2011 Fifth International Conference on Secure Software Integration and Reliability Improvement

Trust Observations in Validation Exercises

F. Amato*, M. Felici†, P. Lanzi*, G. Lotti*, L. Save* and A. Tedeschi*
*Deep Blue S.r.l.
Piazza Buenos Aires, 20
00198 Roma - Italy
URL: http://www.dblue.it/
Email: {paola.lanzi, giulia.lotti, luca.save, alessandra.tedeschi}@dblue.it
†School of Informatics, The University of Edinburgh
10 Crichton Street, Informatics Forum
Edinburgh EH8 9AB, UK
Email: mfelici@staffmail.ed.ac.uk

***Abstract*—This paper is concerned with an operational account of trust. It reports our experience in observing different trust aspects during a validation session for the assessment of a new tool and relevant operational concepts in the Air Traffic Management (ATM) domain. Despite the fact that trust is yet an elusive concept, our results show how monitoring trust can support the validation of alternative system settings and their operational aspects. This paper reports our experimental work on observing trust during validations exercises. Moreover, it provides new insights about the nature and the investigation of trust.**

***Keywords*-Air Traffic Management; Operational Validation; Empirical Analysis; Trust**

I. Introduction

Ongoing developments in the Air Traffic Management (ATM) domain involve the implementations and deployments of new technologies and operational concepts, which change ATM practices. The ATM 2000+ Strategic Agenda [1] and the Single European Sky ATM Research (SESAR) Initiative [2] are concerned with a structural revision of ATM processes, a new ATM concept and a system approach for the ATM Network. This requires ATM services to go through significant structural, operational and cultural changes that will contribute towards SESAR. Validation activities are crucial for the development of new technologies and relevant operational concepts. They are often concerned with different *Key Performance Indicators* (KPIs), such as safety, efficiency, and so on, that are critical within the ATM domain. Validation activities involve subsequent evaluations of technologies with respect to different scenarios. In the ATM context, Air Traffic Controllers (ATCOs) exercise with technologies in order to validate system features [3]. This is useful to assess how new technologies and the operational concepts they implement affect work practices.

Trust is a concept that has been recognized to be critical for the acceptance and adoption of new technologies. Research highlights that trust is critical in the *automation* of various human activities (e.g. see [4] for an account of trust in the ATM domain). Moreover, trust may interact (that is, either support or affect) other critical aspects (e.g. safety, risk perception, user acceptability, etc.). Unfortunately, our understanding how trust relates to other critical dimensions of technologies is still patchy. Most research is still debating on a generally accepted account of trust (and other relevant concepts like trustworthiness) [5].

This paper is concerned with an operational account of trust. Our main interest is to investigate how trust can support validation activities and investigations. Despite any theoretical account of trust, we are seeking to acquire an understanding of trust during operational validation sessions. Our assumption is that trust provides a convenient and alternative viewpoint of analysis that can be combined together with other operational aspects (e.g. situational awareness, teamwork, workload, etc.). This paper is structured as follows. Section II highlights the criticality of trust within the ATM domain. Section III describes the ATM validation case study. Section IV introduces our observational approach to trust adopted during a validation session. Section V discusses our trust observations. In particular, it points out different trust aspects. The discussion is supported by qualitative and quantitative analyses carried out during the validation session. Section VI, finally, draws some concluding remarks.

II. Trust in Air Traffic Management

Trust and trustworthiness [6] capture many diverse aspects that are becoming as critical and relevant as many other aspects (e.g. safety, security, dependability, etc.) for system design and assessment [7], [8]. Research on trust drawn from multidisciplinary domains highlights a continuing debate on its definition and its nature [5], [6], [9], [10]. Yet, it is difficult to derive a definitive and widely-accepted definition of trust. One conclusive remark is that trust is a complex concept. Unveiling its complexity requires us to understand subtle interactions among different aspects of trust (e.g. trust and trustworthiness). Trust is a concept that has been investigated in different disciplines (e.g. economics, social

978-0-7695-4453-3/11 $25.00 © 2011 IEEE
DOI 10.1109/SSIRI.2011.26

science, computer science, etc.) and used in every-day life differently [10]. This situation creates confusion about trust itself. Simply, many people refer to trust, although they mean different things [11]. Research has addressed this problem by seeking for a general account of trust [6], [9], [10], [12], [13], which 'unifies' the different uses of the word trust. However, the generalization of the concept of trust faces the problem of making sense of different perspectives that might contradict each other. Moreover, it is unclear how these perspectives (and their underlying assumptions and models) have shaped our understating of trust — *What is our understanding of trust in situated contexts?*

