

On Trust and Influence

A Computational Red Teaming Game Theoretic Perspective

Eleni Petraki

Faculty of Arts and Design,
University of Canberra,
Canberra, Australia
Eleni.Petraki@canberra.edu.au

Hussein A. Abbass

School of Engineering and Information Technology
University of New South Wales – Australia
Canberra, Australia
h.abbass@adfa.edu.au

Abstract— The concept of trust has attracted the attention of many researchers over the years who studied the impact of trust in many domains. Trust is a ubiquitous concept. It is pervasive in every aspect of our life, from interpersonal relationships to national defence and security applications. However, despite the vast literature on trust, we are not close enough to mastering the dynamics of trust. One reason is that if we define procedural steps for trust, we simultaneously define steps for deception; thus, we simply define a vacuous cycle. Another reason is that, the dynamics of trust change as the world changes. But how can we then study trust?

This paper connects the interdisciplinary literature to synthesize a Computational Red Teaming (CRT) based model of trust that defines opportunities whereby computational intelligence techniques, more specifically, evolutionary game theory researchers, can contribute to this vastly growing research area. We offer a position on the topic by reviewing games for trust and introduce a new theoretic game to study influence and transfer of trust.

Keywords—Trust, Evolutionary Game Theory, Computational Red Teaming

I. INTRODUCTION

Trust is a ubiquitous concept. It underpins every decision made during the reciprocal interaction of two agents¹. Trust impacts all domains relying on interactions, from marketing as for example in studying customer loyalty [8] to relationships in organizations [43], where absence of trust is found to deteriorate the overall performance of an organization [23]. The element of trust seems to be a sought after characteristic in future employees of an organisation, managers, medical practitioners, nurses, teachers and many other professionals as it is considered to be the basis for forming healthy relationships, improving the quality and performance of the organisation/institution.

The literature is abundant with studies on trust, from what it is, its roles in social and human systems [12],[13],[32] and the ethics of trust and antitrust [7], to what the implications of

trust/mistrust are [14],[19],[23], and how we can influence the shaping of trust in a human society [10],[33]. This literature is driven by four main questions:

1. What indicators should we be using to judge on whether trust is present or absent in a situation?
2. How to trace back these indicators to causes? Can we design an Atlas of Causes of Trust?
3. Can we predict if trust or mistrust in a situation will continue to operate, therefore, the indicators will remain stable, in the future? What is the maximum look-ahead time, and in what context (reliance), to obtain a prediction with an acceptable level of confidence?
4. If we have at least correctly discovered partial causes of trust and designed a relatively adequate model to predict trust (i.e. stability of indicators over a fixed time horizon), can we manipulate the causes to change the effects (control)?

Interestingly, the vast literature on trust attempted to answer all of these questions directly and indirectly. There have been many answers which are specific to a number of contexts, thus have limited applicability. Added to that, the concept itself is evolving. Globalization, the internet, the move to massive online courses and availability of information, etc are changing the world and with that, trust itself as a concept is changing.

Despite that the field of computational intelligence can offer a great deal of insight and practical tools to the literature on trust, limited research have been done in this area. In the wider computer science and security literature, there is a vast amount of publications on trust in areas such as computer networks, information assurance and databases. However, most of this work deploys the concept for use within the computer science literature without contributing per se to the epistemology of trust.

The rest of this paper is organized in three main sections. Section II presents a literature review on trust from the humanities and social sciences perspectives to visit the roots of the concept. Section III presents a perspective on trust from a Computational Red Teaming point of view, while Section IV presents opportunities for the field of Computational Intelligence in the area of trust.

¹ A person can be natural - a human - or a juridical – as a corporation. We will reserve the word “agent” to mean both a person and a software that performs some tasks by producing actions. We will use the word ‘entity’ to refer to agents that think/compute and act, and objects that do not think or act.

