

*This Document is issued as an EATMP Guideline. The contents are not mandatory. They provide information and explanation or may indicate best practice.*

## Guidelines for Trust in Future ATM Systems: Measures

Edition Number	:	1.0
Edition Date	:	05.05.2003
Status	:	Released Issue
Intended for	:	EATMP Stakeholders

## DOCUMENT CHARACTERISTICS

TITLE		
<b>Guidelines for Trust in Future ATM Systems: Measures</b>		
<b>Document Identifier</b> HRS/HSP-005-GUI-02	<b>EATMP Infocentre Reference:</b> 030317-02	<b>Edition Number:</b> 1.0
		<b>Edition Date:</b> 05.05.2003
<b>Abstract</b>		
<p>The purpose of this document is to describe the development, evaluation, validation and potential use of a measure of Air Traffic Control (ATC) trust. The measure, named 'SATI' for 'SHAPE ATM Trust Index', is primarily concerned with human trust of ATC computer-assistance tools and other forms of automation support, which are expected to be major components of future Air Traffic Management (ATM) systems.</p> <p>This deliverable is the second one developed within the 'Solutions for Human-Automation Partnerships in European ATM (SHAPE)' Project. A related deliverable provides a set of human factors guidelines for facilitating and fostering human trust in ATM systems (see EATMP, 2003a). A subsequent deliverable on the trust issue provides detailed information on trust principles (see EATMP, 2003b).</p>		
<b>Keywords</b>		
Air Traffic Control (ATC)	Air Traffic Management (ATM) system	Automation
Computer assistance	Distributed cognition	Evaluation
Human-Computer Interaction (HCI) trust	Human factors	Human-machine
Rating scale	Real-time simulation	SHAPE ATM Trust Index (SATI)
Solutions for Human-Automation Partnerships in European ATM (SHAPE)	Trust	Usability
Validation		
<b>Contact Persons</b>		
Oliver STRAETER, SHAPE Project Leader	<b>Tel</b> +32 2 7295054	<b>Unit</b> Human Factors & Manpower Unit (DIS/HUM)
Michiel WOLDRING, Manager, HRS Human Factors Sub-Programme (HSP)	<b>Tel</b> +32 2 7293566	<b>Unit</b> Human Factors & Manpower Unit (DIS/HUM)
<b>Authors</b>		
P. Goillau, C. Kelly, M. Boardman and E. Jeannot		

STATUS, AUDIENCE AND ACCESSIBILITY			
Status	Intended for	Accessible via	
Working Draft	<input type="checkbox"/> General Public	<input type="checkbox"/> Intranet	<input type="checkbox"/>
Draft	<input type="checkbox"/> EATMP Stakeholders	<input checked="" type="checkbox"/> Extranet	<input type="checkbox"/>
Proposed Issue	<input type="checkbox"/> Restricted Audience	<input type="checkbox"/> Internet (www.eurocontrol.int)	<input checked="" type="checkbox"/>
Released Issue	<input checked="" type="checkbox"/> <i>Printed &amp; electronic copies of the document can be obtained from the EATMP Infocentre (see page iii)</i>		

ELECTRONIC SOURCE		
Path:	Software	Size
G:\Deliverables\HUM Deliverable pdf Library\	Microsoft Word 8.0b	
Host System Windows_NT		

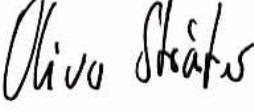
**EATMP Infocentre**  
 EUROCONTROL Headquarters  
 96 Rue de la Fusée  
 B-1130 BRUSSELS

Tel: +32 (0)2 729 51 51  
 Fax: +32 (0)2 729 99 84  
 E-mail: [eatmp.infocentre@eurocontrol.int](mailto:eatmp.infocentre@eurocontrol.int)

Open on 08:00 - 15:00 UTC from Monday to Thursday, incl.

## DOCUMENT APPROVAL

The following table identifies all management authorities who have successively approved the present issue of this document.

AUTHORITY	NAME AND SIGNATURE	DATE
<i>Please make sure that the EATMP Infocentre Reference is present on page ii.</i>		
SHAPE Project Leader	 O. STRAETER	14.05.03
Chairman HRT Human Factors Sub-Group (HFSG)	 V.S.M. WOLDRING	16 may 2003
Manager EATMP Human Resources Programme (HRS-PM)	 M. BARBARINO	16/05/03
Chairman EATMP Human Resources Team (HRT)	 A. SKONIECKI	16/05/2003
Senior Director Principal EATMP Directorate (SDE)	 W. PHILIPP	14.05.03

## DOCUMENT CHANGE RECORD

The following table records the complete history of the successive editions of the present document.

EDITION NUMBER	EDITION DATE	INFOCENTRE REFERENCE	REASON FOR CHANGE	PAGES AFFECTED
0.1	24.04.2001		Working draft	All
0.2	31.08.2001		First Draft	All
0.3	13.02.2002		Approval by HFSG7	All
0.4	30.08.2002		Proposed Issue for HRT18	All (document configuration)
1.0	05.05.2003	030317-02	Released Issue	All (document configuration)

# CONTENTS

<b>DOCUMENT CHARACTERISTICS .....</b>	<b>ii</b>
<b>DOCUMENT APPROVAL .....</b>	<b>iii</b>
<b>DOCUMENT CHANGE RECORD .....</b>	<b>iv</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>1</b>
<b>1. INTRODUCTION .....</b>	<b>3</b>
1.1 Purpose .....	3
1.2 Scope.....	3
1.3 Background .....	3
1.4 Structure .....	5
<b>2. TRUST BACKGROUND .....</b>	<b>7</b>
2.1 What is Trust? .....	7
2.2 Previous Work .....	7
2.3 Measurement of Trust .....	9
<b>3. DEVELOPMENT OF A TRUST MEASURE.....</b>	<b>11</b>
3.1 Overlap between SHAPE Measures .....	11
3.2 Development Process for SATI Measure .....	12
3.3 SATI Theoretical Frameworks and Assumptions .....	14
3.4 SATI Structure and Content .....	16
3.5 SATI Usage .....	17
<b>4. SATI EVALUATION AND VALIDATION .....</b>	<b>19</b>
4.1 SATI Usability Evaluations .....	19
4.2 SATI Evaluation Results.....	20
4.3 SATI Validation.....	24
4.4 Problems of Measuring Trust in Simulations.....	29
4.5 Empirical Validation .....	29
4.6 Future Developments .....	30
<b>5. CONCLUSIONS.....</b>	<b>33</b>
<b>6. RECOMMENDATIONS .....</b>	<b>35</b>
<b>GLOSSARY OF TRUST DIMENSIONS .....</b>	<b>37</b>
<b>REFERENCES .....</b>	<b>39</b>
<b>ABBREVIATIONS AND ACRONYMS.....</b>	<b>43</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>45</b>

<b>APPENDICES .....</b>	<b>47</b>
<b>Appendix A - SATI Questionnaire v0.2a.....</b>	<b>49</b>
<b>Appendix B - SATI supplement v0.2a.....</b>	<b>57</b>
<b>Appendix C - SATI Questionnaire v0.3 - proposed final version .....</b>	<b>63</b>

## EXECUTIVE SUMMARY

This guideline document describes the development and evaluation of a human factors technique for measuring human trust in ATM systems. The measure is primarily concerned with human trust of ATC computer-assistance tools and other forms of automation support, which are expected to be major components of future ATM systems.

The document contributes the first part of a larger project entitled 'Solutions for Human-Automation Partnerships in European ATM (SHAPE)' being carried out by the ATM Human Resources Unit of EUROCONTROL, which has later become the Human Factors and Manpower Unit (DIS/HUM).

The former UK Defence Evaluation and Research Agency (DERA), now known as QinetiQ, was awarded the investigation of three specific human factors topics concerned with trust (see EATMP, 2003a, 2003b, and this document), situation awareness (see EATMP, 2003c), and teamworking (currently under preparation).

Four additional human factors issues are also in the SHAPE overall objectives: recovery from system failure, workload and automation, future controller skill-set requirements, and experience and age (see EATMP, 2003d).

This deliverable, on the subject of trust measurement, is the second one developed within the SHAPE Project. A related deliverable provides a set of human factors guidelines for facilitating and fostering human trust in ATM systems (see EATMP, 2003a). A subsequent deliverable provides detailed information on trust principles (see EATMP, 2003b).

Section 1, 'Introduction', outlines the background to the project, and the objectives and scope of the document.

Section 2, 'Trust Background', recaps what is meant by trust and briefly summarises previous work on trust, trust dimensions and trust measurement.

Section 3, 'Development of a Trust Measure', begins by examining the potential overlap between the SHAPE trust measure named 'SATI' for 'SHAPE ATM Trust Index' and other SHAPE measures. The development process for SATI is explained, its theoretical framework and assumptions are outlined, and the structure and contents of SATI are described.

Section 4, 'SATI Evaluation and Validation', describes the process and findings of SATI evaluation through its usability assessment in real-time simulation experiments. The construct validity of the SATI Measure is also determined using feedback from controllers and assessment of the technique against a set of validation/success criteria. Problems in assessing trust in simulation environments and of establishing empirical validity are covered.

A rich set of controller comments has been obtained. A key finding is that controllers regard ATC trust as a discrete binary (Yes/No) concept, linked to their usage of any automation tool. Confidence, on the other hand, is a finer-grained continuous variable. This finding is in direct contradiction to the bulk of the previous, process-control derived, research literature on trust.

Finally, a number of proposed changes to the SATI Measure are recommended and some future work for empirical validation and modelling of SATI scores is proposed.

Section 5, 'Conclusions', summarises the findings concerning SATI and provides guidelines for deploying SATI to measure controllers' evolving trust in the design and development of future ATM systems.

Section 6, 'Recommendations', lists a number of recommendations for the further development of the SATI Measure.

A Glossary of Trust Dimensions, References, a list of the Abbreviations and Acronyms used in these guidelines and their full designations, and Acknowledgements can be found at annex.

Copies of the current SATI Measure and its supplementary questionnaire are appended (see Appendices A and B). Taking into account the usability and validation feedback from controllers, a proposed revised version of SATI is also appended for discussion purposes (see Appendix C).

## 1. INTRODUCTION

### 1.1 Purpose

The purpose of this document is to provide a human factors technique for measuring human trust in ATC systems. The measure is primarily concerned with trust in ATC computer-assistance tools and other forms of automation support, which are expected to be major components of future ATM systems.

Trust is important because from the controllers' point of view it means, ultimately, accepting information, advice and decisions from the automation, and possibly accepting system intervention too.

### 1.2 Scope

The document is the second of a series within the SHAPE Project. It is intended to provide a description of the development and evaluation of a human factors technique for measuring human trust in ATM systems, particularly those incorporating computer-assistance tools and other forms of automation support.

In addition, the deliverable aims to provide a resource in the form of a practical trust measurement technique for EUROCONTROL project leaders and other project staff who are concerned with measuring trust. The trust measure is intended principally for deployment in real-time simulations of future ATM systems.

### 1.3 Background

The work on trust presented in this module is embedded in a larger project called 'Solutions for Human-Automation Partnerships in European ATM (SHAPE)'. The SHAPE Project started in 2000 within the Human Factors Sub-Programme (HSP) of the EATMP Human Resources Programme (HRS) conducted by the Human Factors and Manpower Unit (DIS/HUM) of EUROCONTROL, formerly known as the ATM Human Resources Unit (see EATMP, 2000).

SHAPE is dealing with a range of issues raised by the increasing automation in European ATM. Automation can bring success or failure, depending on whether it suits the controller. Experience in the introduction of automation into cockpits has shown that, if human factors are not properly considered, 'automation-assisted accidents' may be the end result.

Seven main interacting factors have been identified in SHAPE that need to be addressed in order to ensure harmonisation between automated support and the controller:

- **Trust:** The use of automated tools will depend on the controllers' trust. Trust is a result of many factors such as reliability of the system and transparency of the functions. Neither mistrust nor complacency are desirable. Within SHAPE guidelines were developed to maintain a correctly calibrated level of trust (see EATMP, 2003a, 20003b, and this document).
- **Situation Awareness (SA):** Automation is likely to have an impact on controllers SA. SHAPE developed a method to measure SA in order to ensure that new systems do not distract controllers' situation awareness of traffic too much (see EATMP, 2003c).
- **Teams:** Team tasks and performance will change when automated technologies are introduced (team structure and composition change, team roles are redefined, interaction and communication patterns are altered). SHAPE has developed a tool to investigate the impact of automation on the overall team performance with a new system (currently under preparation).
- **Skill set requirements:** Automation can lead to both skill degradation and the need for new skills. SHAPE identifies new training needs, obsolete skills, and potential for skill degradation aiming at successful transition training and design support (currently under preparation).
- **Recovery from system failure:** There is a need to consider how the controller will ensure safe recovery should system failures occur within an automated system (currently under preparation).
- **Workload:** With automation human performance shifts from a physical activity to a more cognitive and perceptual activity. SHAPE is developing a measure for mental workload, in order to define whether the induced workload exceeds the overall level of workload a controller can deal with effectively (currently under preparation).
- **Ageing:** The age of controllers is likely to be a factor affecting the successful implementation of automation. Within SHAPE this particular factor of human performance and its influence on controllers' performance are investigated. The purpose of such an investigation is to use the results of it as the basis for the development of tools and guidance for supporting older controllers in successfully doing their job in new automated systems (see EATMP, 2003d). Note that an additional report providing a questionnaire-survey throughout the Member States of EUROCONTROL is currently under preparation.

These measures and methods of SHAPE support the design of new automated systems in ATM and the definition of training needs. It also facilitates the preparation of experimental settings regarding important aspects of human performance such as potential for error recoveries or impacts of human performance on the ATM capacity.

The methods and tools developed in SHAPE will be complied in a framework in order to ease the use of this toolkit in either assessing or evaluating the impact of new systems on the controller performance, efficiency and safety. This framework will be realised as a computerised toolkit and is planned to be available end of 2003.

