

Trust Strategies and Policies in Complex Socio-technical Safety-Critical Domains: An Analysis of the Air Traffic Management Domain

Massimo Felici

School of Informatics, The University of Edinburgh
Edinburgh EH9 3JZ, UK
mfelici@inf.ed.ac.uk
<http://homepages.inf.ed.ac.uk/mfelici/>

Abstract. The future development of Air Traffic Management (ATM), set by the ATM 2000+ Strategy, involves 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 the ATM 2000+ Strategy. Moreover, from a technology viewpoint, future ATM services will employ new systems forming the emergent ATM architecture underlying and supporting the European Commission's Single European Sky Initiative. Introducing safety relevant systems in ATM contexts requires us to understand the risk involved in order to mitigate the impact of possible failures. This paper is concerned with trust in technology. Technology innovation supports further (e.g., safety or performance) improvements, although there is often a lack of trust in changes. This paper argues that organizations need to identify trust strategies and policies supporting the delivery of technology innovation. Moreover, the identification of trust strategies and policies supports the understanding of subtle interactions between diverse, often competing, system objectives.

1 Introduction

Computer systems support diverse human activities (e.g., monitoring, decision making, etc.). The introduction of new computer systems, or the upgrade of existing ones, in any environment often modifies work practice. For instance, system operators often need to adjust their procedures around new systems. Moreover, systems may act as a means of communication/mediation between human beings. Complex interactions [17] emerge as results of changes (e.g., environmental changes, new computer systems, adjusted work practices, etc.). The introduction of new technology often requires the re-negotiation of social organizations (e.g., responsibility and accountability) as well as overall system features (e.g., safety). Change gives rise to uncertainties with respect to computer systems. For instance, in the Air Traffic Management (ATM) domain, air traffic controllers often react to system changes or failures by managing less traffic in their air spaces. Uncertainties require of us an extent of *trust* (e.g., with respect

to computer systems). Unfortunately, changes often trigger mistrust. Norman, for instance, reports how the introduction of questioning between pilots in work practice, initially, triggered a lack of trust in the commercial aviation community [46]. However, the new practice, eventually, produced increased safety¹. Similarly, empirical studies point out the relationship between trust in automation and effectiveness of human intervention in continuous process control [43]. Human Reliability Analysis (HRA) highlights how the “*human component*” affects the overall performance and reliability of heterogeneous systems [29].

Technology involves an extent of *risk* [50], regardless our knowledge or trust in it. Any time we use or rely on technologies we take risks. Understanding trust is very important in presence of uncertainties with respect to computer systems and, generally speaking, socio-technical systems. On the one hand technology supports human activities. On the other hand it is a source of harm. Engineering safety-critical systems involves risk analysis [34,55] as part of safety analysis in order to identify safety requirements, although assessing the benefits of technology exposes the limits of pure technical arguments [25]. Whatever is the risk associated with technology, social aspects constrain risk perception - “*Acceptable risk is a matter of judgement*” [10]. However, social and cultural aspects affect judgement [10]. For instance, MacKenzie analyzes how social connectivities affect global financial markets [37]. In particular, the study highlights how, even, electronic mediated trading relies on trust between traders communicating by computers [37]. This further points out contingencies between cooperation (competition) [44] and emergent trust (mistrust).

This paper analyzes trust in the context of Air Traffic Management (ATM). The future development of ATM, set by the ATM 2000+ Strategy [13], involves a structural revision of ATM processes, a new ATM concept and a systems approach for the ATM network. Despite the overall objectives [13], emerging lack of trust may undermine any improvement in the aviation domain (e.g., increased safety and performance). Ongoing research (see, Section 2) is debating and addressing the notion of trust: *What is trust? How to model trust?* Section 2 acknowledges that it is important to understand trust. But, it argues, too, that it is important to investigate trust dynamics. *Trust strategies and policies* should capture how socially constructed risk and knowledge (e.g., system reliability) interact each other. This paper stresses trust strategies (in terms of game theory) and trust policies for the investigation of interaction between *trust*, *risk* and *knowledge*. This paper is structured as follows. Section 2 reviews models of trust. Section 3 highlights the current developments in ATM. Section 4 introduces trust games and elaborates the motivations for trust strategies and policies in ATM.

¹ “Obviously, getting this process in place was difficult, for it involved major changes in the culture, especially when one pilot was junior. After all, when one person questions another’s behavior, it implies a lack of trust; and when two people are supposed to work together, especially when one is superior to the other, trust is essential. It took a while before the aviation community learned to take the questioning as a mark of respect, rather than a lack of trust, and for senior pilots to insist that junior ones question all of their actions. The result has been increased safety”, p. 145, [46].

Trust games capture processes of negotiating, may be competing, over different objectives (e.g., increased safety or performance) limiting phenomena of *risk homeostasis* (e.g., increased system reliability or safety may imply a decreased risk perception favoring risk-taking strategies or behaviors) [29,30] - *Is trust in technology appropriate to the risk?* Section 5, finally, draws some conclusions.

