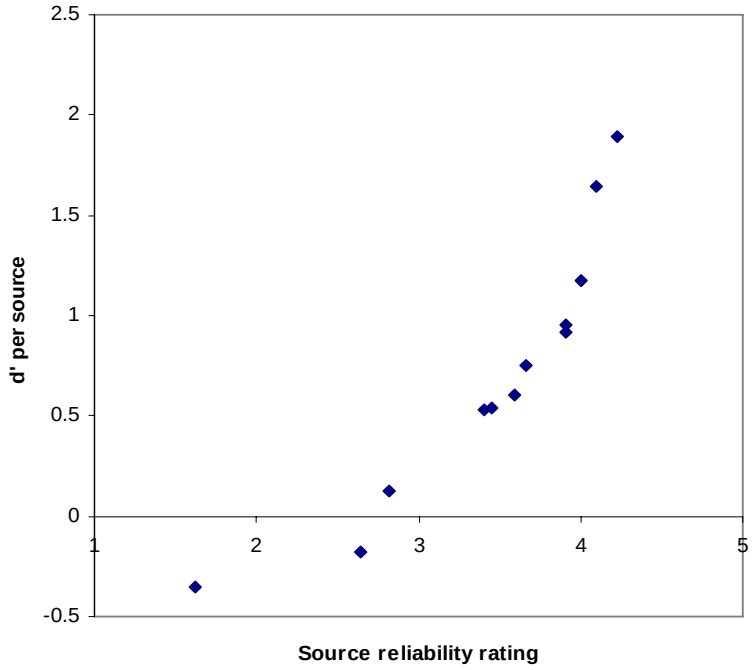

Figure 5: Scatter diagram showing correlation between source reliability ratings and *d'*

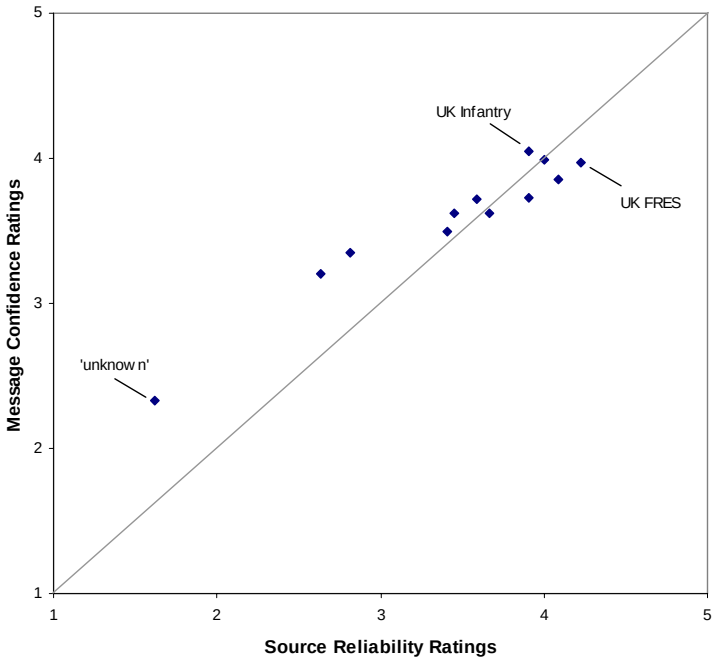

Figure 6: Scatter diagram showing correlation between source reliability ratings and information confidence ratings. Some source identities are indicated as examples.

It was then found by analysis of variance that the probability of type 1 errors (incorrect acceptance) varied greatly according to the identity of the source of information ($F_{1,21}$ = 38.979; $p < 0.001$). As we can see in Figure 7, the subjects were far more likely to accept

‘flawed’ information from some sources (notably the US strategic and French UAV sources) than from others (such as the UK FRES battlegroup and the host nation HQ). Likewise, the rate of type 2 errors (incorrect rejection) also varied according to the identity of the source of information ($F_{1,21} = 9.59$; $p < 0.01$). In this case, the subjects were most likely to reject good information from the host nation (‘Kumbiba’) source and the ‘unknown’ source (Figure 8).

