# ETIENNE AMÉRICO CARTOLANO JÚNIOR A model for trust under a suitcase word perspective

# ETIENNE AMÉRICO CARTOLANO JÚNIOR

# A model for trust under a suitcase word perspective

Thesis presented to Escola Politécnica da Universidade de São Paulo to obtain the title of Doctor of Sciences

Supervisor: Antonio Mauro Saraiva

Co-Supervisor: Robert David Stevenson

Este exemplar foi revisado e corrigido em re responsabilidade única do autor e com a ar	
São Paulo, de	de
Assinatura do autor:	
Assinatura do orientador:	

### Catalogação-na-publicação

Cartolano Júnior, Etienne Américo

A model for trust under a suitcase word perspective / E. A. Cartolano Júnior -- versão corr. -- São Paulo, 2017. 139 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.trust 2.trust model 3.suitcase word 4.simulation 5.behavioral decision I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

Dedicated to my beloved wife Flávia and to all who trust me

### **AGRADECIMENTOS**

Agradeço ao meu orientador Prof. Antônio Mauro Saraiva por seu o meu mentor e amigo por tantos anos, desde os primeiros dias da minha graduação. Sob sua orientação e liderança tive muitas oportunidades para descobrir e crescer no mundo acadêmico e na minha vida pessoal.

Agradeço também ao Prof. Robert Stevenson por ter sido meu segundo orientador e um anfitrião paciente e interessado. Minha estadia em Boston não teria sido tão incrível, agradável e produtiva sem o seu suporte.

Além disso, agradeço ao Dr.(!) Allan Koch Veiga, sem quem eu não poderia andar na vida acadêmica. Pelo menos sem tropeçar! Certamente, ele é um dos grandes responsáveis pelo meu sucesso acadêmico.

Aos funcionários e professores da (gloriosa) Escola Politécnica da USP, agradeço por seu trabalho, conhecimento e paixão, que preservam a excelência e a tradição da nossa Escola.

Aos colegas do Itaú-Unibanco, agradeço pela paciência e incentivo durante esta jornada, especialmente, àqueles que, no dia-a-dia da nossa equipe, compartilharam os melhores e os piores momentos desta minha vida dupla.

Aos amigos de sempre que me ajudam a qualquer momento e por qualquer razão, e que contribuíram indiretamente com esse trabalho, e aos amigos de Boston, que fizeram daquela jornada algo tão especial.

Agradeço a minha família (Cartolano, Borges, e De Conti), para quem eu vivo, e que me faz quem sou. E aos meus futuros engenheiros: Bruno, Lucas, Melissa, Benjamin (Tet) e Felipe, digo: vão estudar!

E, de maneira especial, agradeço a quem tem estado ao meu lado a cada etapa da minha vida. Quem confia e faz de mim uma pessoa melhor todas as manhãs. Que vida sem amor eu teria sem a minha esposa Flávia ao meu lado!

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil

In God we trust (unknown)

