NIRSA Working Paper Series
No 40 – September 2008

NIRSA
National Institute For Regional and Spatial Analysis
An Institiúid Náisiúnta um Anailís Réigiúnach agus Spásúil

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Conceptualising Trust: A Literature Review

Stefano De Paoli
Aphra Kerr

John Hume Building, National University of Ireland, Maynooth,
Maynooth, Co Kildare, Ireland.
Áras John Hume, Ollscoil na hÉireann, Má Nuad,
Má Nuad, Co Chill Dara, Éire.
Tel: + 353 (0) 1 708 3350 Fax: + 353 (0) 1 7086456
Email: nirsa@nuim.ie Web: http://www.nuim.ie/nirsa

HEA
Higher Education Authority
An tÚdarás um Ard-Oideachas
Funded under the
Programme for Research
in Third Level Institutions (PRTLI),
administered by the HEA

NDP
NATIONAL DEVELOPMENT PLAN

CONCEPTUALISING TRUST : A LITERATURE REVIEW

By Stefano De Paoli and Aphra Kerr
Department of Sociology and NIRSA
National University of Ireland, Maynooth
Stefano.DePaoli@nuim.ie ; Aphra.Kerr@nuim.ie
June-August 2008.

Much of the current research on the internet that focuses on security and trust issues proposes that we can construct technological and legal solutions to increase trust and improve the user experience in online environments. We argue in this paper that these discussions and the solutions proposed may have serious implications for our online experiences, for user rights and indeed for user driven innovation. Further, they tend to ignore a whole range of user practices which threaten to undermine our trust and use of the internet more than hackers, cheaters, lurkers, spam and illegal content (Nissenbaum, 2001). Such practices include the behaviour of commercial operators who routinely and implicitly datamine their users for commercial purposes, track user behaviour for copyright and IP infringements, filter out 'dangerous' information or, in the case of some public sector bodies, 'misplace' whole databases of non-encrypted information on citizens. As Ireland becomes a post-construction economy it is apparent that only certain user practices get socially constructed as 'harmful' and 'risky' in the knowledge society while others do not.

In this paper we lay some theoretical foundations for examining governance, users and trust in online environments. Trust has been a central concern in the social sciences since, at least, the pioneering work of Georg Simmel's "The Philosophy of Money", in which the author described trust as fundamental for the integration of the society. Later sociologists – like Niklas Luhmann and Anthony Giddens -, have attended to the problem of trust, clearly relating the concept to the issue of "risk in modern societies". The goal of this paper is not to provide an exhaustive survey of

the sociological literature in this area but rather to begin to explore different disciplinary conceptions of trust. The PRTLI funded project from which this paper has emerged is an interdisciplinary team of mathematicians, computer scientists, engineers, designers and sociologists and it emerged early on that people had different understandings of the key challenges facing the internet, particularly as they related to security and trust. Of particular concern to the authors is the extent to which we can delegate security maintenance and governance to technology, what decisions get coded into these solutions and what the implications might be for end users.

This paper is then a first contribution to building bridges between different disciplines and a contribution to the project's mapping of key social and economic issues surrounding the future development of the internet. In this paper we compare different notions of trust: the sociological, the definition existing in automation and the definition of trust developed in early computer science (CS) literature. Our goal is to outline similarities and differences in these notions, as a starting point for a forthcoming review on the contemporary conception of trust in computer science. For this reason, at the end of the paper, we identify some interpretative concepts – in the light of the analysis provided in this manuscript - that we will use in our future work. We begin however by outlining the importance of trust as an issue in relation to the internet and the knowledge society in post-celtic tiger Ireland.

2.THE INTERNET AND THE KNOWLEDGE SOCIETY IN POST-CELTIC TIGER IRELAND.

The Internet grew out of the libertarian counter culture of the 1960s and of course the military-academic research complex in the United States. It outgrew proprietary networks in the 1980s and 1990s in large part due to its openness, or what Zittrain calls its generativity, i.e. the fact that anyone can write an application to operate over the internet (Zittrain, 2008). It became, along with the personal computer, the defining technology of the information or knowledge society and increasingly a key driver in the growth of the American and subsequently the Irish ICT and software industries and economies in the 1990s. Metaphors like the 'information

superhighway', the 'world wide web' and 'surfing' were used to domesticate an otherwise rather unattractive technology of networks, switches and software programmes.

It is over ten years since Ireland published a strategy and vision of its information society (ISSC, 1996). The strategy focussed on five key areas: awareness, infrastructure, learning, enterprise and government. In reality the focus was on developing a high speed telecommunications infrastructure which schools, companies, public institutions and individuals could 'exploit' to generate, share and access knowledge. Bill Melody noted a year later that the focus in this report was on the supply side, the networks and the hardware and software suppliers rather than on the demand side (Melody, 1997). Indeed Ireland's development of a broadband information infrastructure has been slow during the period now referred to as the Celtic Tiger years and by 2008 just over thirty percent of households in Ireland had a broadband connection to the internet with two thirds of this growth occurring in 2006/07. The EU27 average was 77 percent. Much of this growth has occurred in the Eastern part of the country and regional variations in access still exist (CSO, 2008). Meanwhile the political impetus behind the information society vision has dissipated. The annual surveys are no more and the Information Society Commission expired at the end of 2004[1]. Before it disappeared the ISC itself was commissioning reports on the knowledge society and while access and inclusion remained important issues, a new interest emerged and included data protection, unsolicited e-mail, network security and recommended a study of cyber-crime in Ireland (ISC, 2002). Building the network has clearly given way to concerns with what people are doing on that network.