## 1.4 Structure

The document is divided into six sections, following this Introduction, as shown in Figure 1.

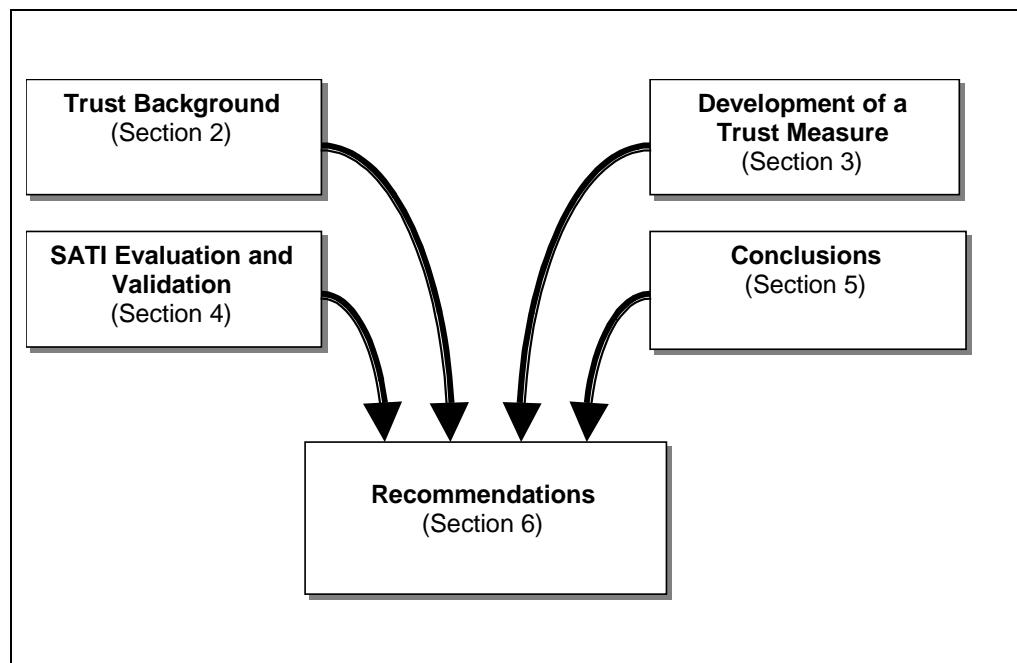


Figure 1: Structure of the guideline document

Page intentionally left blank

## 2. TRUST BACKGROUND

### 2.1 What is Trust?

Trust is a familiar term in everyday language but conveys a variety of meanings. The first deliverable of the SHAPE Project (EATMP, 2003a) notes that psychological trust is an internal state manifest as a subjective experience, an *intervening* variable between particular external conditions and observable human behaviours.

However, in order to measure controllers' trust in the context of complex, human-machine ATM systems, it is necessary to reach an operational definition. Moray (2001) notes that trust is centrally important in ATC because system designers want controllers to actually use their automation tools where such tools are reliable and useful – to use the information, to accept the advice, decisions or interventions from the automation. Some ATC computerised tools provide 'just' information (e.g. STCA, MTCD), whilst other more advanced tools provide advice or recommendations (e.g. CORA). To include both scenarios, the definition from EATMP (2003a), based on Madsen and Gregor (2000), may be extended as follows:

**Trust is the extent to which a user is willing to act on the basis of the information, recommendations, actions, and decisions of a computer-based 'tool' or decision aid.**

### 2.2 Previous Work

The first guideline document of SHAPE (EATMP, *op. cit.*) also reviews relevant research literature in the fields of automation, trust dimensions and trust in human machine systems. The literature mainly concerns industrial process control systems and their faults.

Research on trust in ATM systems, as opposed to controller workload, was found to be surprisingly limited and somewhat anecdotal, given its undoubtedly importance (Kelly *et al.*, 1995; Kelly & Goillau, 1996; Graham *et al.*, 1994; Whitaker & Marsh, 1997; Reichmuth *et al.*, 1998; DERA, 1997; Goillau *et al.*, 1998, Nijhuis *et al.*, 1999; Masalonis *et al.*, 1998, 1999; Chabrol *et al.*, 1999; EUROCONTROL, 2000a, 2000b). The possibility of an operator's over-reliance on automation, or *complacency*, was also raised (Parasuraman *et al.*, 1993; Parasuraman & Riley, 1997; Moray, 1999; Moray & Inagaki, 2001).

The EATMP (*op. cit.*) document finally surveys previous attempts to measure trust before synthesising a set of guidelines for developing trust in ATM

systems. A number of dimensions of trust (see '[Glossary of Trust Dimensions](#)') are also noted, based on the work of authors such as Rempel, Holmes and Zanna (1985), Sheridan (1988), Muir (1994), Muir and Moray (1996), Lee and Moray (1992, 1994), Jian *et al.* (1998, 2000), and Madsen and Gregor (2000).

**Trust is a construct composed of several elements or dimensions. The main dimensions identified in the research literature are:**

- ***Predictability***
- ***Dependability***
- ***Faith***
- ***Reliability***
- ***Robustness***
- ***Familiarity***
- ***Understandability***
- ***Explication of intention***
- ***Usefulness***
- ***Competence***
- ***Self-confidence***
- ***Reputation***

A simple influence diagram model of trust is further proposed to allow trade-offs between the different influencing factors. This is shown in [Figure 2](#). The trust model distinguishes between automation attributes and human properties, both cognitive and attitudinal or emotional.

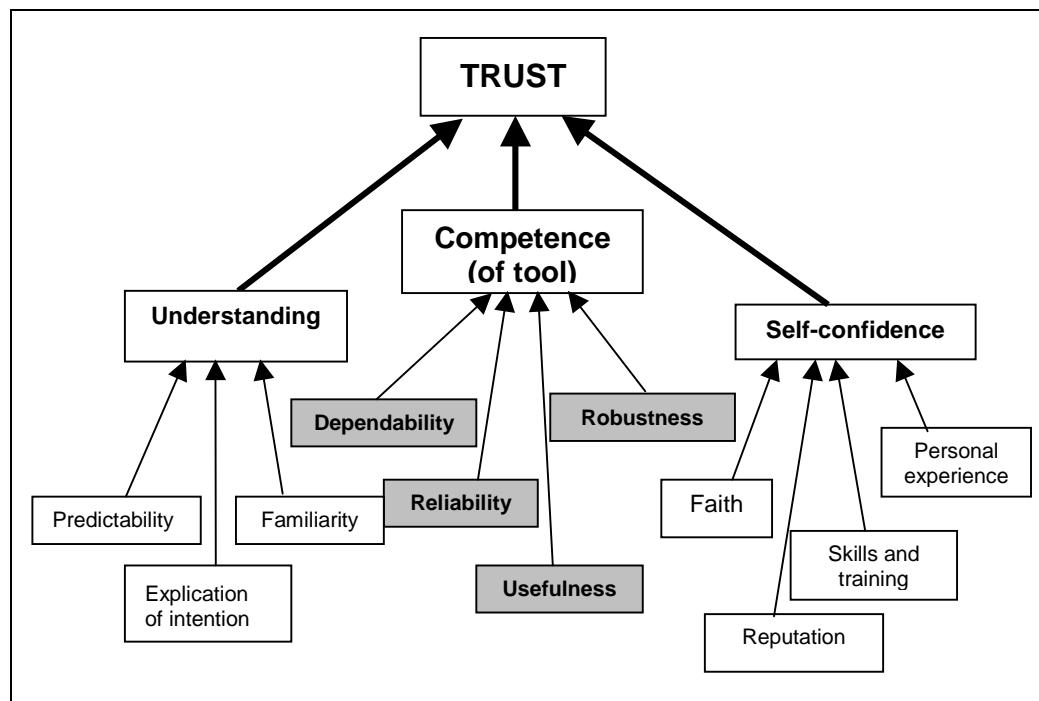


Figure 2: Simple model of trust and the relationship between factors

It is noted that a fundamental premise of the SHAPE Project is that the concept of trust and the dimensions of trust are equally applicable to the domain of ATM as they are for other domains such as industrial process control. Whereas it is accepted that controllers and pilots must trust each other, their procedures and their equipment (Hopkin, 1975, 1998), it is observed that controllers thoughts about trust are couched more in terms of automation reliability and benefits, or at least their understanding and knowing the limitations of their systems.

**Trust is an intrinsic part of air traffic control. Controllers must trust their equipment, and trust colleagues, and trust pilots to implement the instructions they are given.**

**Controllers' trust in automation is a key determinant in the development and implementation of new ATM systems. In order to develop that trust at an appropriate level, and to avoid inappropriate distrust, it is essential that:**

- controllers understand the functionality of the automation, and its limitations;
- controllers are given proper and sufficient training;
- the simulation system in general, and the automation in particular, are highly reliable.

## 2.3 Measurement of Trust

The SHAPE Work Package (WP) 1 deliverable (EATMP, *op. cit.*) says that evidence from many empirical studies indicates the use of subjective questionnaire-based rating scales is the most common means of measuring trust. Muir and Moray (1996) had concluded that trust is not a discrete variable, but that variable levels of trust can exist between none and total. Jian *et al.* (2000) also provided the first empirical evidence that the concepts of trust and distrust could be treated as opposite ends of a trust continuum. In practical terms, this implies that trust and distrust can be measured using the same rating scale.

Five rating scale approaches to trust measurement were reviewed. The simplest comprised a single rating scale to evaluate operators' overall trust (Lee & Moray, 1992, 1994). There are clear analogies to the use of the Instantaneous Self-Assessment (ISA) workload measure for ATC. More sophisticated techniques used multiple rating scales to elicit dimensions of trust (Muir, 1994; Muir & Moray, 1996; Taylor, 1988; Taylor, Shadrake & Haugh, 1995). Other approaches used multiple scales to rate the degree of agreement/disagreement with a number of trust-related statements (Madsen & Gregor, 2000; Jian, Bizantz & Drury, 2000). A subjective rating scale approach was therefore recommended as most appropriate for SHAPE trust measurement.

Page intentionally left blank

### 3. DEVELOPMENT OF A TRUST MEASURE

#### 3.1 Overlap between SHAPE Measures

The focus of this document is on developing and evaluating a measure of ATC trust. The trust measure has been termed 'SATI' (for 'SHAPE Automation Trust Index'). The goal of SATI is to provide a means of measuring trust at some level, so leading to the identification of trusted and usable ATC automation tools and ultimately to effective combined human-automation ATM system performance.

However, it is possible that trust will not be independent of the other proposed SHAPE measures, namely Situation Awareness (SA) and teamwork. It can be hypothesised that if the controller has a good SA then he/she may also have a high level of trust in the system being operated. Also, if the controller team works well together and has good task and social 'cohesion', there may be a high level of trust distributed between team members. The potential overlap between the SHAPE measures is shown in Figure 3.

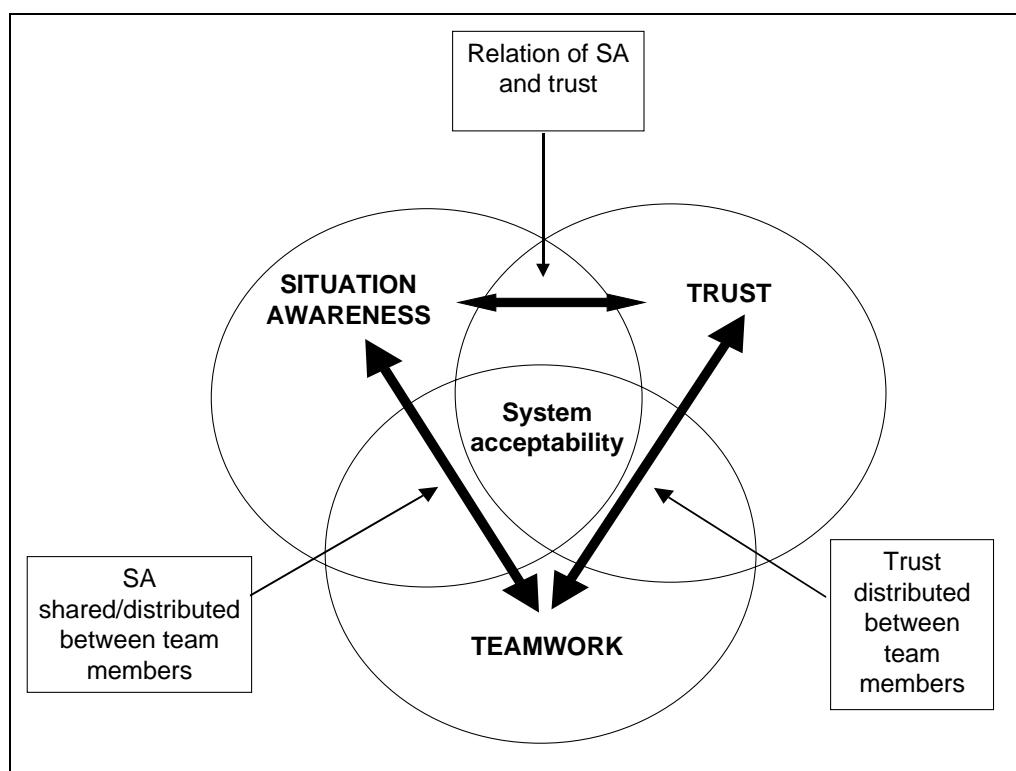


Figure 3: Potential overlap between SHAPE measures of trust, SA and teamwork

Moray (2001) remarks on the overlap between measures. It may be that the relation between SA and trust is unidirectional. Good SA of a reliable system

leads to greater trust, but high trust leads to less SA due to less frequent monitoring. This is what has been called 'complacency'. Indeed, Muir and Moray (1996) found that the more reliable an automated subsystem and the more it was trusted, the less frequently it was observed. Moray and Inagaki (2001) even demonstrate that a rational strategy for someone using a 100% reliable system is never to monitor it!

In future work it may be possible to combine, or embed, the SATI Measure components within the other SHAPE measures of SA or teamwork. As the latter are also in development, this remains an issue for subsequent exploration.