II. ON TRUST

A. *The What ... Defining Trust*

Our discussion will begin by reviewing the definitions of trust in different fields so as to be able to dissect its constituents and synthesise a clear picture of the concept. Each of these definitions reflect a particular social, political, or psychological perspective, which will be examined in detail in the following sections. One of the first attempts at defining the concepts includes “willingness to be vulnerable to another based on the expectation of favourable outcomes for the trusting party” [33]. Mayer et al go even further to mention that trust takes place when the trustee displays specific characteristics. These include the trustee’s ability to carry out a given task, benevolence toward the truster and integrity such as fairness and honesty. Robinson and Morrison define trusting someone as having ‘expectations, assumptions or beliefs about the likelihood that another’s future actions would be beneficial, favourable or at least not detrimental to one’s interests’ [36] (p. 238). The above definitions place emphasis on the concept of trust as a psychological state involving willingness and positive attitude towards another party’s future favourable actions. Kipnis [24] suggests that trust “introduces unwanted uncertainty into our lives”. It means that other people control outcomes that we value. It gives people “power over us” (p. 40). We can conclude from the above that trust can be a powerful and political issue in human relationships, placing both truster and trustee in unwanted positions. This concept of uncertainty resembles the notion of ‘risk’ identified in the definition of trust by some other scholars; trust is likened to a ‘risk-taking act’ [37][28].

It is apparent that the notion of trust integrates a complex array of feelings/emotions/attitudes/beliefs combined with elements of control, risk and power. This complexity in the definition led to research to identify causes or contexts of expressions of trust and distrust. This is explained in Kim who argues that interpersonal trust is based upon listeners’ perceptions of a speaker’s expertness, reliability, intentions, activeness, personal attractiveness, and the majority opinion of the listener’s associates [24].

However, the literature on trust has not matched the definitions of distrust. Distrust and/or mistrust have been initially theorised as antithetical to trust. Only recently has there been an effort to distil and contrast the two concepts of trust and distrust. Lewicki et al [28] have argued that these are independent constructs with a different set of expectations and can co-exist. For example, someone can trust other persons for a particular purpose but also hold feelings of distrust in that person. Thus in Lewicki’s work, trust and distrust entail different expectations, and occupy elements in a continuum. He specifically hypothesises 4 different relationships of trust mistrust, low trust/low distrust, high trust/low distrust, low trust/high distrust, high trust/high distrust. These definitions have not been tested or researched until very recently, when the research in the area has seen an upsurge.

Saunders et al [38] provide evidence on the expectations and categorisations in definitions of trust and distrust. They suggest that trust has cognitive and affective elements in its definitions and they are two different concepts which can be

present. They conclude that trust is a multifaceted phenomenon which is highly dependent on the context and the task, thus more research is necessary.

B. *The Why ... Trust in Society*

Revisiting the roots, social scientists see trust as the backbone that glues a social system and acts as a complexity reduction/management mechanism [32]. Without trust, the social system will be too complex for an individual to live and integrate. Trust, therefore, acts to balance the massive complexity of a social system and the limited cognition of the individuals. Researchers such as Luhmann [30] believe that trust exists to reduce complexity. For individuals to adapt, they need to use trust as a mechanism to reduce complexity to allow for relationships to take place. Consequently, trust has a positive effect on social grouping because it offers opportunities for experiences, actions, and relationships to be established. Notwithstanding, these opportunities will increase the complexity of the social system, and consequently it will reinforce the need for trust, creating a self-reinforcement or positive-feedback cycle. Luhmann sees trust as a facilitator for adaptation to occur in a social system, and therefore trust achieves in a social system the equivalent of adaptation in biological systems [20].

C. *The How ... The Working of Trust in the Person*

Behavioural psychologists see Trust from an individual (the ‘person’) perspective. The psychological school emphasizes the human person with its behavioural traits as in the case of Deutsch [11] who sets constraints on trust in situations where a person is faced with a path with ambiguity. The definition of trust presented in work as Deutch [11],[13] is consistent with the design of most behavioural psychology experiments on trust. A person perceives a situation will lead to two events, one she perceives to have negative valency that is greater than the positive valency she perceives to be associated with the second. However, which event will occur is reliant on a second person. If the first person chooses this path, she is said to trust the second; otherwise she distrusts the second person.