2 On Trust

Modelling has steadily acquired an important role in presence of uncertainty of software-intensive systems [35]. On the one hand, modelling addresses uncertainty of software-intensive systems. On the other hand, it is necessary to contextualize the trust in modelling, that is, acquire trust in models in context. This section reviews diverse models of trust. The diverse models highlight an ongoing debate on the nature of trust. This points out the complexity of trust. Although it is unfeasible, and may be unnecessary, to take a definitive model of trust, models further support the understanding of underlying mechanisms of trust. McKnight and Cherwany propose a *typology of trust* [39]. The typology consists of six trust constructs: *Situational Decision to Trust*, *Dispositional Trust*, *System Trust*, *Trusting Beliefs*, *Trusting Intention* and *Trusting Behavior*. Later, McKnight and Cherbany [41] extend the typology of trust to the notion of *distrust*, as opposed to trust. Although the typology addresses the lack of a unified trust definition, it provides limited support to understand the dynamics of trust formation [42].

The shortcomings of security mechanisms [2] have motivated the increasing interest for the formalization of trust in global computing scenarios [6]. Recent research [6,45] proposes formal models that capture to some extent the typology of trust [39]. Trust constructs, therefore, allow believes to emerge [39]. Other formal models [1,59] exploit the trust constructs and the belief formation processes in order to stress trust into design [46]. Furthermore, formal representations investigate the dynamics of trust [16]. In particular, formal models capture how social connectivities [37] influence the formation of trust in situated relationships (or interactions) between peers [16]. Recent research has exploited similar trust models in order to investigate trust in e-commerce [22,24,40] or other domains involving human-machine interactions [8]. Other research has, instead, investigated quantitative aspects of trust [22,24,56]. However, experimental results expose the limits of extending quantitative approaches to human behaviors [48].

Another aspect of trust is related to its role at the organizational level [23,32,36,49]. It is evident how the formation and perception of trust within, and between, organizations follow mechanisms grounded in the social and cultural nature of trust and risk perception [10,54]. Douglas and Wildavsky elaborate risk perception from a social viewpoint [10]. They analyze how social organizations perceive risk differently [10]. They initially take into account four problems of risk [10]. The four problems consider risk as a joint product of *knowledge* about the future and *consent* about the most desired prospects. It is possible to identify the best solution when knowledge is certain and consent complete. The problem,

in this case, is technical and the solution is one of calculation. By contrast, if consent is contested, the problem is one of disagreement about how to assess consequences. In this case the solution requires further coercion or discussion. In the case in which the consent is complete and the knowledge is uncertain, the risk is related to insufficient information. Therefore, the solution involves research. The last case (i.e., knowledge is uncertain and consent is contested) is how any informed person would characterize risk assessment. In safety-critical systems [34,55], for instance, safety analysis relies on assessment methodology (e.g., FMEA, HAZOP, FTA, etc.) in order to solve the problem of knowledge and consent. Safety assessment gathers evidence in order to acquire consent and confidence over safety arguments and past experiences. Note that diverse arguments may affect each other (e.g., a negated reliability argument of fault free may invalidate a formal argument of correctness) [4,5].

Trust in technology is therefore an emergent judgement depending of knowledge becoming available eventually. Trust in technology mediates different perspectives (e.g., engineering knowledge or safety arguments) and stakeholders (often interacting by technological artifacts). Trust in technology as emergent socially influenced judgement relates (whether directly or not) knowledge and assessed risk influencing its perception. On the other hand, trust in technology is the result of *complex interactions* [50] shaping (e.g., negotiating) knowledge. This paper pinpoints basic mechanisms capturing emergent trust (strategies and policies) relating technological knowledge and risk.

3 Safety, Risk and Trust in ATM

The ATM 2000+ Strategy [13] involves a structural revision of ATM processes, a new ATM concept and a system approach for the ATM network. The overall objective [13] is, *for all phases of flight, to enable the safe, economic, expeditious and orderly flow of traffic through the provision of ATM services, which are adaptable and scalable to the requirements of all users and areas of European airspace*. This requires ATM services to go through significant structural, operational and cultural changes that will contribute towards the ATM 2000+ Strategy. Moreover, from a technology viewpoint, future ATM services will employ new systems forming the emergent ATM architecture underlying and supporting the European Commission's Single European Sky Initiative.

ATM services, it is foreseen, will need to accommodate an increasing traffic, as many as twice number of flights, by 2020. This challenging target will require the cost-effectively gaining of extra capacity together with the increase of safety levels [38,47]. Enhancing safety levels affects the ability to accommodate increased traffic demand as well as the operational efficiency of ensuring safe separation between aircrafts [50]. Unfortunately, even maintaining the same safety levels across the European airspace would be insufficient to accommodate an increasing traffic without affecting the overall safety of the ATM system [11]. Suitable safe conditions (e.g., increased safety levels) shall precede the achievement of increased capacity (in terms of accommodated flights).

