

Trust strategies: motivations from the Air Traffic Management domain

M. Felici

School of Informatics, The University of Edinburgh, UK

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. Although technological innovation supports further (e.g., safety or performance) improvements, there is often a lack of trust in changes. This paper argues that organizations need to identify trust strategies supporting the delivery of technological innovation. Moreover, the identification of strategies for building on trust supports the understating 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. Another aspect is that computers act as a means of communication or mediation between human beings. Complex interactions (Felici 2005; Felici 2006) emerge as results of changes (e.g., environmental changes, new computer systems, adjusted work practice, 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 (Norman 2004) reports how the introduction of questioning between pilots in work practice, initially, triggered a lack of trust in the commercial

aviation community¹. The new practice, eventually, produced increased safety (Norman 2004).

Technologies involve an extent of *risk* (Perrow 1999), regardless our knowledge or trust in them. 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. Although engineering safety-critical systems involves risk analysis (Leveson 1995; Storey 1996) as part of safety analysis in order to identify safety requirements, whatever is the risk associated with technology, social aspects constrain risk perception - "*Acceptable risk is a matter of judgment*" (Douglas and Wildavsky 1982). Social and cultural aspects affect judgment. For instance, MacKenzie analyzes

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, (Norman 2004).

how social connectivity affects global financial markets (MacKenzie 2004). In particular, the study reports how, even, electronic mediated trading relies on trust between traders communicating by computers. This further highlights how cooperation relies on emergent trust.

This paper analyzes trust in the context of ATM. The future development of ATM, set by the ATM 2000+ Strategy (EUROCONTROL 2003a), involves a structural revision of ATM processes, a new ATM concept and a systems approach for the ATM network. In spite the overall objectives, emerging lack of trust may undermine any improvement in the aviation domain (e.g., increased safety and performance). Ongoing research debates (see, Section 3) are addressing the notion of trust: *What is trust? How to model trust?* This paper acknowledges that it is important to understand trust. Moreover, it argues, too, that it is important to investigate the dynamics of trust. *Trust strategies* allow the analysis of how socially constructed risk and knowledge (e.g., system reliability) interact each other. This paper introduces a game as a decision support tool for the investigation of trust strategies with respect to risk and knowledge. This paper is structured as follows. Section 2 highlights the current developments in ATM. Section 3 and Section 4 review models of trust and trust in ATM, respectively. Section 5 introduces a trust game and elaborates the motivations for trust strategies in ATM. The game captures processes negotiating, may be competing, over different objectives (e.g., trust or risk): *Is trust in technology appropriate to the risk?* Section 6, finally, draws some conclusions.

2 SAFETY AND RISK IN ATM

Recent safety requirements, defined by EUROCONTROL (European organization for the safety of air navigation), imply the adoption of safety analysis for the introduction of new systems and their related procedures in the ATM domain (EUROCONTROL 2001). The EUROCONTROL Safety Regulatory Requirements (EUROCONTROL 2001), ESARR4, require the use of a risk based-approach in ATM when introducing and/or planning changes to any (ground as well as onboard) part of the ATM system. This concerns the human, procedural and equipment (i.e., hardware or software) elements of the ATM system as well as its operational environment at any stage of the lifecycle of the ATM system. The ESARR4 (EUROCONTROL 2001) requires that ATM service providers systematically identify any hazard for any change into the ATM system (parts). Moreover, they have to assess any related risk and identify relevant mitigation actions. In order to provide guidelines for and standardize safety analysis, EUROCONTROL has developed the

EATMP Safety Assessment Methodology (SAM) (EUROCONTROL 2004) reflecting best practices for safety assessment of Air Navigation Systems. The SAM methodology provides a means to demonstrate compliance to ESARR4. The objective of the methodology is to define the means for providing assurance that Air Navigation Systems are safe for operational use. The SAM methodology describes a generic process for the safety assessment of Air Navigation Systems. The SAM methodology consists of three major steps: *Functional Hazard Assessment (FHA)*, *Preliminary System Safety Assessment (PSSA)* and *System Safety Assessment (SSA)*. The process covers the complete lifecycle of an Air Navigation System, from initial system definition, through design, implementation, integration, transfer to operations, to operations and maintenance. Moreover, it takes into account three different types of system elements (human, procedure and equipment elements), the interactions between these elements and the interactions between the system and its environment.