### 3.2

### Development Process for SATI Measure

Development of the SATI Measure followed a defined process. The process can be summarised as rapid prototyping and iterative refinement based on usability feedback and informed user comments (as illustrated in [Figure 4](#)).

Starting point: The starting point for the development of SATI was the literature review undertaken in the first SHAPE guideline document (EATMP, 2003a). SATI development was particularly informed by the efforts of previous researchers to measure generic trust in HMI systems, notably the work of Madsen and Gregor (2000). The ATM systems experience of the present authors, especially in ATC human factors evaluation trials and interviewing controllers, also played an important role in establishing an initial prototype.

EUROCONTROL requirements for use of the trust measure in the context of real-time simulations were also a strong motivating factor. Trust can be thought of as an 'enabler' to the successful introduction of new ATM systems. The aim is to find some diagnostic indicators that can help in finding solutions to optimise trust in new ATM systems. It is useful, therefore, to measure controllers' trust during real time simulations. The requirement was for SATI to be relatively easy to apply without being intrusive (comparable with the ISA workload measure mentioned earlier). It was also to provide a deeper and broader level of contextual diagnostic information concerning aspects and dimensions of trust, comparable with the NASA Task Load Index (TLX) workload diagnostic measure.

Theoretical frameworks and assumptions: The development process was underpinned by a number of theoretical frameworks, including Distributed Cognition, and by a number of assumptions (discussed in the next section).

Iterative refinement: The initial SATI prototype trust measure was subject to a process of iterative refinement. Feedback from two informal usability evaluations was used to refine the measure. Initial and successive versions of SATI were tried out at the EUROCONTROL Experimental Centre (EEC, Brétigny, France) in real-time simulation experiments. Similarly, the psychological construct validity of the SATI Questionnaire was assessed using feedback from controllers in a separate consultation exercise. The evaluation and validation results will be reported in [Section 4](#).

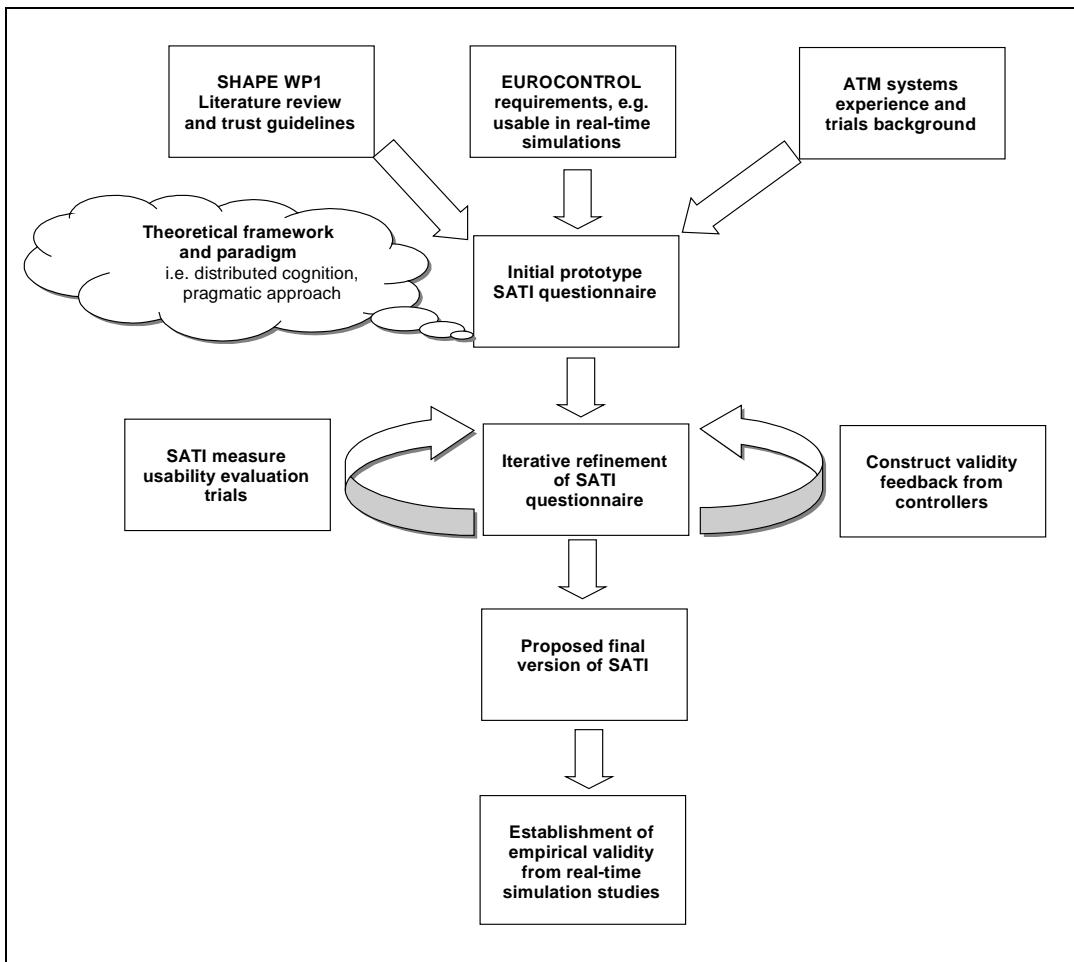


Figure 4: Development process for the SATI Measure

Final version of SATI: Taking into account all the usability trial findings and feedback comments, recommendations can be made for a proposed final (v0.3, as yet untested) version of SATI. These will be covered in Section 6. It remains to establish empirical validity of the final SATI Measure. This could be undertaken by a trial of SATI in a large-scale simulation experiment, and correlating the trust measure scores against available objective system and performance data obtained from the simulation trials. These data might include ATC traffic throughput or the number of measured interactions with an automation tool. Validation is important to establish not only that SATI measures the presence of trust at some level, but that the consequent performance of the combined human-automation system is effective.

### 3.3

## SATI Theoretical Frameworks and Assumptions

### 3.3.1

#### Dimensions of trust

The main influence of the trust dimensions adopted in SATI is the work of Madsen and Gregor (2000), reported in EATMP (2003a).

For the SHAPE trust measure the development of a subjective measure using a rating scale appears to be a simple and straightforward approach that has been used successfully in other domains. A rating scale to measure controllers' overall level of trust would seem to be an appropriate approach. Of the scales reviewed in EATMP (2003a) the one developed by Madsen and Gregor (2000) looks most promising, particularly as the chosen constructs have been shown to have a high degree of empirical validity. Their scales also avoid the emotionality and potential to negatively influence of alternative measures (e.g. Jian, Bisantz & Drury, 2000).

Drawing on the earlier work of Rempel *et al.* (1985), Sheridan (1988), Muir and Moray (1996), and others, Madsen and Gregor (2000) developed a subjective measure for measuring trust of computers. The measure, called the Human-Computer Trust (HCT) scale, consists of five main constructs each with five sub-items. These five items (see '[Glossary of Trust Dimensions](#)') are drawn from an original list of ten trust constructs as having the most predictive validity. Madsen and Gregor claim that the HCT scale has been empirically shown to be valid and reliable. The relationship between the five Madsen and Gregor (*op. cit.*) constructs is shown diagrammatically in [Figure 5](#), in terms of cognitive- and affect-based trust components.

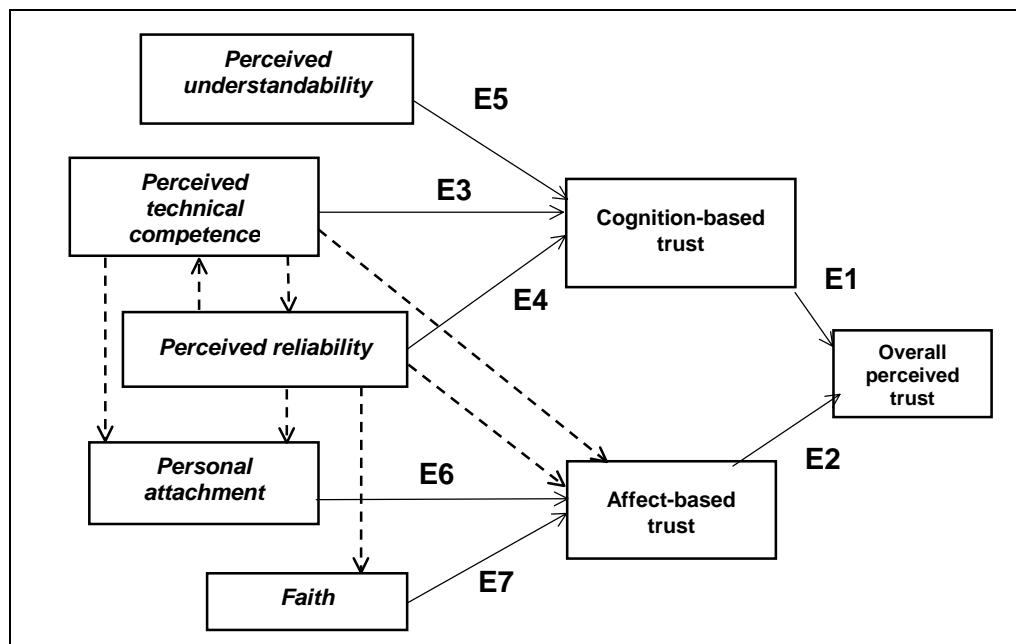


Figure 5: Model of Human-Computer Trust (HCT) components (from Madsen & Gregor, 2000)

Of the original ten trust dimensions identified by Madsen and Gregor the most appropriate to ATM automation, through minimising the possible ambiguity of terms, are judged to be:

- reliability,
- accuracy,
- understanding,
- faith,
- liking,
- familiarity,
- robustness.

Building on previous research these seven factors (as defined in the ['Glossary of Trust Dimensions'](#)) were therefore adopted as the trust dimension framework underpinning SATI.

### **3.3.2      Distributed cognition**

As already mentioned, a second theoretical framework employed is based on Distributed Cognition. Originally developed by Hutchins *et al.* (1991) for modelling cognitive activity in aircraft cockpits, the distributed cognition metaphor has been successfully applied to ATC and teamwork as part of the MEFISTO Project (Fairburn & Wright, 2000). Basically, distributed cognition states that cognitive processes may be manifest internally in a controller's head, or may be partially held externally in a number of outside 'artefacts' such as flight progress strips, radar displays, etc. Whilst usually applied to cognitive processes such as memory, it is hypothesised that 'trust' may also be distributed between external systems and automation tools, colleagues in the controller team, aircraft pilots, a controller's own self-confidence, and any remaining external artefacts. The internal 'self-confidence' factor accords with Muir and Moray's (1996) work.

A 'Rich Picture' (Checkland, 1981) influence diagram showing a potential distribution of ATC trust is shown in [Figure 6](#). Whilst it remains to be tested, this metaphor gives useful leverage for representing and enquiring about the nature and location of trust.

### **3.3.3      Pragmatic measurement**

A final practical assumption is that each of the above frameworks can be translated into subjective questions using a practical What/How/Who/Where/When/Why paradigm. This paradigm may be grounded as separate components or 'modules' of trust measurement:

- What is the overall level of trust? (including "Is this level appropriate in relation to some 'best' level of trust?")
- How is that trust level decomposed into trust dimensions?
- Who/where is the trust distributed between?

- When does the trust vary over time, for example during or between simulation runs?

A final consideration is:

- Why should a particular level or configuration of trust be so for a given ATM automation scenario? Reasons may be elicited by asking controllers for their comments and feedback in a supplementary semi-structured questionnaire format. Such diagnostic information may inform strategies for instilling an appropriate level of trust.

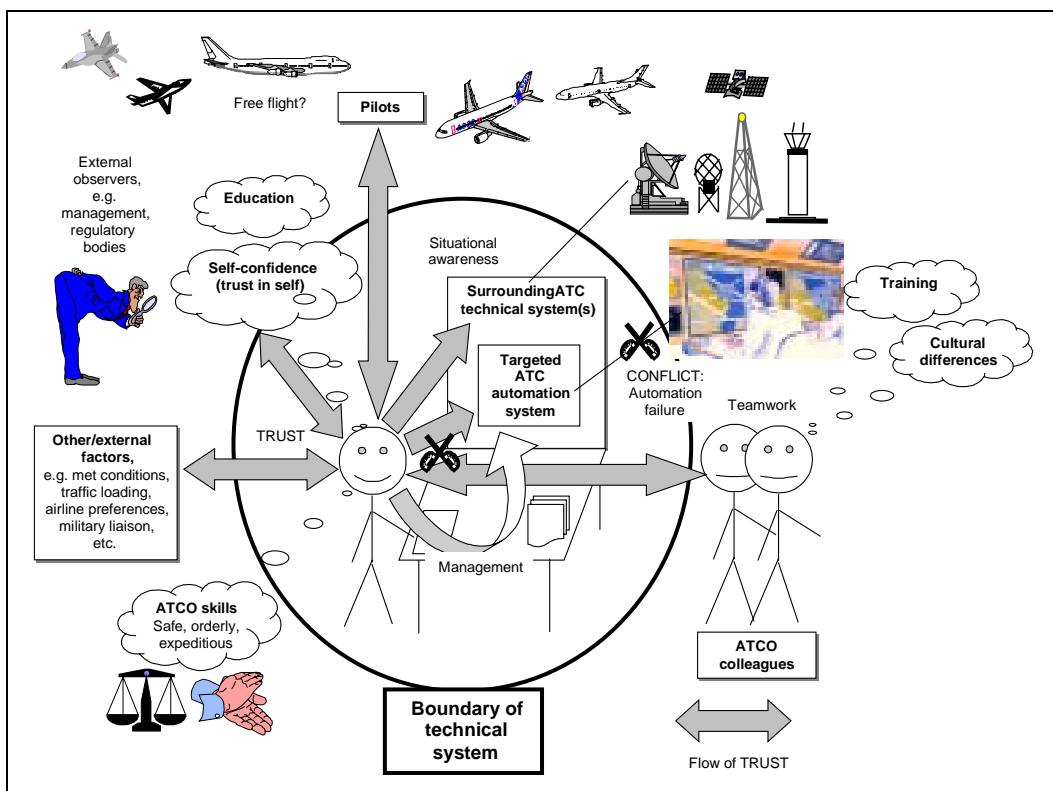


Figure 6: ‘Rich picture’ of trust in distributed cognition model

3.4

## SATI Structure and Content

A copy of the latest tested version of the trust measure, SATI v0.2a, is included in [Appendix A](#). SATI comprises a number of components or ‘modules’ that come together to assess different aspects of trust in ATM systems. These are:

Module	Detail
Overall amount of trust	In the simulated system / automation tool
Variation over time	Trust level at the beginning and end of a time period
Decomposition of trust	Decomposition into dimensions of trust, with ratings of the relative importance of each dimension
Distribution of trust	Between key equipments, actors and artefacts
General comments	Remarks on factors that have influenced trust

In addition, there are introductory explanatory sections and records of 'housekeeping' information to maintain configuration control of the trust data.