### **ABSTRACT**

Trust is a social phenomenon fundamental to relationships and a building block of our society. People experience it daily, such as in a borrowing between friends, in an e-commerce transaction, in a mother-son relationship, in a connection between autonomous agents, or to show faith in God ("In God we trust"). In the specific case of Biodiversity domain, trust is one of the pillars of the Citizen Science projects, which are helping to solve the lack of biodiversity data by engaging citizens to work as volunteers to address this problem. Measuring and simulating levels of trust on these projects might reveal or anticipate losses; for example, the disposal of data because a deficit of trust on the technical capacity of the volunteers, opening an opportunity to manage and improve it. However, trust is a hard concept to define. The word 'trust' may carry different meanings, such as honesty, security, integrity, competence, etc. and this is an attribute of the 'suitcase words'. Adopting the 'suitcase' perspective would change the way as we define, model, and simulate trust, once people would identify, decode, and simulate many meanings of trust with a single approach. In this scenario, the main objective of our research was to verify the hypotheses 1) that trust is a suitcase word, and 2) that trust can be modeled and simulated under a suitcase word perspective. A network analysis of the Web of Science citation database was able to confirm the hypothesis that trust is a suitcase word, since a distribution analysis of articles showed that trust occurs across a wide range of disciplines, and since co-occurrence maps of keywords showed that trust meanings from these disciplines may be significantly different. To verify the second hypothesis, we proposed a framework to manage trust with three components: 1) a suitcase model to identify different meanings of trust, which is the main purpose of this work, 2) a procedure to detail trust situations in terms of the suitcase model, and 3) a behavioral decision model of confidence, which was required for our simulation, since trust and control play complementary roles in the development of confidence, and consequently, to generate a confident behavior to cooperate. In our suitcase model the decision to trust (or distrust) the trustee depends on the trustors' general capacity to take risks (= trustfulness) and on the assessment of trustee's interests and capacity to behavior as the trustor expects (= trustworthiness). In a practical and workable way, trustworthiness was considered a function of the trustor's expectations (expected evidence) and the trustee's previous

behavior (collected evidence) for each situation. We proposed a formalism to the suitcase model, and then replicated the PlayGround simulator to modify it and incorporate our model. The new simulator, the PlayGround 2.0, was used to run a case study using trust situations from Citizen Science projects. Our main goal with this case study was to test the hypothesis that trust can be simulated under a suitcase perspective. A successful simulation would plot agents in the field reacting differently according to each situation. Results were as expected, what demonstrated the comprehensive utility of our model, with potential to handle different meanings of trust in the context of Citizen Science in the Biodiversity domain.

**Keywords**: trust, trust model, suitcase word, simulation, behavioral decision.

### **RESUMO**

Trust, que em português não possui uma tradução que reflita a sua complexidade, é um fenômeno fundamental para os relacionamentos e uma peça fundamental da nossa sociedade. Trust é vivenciada diariamente pelas pessoas, seja em um empréstimo entre amigos, em uma transação de comércio eletrônico, na relação entra uma mãe e um filho, na conexão entre agentes autônomos, ou para demonstrar a nossa fé. No caso específico da ciência da Biodiversidade, trust é um dos pilares dos projetos de Ciência Cidadã, que ajudam a resolver a falta de dados de biodiversidade engajando cidadãos para trabalhar como voluntários para solucionar o problema. Medir e simular *trust* nestes projetos pode revelar ou antecipar perdas, por exemplo, o descarte de dados devido à falta de trust na capacidade técnica dos voluntários, abrindo oportunidades para sua gestão e incentivo. Contudo, trust é um conceito difícil de definir. A palavra pode carregar muitos significados, tais como honestidade, segurança, integridade, competência, etc., e esta é uma característica das suitcase words (palavras-mala). Adotar uma perspectiva de suitcase pode mudar a maneira como definimos, modelamos e simulamos trust, pois as pessoas poderiam identificar, decodificar, e simular vários significados de trust com uma única abordagem. Neste cenário, o objetivo principal desta pesquisa foi verificar as hipóteses de que 1) trust é uma suitcase word e de que 2) trust pode ser modelada e simulada em uma perspectiva de suitcase com potencial para manipulação de diferentes significados. Uma análise de distribuição utilizando o banco de dados Web of Science foi suficiente para confirmar a primeira hipótese de que trust é uma suitcase word, pois uma análise dos seus registros de citações mostrou que trust é estudado por muitas disciplinas da ciência, e além disso, mapas de co-ocorrência de palavraschaves mostraram que os significados de trust nas disciplinas podem ser diferentes. Para verificar a segunda hipótese, nós propusemos um framework com três componentes: 1) um 'modelo suitcase' para identificar diferentes significados de trust, que é o objetivo principal deste trabalho, 2) um procedimento para detalhar as situações de trust em componentes do modelo suitcase, e 3) um modelo de decisão comportamental sobre confiança, que foi necessário para nossa simulação, uma vez que trust e controle tem papéis complementares no desenvolvimento da confiança, e consequentemente, na exibição de um comportamento de confiança que pode levar à cooperação. No nosso modelo, a decisão de *trust* o *trustee* (aquele a quem *trust* é direcionada) depende da capacidade do *trustor* (aquele que direciona *trust*) em aceitar riscos (= *trustfulness*), e da avaliação do interesse e da capacidade do *trustee* em agir como o esperado (= *trustworthiness*). De uma maneira prática, *trustworthiness* foi considerada como uma função das evidências esperadas e coletadas em cada situação. Nós propusemos um formalismo para o novo modelo, e depois replicamos um simulador chamado PlayGround para depois modificá-lo e incorporar o modelo *suitcase*. O novo simulador, o PlayGround 2.0, foi utilizado para rodar um estudo de caso utilizando situações comuns em projetos de *Citizen Science*. Nosso maior objetivo foi testar a hipótese de que *trust* poderia ser simulada em uma perspectiva de *suitcase*. Uma simulação bem-sucedida plotaria os agentes em campo reagindo diferentemente de acordo com as situações apresentadas. Os resultados foram como esperados, o que demonstrou a utilidade abrangente do nosso modelo, com potencial para lidar com diferentes significados de *trust* no contexto dos projetos de Ciência Cidadã para Biodiversidade, ou em outros contextos.