The introduction of new safety relevant systems in ATM contexts requires us to understand the risk involved in order to mitigate the impact of possible failures. Safety analysis involves the activities, i.e., definition and identification of system(s) under analysis, risk analysis in terms of tolerable severity and frequency, definition of mitigation actions, that allow the systematic identification of hazards, risk assessment and mitigation processes in safety-critical systems (Leveson 1995; Storey 1996). The unproblematic application of conventional safety analysis is feasible in some safety-critical domains (e.g., nuclear and chemical plants). Unfortunately, ATM systems and procedures exhibit distinct characteristics (e.g., openness, volatility, etc.) that expose limitations of the approach (Felici 2005; Felici 2006). 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). 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 even increase the overall ATM safety as envisaged by the ATM 2000+ Strategy (Felici 2005; Felici 2006).

3 ON TRUST

Modeling has steadily acquired an important role in presence of uncertainty of software-intensive systems. On the one hand, modeling addresses uncertainty of software-intensive systems. On the other hand, it is necessary to contextualize the trust in modeling, that is, acquire confidence in contextual-

ized models. This section reviews diverse models of trust. The diverse models highlight an ongoing debate on the nature of trust. It 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 Chervany propose a *typology of trust* (McKnight and Chervany 1996). The typology consists of six trust constructs: *Situational Decision to Trust*, *Dispositional Trust*, *System Trust*, *Trusting Beliefs*, *Trusting Intention* and *Trusting Behavior*. The typology originates from an extensive multidisciplinary review of diverse literatures related to trust. Later, McKnight and Chervany (McKnight and Chervany 2001b) extended the typology of trust to the notion of *distrust*, as apposed to trust. Although the typology addresses the lack of a general definition of trust, it provides limited support to understand the dynamics of trust formation (McKnight et al. 1996). Figure 1 shows the typology of trust and the relationships between the trust constructs (McKnight and Chervany 1996). Table 1 describes the trust constructs.

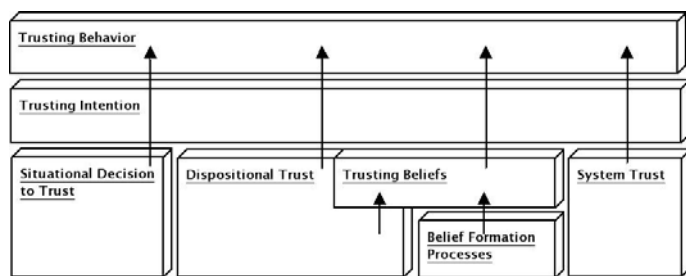


Figure 1. A typology of trust.

Table 1. Trust constructs.

| Construct | Definition |
|-------------------------------|---|
| Trusting Intention | The extent to which one party is willing to depend on the other party in a given situation with a feeling of relative security, even though negative consequences are possible. |
| Trusting Behavior | The extent to which one person voluntarily depends on another person in a specific situation with a feeling of relative security, even though negative consequences are possible. |
| Trusting Beliefs | The extent to which one believes (and feels confident in believing) that another person is trustworthy in the situation. |
| System Trust | The extent to which one believes that proper impersonal structures are in place to enable one to anticipate a successful future endeavor. |
| Dispositional Trust | The recognition that people develop, over the course of their lives, generalized expectations about the trustworthiness of other people. |
| Situational Decision to Trust | The extent to which one intends to depend on a non-specific other party in a given situation. |