The introduction of new safety relevant systems in ATM contexts requires us to understand involved hazards in order to assess the risk and mitigate the impact of possible failures. Diverse domains (e.g., nuclear, chemical or transportation) adopt safety analysis that originates from a general approach [34,55]. The unproblematic application of conventional safety analysis is feasible in some safety-critical domains (e.g., nuclear and chemical plants). In such domains, physical design structures constrain system's interactions and stress the separation of safety related components from other system parts. This ensures to some extent the independence of failures. Unfortunately, ATM systems and procedures have distinct characteristics (e.g., openness, volatility, etc.) that expose limitations of the approach [17,18,20,19]. ATM systems operate in open and dynamic environments where it is difficult completely to identify system interactions (e.g., between aircraft systems and ATM safety relevant systems) [17,18,20,19]. Unfortunately, these complex interactions may give rise to catastrophic failures. Hence, safety analysis has to take into account these complex interaction mechanisms (e.g., failure dependence, reliance in ATM, etc.) in order to guarantee and, possibly, increase the overall ATM safety as envisaged by the ATM 2000+ Strategy [17,18,20,19].

Trust is steadily acquiring an important role in the design of socio-technical systems [46]. This is also driving recent research in ATM [14]. 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 contextualization of trust in ATM [14] 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, defined as [14], *a device or system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator*. The notion of automation influences the understanding of trust in the ATM context. Trust is, defined as [14], *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*. This definition of trust originates from general models of trust. *Complacency*, may be, distinguishes the ATM domain from others. Complacency is a kind of automation misuse, 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 [14]. Although trust and reliability have an important role in ATM², air traffic controllers accept (unreliable) tools as far as they understand the failure modes [14]. Note that the *competence of tool* contributes to the overall trust according to a simple model identified in [14]. Similarly to other domains, ATM is seeking to understand the conceptualization, as well as the quantification, of trust.

² "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", [14].

4 Trust Strategies and Policies

The ATM context provides many examples in which trust and risk may exhibit competing behaviors. For instance, the introduction of new ATM tools aims to support air traffic controllers as well as to increase system performance. However, regardless the (safety) assurance given to the controllers, they often exhibit an initial lack of trust³ (in system evolution) by managing less traffic than planned. This results in economic pressures on the ATM system and customer dissatisfaction. The Short-Term Conflict Detection (STCD) system provides an instance of accepted technology innovation that may result in mistrust or, worst, unsafe behaviors⁴. Figure 1 shows a *Value Net* [44] for the ATM domain.

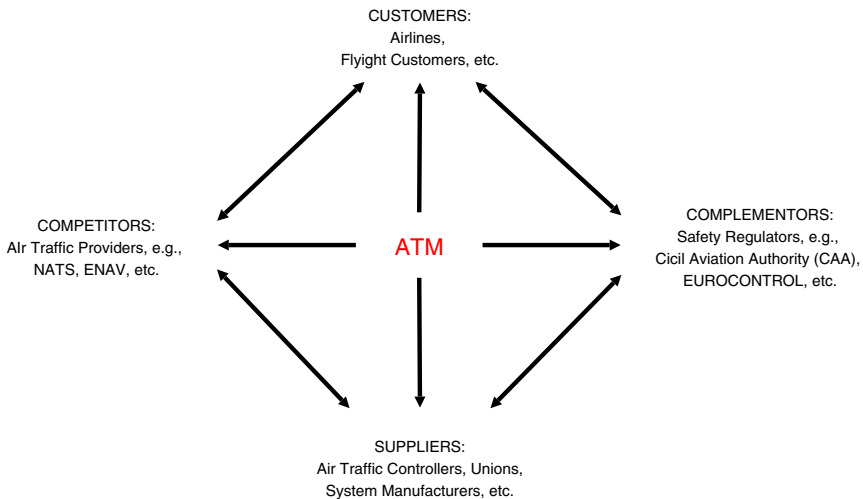


Fig. 1. The value net for ATM

The value net represents all the players and the interdependencies among them. Along the vertical dimension of the value net are *customers* and *suppliers* [44]. Along the horizontal dimension are *competitors* and *complementors* [44]. This section articulates the motivations for trust strategies and policies.

³ “A well-known problem connected with the introduction of a new system (or even changes to an existing system) is that people in the workplace may feel threatened, alienated or otherwise uncomfortable with the change”, p. 19, [12].

⁴ “Mistrust in automation may develop from annoyance about false alarms, for example. While system tools as Short-Term Conflict Detection (STCD) have generally received widespread acceptance among operators, it is crucial for the operator to develop trust in the system. High trust (overtrust or complacency) in automation may on the other hand lead operators to abandon vigilant monitoring of their displays and instruments.”, p. 37, [12].