An initial examination of Irish newspapers found regular stories about computer and internet security and calls for more control over the use and handling of information (i.e. the demand for more Trust). For example, recently one of the biggest cyber-frauds occurred, when a hacker was able to obtain the personal details of every single customer that has booked into one of Best Western's European hotels since

1 http://www.isc.ie/

2007[2]. It is estimated that the identities (including financial details) of 8 millions people have been stolen. On a more local level, issues related to computer security have recently affected government and public bodies,. For example, the banking details and Personal Public Service numbers of at least 2,000 employees from seven different public bodies which were stored on laptop computers were stolen in recent years[3]. The Comptroller Auditor General's (CAG) office, responsible for the stolen laptop has, hence, announced its intention to start a process for improving data security procedures[4]. Similar cases, of stolen laptop from public and governmental bodies, urge the Government to beef up its computer security systems[5] . Another recent example is related to public website attacks. During May 2008 more than a hundred Irish websites fell victim to hacker attacks, including one for the Irish presidency of the EU[6]. Even if the Department of Foreign affairs confirmed that there was no loss of sensitive data from its website, it is clear that the delay in reporting such events to users and their frequency is likely to contribute to a demand for more computer security and reliable data protection solutions. Another consequence of this is that words like firewall, password, privacy, identity theft, malicious hackers, computer worms, cyber crime and so on, are widely becoming part of our common vocabulary. But it is not clear what these terms mean for citizens.

Not all of the issues related to cybersecurity are related to hackers and poor security procedures. The recent case at the European Court of Human Rights taken by the

---

2 See http://www.sundayherald.com/news/heraldnews/display.var.2432225.0.0.php

3 See http://www.irishtimes.com/newspaper/ireland/2008/0812/1218477342200.html

4 See CAG Press Release http://audgen.gov.ie/viewdoc.asp?DocID=1106 and http://www.irishtimes.com/newspaper/ireland/2008/0802/1217368897558.html

5 See http://www.independent.ie/breaking-news/national-news/politics/govt-urged-to-beef-up-computer-security-systems-1285260.html

6 See http://www.irishtimes.com/newspaper/finance/2008/0606/1212677071145.html

Irish Council for Civil Liberties and Rights Watch in the UK highlighted just how extensive and covert the data surveillance and retention activities of governments can be. This case found that the British Ministry of Defence was storing all telephone, fax, e-mail and data communications between the UK and Ireland for the past seven years. The court found that this infringed on people's privacy under the European Convention of Human Rights[7]. Further, academic work points to the covert monitoring of user behaviour on commercial sites and dataveillance whereby personal profiles are matched to user behaviour online and used to target or personalize commercial services (Humphries, 2008, Jarrett, 1999, Jarrett, 2008). It is difficult at this stage to track how extensive such behaviour is but it is clear that internet based technologies and new applications like social networking sites are facilitating easy tracking, monitoring and targeting of users. Manuel Castells has argued that 'The transformation of liberty and privacy on the Internet is a direct result of its commercialisation. The need to secure and identify communication on the Internet to make money out of it, and the need to protect intellectual property rights' (Castells, 2001). This has lead to on the one hand, the increasing use of surveillance and monitoring technologies to track user behaviour and on the other to the development of a discourse of risk, security and threat. This trend has been compounded post 9/11 with the changing geo-political climate.

At present moves to improve security of data and use by institutions and individuals of private information are largely voluntary and focus on improving user practice and raising awareness. Thus recently the Irish Computer Society (ICS) – as an answer to the need for more security - has launched the first forum focusing on data protection and privacy. The forum will offer seminars, workshops and online resources, and encourage computer specialists to become certified on the ICS Data Protection Practitioner course[8]. This will – according to ICS - help Ireland in improving its – at the moment very low - performance in the field of computer

---

7 http://iccl.ie/DB_Data/news/CallsforsurveillancelawreformafterStrasbourgcourtvictory_80.htm

8 See http://www.irishtimes.com/newspaper/finance/2008/0530/1212048843820.html and http://www.ics.ie/dp/

security and privacy. In the same vein the Internet Advisory Board[9] was launched by the Irish Ministry for Justice Equality and Law Reform to promote Internet awareness including that of security and protection. The office of the Data Commissioner is another body which highlights abuses of data and in some cases pursues the perpetrators to court. Much of the current policy in this area places the burden of security on the end user while corporations, public sector bodies, internet service providers and other operators largely operate under voluntary codes of conduct.

The openness of the internet is increasingly seen as a weakness on the international stage. A consortium of regulatory, industrial and trade organisations are pushing for greater control and surveillance of users on the internet, perhaps even turning the internet into a black box, a locked down system to which only a few have the keys. The problem can be understood referring, to *The Future of the Internet* by Zittrain (2008). In this book the author praises what he calls the "generativity" of the Internet + PC architecture (i.e. the openness of this infrastructure). According to Zittrain[10], *Generativity* is "*a system's capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences* ". However Zittrain notes that many of today's major "internet players[11]" advocate a more closed approach. This could lead to a situation in which an "*unthinkable level of enclosure is likely to be the rule from which we must negotiate and justify exceptions*" (pg. 5). In this light, that of trust seems to be one of the key concepts of this process of enclosure and 'Trusted Systems', as currently envisaged by technologists are "*systems that can be trusted as against the people who use them*" (pg. 43). Hence, Zittrain argues, Trusted Systems are one of the ways thorough which the generativity of the internet is threatened by more lockdown: the more of this type of trust we have, the less the Internet+PC users will be able to produce innovative things. A similar argument is maintained by Lawrence Lessig (1999) where he points out that Trusted Systems will enforce the Digital Rights Management at a level in which technology will grant the owner of copyright more control over content than the one prescribed by the law.

---

9 See http://www.iab.ie

10 Jonathan Zittrain Interview http://www.blauexchange.org/int_jzittrain.html

11 Among these players Zittrain (2005, pg. 41-42) quotes regulators, the technology industry and mainstream consumers/users.

Thus it is clear that there are a range of actors currently concerned with negotiating a range of approaches to developing, maintaining and increasing Trust and Trusted Systems on the internet. Understanding the problem of Trust and Trusted Systems seems to us an interesting and challenging sociological problem, especially from a Science and Technology Studies (STS) point of view. We argue that current conceptions of *trust* should include insights from sociology, and not be confined to the knowledge domain of computer security experts, lawyers or mastered exclusively by global IT corporations. We should instead make an attempt to understand this notion using sociological concepts. In fact, it is important that these issues are judged and studied not only from the technical point of view but also from the sociological one: assuming a techno-centric vision there is always the risk to miss crucial attributes and actors in the knowledge society (Akrich and Miller, 2006).