D. *Causes ... On Trust and the Organisation*

Research on the causes and impact of trust in organisations has seen huge growth. A number of studies have identified a link between trust in management and increase in job satisfaction, employee motivation and overall job performance [23]. Whitener, Brodt, Korsgaard, & Werner found that trust in the workplace is linked to team cooperation, performance, and quality of communication in organizations. Spector’s and Jones’ study [41] provided evidence on the fact that individuals with a high trusting stance are more likely to have a higher initial trust for co-workers than are individuals with a low trusting stance. Male’s initial trust level was higher for a new male team member and lower for a new female team member, while for female respondents there was no such similar trend. A similar research project employed a survey of managers and subordinates who were asked to rate a person they trust in the company and provide a rationale [43]. They recruited participants from a range of US states and companies, of mixed gender and mixed cultural backgrounds. Participants were not

found to be affected by race in their expression of trust, however the study unveiled gender differences in the perceptions of trust. Specifically, female employees showed a high preference for female managers due to their personal approach, while there were wide differences in the male and female explanations of trust and distrust.

Due to the inextricable link between trust and workplace performance and group cohesiveness, more research is mandatory to understand the constraints and limitations of trust. More recently, trust has been studied in its relationship with overall institutional well being and employee happiness [20]. It has been suggested that the more the employees trust their managers, the more satisfied they are with their jobs and this leads to their general good health and wellbeing.

E. Influences ... On Trust and Language

Understanding how to influence and reshape trust is a critical skill to promote trust in a social system. Despite the immense literature on trust in sociology, psychology and organisation behaviour, limited research exists on the use of linguistics strategies in conveying trust [27]. One of the very few existing articles examined the use of linguistic strategies in spam emails including requests for money transfers which employ strategies that aim to create rapport and trust and which claim a large number of victims every year despite their widespread use [38]. The second study focused on the effect of managers' use of linguistic politeness strategies in perceptions of trust [27]. Based on a perception survey of 115 students who received emails from a range of team leaders, Lam unveiled that "participants trusted leaders who used linguistic politeness strategies in their emails, as opposed to those who failed to include mitigating strategies". While this study was done on university students, Lam suggested further follow up investigation in workplace contexts with a range of different contexts and purposes. Moreover, strategies on the influential nature of trust have revealed that issuing of superfluous apologies can be effective in promoting people sense of trust towards the apologist [10].

F. A Critical Perspective on Trust

Despite the abundance of studies on trust demonstrated with the sampled literature summarized above, the area suffers from significant fundamental gaps. From a sociological research perspective, the literature is unbalanced when we compare the massive number of studies in the western world and the data that have been collected in these studies, to the less researched societies in Asia, Africa and Latin America. More importantly, with globalization and the increased level of connectivity in the world of today, multi-cultural interactions continuously entail and create new strategies which can reshape what trust is and is not.

From a perspective on management and organisational behaviour, it is evident that there is absence of research on conceptualisations of trust and mistrust, the affective and cognitive dimensions and how they characterise human understanding. Experimental research is necessary to provide support for the hypothesised categories and classifications of trust and distrust and for illuminating the intricate relationship and co-existence of such concepts.

From a psychological research perspective, the same factors discussed above impacting the social system can impact the individual. Within behavioural and organizational psychology research, despite the diversity of articles available on the topic, most research concludes with the need for additional studies into the topic, especially on the study of authentic linguistic data, and the analysis of linguistic strategies in naturally occurring interactions that evoke trust.

With the gaps identified above, we can clearly see a larger gap; the lack of linguistic and computational environments to support studies on trust. With the multi-dimensional nature of the topic, computational environments that can synthesize and blend the multidimensional perspectives on trust would offer a comprehensive picture that is much greater than the sum of its parts.

In neuroscience, there is definitely lack of research on trust. The research is sparse compared to the previous fields we discussed above, and even compared to classical studies on attention and memory. Research on trust in neuroscience is nowhere as mature as these classical fields. Only recently some researchers started to uncover new results in this domain [31].

III. RED TEAMING AND TRUST

A. Computational Red Teaming

Red Teaming (RT) is an ancient military concept of playing the devil's advocate to assess one's own plans, concepts, strategies and ideas. Computational Red Teaming (CRT) [1] attempts to transform the red teaming exercise into systemic and computable steps that guide the design of synthetic red teaming environments in a disciplined manner.