An additional 'SATI supplement', shown in [Appendix B](#), attempts to fulfil the need to understand the reasons why a particular trust level and configuration should be so. A definition of trust is sought, along with factors increasing or decreasing trust in each automation tool and more generally in the human elements of the system i.e. fellow controllers and pilots.

### 3.5 SATI Usage

The intention is that SATI should be available as a flexible framework of trust measurement 'modules' that can be tailored to a particular trust measurement requirement in a given real-time simulation.

The first part of SATI could be used alone to measure overall trust levels at intervals during or at the beginning and end of a simulation run. Alternatively, a full or sub-set of SATI modules could be used in a more diagnostic mode to measure trust components and track their changes. That is, by applying SATI at the end of each simulation run, at the end of each day's runs, or at discrete intervals during the duration of a simulation experiment.

Moray (2001) notes that the SATI scores may be interpreted as a measure of the inherent 'trustworthiness' of the system, which may therefore need to be modified. Alternatively, the SATI scores may indicate that further training on the system is required to give proper opportunity for appropriate controller trust to develop, particularly if there is an apparent mismatch between measured controller trust levels and the known reliability of the system. In practice, both viewpoints may be valid to some extent in real-time simulations

Page intentionally left blank

## 4. SATI EVALUATION AND VALIDATION

### 4.1 SATI Usability Evaluations

#### 4.1.1 Objectives

Evolving versions of SATI were tested as part of its iterative refinement development process. The objective was to assess the effectiveness of SATI as a practical measure of measuring controllers' trust in ATM automation tools, and to obtain feedback on SATI's usability from which future versions of the measure could be improved.

#### 4.1.2 Evaluation process

Testing took place during two real-time simulations at the EUROCONTROL Experimental Centre (EEC), during the last months of 2000. The real-time simulations were:

1. Conflict Resolution Assistant Level 1 (CORA1) EATMP Validation Platform (EVP) simulation.
2. Free Route Airspace Project 5 (FRAP5) simulation.

An early working version of the trust measure, SATI v0.1, was first tested early in its development process at the EEC on the 23<sup>rd</sup>, 24<sup>th</sup> and 27<sup>th</sup> November 2000. Testing took place as an adjunct to the CORA1/EVP real-time simulation of Reims airspace. CORA1 tools included system-assisted coordination, new graphical displays, and interaction modes, with decision support tools for detecting and managing conflict and problem information. These were Medium-Term Conflict Detection (MTCD), Monitoring Aids (MONA) and Vertical Assistance Window (VAW), as well as Short-Term Conflict Alert (STCA). The goal was to assess the utility of the SATI Questionnaire for measurement of trust, rather than the CORA1/EVP tools themselves. SATI was administered to the controllers each morning and after the third exercise on each of the three days. A 'trust-oriented' debrief session was held, and proved extremely informative. A EUROCONTROL representative was also present. The two Irish and two Romanian controllers who acted as subjects were able to complete all sections of the SATI Questionnaire, though there were some problems in rating the relative importance of the trust dimensions. A useful set of comments was elicited. As a result of this usability feedback, refinements were made to SATI v0.1, which were incorporated in SATI v0.2.

SATI v0.2 was next tested as an adjunct to the FRAP5 real-time simulation at the EEC on the 11<sup>th</sup>–13<sup>th</sup> December 2000. FRAP5 tools also included the MTCD. The goal was again to assess the utility of the SATI Questionnaire for measurement of trust, rather than the FRAP5 tools themselves. However, it

should be noted that there had been a number of technical problems during the preceding week of the FRAP5 simulation experiment, which caused the controllers' perceived level of trust in the simulation and tools to be initially very low.

SATI was administered to the controllers each morning and completed after the second exercise on each of the first two days. The 11<sup>th</sup> December simulation condition comprised the 'no tools' condition (i.e. only System Supported Coordination (SYSCO), Civil Military Coordination and On-line Data Interchange (OLDI) estimates passed between controllers). The 12<sup>th</sup> December runs comprised the 'with tools' condition, adding the MTCD tool. The twenty Scandinavian and German controllers who acted as subjects were able to complete all parts of SATI as instructed, but protested that their mental model interpreted the questions in terms of degree of confidence rather than of trust. Controller trust was a binary Yes/No construct. This point will be elaborated further. Averaged SATI scores were again sensitive to the (low) reliability of the simulation and the tools, which was reflected in the (low) mean trust levels obtained. 'Trust-oriented' debrief sessions were also held at the end of each of the first two days, and proved extremely informative.

In view of the issues raised at these debrief sessions, and EUROCONTROL's comments from the CORA1/EVP simulation, a SATI supplement was produced and administered on the third day, together with a final trust-oriented debriefing session. Given the large number of controllers present, the aim was to use this valuable opportunity to get a better handle on some of the underlying issues concerning controllers' views of trust and its influencing factors. The twenty controllers were able to complete all sections of the SATI supplement. A useful set of comments was elicited, which led to recommendations for refinement of the SATI v0.2 Questionnaire.

## 4.2 SATI Evaluation Results

The comments from the CORA1/EVP and FRAP5 controllers regarding SATI's usability may be clustered into three categories:

1. SATI feasibility for use in real-time simulations.
2. SATI quality improvements.
3. Controllers' views on trust and confidence concepts.

In addition, supplementary feedback was obtained from the FRAP5 controllers using the SATI Supplement Questionnaire.

### 4.2.1 SATI feasibility and usability in real-time simulations

- The 'before-after' principle for measuring overall trust levels worked well.
- SATI seemed simple and easy to use. The 'smiley' was appreciated.

- Account needed to be taken of the practice of controllers rotating between simulation 'seats' during the EEC simulations.
- It was useful to complete the trust dimension rating scales after each simulation run, but it was not necessary to rate the importance of these dimensions each time.

#### **4.2.2 SATI quality improvements**

- The SATI Questionnaire wording needs to be kept simple and concise, particularly for a multicultural audience. The questionnaire wording needs to be adapted and made specific to the particular simulation (e.g. CORA1) and its component automation tools.
- Some specific wording of the SATI items needed improving, for example the terms 'accuracy' and 'reliability' seemed identical to the CORA1/EVP controllers. Single rather than multiple adjectives should be used on the scales.
- The dimensions of the trust distribution 'spidergram' needed to be made more concrete and specific to each simulation.
- Controllers proposed a number of additional key questions they would ask of other controllers to measure their trust in the automation tools: For example: "Would you work live traffic with these tools?"; "Can you see these tools in the room in five years time?"; "How would you change the system before putting it in the control room?". Note that these questions all imply a binary Yes/No view of trust.
- Concerning the SATI Questionnaire the meaning of certain SATI questions (e.g. the first question) was unclear and required rewording to make it specific to the simulation. Controller's main concerns centred around the use of the word 'trust' itself.
- SATI questions need to be aimed at specific components of a system, otherwise the system will be rated at the level of the least trusted component. This implies a separate SATI Questionnaire for each automation tool within a simulation.

#### **4.2.3 Controllers' views on trust and confidence**

- As already noted, controllers had problems with the word 'trust'. They held the view that ATC trust was a discrete variable, i.e. either the controllers trusted an automation tool or they did not. Therefore, though they completed the questionnaire, the scientific concept of a '57% trust level' was meaningless to them. This finding that ATC trust is a discrete binary (Yes or No) construct directly contradicts the bulk of the research literature reviewed in EATMP (2003a). Note that the latter is mostly based on industrial process control, laboratory simulation studies that may be less relevant to ATC.

- The controllers further avowed that although their level of **trust** was binary, their level of **confidence** in each referent was variable. SATI questions were interpreted in this light. Therefore, it would seem that controllers regard trust and confidence as different things. What authors of scientific research literature consider to be 'trust', controllers consider to be 'confidence'. This is an important distinction.
- Controllers also linked this view of trust into the use of automation tools. If they do not trust a tool they will not use it, if they do trust it they will use it - provided that their level of confidence in the tool for the specific situation is above a certain criterion level, defined by prior experience. Therefore confidence is situation dependent, and so is tool usage.
- In addition, how does perceived usefulness or utility impact on use? If a system is trustworthy, but perceived by controllers to be of no practical use, then it will not be used. There was an example of this scenario during the FRAP5 simulation: the MTCD tool was used by some controllers and not by others. Some did not use it because they did not trust its reliability, some did not use it simply because they thought it was not of practical use or did not help them.
- Controllers' lower trust/confidence limit was more in relation to the nature of the tool 'failure' rather than in terms of number or frequency of failures. Controllers avowed that they would be ready to trust the system if it failed to alert complex conflicts (e.g. two aircraft merging), but trust would be lost if a single simple head-on conflict were missed.
- Following both group and individual discussions with controllers, it has become apparent that ATC trust is more complex and emotive than is suggested in the existing literature. This is especially true when discussing trust in an ATCO colleague. It may not, therefore, be appropriate to use the existing academic models of trust with reference to air traffic controllers. Instead, another model is needed that conforms to controllers' attitudes and conceptions of trust and confidence. This is especially important in the generation of questions for the questionnaire. Future work might usefully consider the construction of a comprehensive ATCO-specific trust and confidence model, though this is beyond the scope of the present study.
- However, when asked if they monitored certain colleagues, systems or pilots more than others, they said that they did. Given the link suggested in the literature between monitoring and trust, this would imply that either controllers perceive trust differently to that of the authors or there is another intermediate factor or factors involved (such as confidence, reputation, etc.).

The following illustrative comments from the FRAP5 controllers are particularly informative:

- “A tool is more likely to be trusted and used if accurate and simple to use, this is especially important when the user is under stress (which is actually when the tool is likely to be of most use).”
- “False alarms are annoying, especially obvious errors (like flagging a conflict alert for two aircraft which will miss by sixty miles), but not as annoying as missing conflicts completely or alerting the user after they have happened.” As previously noted, there were a number of technical problems in the simulation experiments.
- “Speed of system response is very important.”
- “Measures and display of the system’s confidence that its answers are right are very useful.”
- “Number, severity and frequency of mistakes/failures influence a controller’s confidence and monitoring behaviour.”
- “Trust is initially earned then remains relatively constant despite subsequent errors (automation or human). All systems fail from time to time and this is to be expected, so when a system fails it is not totally unexpected and trust/confidence returns to its previous levels immediately.”
- “Controllers’ skills, and their abilities to compensate for poor automated systems, mean that an unreliable system can still produce good ATC results.”

#### 4.2.4

#### SATI supplement

Information was sought in the SATI supplement, at EUROCONTROL’s suggestion following CORA1/EVP, concerning the FRAP5 controllers’ perceptions of acceptable ‘bands’ of trust and acceptable levels or reliability and frequency of system failure. Results, expressed as percentage agreements with the following statements, were as follows:

- Statement: “I either trust, or I do not trust, ATC automation.” 78% agreed.
- “But I have various degrees of confidence in ATC automation.” 78% agreed.
- “There is a minimum level of trust for me to use ATC automation.” 56% agreed, giving a very high level of 99.95% to 100% required reliability. 44% disagreed, being unable to define any such level.
- “There are acceptable failure rates for ATC automation” were stated as NEVER / Once per year.

- “There are acceptable failure frequencies for ATC automation” were similarly stated as NEVER / One in 100,000 interactions.

It is impossible to emphasise this point strongly enough. As one controller remarked, “We have to rely absolutely on our automation tools – not just to keep aircraft safe, but to protect our jobs. If there were a potential accident, we could be held liable – a judge would ask why we used and relied on a tool if we didn’t trust it. If it didn’t work we should have said so.” However, Moray (2001) stresses the need to ask controllers how many system failures they have actually experienced in recent times, by way of grounding their estimates of acceptable reliability levels.

Overall, the controller’s comments proved a rich source of information on their perceptions of trust. Automation tools were used if they were stable, accurate, reliable, simple, not too sophisticated, available and had a nice HMI. Anything else was a bonus! In particular, the controllers needed to know when the system was wrong (by making system failures clear, salient and understandable). Otherwise, performance of the human-machine system would be reduced because the controller would have to spend time monitoring system performance as well as doing their job – maintaining compensation strategies for potential automation system failure is very time-consuming when made necessary.

## 4.3 SATI Validation

Various forms of psychological instrument validity exist (Cook, 1998). Within the concept of ‘construct validity’, these include ‘face validity’ (i.e. does the overall psychological instrument appear reasonable) and Item validity (i.e. does each item within a psychological instrument measure some dimension of the construct being addressed). These issues are best determined by asking informed users, the ‘Subject Matter Experts’, i.e. controllers. Empirical validity, on the other hand, is concerned with whether the psychological instrument scores are statistically well correlated with current objective task measures (‘concurrent validity’), or future objective task measures (‘predictive validity’).