Trust is steadily acquiring an important role in the deployment of ATM systems [4], [14], [15]. The interaction of trust with system features (e.g., system reliability) highlights contingencies in understanding the role of trust with respect to system dependability and risk perception. The contextualizing of trust in ATM [4], [14], [15] identifies four main relevant aspects: *Automation*, *Understanding Trust*, *Trust and Human-Machine Systems* and *Measuring Trust*. The level of automation takes into account to which extent human and machine cooperate in performing an activity. Automation is *"a device or system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator"* [4], [14], [15]. The notion of automation influences the understanding of trust in the ATM domain. Trust is *"the extent to which a user is willing to act on the basis of the recommendations, actions, and decisions of a computer-based tool or decision aid"* [4], [14], [15]. Note that the *competence of tool* contributes to the overall trust according to a simple model identified in [4], [14], [15]. Although the quoted definition of trust originates from general models of trust, *complacency* may distinguish the ATM domain from others. Complacency is a kind of automation mis-use, which takes into account those situations characterized by an operator's over-reliance on automation resulting in the failure to detect system faults or errors [4], [14], [15]. Although trust and reliability have an important role in ATM — *"Trust is an intrinsic part of air traffic control. Controllers must trust their equipment and trust pilots to implement the instructions they are given. The reliability of new systems is a key determinant of controller trust"* [4], [14], [15] — air traffic controllers accept unreliable tools as far as they understand the failure modes [4], [14]–[16]. Like other industry domains, ATM is seeking to understand trust, qualitatively as well as quantitatively, in order to support system deployments within complex operational settings. This work provides an operational account of trust drawn from our trust observations during validation exercises.

III. Validation Case Study

The case study consists of an empirical validation of a new ATM tool and relevant operational concepts. The validation is concerned with the tool usage and interoperability, hence, how the new operational concepts change current ATM work practice. The validation consists of two subsequent sessions. The first validation session pointed out a limited operational support (e.g. in terms of suggested ATM operations) resulting in *mistrust*. The resulting mistrust inhibits any deployment of the new operational concepts implemented by the tool under analysis in current ATM practice. Such problems arose because operational expectations (in terms of available functionalities and resulting behavior) exposed the limitations of the system settings and configurations, while Air Traffic Controllers (ATCOs) highlighting complex interactions between system functionalities, operational constraints and expert judgments. The results from the first validation session suggested extending the focus of the validation by taking into account *trust* as an additional dimension to be investigated. Figure 1 shows the simulated airspace and sectors.

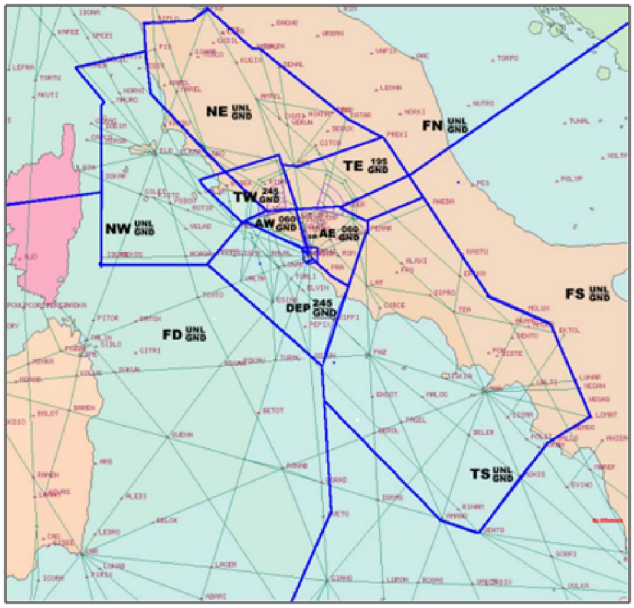


Figure 1. Simulated airspace and sectors

This paper reports an analysis of trust drawn from the second validation session. Both validation sessions address the empirical assessment of the system's operational impact on ATM practices. The validation activities are concerned with various *Key Performance Areas* (KPAs), among which, *Operability*, *Flight Efficiency*, *Capacity* and *Safety*. Our empirical analysis builds on the validation of relevant operational aspects: *Teamwork*, *Situational Awareness* and *Workload*. Moreover, it takes into account trust as a relevant analysis viewpoint alongside the other operational aspects investigated. The validation sessions validated relevant operational concepts in simulated sectors controlled by an Area Control Centre (ACC). The validation sessions simulated different air traffic scenarios: *nominal scenarios* involving just different

levels of traffic, and *non-nominal ('exceptional') scenarios* involving unforeseen events (e.g. reduced system support due to radio failure).