3. A SOCIOLOGICAL APPROACH TO TRUST

In order to understand the conceptualisation of trust in the computer science domain, we must first understand how sociologists approach the concept. This is because we aim to outline the existence of differences and similarities between the two "types" of trust. It must be stated, however, that we do not aim to provide a broad review of sociological trust literature, for which the reader is directed to work by McKnight and Chervany (1996). More modestly our intention is to sketch some interesting issues. For this reason we take into account a restricted but significant set of contributions.

Trust is usually seen by sociologists as a three part relationship: "*A trusts B to do, or with respect to X*" (Hardin, 2006, pg. 19). In this definition we have the **trustor** (A) – the actor who is willing to act, placing confidence, reliability or its own vulnerability, in the hands of others -, the **trustee** (B) – the actor in which trust is placed, who can fulfill his/her role as well as betray the trustor -, and the **object or result** toward which trust is directed (X). For example, "*I trust (trustor) the baby-sitter (trustee) to keep safe my child (the object of trust)*" or "*we (trustor) trust the government (trustee) to fulfill its electoral promises (the object of trust)*" or "*the mother (trustor) trust her daughter (trustee)*

not to return late on night". Of course not all trust relationships are considered equal. Referring to the three examples above Sztompka distinguishes between "responsive[12]", "anticipatory[13]" and "evocative[14]" trust. There is no need to enter in the problem of the various forms of trust, but for our goal it is important to keep in mind the three part relationship of trust, with particular emphasis on the relation between the trustor and the trustee. This tells us that a trust relation involves somehow a *mutual commitment* (Deutsch, 1958; Sztompka, 1999; Jalava, 2003) – a close inter-relation – between the trustor and the trustee.

It is worth noting that many major sociologists have addressed the issue of trust. Historically, Simmel was probably one of the first to take trust into account when, in *The Philosophy of Money* (1978), he saw trust as one of the basic element for economic/monetary transactions and ultimately for the integration of society. However, the main contributions to the problem are probably the works by Niklas Luhman *Trust and Power* (1979) and *Familiarity, Confidence, Trust* (1988), the latter a chapter contained in the often quoted book *Trust: Making and Breaking Cooperative Relations* by Gambetta (1988b).

Luhmann defined trust as "*an attitude which allows for risk-taking decisions*" (Luhmann, 1988, pg. 103) or "*as a gamble, a risky investment*" (Luhmann, 1980, pg. 24). Trust, then, has been directly linked by Luhmann in a circular relation with risk, considered as both external and internal to actions. Risk "*represents a re-entry of the difference between controllable and uncontrollable into the controllable*" (1988, pg. 100). Therefore, according to Luhmann, trust is directly related with the subjective ability to assume risk-taking decisions. As such trust helps also in the reduction of complexity in the presence of several alternatives of action. On this basis Luhmann distinguished between trust and confidence, the latter seen as a situation in which "*If you do not consider alternatives […] you are in a situation of confidence*" (pg. 97-98). Another major Luhmannian distinction is between trust and rational calculation, the former related

---

12 Responsive trust "involve the act of entrusting some valuable object to somebody else, with his or her consent" (Sztompka, 1999, pg. 25)

13 Anticipatory trust is when "I act toward others because I believe that the actions which they carry out anyway will be favorable to my interests, needs and expectations" (Sztompka, 1999, pg. 25)

14 Evocative trust is when "we act on the belief that the other person will reciprocate with trust toward ourselves" (Sztompka, 1999, pg. 25)

to an action's result that could lead to a damage greater than the possible advantages (Deutsch, 1958), whereas instead in rational calculation the relations between advantage and damage are always in favour of the advantages.

For several reasons the work by Luhmann seems to be the basis for understanding the sociological notion of trust. Many other authors have adopted a perspective on trust very close to that of Luhmann or at least using the concepts, outlined by him, in a similar fashion. For example Lewis and Weigert (1985) and Moellering (2001) saw trust as a functional alternative to rational prediction for the reduction of complexity. Giddens (1990) too seemed to adopt a Luhmanian perspective, linking trust with risk. However, Giddens also partly criticized Luhmann in relation to the role of institutionalized risk as part of the *Risk Society*[15] notion (Beck, Giddens and Lash, 1994) and also in relation to a weaker separation between confidence and trust. McAllister (1995) in his inquiry on the interpersonal trust saw trust as enablers for people to take risks.

Trust has been related to the notion of social capital (Putnam, 2000; Fukuyama, 1995), whereas trust seems to have a functional[16] role in supporting cooperative relations in communities or among people (Gambetta, 1998b; Axelrod, 1981). Vulnerability has also often been portrayed as one of the main features of trust (Baier, 1986; Mayer et al, 1995), and considered as the situation where the trustor depends on the trustee good will, hence the trustor is vulnerable to the limits of that good will. Reduction of uncertainty for action has also been related to trust (Coleman, 1990; Gambetta, 1998a; Sztompka, 1999). Barber (1983) and Gambetta (1988) linked trust with expectations that people have of each other and to the possibility to a betrayal in such expectations. Another major issue is that of reputation (Dasgupta, 1988; Good, 1988), which emerged in particular in relation to risky economic transactions and described as a fundamental pre-condition for placing trust. Also weak knowledge or "quasi-faith" has been related to trust (Simmel, 1978; Giddens, 1991; Moellering, 2001), meaning that we trust basing our judgment on unknowable situations or on "technical principles" we ignore.

---

15 Society increasingly preoccupied with the future (and also with safety), which generates the notion of risk.

16 According to Luhmann (1988, pg. 95) his main goal was to address the problem of the "function of trust" in the social system.