### 4.3.1 SATI construct validation

A brainstorming meeting was held with two very experienced, ex-operational UK controllers. The goals were to determine:

- What is trust/confidence from a UK controllers’ perspective?
- SATI face validity – does the questionnaire look reasonable overall?
- Item validity – does each SATI item usefully reflect some component of trust? Should any items be omitted, reworded or additional items included?

- Ratings against a set of EUROCONTROL-agreed validation/success criteria.
- Consideration of any nationality / ATC-cultural differences.

#### 4.3.2

#### General comments on trust

Using the SATI supplement (see [Appendix B](#)) as a springboard, the two controllers' views were elicited on the general constructs of 'trust' and 'confidence'. Trust was defined as "reliable performance in terms of behaviour". It was confirmed that ATC trust is viewed as a discrete concept: equipment is used and trusted, or is not used so not trusted. A three-category 'Bad/OK/Good' metric was suggested. However, confidence as a metric could be more fine-grained than trust.

Trust was considered to be an implicit thing – earned, but rarely spoken of. Most controllers would never have considered whether they 'trust' equipment - just whether it works, is useful and they can use it. The visibility of equipment, or automation tool, failure is an important consideration – if it breaks obviously, then controllers can make allowances and work round it. If it is faulty but this is not apparent, the situation becomes far more dangerous because they do not know to look out for it and compensate. Moray (2001) remarks that this is similar to mode errors in aircraft – the system must always make clear its state.

A distinction was made between trust in equipment and trust in people. Confidence in colleagues starts from a baseline of their known training and experience, but is won gradually over time from working alongside them on a sector and experiencing them in action. "Trust is the baseline you have to establish with the people you work with." Trusting a colleague over a protracted period of time is a moral construct based on their integrity, ability and motives. "They say they'll do it, and you know they will." Trust is not necessarily shattered if a colleague 'messes up'; allowances can be made. Though all controllers have 'graduated' from an ATC training college, not all are cut out for and capable of validating at, say, the busy Heathrow approach or on the demanding LATCC sectors. People can still be valuable team members in other ways.

#### 4.3.3

#### Construct validity - face validity

The controllers remarked that the SATI Measure v0.2 looked OK, but they were concerned that it was too complex. They had also remarked that a real-time simulation environment was very different from the operational one, the effects of which should not be overlooked on measurements of trust. This point is revisited later.

In general they considered the wording of SATI difficult for controllers to follow, it was not "controller friendly". Changes were proposed to simplify the questionnaire wording, and these are noted below under item validity.

#### 4.3.4 Construct validity - item validity

Each individual SATI component was considered on its merits. The controllers suggested replacing the 0-100% 'trust meter' scale with the question "What do you think of the simulation?" and just three-response categories: "Bad/OK/Good". The ruler scale could be retained against these categories if desired. Trust itself was a 'Yes/No' variable, though confidence could be a more fine-grained measure.

In general the SATI questions needed to be simplified and made specific to the particular simulation and automation tools. It was proposed to add one additional question and to delete, combine or reword certain existing items. A key additional question under trust dimensions was the practical usefulness of a tool – "Does it make your task easier?". The reliability and robustness questions were almost the same so could be merged. The faith question could be deleted, because controllers would not know (about tool performance in unknown situations) so could not answer. Moray (2001) comments that for Muir (1994), faith is the answer to the question "Do you believe that it will deal with all (or at least most) of the situations which you and it have not yet encountered?". The controllers thought there was no place for faith in the unknown in the ATC world.

The spidergram format for distribution of trust was confusing and needed either explaining or re-formatting into conventional horizontal scales. The spidergram format would, however, still be fine for representing the results.

The previous spidergram trust distribution questions could be re-worded as "Did you like the automation tool?"; "Did the simulator work properly?"; "Rate your own performance.". On simulators, trust in (pseudo-)pilots is difficult to answer (it is policy always to be polite to the 'blip drivers'). Concerning the question on trust in local colleagues, not many controllers would give a true reply. The controllers suggested using the word 'confidence' and relating this back to trust.

In the SATI supplement questions such as 'acceptable level of trust' or 'acceptable failure rate' were meaningless and impossible for controllers to answer, because they implied acknowledging and endorsing a system that may fail. One controller commented "If equipment is 100% reliable, useful and friendly, we use it. If it's useful but difficult to use, we use it occasionally. If it's unreliable or doesn't help in the job, it won't be used.".

The question on whether controllers had equal confidence in pilots was valid. Confidence was unequal, depending on the pilot's command of English and the originating country of the airline. However, asking controllers to rate their trust in colleagues was considered unfair in that it placed them in a difficult position – "this question tests the veracity of the supplicant!".

#### 4.3.5 SATI validation/success criteria

SATI Questionnaire v0.2a was finally rated against a number of validation/success criteria previously agreed with EUROCONTROL. Each criterion could be assessed as to whether it had been broadly met, and rated on a 1-5 scale (with 1= SATI does not currently meet the criterion, up to 5= SATI fully meets the criterion). Some of the criteria are conflicting (e.g. concise vs detailed enough), and the goal was for SATI to achieve an acceptable compromise between the various criteria. It had been hoped to further allow the weighting of the individual criteria, but time did not permit this refinement.

The criteria used were:

- usable in real-time simulations,
- practical,
- simple,
- concise,
- easy/quick to use,
- acceptable to controllers,
- non-intrusive (ATCOs losing picture),
- contextual questions being included,
- diagnostic – for designers / project managers,
- predictive,
- agreement with theoretical models,
- agreement with ATCOs' perception of trust,
- psychological/construct validity,
- non-interfering with trust,
- detailed enough (opposite of concise),
- understandable (to all nationalities).

The results are shown for each criterion in Table 1 overleaf. Where possible, criteria were assessed against the CORA1/EVP and FRAP5 experience. For other criteria ratings were made by the ex-operational controllers for SATI, both in its existing form of v0.2a (X = v0.2a, ratings were moderately low) and in its proposed modified form (R = revised v0.3, ratings would be improved).

Based on these comments recommendations can be made for an improved version of the SATI Questionnaire (version v0.3), taking the proposed changes into account. These are summarised in Section 4.6.

Table 1: Rating of SATI against validation/success criteria

	Subjective validation criterion/goal	Criterion source (science / EHQ / SHAPE Team / ATCO / system designers)	Criterion met? (Y/N)	Criterion Rating					Evidence source	Notes
				1	2	3	4	5		
1	Usable in real-time simulations	EHQ <sup>1</sup>	Yes					X	CORA1/EVP, FRAP5	Q wording
2	Practical	EHQ <sup>1</sup> / SHAPE Team	See rating		X	R			Controllers	
3	Simple	EHQ <sup>1</sup> / SHAPE Team	See rating		X	R			Controllers	
4	Concise	EHQ <sup>1</sup> / SHAPE Team	See rating		X	R			Controllers	
5	Easy/quick to use	EHQ <sup>1</sup> / SHAPE Team	See rating		X	R			Controllers	
6	Acceptable to controllers	EHQ <sup>1</sup> / SHAPE Team	See rating		X	R			Controllers	Simplify
7	Non-intrusive / ATCOs lose picture?	EHQ <sup>1</sup> / SHAPE Team	tbc <sup>1</sup>						CORA1/EVP, FRAP5 After simulations OK	
8	Contextual questions	EHQ <sup>1</sup>	Yes				X		SATI revisions	
9	Diagnostic – for designers/PMs	EHQ <sup>1</sup>	Yes				X		CORA1/EVP, FRAP5	
10	Predictive	EHQ <sup>1</sup> / SHAPE Team	tbc <sup>1</sup>							
11	Fits theoretical models	SHAPE Team	?							
12	Ask ATCOs – measured trust?	EHQ <sup>1</sup> / SHAPE Team	Yes, rev. SATI tbc <sup>1</sup>				X		CORA1/EVP, FRAP5	Sensitive
13	Questionnaire construct validity?	SHAPE Team	See rating		X	R			Controllers	Simplify
14	Non-interference with trust?	EHQ <sup>1</sup> / SHAPE Team	tbc <sup>1</sup>							
15	Detailed enough? (opposite of concise)	EHQ <sup>1</sup>	Yes (too detailed?)		X	R			Controllers	Simplify
16	Understandable (by all nationalities)?	EHQ <sup>1</sup> / SHAPE Team	See rating		X	R			Controllers	Simplify wording

<sup>1</sup> EHQ: EUROCONTROL Headquarters – tbc: to be confirmed

#### 4.4

### Problems of Measuring Trust in Simulations

A number of problems have become evident in carrying out trust measurements in real-time simulations, rather than in an operational environment. The principal philosophical difference is that in simulations the aircraft are virtual – an actual collision involving real loss of life is not a possibility, unlike in the operational environment. The metaphor of attending a play at the theatre is relevant here. Due to the skill of the director and actors (the simulation designers), the audience (controllers) may become so involved in the play (simulation) that ‘disbelief is suspended’ and they become immersed in the plot (air traffic scenario). With some other plays (simulations), a level of detachment remains and true audience (controller) involvement is never attained. So controllers may be willing to accept and make allowances for (i.e. they may trust) automation tools that they would be unwilling to actually rely on in real-life, where human life and reputation are at stake. As one controller noted, the results regarding trust are likely to be different – simulations are microcosms of, but are **not** the real world. This perspective contradicts that of Moray (2001), whose work with UK military fighter controllers detected no difference in how they worked on the simulator and with real aircraft.

A second, more practical, problem concerns issues of training and study duration. In real-time simulations, controllers will have received a period of theoretical and practical training on any new automation tools, but will not have deployed the tool in continuous operational use for many months or years. Similarly, the duration of many simulation experiments – in the order of several weeks – precludes extended study of any automation tools. The problem is compounded if, as in the informal usability studies reported here, practical considerations restrict simulation access to that obtained over several days. In these circumstances, trust assessment is based on a relatively informal ‘snapshot’ of behaviours, attitudes, opinions and beliefs which may not be representative of a longer period of use. For these reasons caution should be exercised in the interpretation of the results reported here. Further, more extensive, study of the measurement technique is suggested.

#### 4.5

### Empirical Validation

Continuing the above theme, at the time of writing it has not been possible to establish empirical validity of the SATI Measure, by correlating SATI trust scores with objective system performance measures such as traffic throughput or number of times a tool window was accessed. Obtaining such data was not possible at the CORA1/EVP simulation (where SATI was initially tested) due to small numbers of available controllers, or at the FRAP5 simulation due to technical problems with the simulation itself. However, it is recommended that such an exercise be undertaken, if practically possible, at a future real-time simulation.

This empirical validation approach can be seen to be closely related to the work of Moray and his colleagues, reviewed in the first SHAPE guideline

document (EATMP, 2003a) on empirically modelling and mathematically predicting trust (e.g. Moray, Inagaki & Itoh, 2000; Moray, 1999; Muir & Moray, 1996; Lee & Moray, 1992, 1994). Moray has shown that it is possible to develop an empirical model of trust, at least in process control simulations, and that the model equations are highly predictive. Moray suggests that trust in automation can be:

- a) measured directly by asking the operators/controllers;
- b) modelled on the basis of measurements of physical objective properties of the system in real-time;
- c) modelled dynamically to predict trust, self-confidence, and the probability of intervention by operators in automated systems.

In the context of SHAPE the degree of trust in automation could, theoretically at least, be inferred *post hoc* from objective measures of controller performance (e.g. frequency, accuracy or speed of interaction), if the relationship between these measures and the automation could be unequivocally established.

This approach to ATC trust is theoretically entirely feasible. A very simple, but crude measure could be whether or not the controller has activated a particular tool. A more sophisticated measure would be, in the case of a conflict advisory tool such as MTCD, the type of data entered and the controller's measured speed of response. These measures could be used, theoretically at least, to indicate the controller's level of trust.

However, the controllers' perspectives on trust as a binary (present / not present) construct, and confidence as a continuous variable, should be borne in mind. It may be that Moray's approach, described above, could be applicable to modelling ATC **confidence** rather than to ATC **trust** *per se*. This requires further examination.

## 4.6 Future Developments

Based on the comments from the two informal SATI usability evaluation trials, and the construct validation exercise, a number of recommendations can be made for a further refinement of the SATI Measure. These are subject to further discussion and approval by EUROCONTROL, but may include:

- Asking controllers what they thought of the simulation (using a three-point scale: Bad/OK/Good).
- Reframing questions regarding trust into a binary (Yes/No) format.
- Reframing continuous percentage trust scales as appropriately-defined confidence scales.

- Making the generic version of SATI specific to each simulation and its component computer-assistance or automation tools.
- Administering the trust/confidence dimensions rating scales separately for each individual automation tool that is available in a simulation. An additional question on the practical usefulness of each automation tool is required, together with rewording other scales to simplify them and make their language 'controller-friendly'.
- Reframing the trust/confidence dimensions 'spidergram' into a more conventional linear format (the spidergram pattern is informative and may be retained for presentation purposes), and again rewording the scales to simplify them and make their language 'controller-friendly'.
- In addition, Moray (2001) has suggested that controllers should rate their overall level of ability to trust / amount of available confidence at the start of each day of a simulation. Absolute confidence values could then be derived from the percentage scales relative to this value of available confidence, in a similar manner to the Malvern Capacity Estimate (MACE) Technique for deriving absolute capacity from relative workload estimates (Goillau & Kelly, 1997).
- Moray (2001) further recommends, for the SATI, supplement asking controllers how many system failures they have actually experienced in the previous month – to establish whether their acceptable failure rates are based on experience or wishful thinking.

Taking all these points into account, a revised version of SATI (version v0.3) is included at [Appendix C](#). It must be stressed that this version of SATI is as yet untested, but is included as a potential working version for discussion and further evaluation.

As noted earlier, it remains to establish empirical validity of the final SATI Measure by trialling it in a future reliable large scale simulation experiment, and correlating the trust measure scores *post-hoc* against available objective system and performance data obtained from the simulation trials. These data might include ATC traffic throughput, the number of measured interactions with an automation tool or the number of times a colleague is asked to help. It is clearly important to distinguish between trust in the simulation itself and trust in the advice of the automation tools. This activity is recommended, and could possibly take place in conjunction with future validation of the teamwork and SA measures currently being developed in other SHAPE work packages. The possibility remains to explore including or embedding components of the modular SATI trust index within the latter measures. It would also be necessary at some point to establish and maintain SATI trust score population and sub-population norms (Cook, 1998), but that activity is beyond the scope of the present study.