4.1 Trust, Risk and Knowledge: A Game

Various models capture to some extent the notion of trust, although there has been little attention in the investigation of the dynamics of trust. Social aspects of trust and risk perception [10] stress the interaction between trust, risk and knowledge [23]. Therefore, a social viewpoint provides a convenient intersection between risk, trust and technology. The different relationships (e.g., independence, mediation and moderation) between trust and risk affect emergent behaviors [23]. These relationships between risk and trust highlight different behaviors. The interaction between trust and risk perception finds grounds in the social aspects of technology [10]. The characterization of trust and risk [23] suggests that the underlying constructs interact in the formation of trust and the perception of risk. This interaction originates from the social aspects of trust and risk [10]. Many models address the understanding of trust and risk, although they often treat these aspects in isolation. Whereas, social aspects stress their interdependency. This section presents the interaction between risk, trust and knowledge as a game (in terms of game theory). The underlying idea is to contextualize (i.e., put the risk and trust interdependency into perspective) the conceptualization of risk, with respect to knowledge and consent, in the case of trust in ATM, with respect to system reliability. It is possible to capture the interactions between trust and risk as trust games extending the *Prisoners' Dilemma*.

The Prisoners' Dilemma is a (decision support) game that captures those situations in which there might be competing or cooperative stakeholders having different viewpoints. The prisoners' dilemma has been extensively investigated and used in social, economic, and political contexts [3,9,33,44]. In the Prisoners' Dilemma, two prisoners are placed in separate cells. Both prisoners care much more about their personal freedom than about the welfare of their accomplice. They may choose to confess or remain silent. If they both confess, they will receive reduced convictions (i.e., reward for mutual cooperation). If they both remain silent, they will receive minimal convictions (i.e., punishment for mutual defection). However, if they disagree (i.e., a prisoner confesses and the other remains silent, and vice versa), the silent one will receive the full conviction. Whereas, the one who confessed will be freed. The dilemma here is that, whatever the other does, each is better off confessing than remaining silent. But the outcome obtained when both confess is worse for each than the outcome they would have obtained had both remained silent. Note that different matrices and different rules identify different characterizations (e.g., symmetric, asymmetric, iterative, etc.) of the prisoners' dilemma [33]. The prisoners' dilemma captures those situations in which two players have conflicting interests. Although the two players have their own interests in winning the game, the better strategy corresponds to cooperation [3]. It is possible to identify different heuristics depending on whether or not *dominant strategies* exist [9]. Therefore, the prisoners' dilemma captures those situations that may result in cooperation or competition (i.e., *co-opetition* [44]). The prisoners' dilemma captures trust between individuals (or groups of individuals). People have to collaborate in order to improve their situations. If they trust each other, they have a cooperative strategy.

Trust games as extensions of the Prisoners Dilemma enable the modeling of realistic scenarios [52]. Several studies use the prisoners' dilemma in order to characterize trust, e.g., in computer-mediated communications [52]. Unfortunately, social connectivity [37] exposes the limitations of interpreting the rate of cooperation (measured in terms of collective pay-off) as the level of trust in computer-mediated communications [52]. Characterizations of trust based on the basic prisoners' dilemma partially capture trust complexity. Trust games extend the prisoners' dilemma in order to overcome some of its practical limitations [52]. Trust games capture real scenarios that exhibit asynchronous and asymmetric properties, which expose the limitations of the prisoners' dilemma [52]. In particular, asymmetric games capture differences of risk perceptions among individuals, actors or agents (e.g., systems, business competitors, users, etc.). Risk perception affects interactions. Figure 2 shows a representation of a trust game.

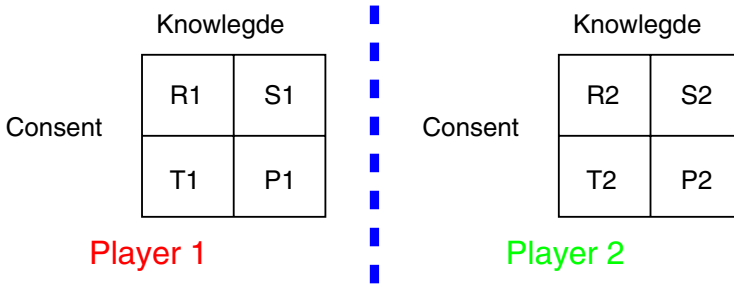


Fig. 2. A Trust Game

The game involves two players: **Player 1** and **Player 2**. The two players have some *common knowledge* [51] about the system (e.g., system reliability). The two players have different strategies according to their expected pay-offs (or convictions). For instance, **Player 1** (i.e., *complete-certain*) can have complete consent and being certain of the system reliability. That is, **Player 1** trusts the common knowledge and expect a similar behavior from **Player 2**. This corresponds to $R1$ in the pay-offs matrix (see, Figure 2). The other pay-offs, i.e., $T1$, $S1$ and $P1$, correspond to the different combinations of consent and certainty about knowledge, i.e., *contested-certain*, *complete-uncertain* and *contested-uncertain*, respectively.