3.2 Humans don't trust machines

Even if the concept of trust has, as we saw above, many "features", there is one which has attracted our attention: the trust relationships between human/social and natural/technological actors. This feature of trust is well described in the work by Sztompka (1999) :

> "Intuitively we feel that trust must be vested in people, rather than natural objects or event. Even if we seemingly confer trust on objects, such as saying "I trust Japanese cars" or "I trust Swiss watches," or " I trust French Rapid trains", we in fact refer to humanly created systems and indirectly we trust the designers, producers, and operators whose ingenuity and labor are somehow encrypted in the objects." (pg. 19-20)
>
> "Trust is not the orientation we would take toward the natural world. It does not sound proper to say "I trust the rain to fall," or "I trust these flowers to grow," but it is quite normal to say "I trust the meteorologists to predict the rain," or " I trust the gardener to tend for flowers well,". Trust belongs to human and not to natural discourse." (pg. 21)

Trust seems then, according to Sztompka, to be exclusively a human affair: it does not belong at all to natural discourse. Trust cannot be conferred at all to natural object such as the sun or the rain. Even when it seems to be quite acceptable to confer trust to things, like the car or the train, according to Sztompka we are conferring trust not to things themselves but to the designer, operators, producers and finally toward the institutionalized systems of knowledge encrypted into those things. We can't place trust in things or natural objects because – according to Sztompka - this doesn't help us in making the social world controllable or less uncertain.

It is interesting to note that the argument by Sztompka echoes the well known contribution by Giddens (1990). Giddens argues that lay people trust expert systems "*of technical accomplishment or professional expertise that organize large areas of the material and social environment in which we live today*" (pg. 27). According to Giddens it is a condition of what he called Reflexive Modernity for lay persons (trustor) to confer trust, based on "faith" and "weak knowledge", to expert systems (trustee). However the trust relation in this case is mediated by what Giddens call the access point "*Although everyone is aware that the real repository of trust is in the abstract system, rather than in the individuals who in specific contexts "represent" it, access points carry a reminder that it is flesh-and-blood people (who are potentially fallible) who are its operators.*".

This is clear for example in the case of air-travel, where lay people place their trust in the air crew personnel, more than on the air security statistics. Therefore we see that the access point[17]– interfaces between lay persons and systems of expert and expertise – remind us that technological systems, according to Giddens, are in fact human actions (even if actions are disembedded and distanciated in time and space). Even for Giddens then, who truly touches the point of lay persons' trust in technological systems, trust seems to be only a human affair.

Summarizing, we can conclude that one of the main features of trust, according to sociologists, is that this is a human property only. Both the trustor and the trustee of the three part relations of trust are in the end human[18] actors.

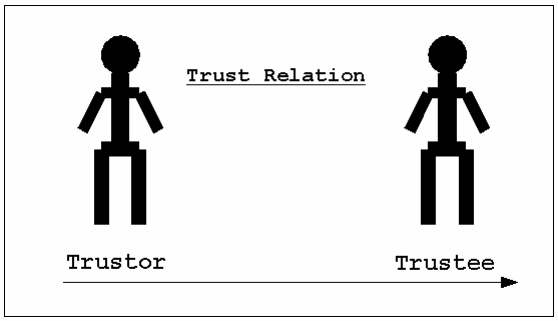

Figure 1 – Trust relation according to sociologists

4. TRUST IN THE MACHINE

Before entering into the CS domain we have to account for another way of conceiving trust: the trust relations between humans and machines in the field of automation[19]. It is interesting to note that the sociological relation between the trustor and the trustee is somehow radically modified if we look at this domain (e.g. Zuboff, 1988; Miur, 1994; Lee and Moray, 1991; Moray and Inagaki, 1999**;** Madsen and

17 That are considered form of faceless commitment.

18 Whether in the form of individual (e.g. McAllister, 1995; Sztompka, 1999), collective (e.g. Hagen and Cohe, 1998) or institutional (e.g. Shapiro, 1987; Farrel and Knight, 2003) actors.

19 Automation: is the use of control systems such as computers to control industrial machinery and processes, reducing the need for human intervention. From Wikipedia, http://en.wikipedia.org/wiki/Automation

Gregor, 2000; Lee and See, 2004). For example Moray and Inagaki (1999) defined trust as follows:

> "We use the term to refer to an attitude of mind towards an agent with whom a human operator is collaborating. The agent may be another human or a machine (automated sensor, automated controller, computer hardware, software programs or software 'agents', etc.). " (pg. 204)

From the above definition we see that the relation expressed by Sztompka - where when we say that we "trust the car to run" we trust its designer and not the car in itself -, is somehow modified. The trustee, according to Moray and Inagaki, may be a human as well as a machine, in various shapes or form. The operators using automation machinery and decision aid systems, in fact, seem to confer their trust in the machines, trusting them to perform the production tasks properly. It is not surprising then that also computers have gained the trust of humans. In human-computer interaction for example, the human-computer trust relation is defined by Madsen and Gregor (2000) as "*the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid.*". An example of the trust placed by humans in machines is: "*the pilot (trustor) trust the auto-pilot system (trustee) to drive the aircraft*".

The issue of trust in automation recall somehow the idea behind the book *The Media Equation* by Reeves and Naas (1996), in which the authors described how people treat various media (such as computers) exactly as they would real people. Lee and See (2004) however warn us that there are important differences between human-human and human-automation trust, such as the lack of intentionality of machines and also the existence of an asymmetry between the human operator (the one who has the ultimate responsibility for the task, especially in production processes) and the machine who performs the task.

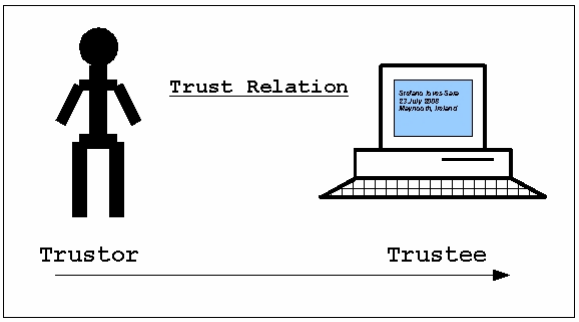

Figure 2 – Trust relation in automation and HCI

What it is important for our discussion is, anyway, to consider the three part relationship of trust. The relationship described by Sztompka is clearly modified in automation (see Figure 1 Vs 2), because the trustee can clearly be a machine, while the trustor still remains a human actor: it is the operator or the user who place his/her trust in the machine and not vice versa. In conclusion while the actors in the sociological accounts don't trust the machines (but trust the designers of the machines), the human actors in automation accounts – operators, users - seem instead to establish a concrete trust (or distrust) relationships with machines. However trust still remain clearly a human affair.