The shortcomings of security mechanisms (Anderson 2001) have motivated the increasing interest for the formalization of trust in global computing scenarios (Carbone et al. 2003). Various proposed formal models (Carbone et al. 2003; Nielsen

and Krukow 2003) capture to some extent the typology of trust (McKnight and Chervany 1996). Trust constructs, therefore, allow beliefs to emerge (McKnight and Chervany 1996). Other formal models (Abdul-Rahman and Halles 1997; Yu and Liu 2001) exploit the trust constructs and the belief formation processes in order to stress trust into design (Norman 2004). Furthermore, formal representations investigate the dynamics of trust (Falcone and Castelfranchi 2001). In particular, formal models capture how social connectivities (MacKenzie 2004) influence the formation of trust in situated relationships (or interactions) between peers (Falcone and Castelfranchi 2001). Recent research has exploited similar trust models in order to investigate trust in e-commerce (Gefen et al. 2003; Gefen and Straub 2004; McKnight and Chervany 2001a) or other domains involving human-machine interactions (Dassonville et al. 96). Other research has, instead, investigated quantitative aspects of trust (Gefen et al. 2003; Gefen and Straub 2004; Uggirala et al. 2004).

Another aspect of trust is related to its role at the organizational level (Gefen et al. 2003; Kasper-Fuehrer and Ashkanasy 2001; Luo 2002; Pavlou et al. 2003). 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 (Kramer 1999) and risk perception (Douglas and Wildavsky 1982; Sorensen 2002). Organizational theory (Kramer 1999) emphasizes different trust perspectives: *trust as a psychological state* and *trust as a choice behavior*. Subtle interactions between diverse trust aspects highlight the complexity of trust, which may result into a *three-part* conceptualization of trust, involving properties of a truster, attributes of a trustee and a specific domain over which trust is conferred (Kramer 1999). A three-part theory of trust captures a combination of both the *calculative and relational underpinnings of trust* (Kramer 1999). Bases of organizational trust (Kramer 1999) and the typology of trust (McKnight and Chervany 1996) have some similarities. Both trust accounts emphasize the benefits of understanding trust mechanisms, although they also highlight the complexity of trust and mistrust mechanisms. In particular, it is still challenging the understanding of trust mechanisms with respect to technology. Although technological innovations aim to improve confidence over critical objectives (e.g., security or safety), technologies could undermine trust (Kramer 1999).

3.1 Risk perception and trust

Douglas and Wildavsky elaborate risk perception from a social viewpoint (Douglas and Wildavsky 1982). They initially take into account four problems of risk (see, Figure 2) (Douglas and Wildavsky

1982). However, they analyze how different social organizations perceive risk differently (Douglas and Wildavsky 1982).

| | | Knowledge | |
|---------|-----------|---|---|
| | | Certain | Uncertain |
| Consent | Complete | Problem: Technical Solution: Calculation | Problem: Information Solution: Research |
| | Contested | Problem: (dis)Agreement Solution: Coercion or Discussion | Problem: Knowledge and Consent Solution: ? |

Figure 2. Four problems of risk.

The four problems (see, Figure 2) 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 (Leveson 1995; Storey 1996), 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.

4 TRUST IN ATM

Trust is steadily acquiring an important role in the design of socio-technical systems (Norman 2004). This is also driving recent research in ATM (EUROCONTROL 2003b). 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. Figure 3, for instance, shows the theoretical assumption of the relationship between trust and system reliability (EUROCONTROL 2003b).

The contextualizing of trust in ATM (EUROCONTROL 2003b) 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 (EUROCONTROL 2003b), *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 (EUROCONTROL 2003b), *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*. Although 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 (EUROCONTROL 2003b). Although trust and reliability have an important role in ATM², air traffic controllers accept unreliable tools as far as they understand the failure modes (EUROCONTROL 2003b). Similarly to other domains, ATM is seeking to understand the conceptualization, as well as the quantification, of trust. Note that the *competence of tool* contributes to the overall trust according to a simple model identified in (EUROCONTROL 2003b).