Although the two players partially have some common knowledge about the system, the two players will normally choose their dominant choice (i.e., defection: $P1$ and $P2$). Thus, each will get less than they both could have gotten if they had cooperated (i.e., cooperation: $R1$ and $R2$) [3]. If they play a known finite number of times, the players would have none incentive to cooperate. By contrast, if the players will interact an indefinite number of times, cooperation can emerge [3]. Each player chooses the preferred strategy independently (that is, without knowing each other strategy). **Player 1** would like to have a

dominant strategy such to have correct trust in technology. However, **Player 2** would prefer to have a dominant strategy such to have complete consent in the risk associated with technology. Once the two players have decided their strategies, **Player 1** exhibits the chosen trust in technology and exhibits relevant evidence (e.g., high reliability or low reliability). **Player 2**, then, according to the chosen strategy (i.e., certain or uncertain knowledge), can have a contested or complete consent of the knowledge exhibited (e.g., high or low reliability). The unfolding of the game identifies different strategies (e.g., trust as well as risk taking). The two players may have different overall objectives or cooperate towards common objectives. The next section shows how the game allows the understanding and the characterization of the relationship between trust, risk and knowledge. Moreover, playing the game identifies trust strategies.

4.2 Trust Strategies

This section highlights that trust games allow the characterization of trust in situated (risk) contexts. Trust games take into account that risk perception and trust may behave as opponent (or competing) forces, regardless the (system) knowledge (e.g., system reliability). Playing trust games shows whether the two players exhibit cooperative or competing strategies [44]. Once the players have chosen their strategies (i.e., trust or mistrust, and certain or uncertain), they both have limited choices for the next move. For instance, if **Player 1** has trust in technology, whatever the knowledge about it. **Player 1** can only exhibit partial knowledge about the system (e.g., high reliability or low reliability). Although, it seems a contradiction there are cases in which people have trust in technology, despite low reliability, because they understand it. Similarly, **Player 2** may have a contested or complete consent over the knowledge in alternative strategies of certain or uncertain knowledge.

Figure 3 shows an example of possible choices (in terms of decision tree) when both players have trust in knowledge about the system. The decision tree shows the different combinations and identifies the different outcomes. In this case, full cooperation between the ATM service provider and Air Traffic Controllers is a possible outcome. Let us assume that both players **Player 1** and **Player 2** (e.g., ATM provider and Air Traffic Controllers) have certain knowledge of system reliability. **Player 1**, therefore, can be in two situations: $R1$ or $T1$. Similarly, **Player 2** can choose $R2$ or $T2$. Figure 3 shows the different cases. For instance, the combination $T1$ and $R2$ may result in a risk taking strategy, because there is a complete consent on over-trusting the system according to certain knowledge. This could be the case, when unreliable technology is still adopted, because the air traffic controllers understand the failure modes. Therefore, they systematically work-around faulty conditions. Another example is the situations in which there is a complete consent in technology trust according to certain available knowledge (i.e., $R1$ and $R2$). This would be the optimal case in practice for trust strategies - people have trust in reliable technology.

The combination of the different conditions allows the identification of potential strategies. However, any strategy may require further commitments in terms

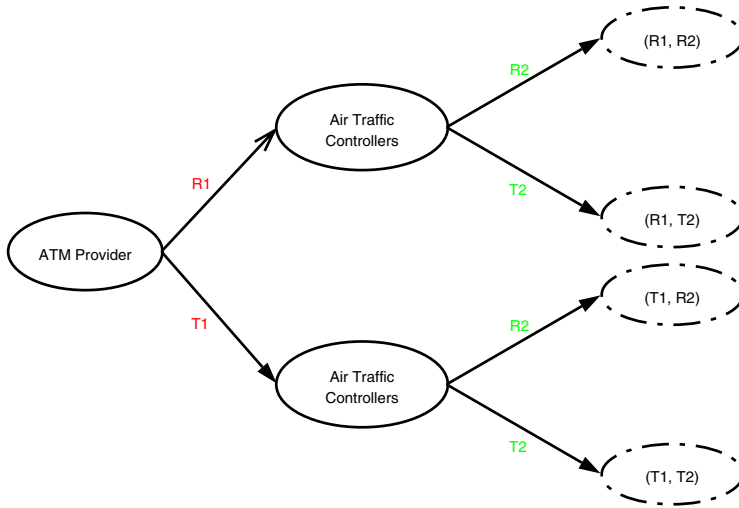


Fig. 3. A decision tree

of resources (e.g., financial investment) and activities (e.g., gathering further evidence). Similarly, it is possible to explain and analyze the other conditions and the cases of mistrust in technology. Note that the players may engage two different types of games: a cooperative game or a competitive game. The cooperative game corresponds to situations in which both players have common objectives, despite that they might have different understanding how to achieve them. The competitive game corresponds to those situations in which both players have different objectives (e.g., customer vs. supplier, regulator vs. service provider, etc.).

4.3 Trust Policies

Cooperation or competition among ATM actors stress the characterization of trust strategies in terms of (multi-agent) trust games. Policy-based frameworks (e.g., KAoS [57,58]) differently support the building on trust (strategies) within organizations. For instance, KAoS policy supports the specification, analysis, disclosure and enforcement for semantic web services [57,58]. KAoS provides a framework in which agents will discover, communicate and cooperate with other agents and services. Therefore, KAoS enables the specification of trust policy-based management systems. Hence, trust policy-based management systems extend trust management systems [31,53]. However, trust policies, in this way, represent and extend security mechanisms within specific virtual organizations. Trust policies would enhance to a certain extent trust, although they provide limited support for organizational trust.