5. THE MACHINES WHO TRUST

The scenario of trust changes radically if we look at the CS domain. It is rather difficult to trace the origin of the notion of trust in CS, however we do not fail in saying that this issue emerged during the '70s in an age where computer security started to become a fundamental issue (Myers, 1980), in particular inside the military environment of the USA. In this context computer security has been defined as "*The protection resulting from all measures to deny unauthorized access and exploitation of friendly computer systems*"[20].

One of the first to use the term *Trusted Computer System* was probably Nibaldi (1979). For him "*Trusted computer systems are operating systems capable of preventing users from accessing more information than that to which they are authorized*" (pg. 1). This prevention of access was based on the access control and related security policies within an

[20] US Department of Defense Dictionary of Military and Associated Words http://www.answers.com/library/Military%20Dictionary

Operating System. The most important contribution to the problem however was probably the so-called DoD Orange Book (NCSC, 1987) which provided a series of criteria as standards and metrics to evaluate "*the degree of trust that can be placed in computer systems*". The Orange Book coupled the definition of a Trusted System with that of Trusted Computing Base considered as "*The totality of protection mechanisms within a computer system -- including hardware, firmware, and software -- the combination of which is responsible for enforcing a security policy*". Some of the concepts of this definition will become clear later on, for the moment we would like to focus on the "origin" of the idea of trusted systems and describe some key concepts.

The first, often quoted, seminal work in this domain is the Anderson Report (1972) which formalized computer security as related to control: "*Explicit control must be established over each user's (programs) access to any system resource which is shared with any other user or (system) program*". The Anderson Report introduced the notion of Reference Monitor – a software module of Operating Systems which has the role of controlling the access to data and devices – and the notions of mandatory and discretionary security policy. Before we define what is a security policy we have to refer to another major contribution in the domain of Computer Security that of Bell and Lapadula (1976) which defined the major goal and main actors of Trusted Computer Systems: "*The essential problem is to control access of active entities to a set of passive (that is, protected) entities, based on some security policy. Active entities are called subjects […]; passive entities are called objects.*" (pg. 9). In the computer security language a subject, within an Operating System, is considered as "*An active entity, generally in the form of a person, process*[21]*, or device that causes information to flow among objects or changes the system state*." (NCSC, Aqua Book, 1988). Whereas an object is "*A passive entity that contains or receives information*" (Aqua Book, 1988) such as certain I/O devices, data files, directories, memory segments and so on.

Basically the problem of a Trusted Computer Systems is that to control how a subject access an object. According to Lee (1999, pg. 2) "*The purpose of access controls is to authorise legitimate access by subjects to objects, and to prohibit illegitimate accesses. The essential notion is that without a legitimate access to an object, the system prohibits anything from happening with or to it. It can not be owned, created, destroyed, observed, altered,*

21 A Process is the execution of a computer program by a system.

appended to, distributed, executed, or searched. ". What distinguishes whether or not a subject may access the object are the security policies. That of security policy is a key concept of trusted system and it is defined as "*a statement of intent with regard to control over access to, dissemination of, and modification of information* " (From NCSC, Neon Orange Book[22], Glossary, 1987). Basically a security policy is a statement (properly implemented within the computer system) which says how subjects (in which context, at what time, for what purpose, and according to what laws, standards, organizational rules and so on) can access a certain piece of information. The access is referred to a set of actions that can be performed, such as for example read, write or delete a piece of information. Certain policies are then considered related to a mandatory access control: "*A means of restricting access to objects based on the sensitivity (as represented by a label) of the information contained in the objects and the formal authorization (i.e., clearance) of subjects to access information of such sensitivity.*" (from NCSC Aqua Book, 1988). Translated in a more accessible language this type of access control is often related to sensitive data (e.g., government classified information or sensitive corporate data). Systems providing mandatory access controls must assign sensitivity labels to subjects and objects in the system. The labels represent the level of trust: the trust that the systems place to a subject and the level of trust that a subject must have to be able to access the object. In summary, mandatory access controls use sensitivity labels ranging from the most to the least sensitive (in the case of object for example from "Top Secret", down to "Unclassified") to determine who can access what information in the system (Russell and Gangemi, 1991).

Access control can also be discretionary: "*A means of restricting access to objects based on the identity and need-to-know of the user, process and/or groups to which they belong. The controls are discretionary in the sense that a subject with a certain access permission is capable of passing that permission (perhaps indirectly) on to any other subject*. "(from Aqua Book, 1988). Let's imagine that a subject A belongs to the research group X, then we can imagine that A is entitled to access all the information related to the research projects (the object) of the group X. In other words discretionary access control is

---

22 See this link where all the DoD books (so called Rainbow Series) are quoted http://en.wikipedia.org/wiki/Rainbow_Series

based on access policies that restrict access to objects based on the identity of subjects and/or the groups to which they belong (Russell and Gangemi, 1991).

One further point that we need to outline are the threats against which the security Trusted Systems here were built. In the Aqua Book (NCSC, 1988)- a DoD Glossary of Trusted Systems terms – there are several entries that describe different kinds of threats and the action these threats may conduct against the system. Of course given that the Trusted System developed within a military context the protection of the system from attacks (think for example about foreign countries attacks) as well as the protection of sensitive information were considered fundamental. According to the glossary a threat may be defined as "*Any circumstance or event with the potential to cause harm to a system in the form of destruction, disclosure, modification of data, and/or denial of service.*". It is interesting to note that in this domain the concept of risk is taken into account in relation to threats. Risk is defined as "*The probability that a particular threat will exploit a particular vulnerability of the system.*". These threats could have been related for example to the existence of back doors, logic bombs or Trojan horses in the code of the system as well as the inclusion of malicious hardware and firmware. All these threats were defined as "malicious logic". Among the possible dangerous actions that could be perpetrated against the system the Aqua Book defined the penetration, the compromise, the tampering and the spoofing. Finally it is interesting to observe that the system security could have been circumvented also by the mean of exploiting design "errors" such as a security flaw "*An error of commission or omission in a system that may allow protection mechanisms to be bypassed.* ".