**Palavras-chave**: *trust*, *suitcase word*, confiança, decisão comportamental, simulação.

# **LIST OF FIGURES**

Figure 1 - Graphical representation of our research method	24
Figure 2 - Adopted hierarchy of categories from Web of Science	30
Figure 3 - Distribution of articles among main areas of Science in Web of Science	30
Figure 4 - Co-occurrence of authors in the Psychology research area	33
Figure 5 - Psychology research area co-occurence map of keywords	35
Figure 6 - Computer Science research area co-occurrence map of keywords	35
Figure 7 – Co-occurrence map of keywords from Computer Science, Psychology, Operation	on
Research Management Science, and Sociology	36
Figure 8 - Ngram distribution of types of trust	43
Figure 9 – The trust continuum, extracted from (CHO; CHAN; ADALI, 2015)	44
Figure 10 - Confidence behavior model, extracted from (COFTA, 2007)	50
Figure 11 - Basic building block for confidence, extracted from (COFTA, 2007)	52
Figure 12 - Building blocks within a horizon of three entities, extracted from (COFTA, 2007)	7)
	53
Figure 13 – Proposed framework to define trust under a suitcase perspective	56
Figure 14 - Suitcase model components and their relations	57
Figure 15 - Confident behavior components (blue) and their relations with the suitcase mo	del
of trust	58
Figure 16 – ABX+C relation, an expanded three-part relation of trust	59
Figure 17 - New basic building block for confidence, modified from (COFTA, 2007)	61
Figure 18 - Alice's timeline of trust situations	63
Figure 19 - Suitcase model version of the basic building block for confidence with	
attenuators, modified from (COFTA, 2007)	65
Figure 20 - Graphical relation among the components of the suitcase model	67
Figure 21 - Alice's timeline of trust situations marked with the suitcase model formalism	68
Figure 22 - Axelrod's Tournament environment	81
Figure 23 - A screenshot of the PlayGround, extracted from (MARSH, 1994)	82
Figure 24 - Initial agent's distribution in experiment F part I replication	90
Figure 25 - Final agent's distribution in experiment F part I replication	90
Figure 26 - Percentage of successful cooperation in time for trusting agents	92
Figure 27 - PlayGround 2.0 screenshot	92
Figure 28 – Relation between entities in the 'Suitcase Trustor' strategy of the PlayGround	
2.0.	95
Figure 29 - Co-occurrence map for Operations & Management Science Research Area in	