CRT relies on two fundamental concepts, the concept of risk and the concept of challenge. Risk is the impact of uncertainty on objectives. This impact can be positive for some objectives while negative for others. As such, decision making in CRT uses risk thinking to understand and make decisions based on risk trade-offs. The concept of a challenge represents a proactive approach by one entity to compete and/or cooperate during a reciprocal interaction with another entity with the objective of the former entity to influence the latter by pushing its behaviour beyond its normal operating envelope.

The most recent use of CRT was in an augmented cognition experiment linking the air traffic controller brain with the air traffic control environment [2], [3]. The CRT system was successful in making risk trade-off decisions and influencing the controller and the environment to manage cognitive and traffic complexity.

To understand the relationship and role of trust within CRT, we will discuss the relationship between trust, risk and perception first. It will then become apparent why trust is essential within CRT.

B. Perception and Risk

Perception and risk are two inter-related concepts that greatly influence trusting decisions. Trust is inherently based on a judgement assessed through the subjective probability that a second person is trustworthy [14]. The two words “subjective” and “probability” represent the links between trust on the one hand, and perception and risk on the other. Deutsch [13] sees the role perception plays in trust with his use of the word “perceive” in his discussions on trust. The subjectivity element connects trust and reliance; the same trustee and truster in different situations may behave differently; either because the context is different or their perception of the context changed. Deutsch asserts that different players would perceive different valencies, and even the same player would perceive different valencies in different situations (reliance). These changes impact the subjective probability associated with the trusting decision; thus, the risk of trusting changes and the decision changes accordingly. This tied coupling between risk and trust can be seen in almost all research on trust [9],[14],[15].

C. Challenging Trust

The thinking process that a person goes through it to make a trusting decision is in effect a red teaming exercise, where both the thesis (blue) and antithesis (red) attempts to compete head to head to evaluate the risk associated with a trusting decision.

CRT relies on computational intelligence techniques in its search for challenges, risk evaluation, explaining the dynamics of the interaction between red team and blue team, and mining for behavioural patterns in blue team’s data to model and exploit the blue team. Modelling trust from computational red teaming perspective offers a great deal of opportunities for researchers in the field of computational intelligence to create new areas of research.

D. Influencing Trust

Less work has been done in the literature on strategies to influence and transfer trust. Most studies in the social sciences and psychology literature would distil the factors that influenced trust from participants. However, understanding how these factors interact and form trust or mistrust requires more theoretical studies in game theory to support those social and psychological theories.

CRT can provide insight into influences of trust and how trust gets transferred among agents. In this paper, we will present a game to model influence, while in our future work we will show how this game can be played using a CRT.

IV. COMPUTATIONAL INTELLIGENCE FOR TRUST

Interdisciplinary studies have the potential to synthesise the diverse literature on trust into a new form that can reduce the gap between the level of abstraction used in each approach in isolation and how trust is used in-situ in real world contexts. Interdisciplinary studies started to appear in areas such as neuroeconomics and social cognitive neuroscience, where social and group-level interactions are blended with the

behavioural and cognitive level traits. For example, we can use multi-agent systems – classically used for social and group modelling - to study a psychological phenomenon (see our work [30] where we adopted Minsky Society of Mind [35] to model the behaviour of drivers, in which we modelled a person’s behaviour as a society of thinking agents). Similarly, a social phenomenon can be studied as a series of repeated games – classically used for behavioural modelling - on a

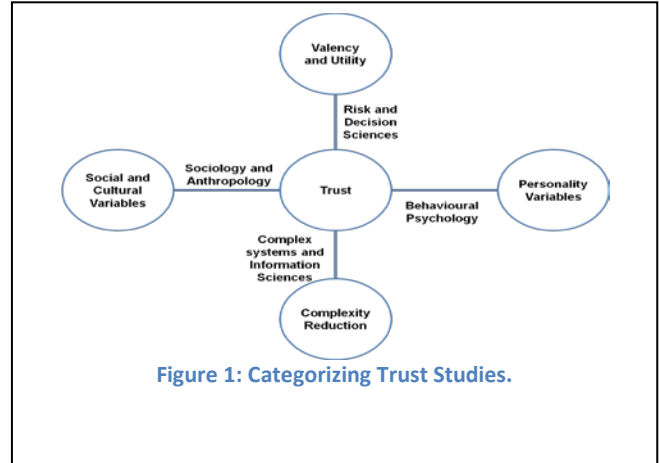


Figure 1: Categorizing Trust Studies.

network using evolutionary game theory (see for example our work on evolutionary game theory to study risk in an interdependent security system [40]).