6. DISCUSSION

Entering the technical domain of trust we encounter several issues that the human-centred approaches to trust, both in sociology and automation, are not able to address. It is clear that in a situation described above it is the Trusted System that places its trust (and so it is the trustor) toward the subject (the trustee – which may be a human user as well as a process or a device) in relation to the access of information (the object of trust). The role of a Trusted System is indeed that of conferring trust to subjects, meanwhile "keeping an eye" on what they do, hence preventing unauthorized accesses to information. Early work in CS clearly outlines a very

different approach to trust, if compared with the sociological (as well as the automation) view of trust. A Trusted Computer Systems is a machine (trustor) who is entitled to trust several other entities (so called subject – the trustee) which are both human and non-human. If the trustor is a machine and the trustee might be in turn a human and a non-human, aren't we radically approaching a different domain of trust?

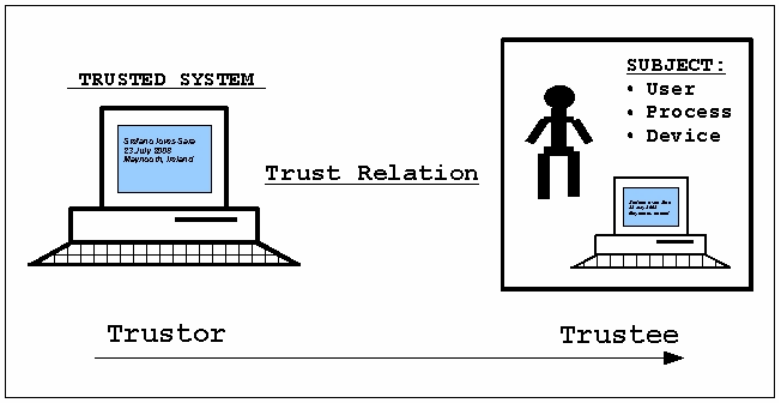

Figure 3 – Trusted Systems: machines who trust

In further developing this literature review we will review contemporary concepts of trust in computer science . In our opinion what we will need – for sociologically approaching it - are concepts that will allow us to break the common assumed division between the social and the technical worlds.

At first, it is worth noticing that, many useful concepts have already been developed and used to understand the work of engineers and scientists (for example Bloor, 1976; Bijker, 1995; Latour, 1987). In our opinion two concepts, drawn from the approach known as Actor-Network Theory (Callon, 1986; Latour, 1987), seem to be useful for dealing with the CS trust notion: *free association* (Callon, 1986) and *delegation* to non-humans (Latour, 1992).

According to Callon (1986) one of the concepts which helps in the study of the role played by science and technology in structuring power relationships, is that of *free associations*. The basic idea behind this concept is that:

> "The observer must abandon all a priori distinctions between natural and social events. […]. Instead of imposing a pre-established grid of analysis upon these, the observer follows the actors in order to identify the manner in which these define and associate the different elements by which they build and explain their world, whether it be social or natural" (pg. 200-201)

The concept of free association captures how computer scientists use the word trust, in its various facets, to refer to both human and non-human actors**.** According to Latour (2005) the basic idea behind the free association concept is that it is not the duty of the sociologists to decide in advance what the social world is made of, because people as well as machines (human and non-human actors) may be part of this world (Latour, 1993).

A second concept we found useful for our purposes is that of *delegation* to non-humans. Latour (1992) described this concept in a didactic example of how to convince people to close the door behind them. The argument by Latour is that artefacts are at once *strongly social and highly moral*. In order to convince people to close the door there are two solutions:

> "either to discipline the people or to substitute for the unreliable people another delegated human character whose only function is to open and close the door. ".

However paying a human porter for his/her unskilled work is quite an expensive solution and here is where engineers came into play with their arte-fact moral and political solution:

> "It is at this point that you have a relatively new choice: either to discipline the people or to substitute for the unreliable humans a delegated non-human character whose only function is to open and close the door. This is called a door-closer or a groom ('Groom' is a French trademark which is now part of the language).".

According to Latour, the ability to perform the task of closing the door has been delegated by engineers to a non-human character, which is able to perform the job of closing the door: the door-closer. This non-human character is very social and moral

because according to Latour "*We have been able to delegate to non-humans not only force but also values, duties and ethics..*".

What it is interesting in the concept of delegation – or moral prescription[23] - to non-human as related to trust in CS is that the act of deciding who is to be trusted or not, is no longer in the hand of the "human character". Computer scientists have, since the early '70s, delegated to Trusted Systems the moral ability to decide who has to be trusted and who has not, and for what reasons (the moral prescription) (see for example NCSC, 1983). But the interest in the concept of delegation goes on as long as Latour noticed that "*How can these prescriptions be brought out? By replacing them by strings of sentences (usually in the imperative) that are uttered (silently and continuously) by the mechanisms for the benefit of those who are mechanized: do this, do that, behave this way, don't go that way. Such sentences look very much like a programming language.*". The image portrayed here by Latour is that mechanisms and in general technological artifacts impose their moral prescription in a disciplined way: they never fail to tell people to behave in certain ways. It is easier to discipline a non-human which silently will pursue the task delegated to it by its designers, rather than convince all users to pursue a certain behavior. Hence, the problem we will need to address in future work will be to understand what is meant by saying that computer systems have been delegated the ability to trust.

CONCLUSION

Discourses around access and inclusion in the information or knowledge society have given way to discourses around security, trust and governance. Much sociological work has seen trust as a human attitude which is placed in another human being in the form of a relationship. However computer science work makes no such distinction and suggests that we can equally trust human or non-human

23 Prescription is the moral and ethical dimension of mechanisms.

actors. In 'Trusted Computing' the system is designed to control and manage access and authorization to information and knowledge. Concepts like free association and delegation from Science and Technology Studies will we believe be useful in bridging the interdisciplinary gaps between members of our research team and can be deployed empirically to study the design, negotiation and contestation of these types of systems by groups of users.

REFERENCES

- **General References**

Akrich, M., 1992. The De-Scription of Technical Objects. In *Shaping Technology/Building Society*. Cambridge, Mass.: MIT Press, pp. 205-224.

Akrich, A. & Miller, R. 2006. The future of key actors in the European research area: synthesis paper report paper for the *Research Expert groups on the Future of Key Actors*

Bijker, W., 1995. *On Bicycles, Backelites, and Bulbs*, Cambridge Mass.: MIT Press.