Policies in safety-critical domains, for instance, are differently decomposed in order to constrain the behavior of a System of Systems (SoS) [27]. A goal-based approach is used to decompose safety policies, which represent a means

for achieving safety. Although structured notations support the decomposition process, they face evolution [20]. However, it is possible to capture emergent complex interactions [17]. Modeling enables the characterization of evolutionary structures [21], although it is still required the identification of change strategies. This stresses the interactions between trust strategies and (structured) trust policies.

4.4 A Matter of Knowledge

This section points out a characterization of trust games in a logical framework for reasoning about knowledge and uncertainty [15,28]. The logical framework allows the characterization of knowledge (uncertainty) in multi-agents systems [15]. Therefore, the framework easily captures trust games. The basics consist of well-established results in modal logic [7,15]. Although the theoretical results in modal logic extend over several levels of expressiveness (e.g., intuitionistic, propositional, first-order, etc.), this section refers to a simple propositional modal logic. Modal logic allows the formalization of the intuitions about necessity and possibility. There exist many different representations that describe modal logic. Most of them are equivalent from a theoretical viewpoint. A semantics for propositional modal logic relies on the *possible worlds* framework, *Kripke structures* or *Kripke frames*. This allows us to define a notion of validity for modal logic, hence *Kripke models*. Intuitively, the Kripke semantics interprets modal formulas like worlds that are related each other by an accessibility relationship.

The basic framework of modal logic allows the modeling of multi-agents systems [15]. For instance, in a group of agents (or players) G , given current information, an agent may not be able to tell which of a number of possible worlds describes the actual state of affairs. An agent is then said to know a fact, if the fact is true at all the possible worlds (according to given knowledge). It is possible to extend the modal logical framework in order to express the notions of *common knowledge* and *distributed knowledge* [15]. To express these notions, the language is extended with the modal operators “*everyone in the group G knows*”, “*it is common knowledge among the agents in G* ” and “*it is distributed knowledge among the agents in G* ” [15]. This allows the modeling of multi-agents systems or trust games.

5 Conclusions

The social aspects of trust and risk perception highlight the interactions between trust, risk and knowledge. These interactions exhibit different behaviors situated in contexts. The analysis of trust with respect to risk perception and knowledge allows the characterization of practical situations in which trust, or mistrust, emerges. This paper presents a trust game that captures the interdependency between trust and risk perception. The trust game is an extension of the prisoners’ dilemma. Unfolding the game corresponds to different trust strategies. Moreover, the game captures the interdependency between trust and

risk perception into contextualized (system) knowledge. Trust games capture the interactions between risk, trust and knowledge that emerge in practice. Organizational (e.g., social and cultural) aspects constrain the game, that is, the movements available to each player. Trust policies may capture these organizational constraints. Therefore, it could be the case that some practical situations lack any achievable solution, that is, none of the player has a dominant strategy. It is possible to formalize the game in a logical framework for reasoning about knowledge (and uncertainty) [15,28].

In conclusions, this paper analyzes the interaction of trust, risk and knowledge in the context of Air Traffic Management (ATM). It is possible to characterize the emergence of trust strategies and policies. Trust games highlight that trust plays a crucial role with respect to risk and knowledge in order to achieve overall objectives [13] in the ATM domain. Although trust games capture the interaction between trust, risk and knowledge, in practice, it is still challenging the instantiation and construction of trust games (e.g., identification of the decision matrix, rules, etc.). However, the paper stresses and justifies future investigations of trust strategies and policies. Moreover, it provides a game-oriented characterization for the analysis of trust strategies and policies. Future formalization of the game in theoretical terms would allow the identification of game conditions. Future work aims to formalize the rules underlying trust games. Moreover, the instantiation of trust games in situated context would allow the identification of heuristics [26]. Future work intends to use trust games in order to investigate relationships between different strategies (e.g., adoption of technology innovation, system testing and validation, etc.) and policies. This would further support the understanding and generalization of the notion of trust. However, organizations may, already, use and instantiate trust games in order to understand and investigate how trust, risk and knowledge interact within their contexts.

Acknowledgements

This work has been supported by the UK EPSRC Interdisciplinary Research Collaboration in Dependability, DIRC - <http://www.dirc.org.uk> - grant GR/N13999.

References

1. Alfarez Abdul-Rahman and Stephen Halles. A distributed model of trust. In *Proceedings of the New Security Paradigms Workshop*, pages 48–60. ACM, 1997.
2. Ross Anderson. *Security Engineering: A Guide to Build Dependable Distributed Systems*. Wiley Computer Publishing, 2001.
3. Robert Axelrod. *The Evolution of Co-operation*. Penguin Books, 1990.
4. Robin Bloomfield and Bev Littlewood. Multi-legged arguments: the impact of diversity upon confidence in dependability arguments. In *Proceedings of the 2003 International Conference on Dependable Systems and Networks, DSN'03*, pages 25–34. IEEE Computer Society, 2003.