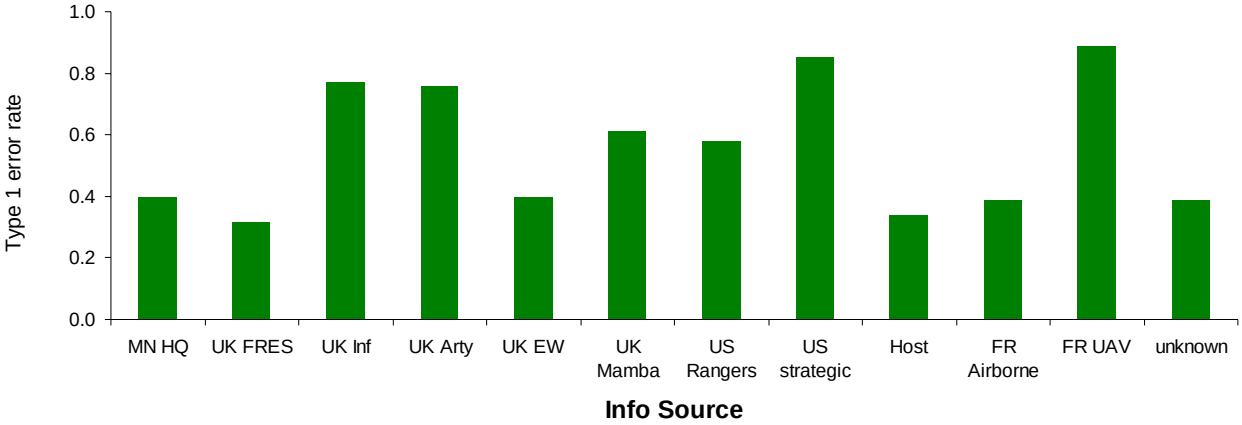

Figure 7: Rate of type 1 error (accepted flawed information) associated with each source

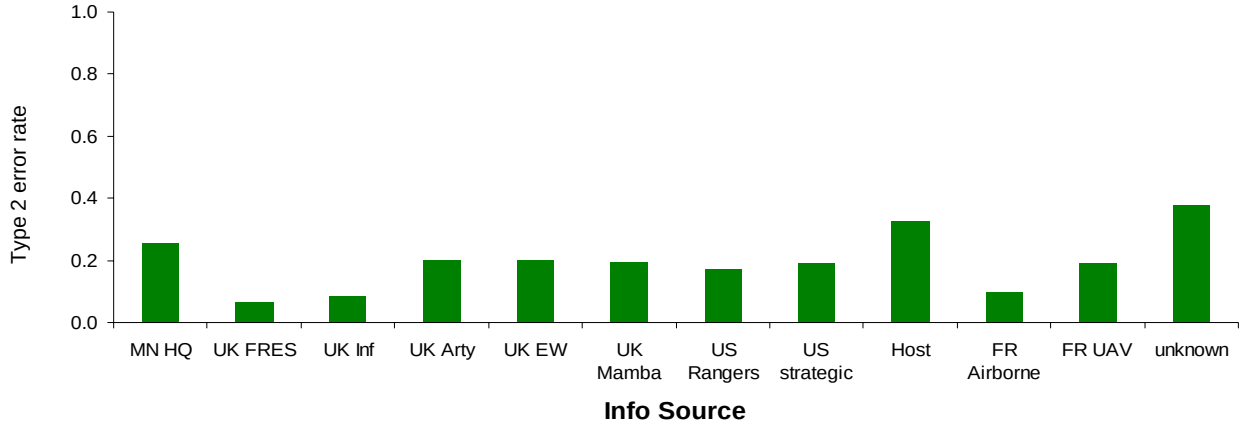

Figure 8: Rate of type 2 error (rejected good information) associated with each source

As for response bias, the average value of *ß* for the subjects was 0.71. This was somewhat more conservative (distrustful) than the optimal value, which for this experiment was calculated to be 0.33. In other words, the subjects were tending to reject more information than was ideal.

Bias was not significantly affected by any of the SA/SU manipulations. Interestingly, however, a very strong negative correlation was found between d' and *ß* ($r = -0.89$; $p < 0.005$). This indicates that those subjects who were less distrusting (i.e. less prone to rejecting information) were better able at discriminating good vs. bad information.

Finally, bias (*ß*) was also found to correlate with a specific trust attitude identified through the questionnaires ($r = 0.38$, $p < 0.05$). This attitude was labelled ‘Caution with information’, and represents in essence an unwillingness to trust information from unknown sources. It was thus found that those who scored higher on ‘Caution with information’ also showed more

conservative bias in their ACCEPT/REJECT responses, i.e. they were tending to err on the side of caution by rejecting information unless certain of its reliability.