Web of Science	129
Figure 30 - Co-occurrence map for Sociology Research Area in Web of Science	130
Figure 31 - Co-occurrence map for Computer Science Information Systems Category in	Web
of Science	131
Figure 32 - Co-occurrence map for Artificial Intelligence Category in Web of Science	132
Figure 33 - Co-occurrence map for Hardware Category in Web of Science	133
Figure 34 - Co-occurrence map for Software Engineering Category in Web of Science	134
Figure 35 - Co-occurrence map for Theory Methods Category in Web of Science	135

## LIST OF ABBREVIATIONS AND ACRONYMS

ABX Three-part relation definition for trust

ABX+C Modified three-part relation definition for trust

BT Basic Trust

EGT Estimated General Trust

GT General Trust

IPD Iterated Prisoner's Dilemma

PD Prisoner's Dilemma

ST Situational Trust

TDWG Biodiversity Information Standards

VOS VOSviewer

WOS Thompson Reuter's Web of Science

# **CONTENTS**

1	INTRODUCTION	21
1.1	Justification	21
1.2	Problem statement	23
1.3	Objectives	23
1.4	Materials and methods	24
1.5	Thesis structure	27
2	SCIENTIFIC LITERATURE DATA MINING	29
2.1	Distribution among areas of knowledge	29
2.2	Co-occurrence of keywords	32
2.3	Considerations about the literature mining	36
3	CONCEPTUAL FOUNDATIONS OF TRUST	38
3.1	Many Definitions of Trust	38
3.2	Actors, Risk & Context	39
3.3	Situation, Belief, Probability, and Behavior Decision	40
3.4	Three-part relation	41
3.5	Group and Cooperation	41
3.6	Types of Trust	42
3.7	Trust dynamics and measurement	43
3.8	Trustfulness and Trustworthiness	45
3.9	Models of Trust	46
4	MODELING TRUST UNDER A SUITCASE WORD PERSPECTIVE	56
4.1	Suitcase model	56
4.1.1	Trust belief	57
4.1.2	Trust and Confidence	57
4.1.3	Trust situation with a behavioral decision of confidence	58
4.1.4	Trustworthiness computation	60
4.1.5	Suitcase model formalization	
4.2	Defining a trust situation	75
4.3	A generalized model of behavioral decision	76
4.4	Considerations about modeling trust under a suitcase perspective	77

5	SIMULATING THE SUITCASE MODEL79	
5.1	The Iterated Prisoner's Dilemma79	
5.2	Replicating the original PlayGround81	
5.2.1	The original PlayGround	
5.2.2	Assumptions about original implementation	
5.2.3	Replication results	
5.3	PlayGround 2.0 - the suitcase version92	
5.4	Considerations about the model simulator95	
6	CASE STUDY IN CITIZEN SCIENCE96	
6.1	Scenario of Citizen Science in Biodiversity96	
6.2	Defining the Citizen Science situations97	
6.3	Reaction under different situations	
6.4	Considerations about the case study 105	
7	DISCUSSION	
8	FINAL REMARKS	
8.1	Conclusions	
8.2	Future work	
BIBLIOGF	RAPHY113	
A.	EXTRACTION OF ARTICLES RECORDS FROM WEB OF SCIENCE 117	
B.	HIERARCHY RELATION AMONG WEB OF SCIENCE CATEGORIES. 119	
C.	VOSVIEWER CONFIGURATION	
D.	MAPS OF CO-OCCURRENCE	
E.	RESULTS OF MARSH'S PLAYGROUND	

# 1 Introduction

This chapter presents the research justification, problem statement, specific objectives, methods, and the structure of this document.