Bijker, W. & Law, J., 1992. *Shaping Technology/Building Society*, Cambridge Mass.: MIT Press.

Bloor, D., 1976. *Knowledge and Social Imagery*, Chicago: University of Chicago Press.

Callon, M., 1986. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St. Brieuc Bay. In *Power, Action and Belief: a New*. London: Boston and Henley, Routledge and, pp. 196-223.

Callon, M. & Latour, B., 1992. Don't Throw the Baby Out with the Bath School! A reply to Collins and Yearley. In *Science as Practice and Culture*. Chicago: Chicago University Press, pp. 343-368.

Castells, M. 2001. The Politics of the Internet II: Privacy and Liberty in Cyberspace. *The Internet Galaxy. Reflections on the Internet, Business and Society.* Oxford:

Oxford University Press

CSO, 2008. Information Society and Telecommunications Cork Central Statistics Office.

Frates, J. & Moldrup, W., 1980. *Introduction to the computer; an integrative approach*, New Jersey: Prentice-Hall.

Humpries, S., 2008. Ruling the Virtual World. Governance in massively multiplayer online games. *European Journal of Cultural Studies,* 11**,** 149-171.

ISC, 2002. Building the Knowledge Society. Report to Government. . Dublin, Information Society Commission.

ISSC, 1996. Information Society Ireland Strategy for Action. Dublin, Forfás

Jarrett, K., 1999. The Perfect Community. Disciplining the eBay User. IN HILLIS, K., PETIT, M. & EPLEY, S., NATHAN (Eds.) *Everyday eBay: Culture, Collecting, Desire.* Routledge.

Jarrett, K., 2008. Interactivity is Evil! A Critical Investigation of Web 2.0. *First Monday,* 13.

Kemmerer, R., 1994. Computer security. In *Encyclopedia of Software Engineering*. John Wiley and Sons, pp. 1153–1164. Available at: http://scholar.google.com/scholar?hl=en&lr=&cluster=2743382650620738002

Latour, B., 1999. *Pandora's Hope*, London: Harvard University Press.

Latour, B., 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford: Oxford University Press.

Latour, B., 1987. *Science in Action, How to Follow Scientists and Engineers throughSociety*, Cambridge: Harvard University Press.

Latour, B., 2003. The power of Fac-Similes: a Turing Test on Science and Literature. Available at: http://www.bruno-latour.fr/articles/article/94-POWERS%20TURING.html.

Latour, B., 1993. *We have never been modern*, Cambridge, Mass.: Harvard University Press.

Latour, B., 1992. Where are the missing masses? The sociology of a few mundane artifacts. In *Shaping Technology/ Building Society*. Cambridge, Mass.: MIT Press, pp. 225-258.

Law, J. (Ed.) , 1991. *A Sociology of Monsters*, London: Routledge and Kegan Paul.

Law, J. (Ed.)., 1986. *Power, Action and Belief : a New Sociology of Knowledge?*, Routledge and Kegan Paul.

Law, J., 1987. Technology and heterogeneous engineering: The case of Portuguese Expansion. In *The Social Construction of Technological Systems : New directions in the*. Cambridge, Mass.: MIT Press, pp. 111-134.

Lee, S.E., 1999. *Essays about Computer Security*, Available at: http://www.cl.cam.ac.uk/~mgk25/lee-essays.pdf.

Lessig, L., 1999. *Code and Other Laws of Cyberspace*, Basic Books

Melody, W., 1997. Identifying Priorities for Building Distinct Information Societies. *The Economic and Social Review,* 28**,** 177-184.

Menezes, A., van Oorschot, P.C. & Vanstone, S.A., 1996. *Handbook of Applied Cryptography* 5th ed., CRC Press.

Nissenbaum, H., 2001. Securing Trust Online: Wisdom of Oxymoron? *Boston University Law Review*, 81(3), 101-131.

Peterson, J.L. & Silberschatz, S., *Operating Systems Concepts*, Reading, Mass: Addison Wesley.

Russell, D. & Gangemi, Sr., G.T., 1991. *Computer Security Basics* First Edition ., O'Reilly.

Zittrain, J., 2008. *The Future of the Internet - and How to Stop it,* London, Yale University Press.

- **General Definitions**:

Oxford English Dictionary trust, n. Available at: http://dictionary.oed.com/cgi/entry/50259124?query_type=word&queryword=trust&first=1&max_to_show=10&sort_type=alpha&result_place=2&search_id=uNov-vP04MA-4819&hilite=50259124 [Accessed June 9, 2008].

Trust. *Stanford Encyclopedia pf Philosophy*. Available at: http://plato.stanford.edu/entries/trust/.

Trust. *Merriam-Webster's Online Dictionary*. Available at: http://www.merriam-webster.com/dictionary/trust.

Trust (social sciences) - Wikipedia, the free encyclopedia. Available at: http://en.wikipedia.org/wiki/Trust_%28social_sciences%29 [Accessed June 9, 2008].

- **General Sociology on Trust:**

Axelrod, R. & Hamilton, W., 1981. The Evolution of Cooperation. *Science*, 211(4489), 1390–1396.

Baier, A., 1986. Trust and Antitrust. *Ethics*, 96(2), 231-260.

Barber, B., 1983. *The Logic and Limits of Trust*, New Brunswick: Rutgers University Press.

Beck, U., Giddens, A. & Lash, S., 1994. *Reflexive Modernization*, Stanford, California: Stanford University Press.

Dasgupta, P., 1988. Trust as a Commodity. In *Trust: Making and Breaking Cooperative*

Relations. Oxford: Blackwell.

Deutsch, M., 1958. Trust and suspicion. *The Journal of Conflicts Resolution*, 2(4), 265-279.

Earle, T.C. & Cvetkovich, G., 1995. *Social Trust: Toward a Cosmopolitan Society*, Greenwood Publishing Group.

Fukuyama, F., 1995. *Trust*, New York: Basic Books.

Gambetta, D., 1988a. Can We TrustTrust? In *Trust: Making and Breaking Cooperative Relations*. Oxford: Blackwell.