The validation sessions simulated the different scenarios for three different system configurations, simply named for anonymity: A, B and C. Configuration A involves current ATM practice. That is, ATCOs worked throughout the scenarios according to current operational procedures. Whereas, configuration B and C involved the integration of the new proposed system and relevant operational concepts. The main difference between configuration B and C is the extent to which ATCOs have to comply and adapt to advisory information. Configuration B allows a greater level of flexibility than configuration C in order to accommodate traffic and to comply with advisory information. Validation sessions involved the required number of ATCOs in order to cover the simulated sectors. ATCOs worked on different sectors, scenarios and configurations in order to minimize individual learning.

IV. Observational Approach to Trust

This section describes the overall process we adopted in order to investigate trust. Our main concern was to make sense of trust in operational terms. In spite of the different trust accounts, our problem has been to take into account trust as an alternative viewpoint of analysis. The rationale is that trust would enable us to critically analyze other operational aspects too. The assumptions underlying our observational approach to trust are:

1) It is possible to monitor trust during validation simulations
2) It is possible to assess trust variations according to different system configurations
3) It is possible to narrow the analysis of trust by taking into account operational information (e.g. sectors, controllers, scenario complexities, etc., in the case of ATM simulations)
4) It is possible to notice deviations in the information flow in correspondence with trust variations at the operational level.

In order to assess the stated assumptions, our observational approach to trust consists of a process investigating trust at different levels of analysis. This process consisted of two main phases: *Macro-Analysis of Trust* and *Micro-Analysis of Trust*. Figure 2 shows the observational phases of our trust analysis.

At the macroscopic level of analysis, we looked at different trends capturing trust information for different system configurations. At the microscopic level of analysis, we looked at different operational aspects (e.g. sector interaction) in order to refine our trust investigation.

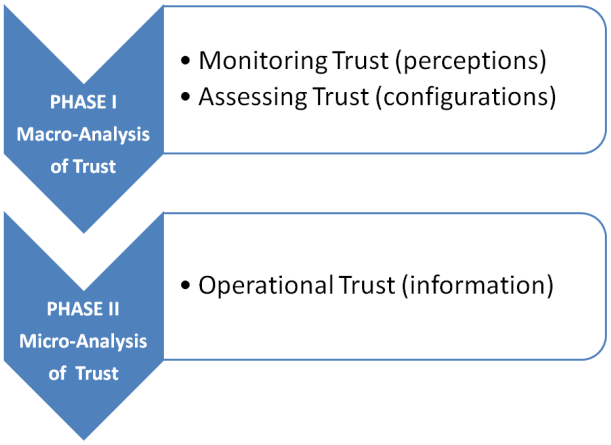


Figure 2. A process for investigating Trust

A. Observational Methods

The empirical investigations involved quantitative and qualitative analyses. The operational scenarios were characterized in terms of traffic (e.g. number of flights and critical events) and other relevant information (e.g. communication between controllers/sectors) that were recorded during each session. After each validation exercise we collected a Post Exercise Questionnaire (PEQ). The PEQ consisted of different parts concerning with each of the operational aspects, i.e. *Teamwork*, *Situational Awareness* and *Workload*, respectively. We also included an additional part concerning with trust. As trust questionnaire we adopted and revised the EUROCONTROL *SHAPE Automation Trust Index* (SATI) questionnaire [17]. The SATI questionnaire is concerned with six different aspects (i.e. *Utility*, *Reliability*, *Accuracy*, *Understanding*, *Robustness* and *Confidence*) that affect trust in system (with respect to the level of support, functionality, etc.). Figure 3 shows the six points (as presented by EUROCONTROL [17]) forming the SATI questionnaire. We tailored the six points mainly to emphasize their relevance with respect to the validation exercises.

	never						always
1) … the system was useful.	0	1	2	3	4	5	6
	never						always
2) … the system was reliable.	0	1	2	3	4	5	6
	never						always
3) … the system worked accurately.	0	1	2	3	4	5	6
	never						always
4) … the system was understandable.	0	1	2	3	4	5	6
	never						always
5) … the system worked robustly (in difficult situations, with invalid inputs, etc.).	0	1	2	3	4	5	6
	never						always
6) … I was confident when working with the system.	0	1	2	3	4	5	6

Figure 3. SATI Questionnaire

V. Trust Observations

This section discusses our trust observations. Trust observations follow our empirical approach to trust. We start from a macroscopic viewpoint of analysis by looking at the trust datasets. We then refine our analysis by taking into account other aspects that allow us to look at detailed trust observations.