### 1.1 Justification

Citizen Science projects engage volunteers to collect and/or process data as part of scientific enquiries (SILVERTOWN, 2009) with diverse purposes, such as water pollution control, birds migration observation, discover of new celestial corps, or social experimentations. Specially in the Biodiversity domain, in the last decade, citizen science projects are helping to solve the issue of lack of data to face problems such as invasive species, habitat degradation, managing ecosystem services, and climate change (BISBY, 2000).

Large scale citizen science projects such as eBird (E-BIRD, 2016), iNaturalist (INATURALIST, 2016), iSpot (ISPOT, 2016), and REEF (REEF, 2016) have rapidly increased data collection because of their ability to recruit large numbers of participants and because people volunteer their time. However, citizens may lack the training about how to make observations and/or deal with situations in which they are uncertain about observations (ALABRI; HUNTER, 2010) (WIGGINS; CROWSTON, 2011) and doubts remain about the quality of citizen science data (SOURCE, 2008) (WIGGINS et al., 2011) (GALLOWAY; TUDOR; HAEGEN, 2006).

A central question then becomes "How can project leaders and users have trust in the data collected by citizens?". Whether or not to have trust in citizen science data is one of the many complex trust behavioral choices experimented in Citizen Science projects for biodiversity.

In fact, trust is an inherent building block of our society (FUKUYAMA, 1996) and it is experienced daily by people (BAIER, 1986). Trust is fundamental to social, interpersonal, and commercial relationships (BAUER, 2015; CHO; CHAN; ADALI, 2015; HARDIN, 2002; MCKNIGHT; CHERVANY, 2001). Societies in which members commonly cheat others and take advantage of them, a system of formal rules and regulations have to be negotiated, agreed to, litigated, and enforced, sometimes by

coercive means. This legal apparatus, serving as a substitute for trust, entails what economists call "transaction costs" (FUKUYAMA, 1996). These costs impact directly the economy by making business difficult to realize, personified in contracts, lawyers, certificates, and other bureaucratic items.

The buyer-seller relationship is a practical example of the economic impact of trust. Hopping to have a product or service of quality, replacement in case of failure, and a fair price, most part of customers rely in trust to buy their products or services, so that companies constantly measure trust in its brands as a portrait of customer loyalty. For instance, on Internet marketplace the brand or reputation (a kind of personal brand) are fundamental, once buyers sometimes do not have direct contact or do not know the sellers.

Generally speaking, we trust when engaging in a (social, interpersonal, economic) relationship, and there is always an associated risk that others (people, computers, societies, governments) may not behave according to our expectations (HARDIN, 2002). If there is no risk, then there is no reason to trust or distrust, because others will meet our expectations.

It is possible to use instruments such as contracts, laws, inspections, and social rules to mitigate risks, and increase the confidence to engage as a trustor in a trust situation. As the Russian proverb, made famous by Ronald Reagan, says, "Trust, but verify" (WIKIPEDIA, 2016). Trust and control (or verification) play complimentary roles in the basis for confidence in relationships (COFTA, 2007). However, we cannot have a relationship with 100% of control. We are not capable to control or verify all risks, so that we as humans would not even be able to face the complexities of the world without resorting to trust (LUHMANN, 1979).

Once trust supports, together with control, our behavioral decision to engage in a relationship, such as volunteers in a Citizen Science project, it is also important to measure it and cultivate it, as we do with control. Measuring levels of trust can reveal or anticipate losses of loyalty, customers, employees, voters, partners, friends or volunteers. Low levels of trust can be changed, and improved trust might well reduce disparities, increase access, improve outcomes (THOM; HALL; PAWLSON, 2004), or improve the use of citizen science data.

### 1.2 Problem statement

Although trust is everywhere in our daily lives, even in Citizen Science projects the word "trust" carries different meanings such as: honesty, security, quality, caring, integrity, benevolence, etc. that is context dependent. This variety of meanings invoked by "trust" may simplify our daily communication, so that people or machines do not have to explain whom they are or what they are saying every time; however, this variety also makes "trust" a complex concept to be defined by people and by science. Both tend to define trust according to their particular or momentary use of the word, or they tend to establish general definitions that leave aside uncountable meanings of the word.