Figure 1 distils the key points in the above brief discussion to service an important purpose for this paper. The x-axis represents the two most common schools for trust (psychological and the sociological schools), while the y-axis represents the corresponding uses of trust (rational justification in ambiguous situations and complexity reduction). The figure demonstrates the links between decision and computational sciences on the one hand, and the psychological and sociological sciences on the other. Game theory, multi-criteria decision making, and classical risk analysis are examples of computational tools used to address trust to work in harmony with the psychological research, while network theory, complexity, and multi-agent systems are the computational tools used to address trust from a sociological perspective.

A. Games and Trust

Within the area of trust, interdisciplinary research created Trusting games that have been used in neuroeconomics experiments [5], [11], [24].

A trusting decision, however, differs from the classical prisoner dilemma in a fundamental characteristic; while decisions in the prisoner dilemma are concurrent, trusting decisions are sequential. One person needs to make a decision whether they trust another or not. Based on this trusting decision, the other person needs to make a subsequent decision, which carries feedback about their trustworthiness.

Utility matrix of the trust game used in [26]

Player 1	Trust		Do not Trust
Player 2	Reciprocate	Defect	5,5
Utility	10,15	0,25	

Another trust game is presented in [5], whereby there are two players. The first player plays the role of an investor. The second player plays the role of the trustee. The game starts with both players having an endowment of 12 points. Each point can be equated to some real money in the real world.

Since a trust game is sequential, the investor makes the first move, whereby the investor transfers any amount of money, X , between 0 and 12 to the trustee. The money is tripled on the way to the trustee; thus the trustee receives $3X$ points. The trustee is informed about this decision. The trustee is then given the chance to transfer back to the investor any amount of money, Y , between 0 and $12+3X$. Once this transfer is made, the game terminates. At the conclusion of the game, each player will have the following balance

	Balance
Player 1	$12 - X + Y$
Player 2	$12 + 3X - Y$

When $Y = 2X$, both players end up with equal amount of money. If the investor decides to transfer all their endowment of 12 points, and the trustee returns $2X$, each player will end up with 24 points (doubling the amount of point they originally started with). If the trustee is not trustworthy and does not return any money to the investor, the trustee will end up with 48 points and the investor bankrupts. On the other hand, if the investor does not trust the trustee and does not transfer any money, both players end up with the same endowment they started with of 12 points. Therefore, X is a measure of trust and Y is a measure of trustworthiness. The relationship between the two is shown in Figure 2.

We have previously looked at the impact of reputation on the dynamics of the prisoner dilemma when played on a network and found that positive reputations can promote altruism. This work was then continued to study the dynamics of formation of alliances [17], impact of information sharing [18], and limited cognition [16] of the agents on the game. We also used game environments to understand the impact of uncertainty and deception on human decision making [41], and conducted research on the impact of perception of information in decision making [22].

All of this previous work was done in the context of classical games. This work can be extended to include the trust games. In general, there is a lack of work to understand the evolutionary dynamics of the trust game in the computational intelligence field to answer research questions such as: How will evolutionary dynamics change because of the sequential nature in the decision making process of the trust game? If we revisit the two games above, what are the necessary conditions for trust to be an evolutionary stable strategy?

B. Influence Game

In the classical prisoner dilemma, the repeated version does not change the nature of the game; mainly because the concurrent nature of how decisions are made. However, in the trust game, the sequential nature of the game and the fixed order of how the truster plays first then the trustee plays second offer a new opportunity to study influence. The primary

question of this new game that we are proposing is, can the truster influence the trustee by transferring trust to increase the trustee trustworthiness value? In other words, can the truster, through a strategy to play the game, change a trustee that is not reliable to become reliable? If so, what are those influence strategies that should be employed?