A. Trust Datasets

The first aspect is to look at the datasets we collected by the trust questionnaires. Figure 4 shows the trust scores (values between 0 and 6) for the three different configurations. The boxplots provide a characterization of trust (values) for each configuration, respectively. The boxplots characterize our trust datasets as follows. The tops and bottoms of each boxplot are the 75^{th} and 25^{th} percentiles of the samples, respectively. The line in the middle of each boxplot is the sample median. Notches display the variability of the median between samples. The width of a notch is computed so that box plots whose notches do not overlap (as in the figure) have different medians with a 5% significance level. The significance level is based on a normal distribution assumption, but comparisons of medians are reasonably robust for other distributions. Comparing boxplot medians is like a visual hypothesis test. That is, since the notches in the boxplots do not overlap, *we can conclude, with* 95% *confidence, that the true medians do differ. We can then conclude that controllers trust the three configurations differently. Hence, it is possible to compare different systems or configurations by their perceived trustworthiness.*

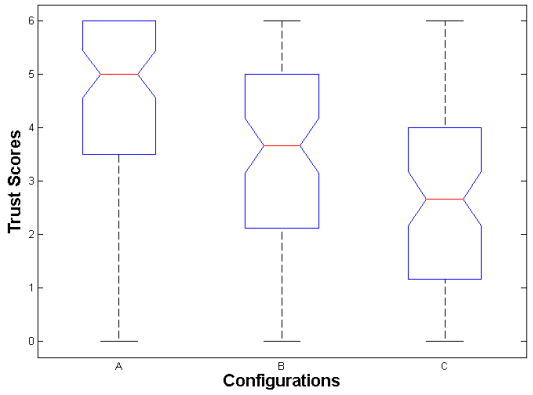
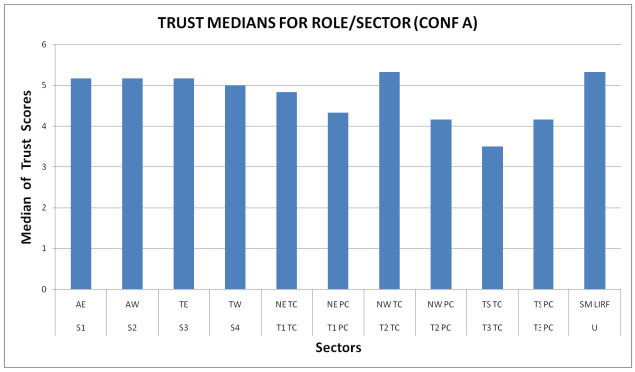


Figure 4. Boxplots of the Trust Scores

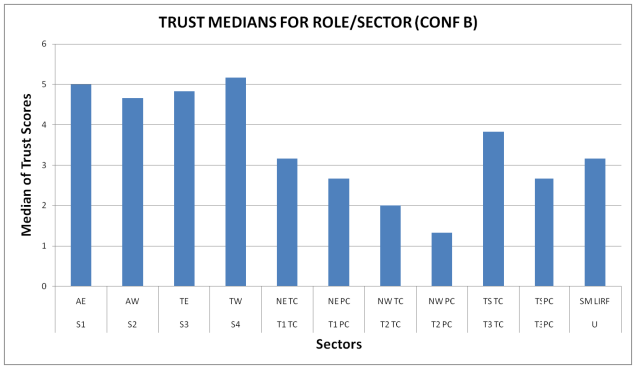
Configuration A is the one that shows the best profile of trustworthiness according to the trust scores. This is somehow not surprising considering the fact that the controllers were working according to current ATM practice and procedures. It is interesting to see how the perceived trustworthiness is lower for the other two configurations, in particular, for configuration C (the one that gives less flexibility to the controllers in order to accommodate advisory information).

B. Trust Medians

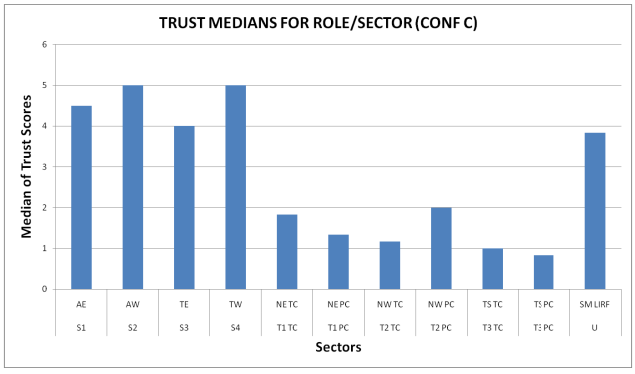
The next step is to look at whether or not different sectors or controllers (having different responsibilities) had different experiences for the three configurations. Figure 5 shows the medians of trust scores registered for each sector/role.