5. Robin Bloomfield and Bev Littlewood. On the use of diverse arguments to increase confidence in dependability claims. In Denis Besnard, Cristina Gacek, and Cliff B. Jones, editors, *Structure for Dependability: Computer-Based Systems from an Interdisciplinary Perspective*, chapter 13, pages 254–268. Springer-Verlag, 2006.
6. Marco Carbone, Mogens Nielsen, and Vladimiro Sassone. A formal model of trust in dynamic networks. In *Proceedings of the First International Conference on Software Engineering and Formal methods (SEFM'03)*. IEEE Computer Society, 2003.
7. Alexander Chagrov and Michael Zakharyashev. *Modal Logic*. Number 35 in Oxford Logic Guides. Oxford University Press, 1997.
8. I. Dasonville, D. Jolly, and A. M. Desodt. Trust between man and machine in a teleoperation system. *Reliability Engineering & System Safety*, 53:319–325, 1996.
9. Avinash K. Dixit and Barry J. Nalebuff. *Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life*. W. W. Norton & Company, 1991.
10. Mary Douglas and Aaron Wildavsky. *Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers*. University of California Press, 1982.
11. John H. Enders, Robert S. Dodd, and Frank Fickeisen. Continuing airworthiness risk evaluation (CARE): An exploratory study. *Flight Safety Digest*, 18(9-10):1–51, September–October 1999.
12. EUROCONTROL. *Human Factor Module - Human Factors in the Development of Air Traffic Management Systems*, 1.0 edition, 1998.
13. EUROCONTROL. *EUROCONTROL Air Traffic Management Strategy for the years 2000+*, 2003.
14. EUROCONTROL. *Guidelines for Trust in Future ATM Systems: A Literature Review*, 1.0 edition, 2003.
15. Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. The MIT Press, 2003.
16. Rino Falcone and Cristiano Castelfranchi. The socio-cognitive dynamics of trust: Does trust create trust? In R. Falcone, M. Singh, and Y.-H. Tan, editors, *Trust in Cyber-societies*, number 2246 in LNAI, pages 55–72. Springer-Verlag, 2001.
17. Massimo Felici. Capturing emerging complex interactions - safety analysis in atm. In Chris Johnson, editor, *Proceedings of the 2nd Workshop on Complexity in Design and Engineering, GIST Technical Report G2005-1*, pages 120–129, 2005.
18. Massimo Felici. Evolutionary safety analysis: Motivations from the air traffic management domain. In R. Winther, B.A. Gran, and G. Dahll, editors, *Proceedings of the 24th International Conference on Computer Safety, Reliability and Security, SAFECOMP 2005*, number 3688 in LNCS, pages 208–221. Springer-Verlag, 2005.
19. Massimo Felici. Capturing emerging complex interactions: Safety analysis in air traffic management. *Reliability Engineering & System Safety*, 91(12):1482–1493, 2006.
20. Massimo Felici. Modeling safety case evolution - examples from the air traffic management domain. In Nicolas Guelfi and Anthony Savidis, editors, *Proceedings of the Second International Workshop on Rapid Integration of Software Engineering Techniques, RISE 2005*, number 3943 in LNCS, pages 81–96. Springer-Verlag, 2006.
21. Massimo Felici. Structuring evolution: on the evolution of socio-technical systems. In Denis Besnard, Cristina Gacek, and Cliff B. Jones, editors, *Structure for Dependability: Computer-based Systems from an Interdisciplinary perspective*, chapter 3, pages 49–73. Springer, 2006.
22. David Gefen, Elena Karahanna, and Detmar W. Straub. Inexperience and experience with online stores: The importance of tam and trust. *IEEE Transactions on Engineering Management*, 50(3):307–321, August 2003.