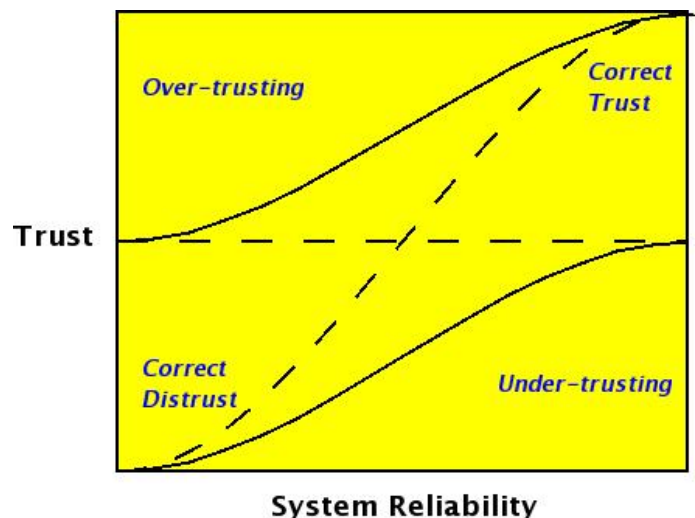


Figure 3. Relationship between trust and reliability.

5 TRUST STRATEGIES

Although many models capture to some extent the notion of trust, there has been little attention in the investigation of the dynamics of trust. In particular, the social aspects of trust and risk perception (Doug-

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.", (EUROCONTROL 2003b).

las and Wildavsky 1982) stress the interaction between trust, risk and knowledge (Gefen et al. 2003). The different relationships (e.g., independence, mediation and moderation) between trust and risk affect emergent behaviors (Gefen et al. 2003). Therefore, these relationships between risk and trust highlight different behaviors. The interaction between trust and risk perception is grounded in the social aspects of technology. Therefore, a social viewpoint provides a convenient intersection between risk, trust and technology. This section articulates the motivations for trust strategies.

5.1 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. This section briefly introduces the Prisoners' Dilemma (Axelrod 1990; Dixit and Nalebuff 1991; Nalebuff and Brandenburger 1996).

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. Figure 4 shows a matrix representation of the prisoners' dilemma. Note that there exists a relationship that specifies the order of the four pay-offs: from best $T > R > P > S$ to worst (Axelrod 1990). Different matrices and different rules identify different characterizations (e.g., symmetric, asymmetric, iterative, etc.) of the prisoners' dilemma.

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 (Axelrod 1990). It is possible to identify different heuristics depending on whether or not *dominant strategies*³ exist (Dixit and Nalebuff

1991). Therefore, the prisoners' dilemma captures those situations that may result in cooperation or competition, i.e., *co-opetition* (Nalebuff and Brandenburger 1996). The prisoners' dilemma captures those situations in which trust emerges as cooperation 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.

| | | Column Player | |
|------------|-----------|--|--|
| | | Cooperate | Defect |
| Row Player | Cooperate | R=3, R=3 Reward for mutual cooperation | S=0, T=5 Sucker's payoff and temptation to defeat |
| | Defect | T=5, S=0 Sucker's payoff and temptation to defeat | P=1, P=1 Punishment for mutual defection |

Figure4. The Prisoners' Dilemma.

Note that social connectivity (MacKenzie 2004) 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 (Riegelsberger et al. 2003). 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 deficiencies (Riegelsberger et al. 2003).

5.2 Trust, risk and knowledge: a game

The characterization of trust and risk (Gefen et al. 2003) suggests that the underlying constructs interact in the formation of trust and the perception of risk. This interaction has origins in the social aspects of trust and risk (Douglas and Wildavsky 1982). Although many models address the understanding of trust and risk, 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 (see, Figure 2), with respect to knowledge and consent, in the case of trust in ATM (see, Figure 3), with respect to system reliability. It is possible to capture the interactions between trust and risk as a trust game, which extends the general prisoners' dilemma. The knowledge of system reliability enables the interaction between the two spaces (i.e., Figure 3 and Figure 2).

³ **Dominant strategy:** one that outperforms all of that side's other strategies, irrespective of the rival's choice (Dixit and Nalebuff 1991).

Figure 5 shows a representation of the game. The game involves two players: *Player 1* (P1) and *Player 2* (P2). The two players have some *common knowledge*⁴ about the system (e.g., system reliability). The two players have different strategies according to their expected pay-offs (or convictions). For instance, P1 (i.e., *complete-certain*) can have complete consent and being certain of the system reliability. That is, P1 trust the common knowledge and expect a similar behavior from P2. This corresponds to *R1* in the pay-offs matrix (see, Figure 5). 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.