(a) Configuration A



(b) Configuration B



(c) Configuration C

Figure 5. Medians of Trust Scores per Role/Sector

It is interesting to notice that configurations B and C affected differently sectors and roles, in particular, for the

configuration C. This points out that the new configurations, i.e. B and C, somehow shift sector/controller responsibilities. Therefore, trust observations for the different configurations and the sector/controller viewpoints allow us to understand the perceived trustworthiness and to identify those sectors and controllers that were mostly affected by the new operational conditions. This is an interesting point because it stresses the relationship between trustworthiness perception and organizational aspects (e.g. organizational structures, roles, responsibilities). That is, *system trustworthiness is perceived differently depending on the role, responsibility, position within an organization.*

We then looked at whether it was possible to identify any relationship between trust and the types of scenarios (i.e. nominal or exceptional scenarios). According to our validation exercises the type of scenario is irrelevant for the perceived trustworthiness. Figure 6 shows the medians of trust scores grouped by types of scenarios and organizations. Configurations B and C show comparable decreases in the trust scores both for nominal as well as exceptional scenarios. This result of course should not be generalized. The interviews after the scenarios allow us to interpret such results. Interviews point out that controllers once started to mistrust the specific configuration, the type of scenario they were dealing with was irrelevant. They adapted their behavior to the system rather than to the scenario. That is, their work practice depends on exhibited (or perceived) system features, e.g. reliability, rather than on particular types of scenarios. They trust or mistrust the system (or the specific configuration) independently from the type of situation — they rely or not on the system functionalities.

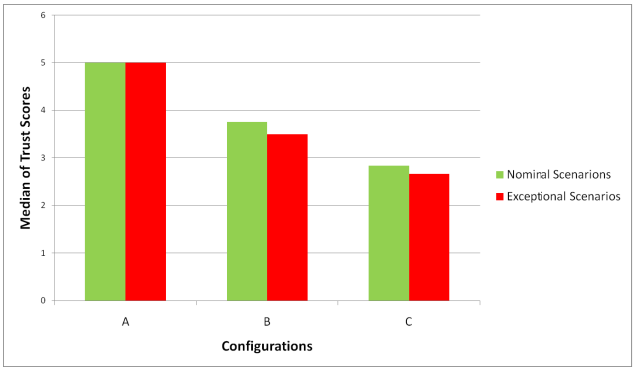
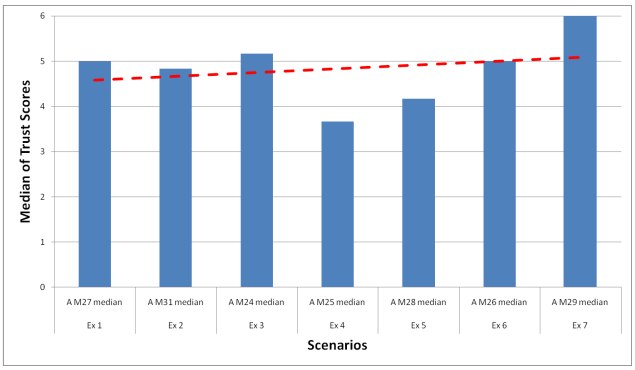
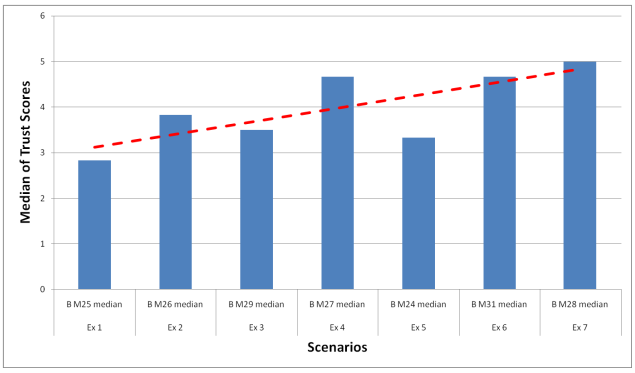