Gambetta, D., 1988b. *Trust: Making and Breaking Cooperative Relations,* Oxford: Blackwell.

Garfinkel, H., 1963. A conception of, and experiments with, Trust as a condition for Concerted Stable Actions. In *Motivation and Social interaction edited by O.J. Harvey*. New York: Ronald press, pp. 187–238.

Giddens, A., 1990. *The Consequences of Modernity*, Cambridge: Polity Press.

Good, D., 1988. Individuals, Interpersonal Relations, and Trust. In *Trust: Making and Breaking Cooperative Relations*. pp. 31-48.

Hardin, R., 2006. *Trust*, Cambridge: Polity Press.

Jalava, J., 2003. From Norms to Trust. The Luhmanian Connections between Trust and System. *European journal of social theory*, 6(2), 173-190.

Lewis, J.D. & Weigert, A., 1985. Trust as a Social Reality. *Social Forces*, 63(4), 967-985.

Luhmann, N., 2000. Familiarity, Confidence, Trust: Problems and Alternatives. In *Trust: Making and Breaking Cooperative Relations*. Oxford: Blackwell.

Luhmann, N., 1980. *Trust and Power*, New York: John Wiley.

Mayer, R.C., Davis, J.H. & Schoorman, F.D., 1995. An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.

McAllister, D.J., 1995. Affect- and Cognition-Based Trust as Foundation for Interpersonal cooperation in Organisations. *The Academy of Management Journal*, 38(1), 24-59.

Meyerson, D., Weick, K.E. & Kramer, R.M., 1996. Swift Trust and Temporary Group. In *Trust in Organisations*. Sage.

Misztal, B., 1996. *Trust in Modern Societies: The Search for the Bases of Social Order*, Cambridge Mass.: Polity Press.

Moellering, G., 2001. The Nature of Trust:From Georg Simmel to a Theory of Expectation, Interpretation and Suspension. *Sociology*, 35(2), 403-420.

Putnam, R., 1995. Bowling Alone: America's Declining Social Capital. *Journal of Democracy*, 6(1), 65-78.

Putnam, R., 2000. *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon and Schuster.

Seligman, A.B., 1997. *The Problem of Trust,* Princeton: Princeton University Press.

Shapiro, S.P., 1987. The Social Control of Impersonal Trust. *American Journal of Sociology*, 93(3), 623-658.

Simmel, G., 1978. *The Philosophy of Money*, London: Routledge and Kegan.

Sztompka, P., 1999. *Trust: A sociological Theory*, Cambridge: Cambridge University Press.

Weber, L. R. & Carter, A., 2003. *The Social Construction of Trust*, New York: Kluwer Academic/Plenum.

- **Trust in Automation:**

Atoyan, H., Duquet , J. & Robert, J., 2006. Trust in new decision aid systems. In *Proceedings of the 18th International Conferenceof the Association Francophone d'Interaction Homme-Machine*. ACM International Conference Proceeding Series;. Montreal, Canada, pp. 115-122.

Duez, P.P., Zuliani, M.J. & Jamieson, G.A., 2006. Trust by Design: Information Requirements for Appropriate Trust in Automation. In *Proceedings of the 2006 conference of the Center for Advanced Studies on Collaborative research*. Toronto, Ontario, Canada.

Itoh, M., Abe, J. & Tanaka, K., 1999. Trust in and Use of Automation: Their Dependence on Occurrence Patterns of Malfunctions. In *Proc.IEEE SMC Conference on Systems, Man, and Cybernetics*. Tokio, pp. 715-720.

Lee, J. & Moray, N., 1991. Trust, Self-Confidence and Supervisory Control in a Process Controlsimulation. In *Systems, Man, and Cybernetics, 1991. 'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference*. pp. 291-295.

Lee, J.D. & See, K.A., 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50-80.

Madsen, M. & Gregor, S., 2000. Measuring Human-Computer Trust. In *Proceedings of the 11th Australasian Conference on Information*.

Moray, N. & Inagaki, T., 1999. Laboratory studies of trust between humans and machines in automated systems. *Transactions of the Institute of Measurement and Control*, 21(4-5), 203-211.

Reeves, B. & Nass, C., 1996. *The media equation: how people treat computers, television, and new media like real people and places*, New York, NY, USA: Cambridge University Press.

Zuboff, S., 1988. *In the Age of the Smart Machine: The Future of Work and Power*, New Brunswick: Basic Books.

- **Classical Work on CS Tust**

Abrams, M.D. & Joyce, M.D., 1995. Trusted Systems Concepts. , 1-19. Available at: http://www.acsac.org/secshelf/papers/trusted_system_concepts.pdf.

Anderson, J.P., 1972. *Computer Security Technology Planning Study*,

Bell, E.D. & LaPadula, L.J., 1976. *Secure Computer Systems: Unified Exposition and MULTICS Interpretation*, Belford, MA: MITRE Corporation.

McHughh, J., 1993. Implementation. In *Handbook for the Computer Secuirty Certification of Trusted Systems*. pp. 1-49 & i-v.

Myers, P., 1980. Subversion: The Neglected Aspect of Computer Security.

National Computer Security Center, 1987. *A Guide to Understanding Discretionary Access Control in Trusted Systems*, Available at: http://www.fas.org/irp/nsa/rainbow/tg003.htm.

National Computer Security Center, 1988. *Glossary of Computer Security Acronyms*, Available at: http://csrc.nist.gov/publications/secpubs/rainbow/tg004.txt.

National Security Institute, 1983. *Departement of defense Trusted Computer System Evaluation Criteria* Online version., Available at: http://nsi.org/Library/Compsec/orangebo.txt.

Nibaldi, G.H., 1979. *Proposed Technical Evaluation Criteria for Trusted Computer Systems*, Belford, MA: MITRE Corporation.

Saltzer, J. & Schroeder, M., 1975. The Protection of Information in Computer Systems. In *Proceedings of the IEEE, 63(9)*.

Thompson, K., Reflections on Trusting Trust. *Communications of the ACM*, 27(8), 761-763.

Walker, S., 1980. The Advent of Trusted Computer Operating Systems. In *National Computer*. pp. pp. 655-665.