The question of whether trust is inherently all-or-none is an interesting one. Moray (2001) contends that fuzzy set measures may be more appropriate,

since the fuzzy set operators often behave like a discrete switch past a given threshold, despite the underlying variables being continuous. This accords with the work of Moray, Inagaki and Itoh (2000) on the relationship between trust in and reliability of process control automation. Further basic research would be needed in this area before such measures could be incorporated within SATI.

An important issue for future SATI application in the multicultural world of ATC is the understanding of terms such as 'trust' and 'confidence' by European controllers whose native language is not English. Some European languages do not distinguish between trust and confidence – for example, does 'confiance', in French, really map one-to-one onto trust? If ever there is discussion about translating SATI from English into other European languages, great care will be needed and further work will be necessary to explore possible confusions in interpreting the various SATI components.

## 5. CONCLUSIONS

1. This document has described the creation, development, evaluation and validation of a measure of controllers' trust in ATM systems. The trust measure has been named SATI for 'SHAPE Automation Trust Index'. SATI is informed by previous literature on trust and trust measurement, and by theoretical underpinning frameworks of trust dimensions and distributed cognition. It adopts a practical, flexible, modular approach to measurement of different elements and aspects of trust, and is intended for deployment in full or in part during real-time ATM simulations.
2. SATI has evolved through a process of rapid prototyping and iterative refinement of the measure. Informal evaluation testing indicates that the measure is usable by ATCOs during real-time simulations, and that the scores are sensitive to the reliability of the simulation and automation tools.
3. Construct validity of the measure has been assessed using informed feedback from Subject Matter Experts. Validation/success criteria have also been established and the current version of SATI assessed against these criteria. It remains to establish empirical validity of the final SATI Measure by testing it in a number of reliable, preferably large-scale, real-time simulations.
4. SATI was usable by the controllers. A rich set of ATCO comments and feedback has been obtained. A key finding is that controllers regard ATC trust as a discrete binary (Yes/No) concept, linked to their usage or otherwise of any automation tool. Confidence, on the other hand, is a finer-grained continuous variable. This finding is in direct contradiction to the bulk of the previous, process-control derived, research literature on trust.
5. In terms of guidance for usage, the intention is that SATI should be available as a flexible framework of trust measurement elements or 'modules' that can be tailored to a particular trust measurement requirement in a given real-time simulation.
6. The first part of SATI could be used alone to measure overall trust levels at the beginning and end of, or at intervals during, a simulation run, analogous to the ISA measurement of workload. Alternatively, a full or sub-set of SATI modules could be used in a more diagnostic mode to measure trust/confidence components and track their changes. That is, by applying SATI at the end of each simulation run, at the end of each day's runs, or at discrete intervals during the duration of a simulation experiment.
7. An important issue for future SATI application in the multicultural world of ATM is the understanding of terms such as 'trust' and 'confidence' by European controllers whose native language is not English. Care will be

needed and further work will be necessary to explore possible confusions in interpreting and translating the various SATI components.

## 6. RECOMMENDATIONS

1. SATI is not seen as static, but rather as an evolving measure. It is recommended that research on SATI's further development, testing and refinement should continue.
2. The latest proposed version of SATI should be empirically assessed and validated in a reliable real-time ATC simulation trial, which makes available objective system performance measures for correlation purposes with SATI scores. This exercise could possibly be undertaken in conjunction with the empirical validation of other SHAPE measures, namely teamwork and SA. It will also be necessary at some point to establish and maintain SATI score population and sub-population norms.
3. The overlap between SHAPE measures could usefully be further considered, exploring the possible embedding of SATI components within other SHAPE measures of teamwork and SA.
4. The interpretation of SATI scores needs further research. Low SATI scores may appropriately indicate an untrustworthy system – which may need to be modified. Alternatively, and particularly in real-time simulations, further controller training may be indicated in order to give proper opportunity for an appropriate level of controller confidence to develop.
5. The relationship between ATC trust and confidence should be further investigated. Future work might usefully consider the construction of a comprehensive ATCO-specific trust and confidence model, encompassing any difference between controllers' trust in automation and their trust in human colleagues and pilots.
6. Specifically, research could be carried out to investigate the application of fuzzy set measures to the measurement of trust. Fuzzy set operators often behave like a discrete switch past a given threshold, despite the underlying variables being continuous.
7. The underlying reasons require investigation into the difference between the concept of ATC trust, as determined in the present study, and the extant research literature on trust in process control and other domains.
8. Language issues in the interpretation of SATI by non-native English speakers warrant further attention, as does the potential translation of SATI into European languages other than English.
9. Finally, it is believed that the findings from the present study regarding controller trust and confidence in ATM automation tools could provide valuable feedback to interested parties such as system designers. Strategies for imparting this information to relevant stakeholders could usefully be explored.

Page intentionally left blank

## GLOSSARY OF TRUST DIMENSIONS

The following definitions are taken from EATMP (2003a).

Trust Dimension	Definition
1. Confidence	Confidence in own ability to successfully complete the tasks with the aid of the adaptive automation
2. Self-confidence	Confidence in own ability to successfully complete the tasks
3. Accuracy	Accuracy of own performance on the tasks with the aid of the adaptive automation
4. Self-accuracy	Accuracy of own performance on tasks
5. Automation confidence	Confidence in ability of the machine to support successful completion of the tasks
6. Automation accuracy	Accuracy of machine in supporting successful completion of tasks
7. Automation dependability	The extent to which you can count on the machine to provide the appropriate support to the tasks
8. Automation reliability	The extent to which you can rely on the machine to consistently support the tasks
9. Predictability	The extent to which you can anticipate and expect the machine to support the tasks
10. Risk	The probability of negative consequences of relying on the machine to support successful completion of the tasks
11. Impact / Survivability	The severity and criticality of adverse or negative consequences of relying on the machine to support successful completion of the tasks
12. Decision complexity	The extent to which the machines' decision on when and how to intervene and support the task can be regarded as a simple and obvious choice
13. Uncertainty / doubt	The extent to which you have confidence in the machines' decision on when and how to intervene and support the task
14. Judgement / awareness	The extent to which the machines' decision on when and how to intervene and support the task requires assessment, knowledge, and understanding of the task
15. Faith	The extent to which you believe that the machine will be able to intervene and support the tasks in other system states in the future
16. Demand for trust	Level of trust required from you when the machine intervenes and supports the task
17. Supply of trust	Level of trust actually provided by you when the machine intervenes and supports task

Page intentionally left blank

## REFERENCES

Abdul-Rahman, A. & Hailes, S. (1999). Relying on trust to find reliable information. *1999 International Symposium on Database, Web and Cooperative Systems (DWACOS'99)*, Baden-Baden, Germany.

Abdul-Rahman, A. & Hailes, S. (2000). Supporting trust in virtual communities. *Hawaii International Conference on System Sciences 33*, Maui, Hawaii, 4-7 January 2000.

Chabrol, C., Vigier, J.C., Garron, J. & Pavet, D. (1999). CENA PD/3 Final Report, PHARE/CENA/PD/3-2.4/FR/2.0.

Checkland, P. (1981). *Systems Thinking, Systems Practice*. Chichester: Wiley.

Cook, M. (1998). *Personnel selection: Adding value through people, 3<sup>rd</sup> edition*. Chichester: Wiley.

DERA (1997). WP6: Application of evaluation techniques. Annex B. Results of DERA cognitive walkthrough activity. EC DGVII RHEA Project, Ref. RHEA/TH/WPR/6/2.0, 30<sup>th</sup> July.

EATMP (2000). Human Resources Programme - Stage 1: Programme Management Plan. Edition 1.0. Brussels: EUROCONTROL.

EATMP Human Resources Team (2003a). *Guidelines for Trust in Future ATM Systems: A Literature Review*. HRS/HSP-005-GUI-01. Edition 1.0. Released Issue. Brussels: EUROCONTROL.

EATMP Human Resources Team (2003b). *Guidelines for Trust in Future ATM Systems: Principles*. HRS/HSP-005-GUI-03. Edition 1.0. Released Issue. Brussels: EUROCONTROL.

EATMP Human Resources Team (2003c). *The Development of Situation Awareness Measures in ATM Systems*. HRS/HSP-005-REP-01. Edition 1.0. Released Issue. Brussels: EUROCONTROL.

EATMP Human Resources Team (2003d). *Age, Experience and Automation in European Air Traffic Control*. HRS/HSP-005-REP-02. Edition 1.0. Released Issue. Brussels: EUROCONTROL.

EUROCONTROL (2000a). Air traffic controller attitudes toward future automation concepts: A literature review. EUROCONTROL Report ASA.01.CORA.2.DEL02-A.RS, 4<sup>th</sup> December.

EUROCONTROL (2000b). Conflict Resolution Assistant level 2 (CORA2). Controller assessments. EUROCONTROL Report ASA.01.CORA.2.DEL02-b.RS, 4<sup>th</sup> December.

Fairburn, C. & Wright, P. (2000). Exploring the Metaphor of "Automation as a Team Player": taking team playing seriously; Paper presented at *10<sup>th</sup> European Conference on Cognitive Ergonomics (ECCE - 10)*, Linkoping, Sweden, 21<sup>st</sup> – 23<sup>rd</sup> August.

Goillau, P. & Kelly, C. (1997). MAIvern Capacity Estimate (MACE) - a proposed cognitive measure for complex systems. In: Harris, D. (Ed) *Engineering Psychology and Cognitive Ergonomics Volume 1: Transportation Systems*, Ashgate Publishers, pp 219-225.

Goillau, P., Woodward, V., Kelly, C. & Banks, G. (1998). Evaluation of virtual prototypes for ATC – the MACAW technique. In: Hanson, M. (Ed) (1998). *Contemporary Ergonomics '98*, London: Taylor & Francis, p. 419-423.

Graham, R., Young, D., Pichancourt, I., Marsden, A. & Irkiz, I. (1994). ODID IV simulation report. EEC Report No. 269/94. Brétigny-sur-Orge, France: EUROCONTROL.

Hopkin, V. David (1975). The controller versus automation. In: AGARD AG-209.

Hopkin, V.D. (1998). The impact of automation on air traffic control specialists. In: M.W. Smolensky & E.S. Stein, *Human Factors in Air Traffic Control*, Academic Press, 391-419.

Hutchins, E. & Klausen, T. (1991). Distributed cognition in the cockpit. In: Y. Engestrom & D. Middleton, *Cognition and communication at work*. Cambridge University Press.

Jian, J.-J., Bisantz, A.M. & Drudy, C.G. (1998). Towards an empirically determined scale of trust in computerized systems: Distinguishing concepts and types of trust. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, Chicago, 501-505.

Jian, J.-J., Bisantz, A.M. & Drudy, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *Int. J. of Cognitive Ergonomics*, 4(1), 53-71.

Kelly, C.J., Goillau, P.J., Finch, W. & Varellas, M. (1995). CAER Future System 1 (FS1) Final trial report. Defence Research and Evaluation Agency, Report No. DRA/LS(LSC4)/CTR/RPT/CD246/1.0, November.

Kelly, C.J. & Goillau, P.J. (1996). Cognitive Aspects of ATC: Experience of the CAER & PHARE Simulations; Paper presented at *8<sup>th</sup> European Conference on Cognitive Ergonomics (ECCE - 8)*, Granada, 10<sup>th</sup> - 13<sup>th</sup> September.

Lee, J. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35, 10, 1243-1270.

Lee, J.D. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *Int. J. Human-Computer Studies*, **40**, 153-184.

Madsen, M. & Gregor, S. (2000). Measuring human-computer trust. In: *Proceedings of Eleventh Australasian Conference on Information Systems*, Brisbane, 6-8 December.

Masalonis, A.J., Duley, J., Galster, S., Castano, D., Metzger, U. & Parasuraman, R. (1998). Air traffic controller trust in a conflict probe during Free Flight. *Proc. of the 42<sup>nd</sup> Annual meeting of the Human Factors and Ergonomics Society*, 1607.

Masalonis, A.J. & Parasuraman, R. (1999). Trust as a construct for evaluation of automated aids: Past and future theory and research. *Proc. of the Human Factors and Ergonomics Society 43<sup>rd</sup> Annual Meeting*, 184-188.

Moray, N. (1999). *Monitoring, complacency, scepticism and eutectic behaviour*. [Proc. of CybErg 1999: The 2nd Int. Cyberspace Conf. on Ergonomics](#). Int. Ergonomics Assoc. Press.

Moray, N. (2001). Personal communication.

Moray, N., Inagaki, T. & Itoh, M. (2000). Adaptive automation, trust and self-confidence in fault management of time-critical tasks; *J. of Experimental Psychol: Applied*, **6**, 1, 44-58.

Moray, N. & Inagaki, T. (2001). Attention and complacency. *In press – Theoretical Issues in Ergonomics*.

Muir, B. (1994). Trust in automation: Part 1. Theoretical issues in the study and human intervention in automated systems. *Ergonomics*, **37**, 1905-1923.

Muir, B. & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, **39**, 3, 429-460.

Nijhuis, H., Buck, S., Kelly, C., Goillau, P., Fassert, C., Maltier, L. & Cowell, P. (1999). [WP8: Summary and consolidation of RHEA results](#). European Commission DGVI, Report RHEA/NL/WPR/8/04, 28<sup>th</sup> Feb.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, **39**, 2, 230-253.

Parasuraman, R., Molloy, R. & Singh, I.L. (1993). Performance consequences of automation-induced "complacency". *Int. J. of Aviation Psychology*, **3**, 1-23.

Reichmuth J., Schick, F., Adam, V., Hobein, A., Link, A., Teegeen, U. & Tenoort, S. (1998). PD/2 Final Report. EUROCONTROL PHARE Report PHARE/DLR/PD/2-10.2/SSR;1.2. February.