Figure 6. Medians of Trust Scores

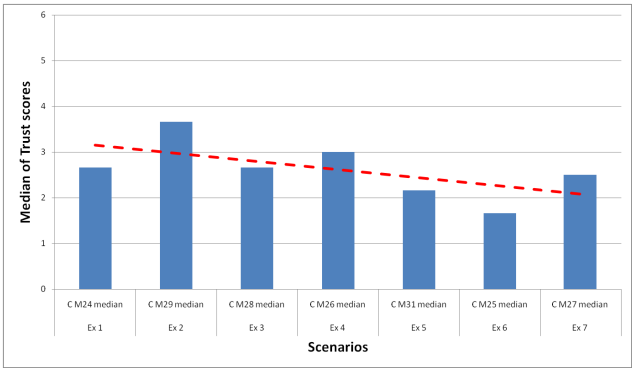
Next, we questioned whether or not there was any *learning* aspect affecting/supporting trust perception — that is, whether or not the perceived trustworthiness increased over subsequent evaluation scenarios. Intuitively, it is desirable that the more experience controllers gain with the system and relevant operational practice, the better. This required to look at how trustworthiness perception changed over the execution of the different scenarios with the same configuration. That is, it is necessary to look at trust decisions over time. Figure 7 shows how the medians of trust scores changed over the different scenarios (ordered by their execution in the validation session).



(a) Configuration A



(b) Configuration B



(c) Configuration C

Figure 7. Trends of Trust Scores

The configuration A (Fig. 7.a) has a trustworthy profile over the exercise scenarios. The configuration B (Fig. 7.b) shows an increasing perceived trustworthiness over the execution of the scenarios. That is, the more controllers work with the configuration B, the more trustworthy they consider the implied operation conditions. However, according to some interviews with controllers, the impression is that

they have managed to adapt their practice and procedures in order to accommodate the new operational conditions implied by the configuration. They learn how to deal with new operational aspects. Whereas, the configuration C (Fig. 7.c) is the one that they disliked mostly. In fact, the medians of trust scores over the executed scenarios show a decreasing trend. The more they work with the system configuration, the more they mistrust it. It is a configuration that they find to be unworkable.

These results show that it is possible to monitor trust decisions (or perceived trustworthiness) over time. This allowed us to confirm our overall perception and the controller's feedback during the validation exercises. Moreover, it is interesting to highlight how the different configurations have different profiles over time. The more controllers experienced each configuration, the clearer their opinions about the different operational conditions. At the end of the validation session, they clearly mistrusted the configuration C. Next, we investigated how trust affects work practice.

C. Operational Aspects of Trust

Finally, we looked at how trust might have affected interactions among sectors and controllers. We looked at quantitative data collected automatically during the validation scenarios. In particular, we analyzed communications, i.e. interactions among sectors and controllers, as relevant aspects of work practice. We questioned whether or not there was any particular difference between the different configurations and their operational aspects. We analyzed for the nominal scenarios different types communications: *Radio Communications* (between sectors and flights) and *Inter-sector Communications*. There are not substantial differences in the way sectors communicate with flights. Regardless the trust in the system, controllers maintained similar communication patterns with flights. Figure 8 shows the (median) number of communications for each sector (there are similar results for their average durations too).

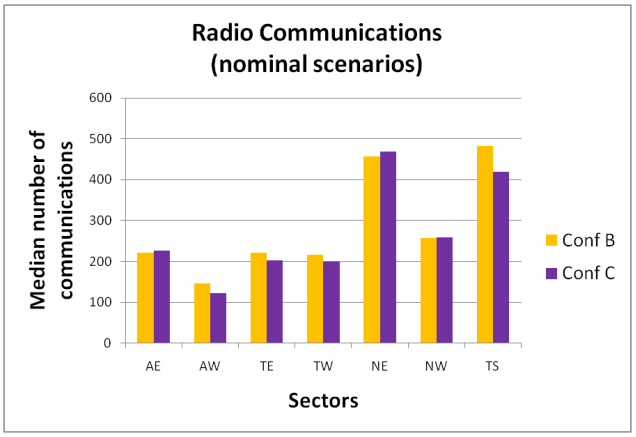
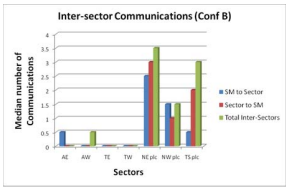
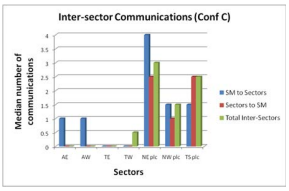


Figure 8. Radio Communications

The inter-sector communications highlight a different situation. Figure 9 shows the median number of Inter-sector Communications for both configurations (i.e. B and C). Although there are no substantial differences, the interactions among the most affected sectors change. Looking at median cumulative durations of Inter-sector Communications highlights such aspect.



(a) Configuration B



(b) Configuration C

Figure 9. Inter-sector Communications

It seems that the sector responsible for the sequencing of flights is mostly affected by the new operational conditions. In practice, the sector tries to overcome the limitations of the configuration by communicating with the other sectors in order to guarantee the orderly management of flights. Figure 10 shows how the cumulative duration of Inter-sector Communications increases substantially for the configuration C.