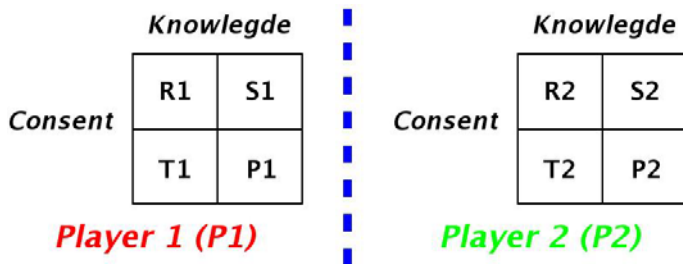


Figure 5. A Trust Game.

Although the two players 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*) (Axelrod 1990). 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 (Axelrod 1990). Each player chooses the preferred strategy independently (that is, without knowing each other strategy). P1 would like to have a dominant strategy such to have correct trust in technology. However, P2 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, P1 exhibits the chosen trust in technology and exhibits relevant evidence (e.g., high reliability or low reliability). P2, 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 trust as well as risk taking strategies. The two players may have different overall objectives or cooperate towards common objectives.

⁴ **Common Knowledge:** Information is common knowledge if it is known to all the players, if each player knows that all players know it, if each player knows that all the players know that all the players know it, and so forth ad infinitum (Rasmusen 1989).

5.3 Trust strategies

This section argues that the proposed trust game allows the characterization of trust in situated (risk) contexts. The game takes into account that risk perception and trust may behave as opponent (or competing) forces, regardless the (system) knowledge (e.g., system reliability). A game play shows whether the two players exhibit cooperative or competing strategies. Once the players have chosen their strategies (i.e., trust or distrust, and certain or uncertain), they both have limited choices for the next move. For instance, if P1 has unconditional trust in technology, whatever the knowledge about it. P1 can only exhibit his 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, P2 may have a contested or complete consent over the knowledge in alternative strategies of certain or uncertain knowledge.

Figure 6 shows a *Value Net* (Nalubeff and Brandenburger 1996) for the ATM domain. The value net represents all the players and the interdependencies among them. Along the vertical dimension of the value net are *customers* and *suppliers* (Nalubeff and Brandenburger 1996). Along the horizontal dimension are *competitors* and *complementors* (Nalubeff and Brandenburger 1996).

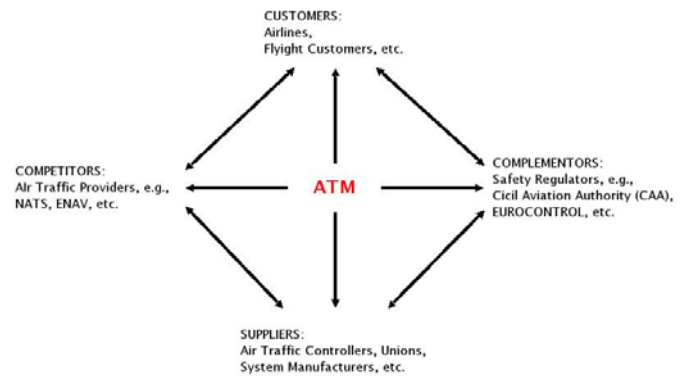


Figure 6. The value net for ATM.

The ATM context provides many examples in which trust and risk may exhibit a competing behavior. 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 trust⁵ (in system evolution) by managing less traffic than planned. This results in economic pressures on the ATM system and

⁵ “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, (EUROCONTROL 1998).

customer dissatisfaction. The Short-Term Conflict Detection (STCD) system provides an instance of accepted technology innovation that may result in mistrust or, worst, unsafe behaviors⁶.

5.4 A matter of knowledge

This section characterizes the trust game in a logical framework for reasoning about knowledge (Fagin et al. 2003). The basics consist of well-established results in modal logic. Modal logic allows the formalization of the intuitions about necessity and possibility. There exist many different representations that describe modal logic. 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. 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.