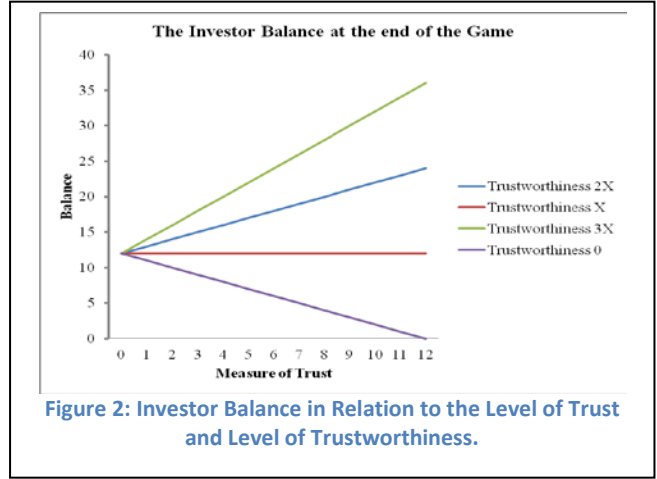


Figure 2: Investor Balance in Relation to the Level of Trust and Level of Trustworthiness.

To enable the study of influence, we will propose below a repeated game by modifying the second trust game presented in the previous section. The Influence Game is as follows:

Two players (Investor and Trustee) have a total budget of $12N$ points each, where N is the number of training rounds. In each training round, t , each player has a budget of 12 points. The investor makes the first move in each of the N rounds of the game. The investor transfers any amount of money, $X(t)$, between 0 and 12 to the trustee. The money is tripled on the way to the trustee; thus the trustee receives $3X(t)$ points. The trustee is informed about this decision. The trustee is then given the chance to transfer back to the investor any amount of money, $Y(t)$, between 0 and $12+3X(t)$. Once this transfer is made, the round terminates and a new round starts till all N training rounds have been played. At the end of the N training rounds, all players lose their money. They are given an endowment of $12N$ points each and are allowed to play a single testing round only. What value for $X(N+1)$ and $Y(N+1)$ will we get?

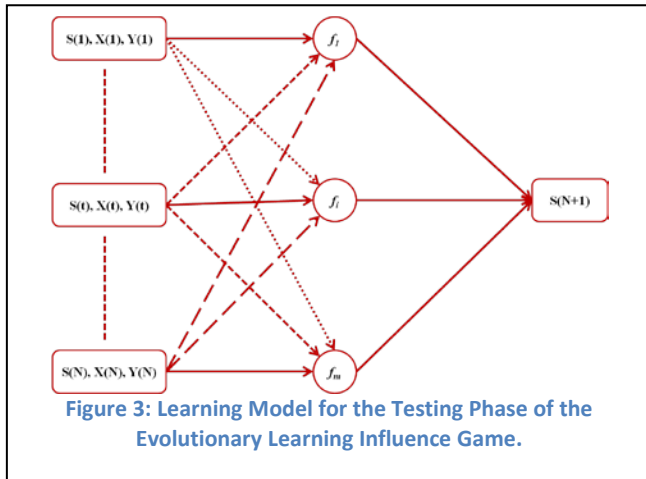
If the investor established complete trust with the trustee over rounds, the investor should be in a position to transfer as much as $12N$ points to the trustee. We would expect that this is correlated with N up to some point; as the number of interactions between the two players increase, they will either establish trust or mistrust.

Meanwhile, the investor has the option to use the first N rounds to establish trust with the trustee, even if the trustee was not trustworthy. The question remains, can the investor influence the trustee to become trustworthy?

C. Evolutionary Influence Learning Game

For evolutionary game theory researchers, it is obvious that a classical representation of a strategy that covers the $N+1$ rounds can be misleading. Moreover, the trust game is a sequential game. The sequential nature of the game makes it

more suitable to study its dynamic through learning models that can either be combined with evolutionary models or other learning models during the training phase.



A strategy representation for this influence game can take the form of a neural network architecture as shown in Figure 3. The input to the network is the measure of trust, trustworthiness, and the strategy followed at each round of the training phase. The output is the strategy and/or action in the testing round. The intermediate transfer functions transform these strategies into the final strategy that will be used during the testing round.