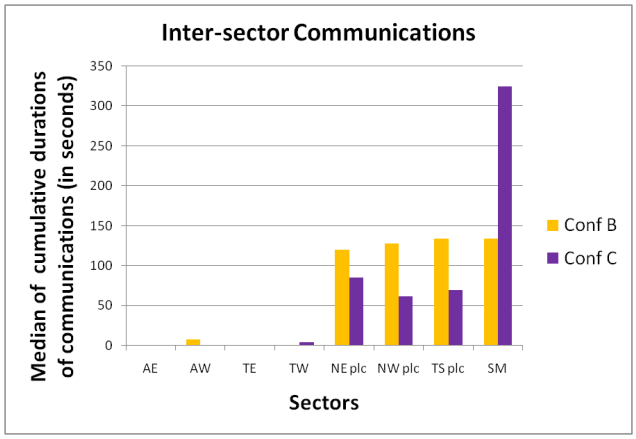


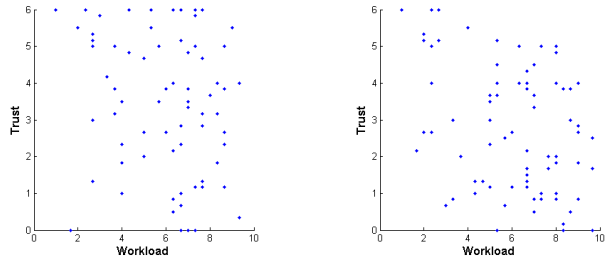
Figure 10. Inter-sector Communications

These results show that different trustworthiness affects not only controllers' perceptions, but also their work practice in terms of Inter-sector Communications.

D. Key Performance Indicators

The final step of our analysis takes an account of trust with respect to the *Key Performance Indicators*, i.e. *Workload*, *Teamwork* and *Situational Awareness*, investigated during the validation session. We looked for any emergent relationship between trust and the other aspects. The confrontations did not highlight any relevant relationship. However,

the overall impression is that a reduced trustworthiness of the system configuration corresponds to a progressive disconnection between trust and with the other performance indicators — that is, a lack of trust causes a disconnect with the other operational aspects. Figure 11, for instance, shows the comparisons between trust and workload for the configurations B and C, respectively. This somehow questions those theoretical accounts that deal with trust and mistrust in an equivalent way.



(a) Configuration B (b) Configuration C

Figure 11. Trust vs. Workload

E. Other Trust Considerations

Finally, we would like to discuss the *'individual'* aspect of trust — that is, how individuals changed their trustworthiness perceptions during the validation session. We are questioning whether or not there was any individual learning aspect. The ATCOs' trust profiles do not highlight any learning curve, or increasing trust. Figure 12, for instance, shows the Trust Scores for the Air Traffic Controllers (ATCOs) working on the most affected sectors over different exercises for the configuration C.

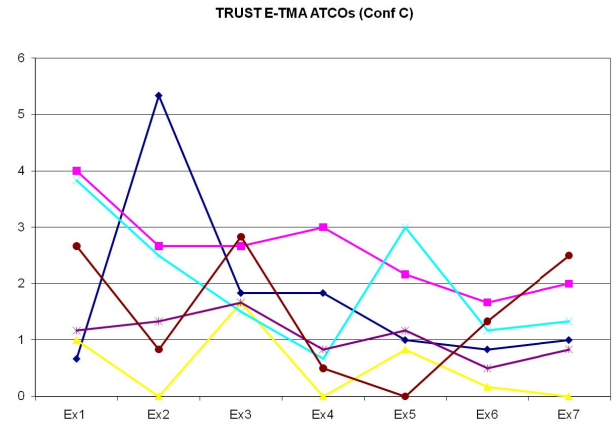


Figure 12. ATCOs' Trust Scores

However, we need to interpret such 'result' in the context of the validation session. ATCOs were systematically rotated across the different sectors and for each validation scenario in order to minimize any learning of the traffic scenarios. The objective was to validate alternative configurations, not to train controllers with specific work practices. In general, objectives of validation sessions are different than the ones of training sessions. Therefore, trust investigations that intend to look at individual behavior need to have different validation organizations.

VI. Discussion and Conclusions

This paper presents an operational account of trust. This operational account builds on our experimental trust observations in a validation case study drawn from the Air Traffic Management (ATM) domain. The case study involved the validation of alternative system configurations and relative operational concepts. The empirical results show how trust observations enable us to analyze the different configurations as perceived by Air Traffic Controllers (ATCOs). Our observations look at trust from different levels of analysis. At the macroscopic level of analysis, we looked at general trends of trustworthiness perceptions. Whereas, at the microscopic level of analysis, we looked at how trust affects operational aspects in terms of sector/controller interactions.