23. David Gefen, V. Srinivasan Rao, and Noam Tractinsky. The conceptualization of trust, risk and their relationship in electronic commerce: The need for clarifications. In *Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS'03)*. IEEE, 2003.
24. David Gefen and Detmar W. Straub. Consumer trust in b2c e-commerce and the importance of social presence: experiments in e-products and e-services. *Omega: The International Journal of Management Science*, 32:407–424, 2004.
25. Gerd Gigerenzer. *Reckoning with Risk: Learning to Live with Uncertainty*. Penguin Books, 2002.
26. Gerd Gigerenzer, Peter M. Todd, and The ABC Research Group, editors. *Simple Heuristics That Make Us Smart*. Oxford University Press, 1999.
27. Martin Hall-May and Tim Kelly. Defining and decomposing safety policy for systems of systems. In R. Winther, B.A. Gran, and G. Dahll, editors, *Proceedings of SAFECOMP 2005*, number 3688 in LNCS, pages 37–51. Springer-Verlag, 2005.
28. Joseph Y. Halpern. *Reasoning about Uncertainty*. The MIT Press, 2003.
29. Erik Hollnagel. *Human Reliability Analysis: Context and Control*. Academic Press, 1993.
30. Chris W. Johnson. *Failure in Safety-Critical Systems: A Handbook of Accident and Incident Reporting*. University of Glasgow Press, Glasgow, Scotland, October 2003.
31. Audun Josang, Claudia Keser, and Theo Dimitrakos. Can we manage trust? In P. Herrmann et al., editors, *Proceedings of iTrust 2005*, number 3477 in LNCS, pages 93–107. Springer-Verlag, 2005.
32. Eva C. Kasper-Fuehrer and Neal M. Ashkanasy. Building trust in cross-cultural collaborations: Toward a contingency perspective. *Journal of Management*, 27:235–254, 2001.
33. Steven Kuhn. Prisoner's dilemma. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/fall2003/entries/prisoner-dilemma/>, 2003.
34. Nancy G. Leveson. *SAFWARE: System Safety and Computers*. Addison-Wesley, 1995.
35. Bev Littlewood, Martin Neil, and Gary Ostrolenk. The role of models in managing the uncertainty of software-intensive systems. *Reliability Engineering & System Safety*, 46:97–95, 1995.
36. Yadong Luo. Building trust in cross-cultural collaborations: Toward a contingency perspective. *Journal of Management*, 28(5):669–694, 2002.
37. Donald MacKenzie. Social connectivities in global financial markets. *Environment and Planning D: Society and Space*, 22:83–101, 2004.
38. Stuart Matthews. Future developments and challenges in aviation safety. *Flight Safety Digest*, 21(11):1–12, November 2002.
39. D. Harrison McKnight and Norman L. Chervany. The meanings of trust. Technical Report 96-04, University of Minnesota, 1996.
40. D. Harrison McKnight and Norman L. Chervany. Conceptualizing trust: A typology and e-commerce customer relationships model. In *Proceedings of the 34th Hawaii International Conference on System Sciences*, pages 1–9. IEEE, 2001.
41. D. Harrison McKnight and Norman L. Chervany. Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, and Y.-H. Tan, editors, *Trust in Cyber-societies*, number 2246 in LNAI, pages 27–54. Springer-Verlag, 2001.
42. D. Harrison McKnight, Larry L. Cummings, and Norman L. Chervany. Trust formation in new organizational relationships. Technical Report 96-01, University of Minnesota, 1996.

43. Neville Moray, Douglas Hiskes, John Lee, and Bonnie M. Muir. Trust and human intervention in automated systems. In Jean-Michel Hoc, Pietro C. Cacciabue, and Erik Hollnagel, editors, *Expertise and Technology: Cognition & Human-Computer Cooperation*, chapter 11, pages 183–194. Lawrence Erlbaum Associates, 1995.
44. Barry J. Nalebuff and Adam M. Brandenburger. *Co-opetition*. HarperCollinsBusiness, 1996.
45. Mogens Nielsen and Karl Krukow. Towards a formal notion of trust. In *Proceedings of PPDP'03*. ACM, 2003.
46. Donald A. Norman. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, 2004.
47. Michael Overall. New pressures on aviation safety challenge safety management systems. *Flight Safety Digest*, 14(3):1–6, March 1995.
48. Alberto Pasquini, Giuliano Pistoiesi, and Antonio Rizzo. Reliability analysis of systems based on software and human resources. *IEEE Transactions on Reliability*, 50(4):337–345, 2001.
49. Paul A. Pavlou, Yao-Hua Tan, and David Gefen. The transitional role of institutional trust in online interorganizational relationships. In *Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS'03)*. IEEE, 2003.
50. Charles Perrow. *Normal Accidents: Living with High-Risk Technologies*. Princeton University Press, 1999.
51. Eric Rasmusen. *Games and Information: An Introduction to Game Theory*. Blackwell, second edition, 1989.
52. Jens Riegelsberger, M. Angela Sasse, and John D. McCarthy. The researcher's dilemma: evaluating trust in computer-mediated communication. *International Journal of Human-Computer Studies*, 58:759–781, 2003.
53. Sini Ruohomaa and Lea Kutvonen. Trust management survey. In P. Herrmann et al., editors, *Proceedings of iTrust 2005*, number 3477 in LNCS, pages 77–92. Springer-Verlag, 2005.
54. J.N. Sorensen. Safety culture: a survey of the state-of-the-art. *Reliability Engineering & System Safety*, 76:189–204, 2002.
55. Neil Storey. *Safety-Critical Computer Systems*. Addison-Wesley, 1996.
56. Ananth Uggirala, Anand K. Gramopadhye, Nrain J. Melloy, and Joe E. Toler. Measurement of trust in complex and dynamic systems using a quantitative approach. *International Journal of Industrial Ergonomics*, 34(3):175–186, 2004.
57. Andrzej Uszok et al. Applying KAOS services to ensure policy compliance for semantic web services workflow composition and enactment. In S.A. McIlraith, editor, *Proceedings of ISWC 2004*, number 3298 in LNCS, pages 425–440. Springer-Verlag, 2004.
58. Andrzej Uszok et al. KAOS policy management for semantic web services. *IEEE Intelligent Systems*, pages 32–41, July/August 2004.
59. Eric Yu and Lin Liu. Modelling trust for system design using the *i** strategic actors framework. In R. Falcone, M. Singh, and Y.-H. Tan, editors, *Trust in Cyber-societies*, number 2246 in LNAI, pages 175–194. Springer-Verlag, 2001.