A Kripke frame consists of two parts: a non-empty set G whose members, generally called possible worlds, and a binary relation R on G generally called the accessibility relation. Thus, a Kripke frame is a pair $\langle G, R \rangle$. Note that although *possible world* is a suggestive terminology, possible worlds are any objects (e.g., numbers, sets or even functions, requirements, set of requirements, etc.) whatsoever in the mathematical treatment of frames. A simple intuitive interpretation considers Kripke frames as graphs. The elements of G are called worlds or points. The accessibility relation R identifies the connections (or edges) in the graph. We will generally use Γ , Δ , etc., to denote possible worlds. If Γ and Δ are in the relation R , we will write $\Gamma R \Delta$, and read this as Δ is *accessible from* Γ (or Δ is an *alternative world* to Γ). A frame is turned into a modal model by specifying which propositional letters are true at which worlds.

The basic framework of modal logic allows the modeling of multi-agents systems (Fagin et al. 2003). 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 φ if φ 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* (Fagin et al. 2003). 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* ” (Fagin et al. 2003). This allows us to capture multi-agents systems or games.

6 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 of 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.

The proposed trust game captures 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. 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. The further formalization of the game in theoretical terms would allow the identification of game conditions. Future work aims to formalize the rules underlying the trust game. Moreover, the instantiation of the trust game in situated context would allow the identification of heuristics.

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. The proposed game highlights that trust plays a crucial role with respect to risk and knowledge in order to achieve overall objectives in the ATM domain. Although the game captures the interaction between trust, risk and knowledge, in practice, it is still challenging the instantiation and construction of the game (e.g., identification of the decision matrix, rules, etc.). However, the paper stresses and justifies future investigations on trust strategies. Moreover, it provides a game-oriented characterization for the analysis of trust strategies. It also highlights a theo-

⁶ “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, (EUROCONTROL 1998).

retical framework for the representation of the trust game.

Future work intends to use the game in order to investigate relationships between different strategies (e.g., adoption of technology innovation, system testing and validation, etc.). This would further support the understanding and generalization of the notion of trust. However, organizations may, already, use and instantiate the game in order to understand and investigate how trust, risk and knowledge interact within their contexts.