V. CONCLUSION

In this paper, we presented a multi-disciplinary review of work on trust from social science, behavioural psychology, organizational psychology, language and communication, and neuroscience perspectives. A computational red teaming perspective on this literature is presented, followed by a discussion of trust games that have been presented in the literature. While there is significant work on trust including proposed game theoretic approaches to study the concept, the dynamics of trust necessitate an understanding of how trust is reinforced in a society and the effective tools and strategies for transferring trust. The proposed influence game addresses this gap. These games open new opportunities for researchers in computational intelligence and games to understand and design strategies for trust and influence.

ACKNOWLEDGMENT

This work has been funded by the Australian Research Council (ARC) discovery grant number, DP140102590: Challenging systems to discover vulnerabilities using computational red teaming.

REFERENCES

[1] Abbass H.A. (2015) Computational Red Teaming: Risk Analytics of Big-Data-to-Decisions Intelligent Systems, Springer, in print.
 [2] Abbass H.A., Tang J., Amin R., Ellejmi M., Kirby S. (2014) The computational air traffic control brain: Computational red teaming and

big data for real-time seamless brain-traffic integration, Journal of Air Traffic Control 56 (2).
 [3] Abbass H.A., Tang J., Amin R., Ellejmi M., Kirby S. (2014) Augmented cognition using real-time EEG-based adaptive strategies for air traffic control, in: International Annual Meeting of the Human Factors and Ergonomic Society, HFES, SAGE.
 [4] Abbass H.A. and Petraki E. (2011) The Causes for No Causation: A Computational Perspective, Information Knowledge Systems Management, IOS Press, Information • Knowledge • Systems Management, 10 (1-4), 51-74.
 [5] Adolphs, R. (2002). Trust in the brain. Nature neuroscience, 5(3).
 [6] Axelrod, Robert. 1984. The Evolution of Cooperation. New York: Basic Books.
 [7] Baier, A. (1986). Trust and antitrust. *Ethics*, 231-260.
 [8] Ball, D., Coelho, P. S., & Machás, A. (2004). The role of communication and trust in explaining customer loyalty: an extension to the ECSI model. *European Journal of Marketing*, 38(9/10), 1272-1293.
 [9] Barber, B. (1983). *The logic and limits of trust* (Vol. 96). New Brunswick, NJ: Rutgers University Press.
 [10] Brooks, A. W., Dai, H., & Schweitzer, M. E. (2014). I'm Sorry About the Rain! Superfluous Apologies Demonstrate Empathic Concern and Increase Trust. *Social Psychological and Personality Science*, 5(4):467-474.
 [11] Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature neuroscience*, 8(11), 1611-1618.
 [12] Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In: Jones, M. R. (ed), *Nebraska Symposium on Motivation*. Nebraska University Press.
 [13] Deutsch, M. (1977). *The resolution of conflict: Constructive and destructive processes*. Yale University Press.
 [14] Gambetta, D. (2000). *Can We Trust Trust? Trust: Making and Breaking Cooperative Relations*, electronic edition, Department of Sociology, University of Oxford, 213-237.
 [15] Garfinkel, H. (2010). A Conception of and Experiments With "Trust" as a Condition of Concerted Stable Actions. *The Production of Reality: Essays and Readings on Social Interaction*, 379.
 [16] Ghoneim A., Greenwood G. and Abbass H.A. (2013) Distributing cognitive resources in one-against-many strategy game, IEEE Congress on Evolutionary Computation, Cancun, Mexico.
 [17] Ghoneim A., Abbass H.A. and Barlow M. (2007) Investigating Alliance Dynamics Using a Co-evolutionary Iterated Prisoner's Dilemma with an Exit Option, IEEE Congress on Evolutionary Computation (IEEE-CEC), Singapore, pp 4065-4072.
 [18] Ghoneim A., Abbass H.A. and Barlow M. (2007) Information Sharing in the Iterated Prisoner Dilemma Game, IEEE Symposium on Computational Intelligence in Games (CIG), Honolulu, 1-4 April.
 [19] Grodzinsky, F. S., Miller, K. W., & Wolf, M. J. (2014). Developing Automated Deceptions and the Impact on Trust. *Philosophy & Technology*, 1-15.
 [20] Helliwell, J.F., and Huang, Haifang (2011). Well-Being and Trust in the Workplace. *Journal of Happiness Studies*, 12:747-767.
 [21] Holland, John H. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
 [22] Iorio, A. W., Abbass, H. A., Gaidow, S., & Bender, A. (2014). Visualizing Cognitive Moves for Assessing Information Perception Biases in Decision Making. arXiv preprint arXiv:1401.7193.
 [23] Karl, K. A. (2000). Trust and betrayal in the workplace: Building effective relationships in your organization. *Academy of Management Executive*, 14, 133-135.
 [24] Kim, G. (1967). The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin*, vol 68(2), 104-120.
 [25] Kipnis, D. (1996). Trust and technology. In R.M. Kramer & T.R. Tyler (eds.). *Trust in organizations. Frontiers of theory and research*. CA: Sage Publications.