Words that carry many meanings which change according to the context were termed "suitcase" words by Minsky (MINSKY, 2006). Conscience, experience, and morality are examples. Adopting this "suitcase" perspective may change the way as we define, model, and simulate trust. Instead of trying to capture all meanings of trust, or to create models that are superficial and cannot be used for certain meanings, the suitcase perspective may establish a model that allows us to handle different meanings of trust according to each situation.

This new approach to model trust intends to support people to identify, define, and simulate trust situations, with potential to handle different meanings of trust in Citizen Science in the Biodiversity domain, or in other daily situations, and this is the main contribution of this work.

# 1.3 Objectives

The main objective of this research is to model trust under a suitcase word perspective and apply it to a case study in the Biodiversity domain.

These objectives may contribute to science by proposing:

- 1. evidence that trust is a suitcase word, and how this can be modeled;
- 2. a comprehensive model with potential to handle many meanings of trust;
- 3. a less subjective method to measure trustworthiness;
- 4. a workable model in the "suitcase" perspective which can be simulated;

### 1.4 Materials and methods

Our research method started by splitting the main objective in two hypotheses to be tested: 1) Trust is a suitcase word and 2) Trust can be modeled and simulated under a suitcase word perspective with potential to handle different meanings.

Figure 1 shows a graphical representation of our research method. From a literature review, we conducted a mining process to collect evidence that trust is a suitcase word, and we collected conceptual foundations to support the formalization of a 'suitcase model' for trust. The literature review was directed by discussions with other scientists, such as Dr. Stephen Marsh and Dr. Piotr Cofta. After proposing the suitcase model, we created a replica of a simulator for trust, the PlayGround (MARSH, 1994), which was modified to carry the new model, originating the PlayGround v2.0. At the end, we ran a case study to simulate how our suitcase model could handle different meanings of trust in Citizen Science in the Biodiversity domain.

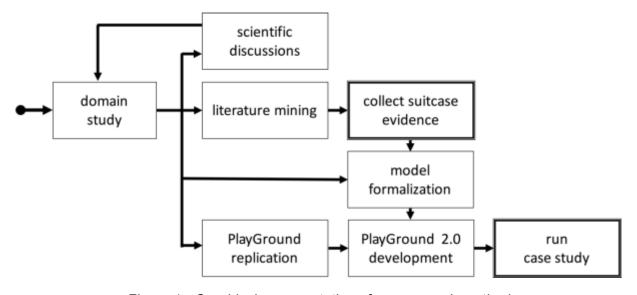


Figure 1 - Graphical representation of our research method

The following four main steps provide more details about our method to test the hypotheses and achieve the objective of this research:

Problem domain study: After initially contextualizing the problem in Citizen
Science in the Biodiversity domain and the objectives, we performed a
literature review across many disciplines of science (Computer Science,

Management, Psychology, Economy, Social Sciences, and others) to collect definitions, models, and conceptual foundations about trust. They helped us to refine our goals, realign our research method, and proceed with a new review. For instance, models from (MARSH, 1994) and (COFTA, 2007) were then more detailed because of their potential to be a good starting point to our suitcase model. After this review stage, we constantly followed trust research to consider additional relevant papers. Our work was influenced by scientific discussions with other scholars, such as Dr. Stephen Marsh and Dr. Piotr Cofta. These discussions took place during an interchange at University of Massachusetts - Boston, and during participations in international events, such as the Biodiversity Information Standards Conference (TDWG Conference, Costa Rica) and the International Federation for Information Processing – Trust Management Working Group (IFIPTM) Graduate Symposium (Germany). TDWG is the most important forum to discuss standards related to biodiversity data, and its 2016 conference took place in Santa Clara de San Carlos, Costa Rica. On another thread, after prior discussions about our work, we were invited by Dr. Marsh to participate in the 2016 IFIPTM Graduate Symposium on Trust and Trust Management (Darmstadt, Germany), which was followed by an opportunity to network with experts and other students in the field. Another important point of validation to our hypotheses was our complete paper being accepted in the IEEE 14th International Conference on Privacy, Security and Trust (Auckland, New Zealand).