We also investigated relationships between trust and other Key Performance Indicators (KPIs). Although we did not find any emergent relationship, the overall impression is that a lack of trust causes a disconnect with different indicators. This suggests that trust and mistrust should probably be investigated and used differently during empirical investigations. However, the empirical results show that trust observations enable the comparison of alternative system configurations and highlight operational differences.

Other trust considerations concerned the lessons learned in terms of trust investigations. We looked at learning aspects during the validation sessions. The results provide no evidence of any learning aspect for the controllers. However, this is in agreement with the validation organization (in terms of sample scenarios and validation exercises) that intentionally rotated controllers across sectors and responsibilities in order to minimize learning factors over subsequent validation exercises. This suggests future trust investigations in different phases of a system life cycle. That is, it would be interesting to analyze trust observations at different developmental phases. We expect that trust observations may give different indications. It is therefore necessary to take into account how trust is affected by other factors, e.g. tool maturity (in terms of reliability), while analyzing operational observations. Interviews conducted after the validation sessions helped us to clarify and to interpret some observed trust behavior. It is therefore critical to have domain expertise in order to conduct and interpret a non-trivial analysis of trust. Our work stresses an *observational approach to trust*, rather than a conceptual or theoretical one.

In conclusion, this paper presents our experience in collecting and analyzing trust observations in validation exercises. The empirical approach to trust we adopted resulted non-intrusive and easily adaptable to the validation context with little effort. The combination of trust observations with other performance indicators allows us to investigate how trust might relate to other critical operational aspects. The empirical nature of our results provides new insights how to investigate trust.

ACKNOWLEDGMENTS

This work has been supported in part by the *Interdisciplinary Design and Evaluation of Dependability* (INDEED) project, UK Engineering and Physical Research Council (EPSRC), Grant EP/E001297/1, and the *Security engineering for lifelong evolvable systems* (SecureChange) project, FP7-EC-GA-231101.

REFERENCES

[1] *EUROCONTROL Air Traffic Management Strategy for the years 2000+*, EUROCONTROL, 2003.

[2] *SESAR D6 — Work Programme for 2008–2013*, SESAR Consortium, 2008.

[3] *European Operational Concept Validation Methodology (E-OCVM)*, 2nd ed., EUROCONTROL, 2007.

[4] *Guidelines for Trust in Future ATM Systems: A Literature Review*, 1.0 ed., EUROCONTROL, 2003.

[5] D. Gambetta, Ed., *Trust: Making and Breaking Cooperative Relations*. Basil Blackwell, 1988.

[6] R. Hardin, *Trust and Trustworthiness*. Russell Sage Foundation, 2002.

[7] L. J. Camp, "Designing for trust," in *Trust, Reputation, and Security: Theories and Practice, Proceedings of AAMAS 2002 International Workshop*, ser. LNAI, R. Falcone *et al.*, Eds., no. 2631. Springer-Verlag, 2003, pp. 15–29.

[8] E. Yu and L. Liu, "Modelling trust for system design using the i^* strategic actors framework," in *Trust in Cyber-societies: Integrating the Human and Artificial Perspectives*, ser. LNAI, R. Falcone, M. Singh, and Y.-H. Tan, Eds., no. 2246. Springer-Verlag, 2001, pp. 175–194.

[9] R. Hardin, *Trust*. Polity Press, 2006.

[10] D. H. McKnight and N. L. Chervany, "The meanings of trust," University of Minnesota, MISRC Working Papers Series 96-04, 1996.

[11] T. W. Guinnane, "Trust: A concept too many," Economic Growth Center, Yale University, Center Discussion Paper 907, 2005.

[12] A. J. Jones, "On the concept of trust," *Decision Support Systems*, vol. 33, no. 3, pp. 225–232, 2002.

[13] R. M. Kramer, "Trust and distrust in organizations: Emerging perspectives, enduring questions," *Annual Review of Psychology*, vol. 50, pp. 569 – 98, 1999.

[14] *Guidelines for Trust in Future ATM Systems: Principles*, 1.0 ed., EUROCONTROL, 2003.

[15] *Guidelines for Trust in Future ATM Systems: Measures*, 1.0 ed., EUROCONTROL, 2003.

[16] H. Bhana, "Trust but verify," *AEROSAFETYWORLD*, pp. 13–17, Jun 2010.

[17] *The new SHAPE questionnaires: A User Guide*, 0.1 ed., EUROCONTROL.