- [26] Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., ... & Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences*, 104(50), 20084-20089.
- [27] Lam, C. (2011). Linguistic politeness in student-team emails: Its impact on trust between leaders and members. *Professional Communication, IEEE Transactions on*, 54(4), 360-375.
- [28] Lewicki RJ, McAllister DJ and Bies RJ (1998) Trust and distrust: New relationships and realities.
- [29] *Academy of Management Review* 23(3): 438-458.
- [30] Leu G., Curtis N., and Abbass H.A. (2012) Modeling and Evolving Human Behaviors and Emotions in Road Traffic Networks, *Procedia Social and Behavioral Sciences*, Elsevier, Vol 54, 999-1009
- [31] Long, Y., Jiang, X., & Zhou, X. (2012). To believe or not to believe: trust choice modulates brain responses in outcome evaluation. *Neuroscience*, 200, 50-58.
- [32] Luhmann, N. Trust and power. 1979. *John Willey & Sons*.
- [33] Mayer, R., J. Davis, and F. Schoorman, "An integrative model of organizational trust," *Acad. Manage. Rev.*, vol. 20, no. 3, pp. 709-734, 1995.
- [34] Miller, C. A. (2014, March). Trust and Delegation: Achieving Robust Interactions with Subordinates. In 2014 AAAI Spring Symposium Series.
- [35] Minsky, M. (1987, April). The society of mind. In *The Personalist Forum* (pp. 19-32).
- [36] Morrison, E., and Robinson, S. (1997). 'When Employees Feel Betrayed: A Model of Psychological Contract Violation Develops,' *Academy of Management Review*, 22, 226-256.
- [37] Rousseau DM, Sitkin SB, Burt RS and Carmerer C (1998) Not so different after all: A crossdiscipline view of trust. *Academy of Management Review* 23(3): 393-404.
- [38] Saunders, M. N.K., Dietz, G. and Thornhill, A. (2014). Trust and distrust: Polar opposites, or independent but co-existing? *Human Relations*, 67(6): 639, DOI: 10.1177/0018726713500831
- [39] Schaffer, D. (2012). The language of scam spams: linguistic features of "Nigerian fraud" e-mails. *ETC: A Review of General Semantics*, 69(2), 157-179.
- [40] Shafi K., Bender A., Zhong W., and Abbass H.A. (2012) Spatio-Temporal Dynamics of Security Investments in an Interdependent Risk Environment, *Physica A: Statistical Mechanics and its Applications*, 391(20), 5004-5017.
- [41] Spector, M.D. and Jones, G.E. (2004). Trust in the Workplace: Factors Affecting Trust Formation Between Team Members. *The Journal of Social Psychology*, 2004, 144(3), 311-321.
- [42] Wang S., Shafi K., Lokan C., and Abbass H.A. (2013) An agent based model to simulate and analyse behaviour under noisy and deceptive information, *Adaptive Behaviour*, Sage, 21(2), 96-117, 2013.
- [43] Wells, C. (2000) Trust, Gender, and Race in the Workplace. *Journal of Social Behavior and Personality*, 2000, Vol. 16, No. 1, 115-126.
- [44] Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. (1998). Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *Academy of Management Review*, 23, 513-530.