Rempel, J.K., Holmes, J.G. & Zanna, M.P. (1985). trust in close relationships. *J. of Personality and Social Psychology*, 49, 1, 95-112.

Sheridan, T.B. (1988). Trustworthiness of command and control systems. *Proc. of Analysis, Design and Evaluation of man-Machine Systems 1988*, 3<sup>rd</sup> IFAC/IFIP/IEA/IFORS Conf., Finland, 14-16 June.

Taylor, R.M. (1988). Trust and awareness in human-electronic crew teamwork. In: *The Human-Electronic Crew: Can They Work Together?* Wright-Patterson AFB, OH, Report WRDC-TR-89-7008.

Taylor, R.M., Shadrake, R. & Haugh, J. (1995). Trust and adaptation failure: An experimental study of unco-operation awareness. R. Taylor & J. Reising (Eds), *The Human-Electronic Crew: Can we Trust the Team? Proc. of the 3<sup>rd</sup> Int. Workshop on Human-Computer Teamwork*. Defence Evaluation and Research Agency, Report No. CHS/HS3/TR95001/02, 93-98.

Whitaker, R. & Marsh, D. (1997). PD/1 Final Report, PHARE Report DOC 96-70-24, PHARE/NATS/PD1-10.2/SSR, 1.1.

## ABBREVIATIONS AND ACRONYMS

For the purposes of this document the following abbreviations and acronyms shall apply:

ATC	Air Traffic Control
ATCO	Air Traffic Control Officer / Air Traffic Controller (UK/US)
ATM	Air Traffic Management
CORA(1)	Conflict Resolution Assistant (1)
DERA	Defence Evaluation and Research Agency (UK; now known as QinetiQ)
DIS	Director(ate) Infrastructure, ATC Systems and Support (EUROCONTROL Headquarters, SDE)
DIS/HUM	See 'HUM (Unit)'
EATCHIP	European Air Traffic Control Harmonisation and Integration Programme (now EATMP)
EATMP	European Air Traffic Management Programme (formerly EATCHIP)
EEC	EUROCONTROL Experimental Centre (Brétigny, France)
EVP	EATMP Validation Platform
FRAP5	5 <sup>th</sup> Free Route Airspace Project
GUI	Guidelines (EATCHIP/EATMP)
HCI	Human-Computer Interaction
HCT	Human-Computer Trust
HFSG	Human Factors Sub-Group (EATMP, HUM, HRT)
HRS	Human Resources Programme (EATMP, HUM)
HRT	Human Resources Team (EATCHIP/EATMP, HUM)
HSP	Human Factors Sub-Programme (EATMP, HUM, HRS)
HUM	Human Resources (Domain) (EATCHIP/EATMP)

HUM Unit	Human Factors and Manpower Unit ( <i>EUROCONTROL Headquarters, SDE, DIS; formerly stood for 'ATM Human Resources Unit'; also known as 'DIS/HUM'</i> )
ISA	Instantaneous Self-Assessment
LATCC	London Air Traffic Control Centre
MACE	Malvern Capacity Estimate
MEFISTO	Modelling, Evaluating and Formalising Interactive Systems using Tasks and interaction Objects
MTCD	Medium-Term Conflict Detection
MONA	Monitoring Aid
NASA	National Aeronautics and Space Administration (US)
ODID	Operational Display and Input Development
OLDI	On-Line Data Interchange
PHARE	Programme for Harmonised Air Traffic Management Research in EUROCONTROL
REP	Report ( <i>EATCHIP/EATMP</i> )
RHEA	Role of the Human in the Evolution of ATM systems
SA	Situation Awareness
SATI	SHAPE Automation Trust Index ( <i>EATMP, HUM, HRS, HSP, SHAPE</i> )
SDE	Senior Director, Principal EATMP Directorate or, <i>in short</i> , Senior Director(ate) EATMP ( <i>EUROCONTROL Headquarters</i> )
SHAPE (Project)	Solutions for Human-Automation Partnerships in European ATM (Project) ( <i>EATMP, HUM, HRS, HSP</i> )
SYSCO	System Supported Coordination
STCA	Short-Term Conflict Alert
TLX	Task Load Index ( <i>NASA, US</i> )
VAW	Vertical Assistance Window

## ACKNOWLEDGEMENTS

The contribution of the Members of the HRT Human Factors Sub-Group to this document through discussions during the group's meetings, and further written comments, were much appreciated.

The contribution of the EUROCONTROL Human Factors and Manpower (DIS/HUM) Unit is gratefully acknowledged, particularly that of Barry Kirwan<sup>2</sup> and Michiel Woldring, for their guidance of this project and their helpful comments.

Many experienced air traffic controllers provided valuable input to the material contained in this deliverable. The authors particularly thank Peter Eriksen and the Members of the Real-time Experimental Simulation Team at the EUROCONTROL Experimental Centre, Brétigny, France, as well as Deirdre Bonini, for their help in facilitating the evaluation of SATI. The document has benefited from the collective experience of two ex-operational controllers, Jon Nias-Cooper and Mike Sargeant.

Finally, the authors express their thanks to Neville Moray, and to Anne Isaac and Dominique Van Damme from DIS/HUM, for their helpful review comments of the document.

### Document Configuration

Carine Hellinckx  
(*External contractor*)

EUROCONTROL Headquarters, DIS/HUM

---

<sup>2</sup> Now works at the EEC

Page intentionally left blank

## **APPENDICES**

APPENDIX A: SATI QUESTIONNAIRE V0.2A

APPENDIX B: SATI SUPPLEMENT V0.2A

APPENDIX C: SATI QUESTIONNAIRE V0.3 – PROPOSED FINAL  
VERSION

Page intentionally left blank

## **Appendix A - SATI Questionnaire v0.2a**

## **Introduction to SHAPE Automation Trust Index (SATI v0.2a)**

Computer-assistance tools and other forms of automation support are being increasingly introduced into today's Air Traffic Management (ATM) systems, and are expected to be fundamental components of systems in the future. The success of such automated tool support will depend in part on the degree to which Human Factors are taken into account in the design and implementation of these tools.

As part of the overall European ATM Programme (EATMP), the Human Factors & Manpower Unit within EUROCONTROL has recently initiated a new programme of work to address the human factors issues of automation in ATM systems. The programme is called SHAPE ('*Solutions for Human-Automation Partnerships in European ATM*'). The present aim of SHAPE is to develop a number of measurement techniques that can be applied during real-time simulations to assess and measure the effectiveness of the automation.

This questionnaire is concerned with one specific measure of human performance called SATI (SHAPE Automation Trust Index) which has been especially developed for measuring the degree of trust that a person (i.e. controller) has in the automated system being operated. The easiest means of measuring trust is to ask a person to say how he or she feels, and to rate or score their degree of trust in the thing in question. This subjective measurement approach is what is used in SATI. More specifically, SATI consists of a set of rating scales to measure your views about how much you trust the automation in the ATM system that you are operating.

There are two parts to SATI:

- Part 1. Each day, *before* starting the simulation runs, you rate your overall level of trust.
- Part 2. Each day, *after* finishing the simulation runs, you rate your strength of feeling about several factors that may contribute to trust, and again you rate your overall level of trust.

Thank you for your assistance and cooperation.

The SHAPE Team

## **SHAPE Automation Trust Index (SATI v0.2a)**

### **SATI Part 1 (please complete before the start of the day's simulation runs)**

Please tell us who you are, and your role in the simulation. Thank you.

#### About you:

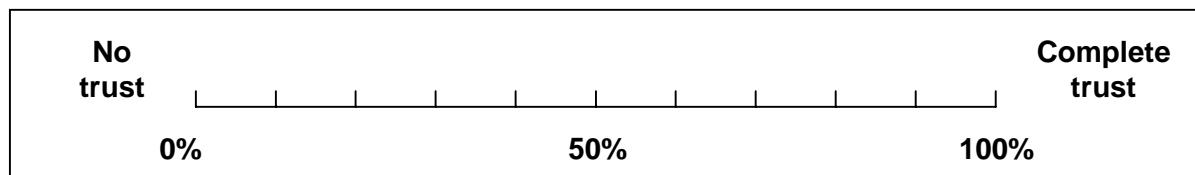
Name:	
Nationality:	
Sex (M/F):	

#### About the simulation:

Date:	
Simulation project:	
Your sector:	
Your role (Planner / Executive Controller)	

### SATI Part 1 (*continued*)

1. Based on your of experience of ATC simulations, either in general or specifically for this system, please indicate your overall amount of trust in the total system. (Please mark the scale with an 'X').



2. In your opinion, what changes would need to be made to the system so that your level of trust would be increased? Would you work live traffic with these automation tools? If not, please explain your reasons.

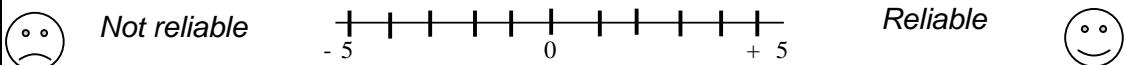
**SATI Part 2 (please complete after the end of the simulation runs)**

Name:	
Date:	

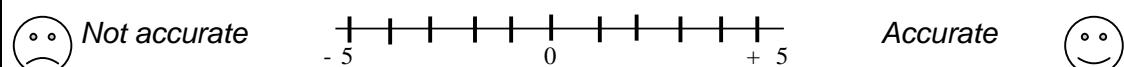
Your last sector:	
Your last role (Planner / Executive controller)	

3. Please indicate the strength of your feelings about the automation tool (\_\_\_\_\_ ) for each of these factors by marking each scale with an 'X'.

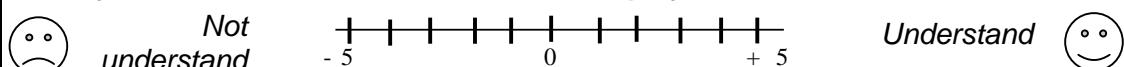
**1. How reliable (in terms of the % of time it is operational) is the automation tool?**



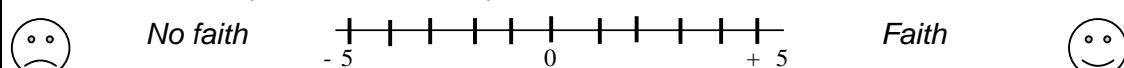
**2. How accurate (in terms of the correctness of displayed data) is the functioning of the automation?**



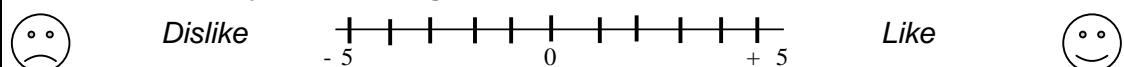
**3. Do you understand the behaviour and displayed intent of the automation?**



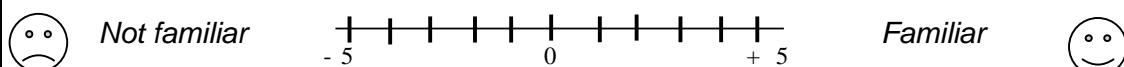
**4. How much do you believe the system in unknown situations?**



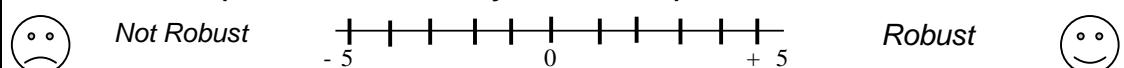
**5. How much do you like using the automation tool?**



**6. How easy, natural and friendly is the automation to use?**



**7. How robust (in terms of recovery from errors) is the automation?**

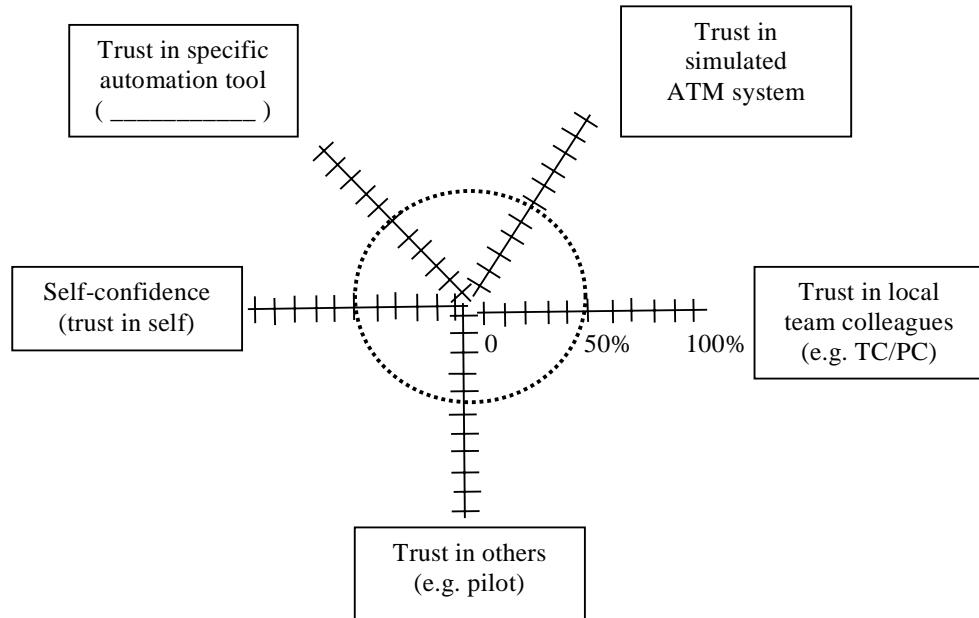


## SATI Part 2 (*continued*)

4. Please rank these factors in order of relative importance, numbering from 1 (*least important*) to 7 (*most important*). Please use each number only once.