2. Scientific literature data mining: We conducted a study with 'trust' articles from Thompson Reuter's Web of Science to collect evidences that trust is a suitcase word. We tested this initial hypothesis by splitting it in two subhypotheses: 1) "Once trust is present in many situations of our daily lives, a wide range of disciplines must study it" and 2) "Trust has different meanings across the disciplines". The first sub-hypothesis was tested with a distribution analysis using an article and journal-based classification from Web of Science. The second one was tested with a co-occurrence analysis of keywords supported by the VOSviewer (VOSVIEWER, 2016). Maps of keywords were created to compare the meanings of trust among disciplines.

Both analyses were then used as evidence to decide if trust could be considered a suitcase word.

- 3. **Modeling the suitcase model**: We proposed a framework to manage trust with three components: 1) a suitcase model to identify different meanings of trust according to each specific situation, which is the main purpose of this work, 2) a procedure to detail trust situations in terms of the suitcase model, and 3) a behavioral decision model of confidence to simulate cooperation, which is required to simulate the new model. We started by creating our suitcase model considering the conceptual foundations found in the literature review. After that, we proposed a sequence of instructions to define a real life trust situation in terms of the suitcase model. The last step uses the notion of the joint role of trust and control to produce a confident behavior (COFTA, 2007), and a modified version of Marsh's computational formalization of trust (MARSH, 1994) to create our generalized model of behavioral decision of confidence. This behavioral decision model is required to simulate trust effectively and verify its performance with different meanings of trust, once trust and control work together to produce a confident behavior and consequently a cooperation action.
- 4. Suitcase model application: We proposed a simulator to apply our suitcase model to a case study. We started by replicating the PlayGround in a web version. Results from this replication were important to test Marsh's formalism and served as basis to our modified version with the suitcase perspective, the PlayGround v2.0. This simulator was used to apply our new model to handle different meanings of trust in Citizen Science in the Biodiversity domain.

### 1.5 Thesis structure

This document is organized into eight chapters:

- Chapter 2 presents a scientific literature data mining aiming to collect evidence that trust is a suitcase word;
- Chapter 3 presents conceptual foundations of trust which were the basis to our proposed model;
- Chapter 4 describes the foundations of the suitcase model and the generalized model of behavioral choice of confidence;
- Chapter 5 presents a replication of the PlayGround simulator and how it was modified to support the suitcase perspective;
- Chapter 6 presents an application of our suitcase model by discussing a case study in Citizen Science in the Biodiversity domain;
- Chapter 7 discusses the results of this research and point the main challenges and contributions of this work;
- **Chapter 8** presents the conclusions and the proposed future steps of our research;

In addition, six appendixes complement this thesis.

- A Extraction of articles records from Web of Science describes how articles were extracted to literature scientific data mining, including download restrictions.
- B Hierarchy relation among Web of Science categories presents a complete hierarchy of categories used in scientific literature data mining;

- **C VOSviewer configuration** presents in details the settings used to produce the co-occurrence maps of keywords.
- D Maps of co-occurrence shows the co-occurrence maps of keywords from Operations & Management Science and Sociology research areas, and from Computer Science categories.
- E Results of Marsh's PlayGround presents the original results from a PlayGround experiment.