REFERENCES

- Abdul-Rahman, A. & Halles, S. 1997. A distributed model of trust. In *Proceedings of the New Security Paradigms Workshop*, pp. 48–60. ACM.
- Anderson, R. 2001. *Security Engineering: A Guide to Build Dependable Distributed Systems*. Wiley Computer Publishing.
- Axelrod, R. 1990. *The Evolution of Cooperation*. Penguin Books.
- Carbone, M., Nielsen, M. & Sassone, V. 2003. 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.
- Dassonville, I., Jolly, D. & Desodt, A.M. 1996. Trust between man and machine in a teleoperation system. *Reliability Engineering & System Safety* 53, 319–325.
- Dixit, A.K. & Nalebuff, B.J. 1991. *Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life*. W.W. Norton & Company.
- Douglas, M. & Wildavsky, A. 1982. *Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers*. University of California Press.
- EUROCONTROL 1998. *Human Factor Module - Human Factors in the Development of Air Traffic Management Systems* (1.0 ed.). EUROCONTROL.
- EUROCONTROL 2001. *EUROCONTROL Safety Regulatory Requirements (ESARR). ESARR 4 - Risk Assessment and Mitigation in ATM* (1.0 ed.). EUROCONTROL.
- EUROCONTROL 2003a. *EUROCONTROL Air Traffic Management Strategy for the years 2000+*. EUROCONTROL
- EUROCONTROL 2003b. *Guidelines for Trust in Future ATM Systems: A Literature Review* (1.0 ed.). EUROCONTROL.
- EUROCONTROL 2004. *EUROCONTROL Air Navigation System Safety Assessment Methodology* (2.0 ed.). EUROCONTROL.
- Fagin, R., Halpern, J.Y., Moses, Y. & Vardi, M.Y. 2003. *Reasoning about Knowledge*. The MIT Press.
- Falcone, R. & Castelfranchi, C. 2001. The socio-cognitive dynamics of trust: Does trust create trust? In R. Falcone, M. Singh, and Y.-H. Tan (Eds.), *Trust in Cyber-societies*, Number 2246 in LNAI, pp. 55–72. Springer-Verlag.
- Felici, M. 2005. Evolutionary safety analysis: Motivations from the air traffic management domain. In R. Winther, B. Gran, and G. Dahll (Eds.), *Proceedings of the 24th International Conference on Computer Safety, Reliability and Security, SAFECOMP 2005*, Number 3688 in LNCS, pp. 208–221. Springer-Verlag.
- Felici, M. 2006. Capturing emerging complex interactions: Safety analysis in air traffic management. *Reliability Engineering & System Safety*.
- Gefen, D., Karahanna, E. & Straub, D.W. 2003. Inexperience and experience with online stores: The importance of tam and trust. *IEEE Transactions on Engineering Management* 50(3), 307–321.
- Gefen, D., Rao, V.S. & Tractinsky, N. 2003. 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.
- Gefen, D. & Straub, D.W. 2004. Consumer trust in b2c e-commerce and the importance of social presence: experiments in e-products and eservices. *Omega: The International Journal of Management Science* 32, 407–424.
- Kasper-Fuehrer, E.C. & Ashkanasy, N.M. 2001. Communicating trustworthiness and building trust in interorganizational virtual organizations. *Journal of Management* 27, 235–254.
- Kramer, R.M. 1999. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology* 50, 569–598.
- Leveson, N. G. 1995. *SAFWARE: System Safety and Computers*. Addison-Wesley.
- Luo, Y. 2002. Building trust in cross-cultural collaborations: Toward a contingency perspective. *Journal of Management* 28(5), 669–694.
- MacKenzie, D. 2004. Social connectivities in global financial markets. *Environment and Planning D: Society and Space* 22, 83–101.
- McKnight, D.H. & Chervany, N.L. 1996. The meanings of trust. Technical Report 96-04, University of Minnesota.
- McKnight, D.H. & Chervany, N.L. 2001a. Conceptualizing trust: A typology and e-commerce customer relationships model. In *Proceedings of the 34th Hawaii International Conference on System Sciences*, pp. 1–9. IEEE.
- McKnight, D.H. & Chervany, N.L. 2001b. Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, and Y.-H. Tan (Eds.), *Trust in Cyber-societies*, Number 2246 in LNAI, pp. 27–54. Springer-Verlag.
- McKnight, D.H., Cummings, L.L. & Chervany, N.L. 1996. Trust formation in new organizational relationships. Technical Report 96-01, University of Minnesota.
- Nalebuff, B.J. & Brandenburger, A.M. 1996. *Co-opetition*. HarperCollinsBusiness.
- Nielsen, M. & Krukow, K. 2003. Towards a formal notion of trust. In *Proceedings of PPDP'03*. ACM.
- Norman, D.A. 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books.
- Pavlou, P.A., Tan, Y.-H. & Gefen, D. 2003. The transitional role of institutional trust in online interorganizational relationships. In *Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS'03)*. IEEE.
- Perrow, C. 1999. *Normal Accidents: Living with High-Risk Technologies*. Princeton University Press.
- Rasmusen, E. 1989. *Games and Information: An Introduction to Game Theory* (Second ed.). Blackwell.
- Riegelsberger, J., Sasse, M.A. & McCarthy, J.D. 2003. The researcher's dilemma: evaluating trust in computer-mediated communication. *International Journal of Human-Computer Studies* 58(6), 759–781.
- Sorensen, J. 2002. Safety culture: a survey of the state-of-the-art. *Reliability Engineering & System Safety* 76, 189–204.
- Storey, N. (1996). *Safety-Critical Computer Systems*. Addison-Wesley.
- Uggirala, A., Gramopadhye, A.K., Melloy, N.J. & Toler, J.E. 2004. Measurement of trust in complex and dynamic systems using a quantitative approach. *International Journal of Industrial Ergonomics* 34(3), 175–186.
- Yu, E. & Liu, L. 2001. Modelling trust for system design using the *ix* strategic actors framework. In R. Falcone, M. Singh, and Y.-H. Tan (Eds.), *Trust in Cyber-societies*, Number 2246 in LNAI, pp. 175–194. Springer-Verlag.