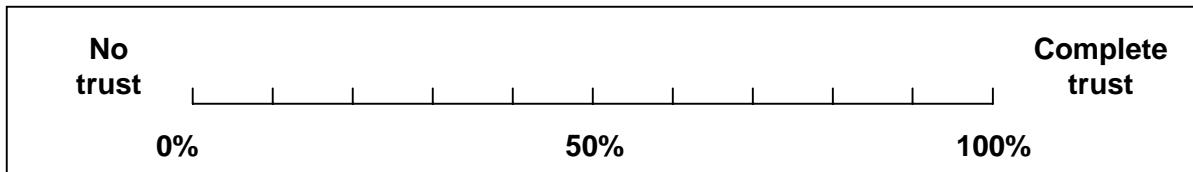
<b>Reliability</b>	<i>ranking:</i>
<b>Accuracy</b>	<i>ranking:</i>
<b>Understanding</b>	<i>ranking:</i>
<b>Faith</b>	<i>ranking:</i>
<b>Liking</b>	<i>ranking:</i>
<b>Familiarity</b>	<i>ranking:</i>
<b>Robustness</b>	<i>ranking:</i>

5. Please indicate your amount of trust for each of the five dimensions of the total system (people and technology) by marking each scale with an 'X'.



## SATI Part 2 (continued)

6. Please indicate your overall amount of trust in the (simulated) total operational system. If your level of trust in the system has changed since the start of the day's simulations, please explain why in the space below.



7. If there are any other factors which influence your trust in an ATC system, or if you have any other general comments about trust, please write them in the space below.

Thank you for completing this questionnaire.

Page intentionally left blank

## **Appendix B - SATI supplement v0.2a**

## **SHAPE Automation Trust Index (SATI v0.2a) - Supplement**

### **SATI Part 3**

Please tell us about yourself.

*About you:*

Name:	
Nationality:	
Sex (M/F):	
Date:	

*About your home ATC centre:*

Your current role (Planner / Executive controller / other)	
---	--

\*\*\*\*\*

As scientists (who are not trained as controllers) we are very interested in your reasons for trusting / having confidence in ATC automation.

Please would you help us by completing the following questions as fully as possible. All replies will be treated in the strictest confidence. Thank you.

1. As an Air Traffic Controller, what do you understand by the word “Trust”? What does it mean to you in ATC operational terms?

**SATI Part 3 (continued): AUTOMATION SYSTEMS**

2. Think of your home ATC centre. Please give examples of ATC automation systems that are present (e.g. telephone system, radar, Code Callsign Conversion system, OLDI, STCA etc). Which of these systems do you actually use? Which systems do you trust?

	Example of ATC automation system	Use? (Yes/No)	Trust? (Yes/No)	Please give your reasons
1.				
2.				
3.				
4.				
5.				
6.				
7.				

3. Concerning ATC automation, which of the following statements do you agree/disagree with?

Either I trust or I do not trust ATC automation **AGREE / DISAGREE**

I have various degrees of confidence in ATC automation **AGREE / DISAGREE**

There is a minimum level of trust for me to use ATC automation **AGREE / DISAGREE**

(If agree, please specify this level of trust \_\_\_\_\_)

There is an optimum range of values of trust for me to use ATC automation. **AGREE / DISAGREE**

Below a certain value I trust too little **AGREE / DISAGREE**

Above a certain value, I trust too much **AGREE / DISAGREE**

(If agree, please specify this range of values \_\_\_\_\_)

There is an acceptable reliability or failure rate for ATC automation **AGREE / DISAGREE**

If agree, please specify by circling one choice from both lists below:

Acceptable failure rate: NEVER / ONCE PER YEAR / ONCE PER MONTH / ONCE PER WEEK / ONCE PER DAY / ONCE PER SESSION / OTHER \_\_\_\_\_

Acceptable failure frequency: NEVER / 1 IN 100,000 INTERACTIONS / 1 IN 10,000 / 1 IN 1,000 / 1 IN 100 / 1 IN 10 / OTHER \_\_\_\_\_

## SATI Part 3 (*continued*) AUTOMATION SYSTEMS

4. What are the positive characteristics of an ATC automation system that will increase your confidence in it? (e.g. reliable, etc.)

5. What are the negative characteristics of an ATC automation system that will reduce your confidence in it? (e.g. not reliable, etc.)

## SATI Part 3 (continued): TEAMWORKING

6. Do you have equal confidence in all the pilots who fly through your airspace? **YES / NO**

What factors determine your degree of confidence in the pilots? (e.g. airline, pilot unfamiliar with airspace, etc.)

7. Do you have equal confidence in all controllers you work with, including adjacent sectors and centres? **YES / NO**

What factors determine your degree of confidence in controllers you work with? (e.g. their experience, whether recently validated on sector, etc.)

8. Are there any other factors which influence your confidence in ATC automation?

Thank you for completing this questionnaire.

Page intentionally left blank

## **Appendix C - SATI Questionnaire v0.3 - proposed final version**

## **SHAPE Automation Trust Index (SATI v0.3)**

### **SATI Part 1 (please complete before the start of the day's simulation runs)**

Please tell us who you are, and your forthcoming role in the simulation. Thank you.

#### About you:

Name:	
Nationality:	
Sex (M/F):	

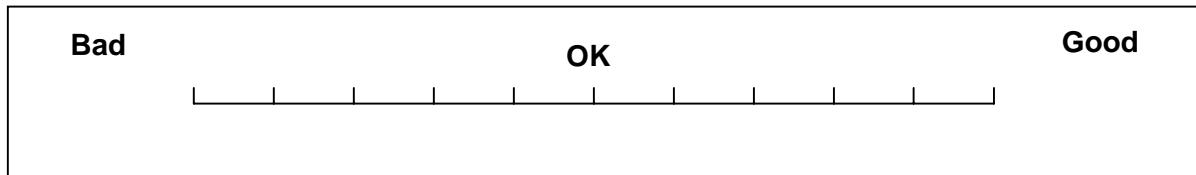
#### About the simulation:

Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1.  2.  3.  4.  5.
Your simulated sector:	
Your role (planner / executive controller)	

## SATI Part 1 (continued)

PLEASE COMPLETE AT THE START OF EACH DAY

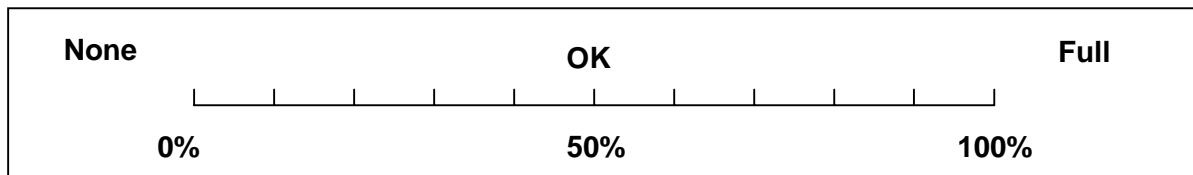
1. What do you think of the simulation so far? (Please mark the scale with an 'X').



2. Are you prepared to trust the simulated system? Please give your reasons.

No	Yes
----	-----

3. How much confidence do you have in the simulated system? (Please mark the scale with an 'X').



4. Please give your reasons

## SATI Part 2 (please complete after the end of the simulation runs)

Please write your name and your last role in the simulation. Thank you.

### About you:

Name:	
-------	--

### About the simulation:

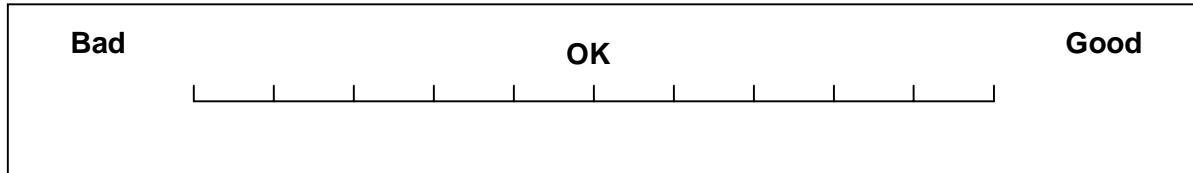
Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1.  2.  3.  4.  5.
Your last simulated sector:	
Your last role (planner / executive controller)	

## SATI Part 2 (continued)

PLEASE COMPLETE AT THE END OF THE DAY'S RUNS

Based on today's runs

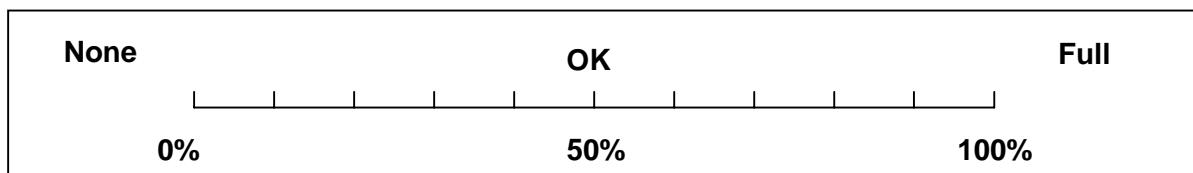
1. What did you think of the simulation? (Please mark the scale with an 'X').



2. Were you prepared to trust the simulated system?



3. How much confidence did you have in the simulated system? (Please mark the scale with an 'X').



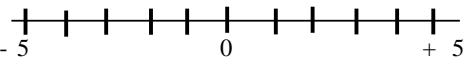
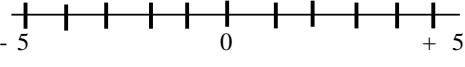
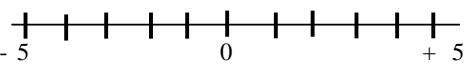
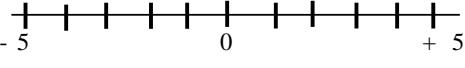
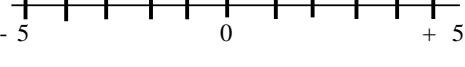
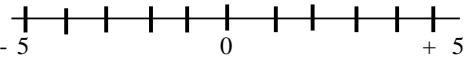
4. Please give your reasons. If your trust or level of confidence in the system has changed since the start of the day, please explain why.

A large empty rectangular box for writing responses to question 4.

## SATI Part 2 (*continued*)

PLEASE COMPLETE A SEPARATE SHEET FOR EACH AVAILABLE AUTOMATION TOOL.

5. Please judge each automation tool against the following factors (mark each scale with an 'X').

Name of automation tool:	
<b>1. Is the automation tool useful?</b>	 <i>Not useful</i>  <i>Useful</i> 
<b>2. How reliable is it?</b>	 <i>Not reliable</i>  <i>Reliable</i> 
<b>3. How accurately does it work?</b>	 <i>Not accurate</i>  <i>Accurate</i> 
<b>4. Can you understand how it works?</b>	 <i>Not understand</i>  <i>Understand</i> 
<b>5. Do you like using it?</b>	 <i>Dislike</i>  <i>Like</i> 
<b>6. How easy is it to use?</b>	 <i>Difficult</i>  <i>Easy</i> 

6. Please rank these factors in order of relative importance. Number them from 1 (*least important*) to 6 (*most important*). Please use each number once only.

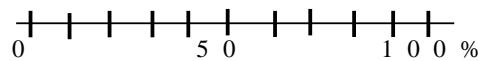
Name of automation tool: _____	
<b>Usefulness</b>	<i>ranking:</i>
<b>Reliability</b>	<i>ranking:</i>
<b>Accuracy</b>	<i>ranking:</i>
<b>Understanding</b>	<i>ranking:</i>
<b>Liking</b>	<i>ranking:</i>
<b>Ease of use</b>	<i>ranking:</i>

## SATI Part 2 (continued)

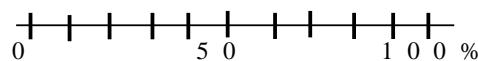
### LOOKING BACK OVER THE DAY'S SIMULATION RUNS:

7. Please rate your amount of confidence in each of these five dimensions.  
Please mark each scale with an 'X'.

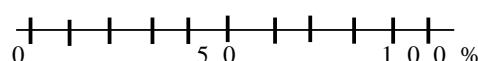
#### **1. Confidence in automation tools**



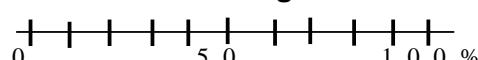
#### **2. Confidence in simulation**



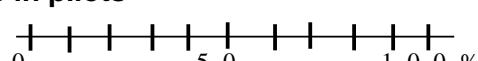
#### **3. Self-confidence**



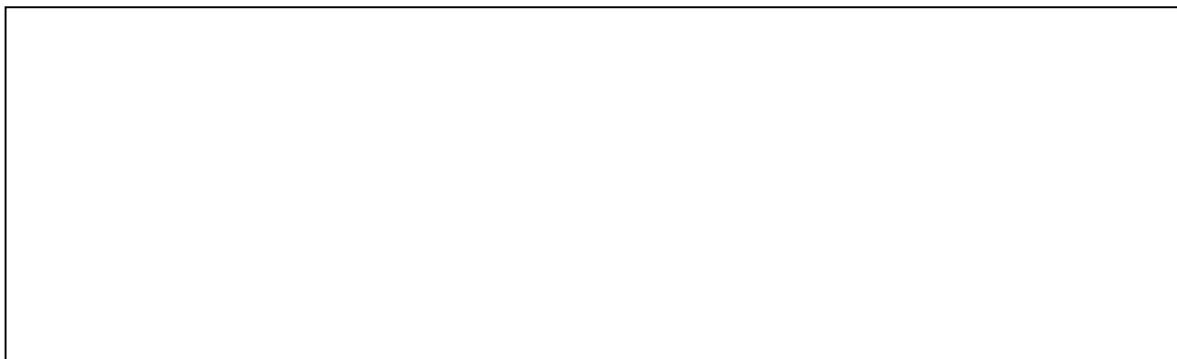
#### **4. Confidence in controller colleagues**



#### **5. Confidence in pilots**



8. Would you work live traffic with the tools? In your opinion, what changes would the automation need so that your trust and confidence would be increased? If there are any other factors which influence your trust in an ATC system, or if you have any general comments, please write them here.



Thank you for completing this questionnaire.

Page intentionally left blank