DISCUSSION

For discussion purposes, the above results were presented to a group of eight military subject matter experts (SMEs), all British Army majors currently studying at the UK Defence Academy, Shrivenham. The SMEs were struck by the extent of the biasing effect of knowledge of information source in the experiment and considered this to be an important result. In seeking the best explanations for the results, the following comments were made.

To begin with, it was suggested that the information items could be assessed in terms of three categories of perceived trustworthiness, each category entailing a different response:

	Assessment	Action
1.	Probably reliable	Immediately use (= provisionally accept)
2.	Probably unreliable	Immediately set aside (= provisionally reject)
3.	Ambiguous	Actively seek verification if time allows, otherwise accept and use

So, when information is deemed "probably reliable", it is used immediately, and when it is deemed "probably unreliable" it is immediately set aside (not deleted, but filed away). Note that an item does not have to be assessed as *definitely* reliable for it to be given immediate use; *probably* reliable is sufficient for practical purposes. This of course entails some risk of type 1 errors, i.e. using false or flawed information. The SMEs explained that in tasks such as this speed is usually of the essence, and under time pressure this risk has to be accepted. Indeed, it seems that "when the heat is on", increasing amounts of ambiguous information will be readily accepted as there is little time or mental capacity to consider trustworthiness issues. (We might predict, therefore, that type 1 error rates in an information trust/distrust judgement task will be found to positively correlate with mental workload.)

Unless there is an obvious reason for not doing so, identity of source is habitually used as the basis for quickly assessing the probable trustworthiness of information. When there is no obvious and immediate sense of source reliability, however, the user will consider the content of the information itself – does it fit the expected pattern, or does it contradict something known for certain? If that cannot provide for a quick assessment, then the user will resort to seeking further evidence to confirm or disconfirm the information (e.g., monitoring how the situation unfolds, or asking around, "How reliable is that source?" or "Could the enemy really have troops over there?"). This is obviously time-consuming and best avoided if possible.

These insights lead us to the following interpretation of the results:

- Some people are, by nature, more willing to trust externally-produced information than others. Those who scored high on 'Caution with information' in this experiment were indeed more cautious in accepting information.

- However, *a substantial amount of information in the experiment was automatically rejected on the basis of it deriving from a source that was prejudged to be unreliable or dubious* (e.g., 'host nation' or 'unknown').
- As a consequence of this bias, type 1 error rates were relatively low for the least trusted sources (host nation and 'unknown') while type 2 error rates for these sources were high.
- It is also possible that *when information was received from a trusted source it was paid more attention*, thereby enabling better discrimination on the basis of information content rather than merely source-related assumptions.

In conclusion, our hypothesis that situational awareness/understanding can influence judgement of information trustworthiness was not given any major support by the data from this experiment. Our SA/SU manipulations (the situation briefing and network alert) were probably not sufficiently impactful to affect performance in any substantial way – certainly not sufficient to dislodge the effect of source awareness. In fact, it was found that awareness of information source tends to have an overriding effect on information trust judgements.

To the extent that there is a risk to information quality, there is a need for individuals to make appropriate information trust/distrust judgements. By better understanding information trust/distrust practice and the factors affecting it, we can seek ways to ensure decision-makers in network-centric operations do not erroneously accept (mistrust) bad information or erroneously reject (misdistrust) good information. It would be interesting, then, to examine information trust in a similar experiment to this one, but with the source identity of messages not given. This would, perhaps, shed light on other factors shaping judgment. Moreover, the removal of assumption-based bias might actually lead to an improvement in performance. This is in fact the focus of the second experiment in this series, reported in our second paper.

REFERENCES

Alberts, D.S., Gartska, J.J., Hayes, R.E. & Signori, D.A. (2001) *Understanding Information Age Warfare*. Washington DC: CCRP Publication Series.

Cox, S., Stimeare, R., Dean, T. & Ashley, B. (1999) Information assurance – the Achilles' heel of Joint Vision 2010? *Air & Space Power Chronicles*, March 1999. http://www.airpower.maxwell.af.mil/airchronicles/cc/ashley.html

Neus, A. (2001) Managing information quality in virtual communities of practice. In: E. Pierce & R. Katz-Haas (Eds.) *Proceedings of the 6th International Conference on Information Quality at MIT* (IQ 2001). Boston, MA: Sloan School of Management.

Waller, D. (1995) 'Onward Cyber Soldiers.' *Time Magazine*, 21 August 1995, p.44.