# 2 Scientific literature data mining

The main goal of this analysis was collect evidence in scientific literature that trust is a suitcase word. Our hypotheses to be tested in this analysis were 1 – if trust is a concept important in many situations of our daily lives, a wide range of disciplines must study it, and 2- trust has different meanings across the disciplines. We selected as source of our analysis the Thomson Reuters' Web of Science (WOS) database<sup>1</sup>. Besides the fact that it indexes respectable journals, from different areas of science, it is able to export complete records (title, abstract, citations, keywords, and references) for local manipulation. We searched for articles written in English language that contained the words "trust", or "trustworthy" or "trustworthiness" in the title, abstract or keywords. Appendix A describes how this search and the extraction of records were conducted.

To test hypothesis 1, the articles were organized in categories, originally assigned by WOS, that allowed us to identify from which discipline was each article was related. It was observed that there is not an authoritative area of knowledge for 'trust', once that articles were classified in a wide range of disciplines. The hypothesis 2 was tested by verifying the words that co-occur with 'trust' in each discipline. A premise for this test was that the different meanings of 'trust' could be identified by the words that occur with it. For instance, Computer Science articles use 'trust' with different meanings from Psychology articles, so that the words that co-occur with 'trust' in both disciplines might be different in each case. This was observed by creating maps of keywords for each discipline, which were different among the disciplines. The method applied and results observed are described in the next items.

# 2.1 Distribution among areas of knowledge

First step to test hypothesis 1 was to understand the categorization used by the Web of Science. Articles are assigned in main areas (Social Sciences, Life Sciences & Biomedicine, Technology, Physical Sciences, and Arts & Humanities), in "Web of

<sup>1</sup> The access to the platform was offered by the University of Sao Paulo, through agreement with the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) of Brazilian Ministry of Education.

Science Research Areas", and in "Web of Science Categories" (Figure 2). Research areas and categories are respectively article-based and journal-based terms assigned by Thompson Reuters. An article can have assigned more than one category and more than one research area if these categories come from different research areas. Web of Science does not offer a complete and explicit hierarchical relation between these terms (area and categories); however, we considered that they were related as presented in Table 21 in Appendix B, where we also provided a description of the method used to relate these terms.



Figure 2 - Adopted hierarchy of categories from Web of Science

From a total number of 55,249 articles selected<sup>2</sup>, trust-related papers occur in roughly equal percentage across Social Sciences (35%), Life Science & Medicine (29%), and Technology (29%), whereas the Physical Sciences and Art & Humanities, make up the remaining 6% of the papers (Figure 3).

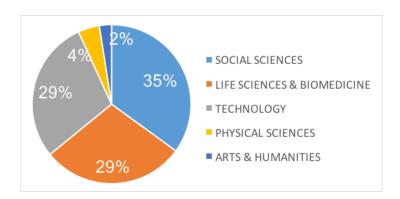


Figure 3 - Distribution of articles among main areas of Science in Web of Science

A more granular analysis (Table 1) shows that interest in trust occurs across 146 research areas and 252 categories. As an article may be classified in more than one category, including from other research areas, the number of appearances

<sup>2</sup> Accessed on November 17th 2016.

(92,765) is greater than the number of articles selected from the database (55,249).

Table 1 - Trust-related articles distributed in Web of Science hierarchy

Main Areas	Research Areas	Categories	Appearances	%
Social Sciences	22	43	32,477	35.0
Life Sciences & Biomedicine	73	95	27,000	29.1
Technology	21	52	26,950	29.1
Physical Sciences	16	36	4,050	4.4
Arts & Humanities	14	26	2,288	2.5
	146	252	92,765	

An analysis over the top 10 categories in terms of appearance shows that they are distributed among the three top main areas (Technology, Social Sciences, and Life Sciences & Biomedicine) and that interest in trust occurs across different categories (Table 2). For instance, in the domain area of Technology the topic of trust is important for Management, Computer Science Information Systems and Computer Science Theory Methods.

Table 2 - Top 10 Web of Science Categories by appearances of trust

#	Web of Science Category	Appeara nces	Main Area
1	Management	4,594	Technology
2	Business	4,236	Social Sciences
3	Computer Science Information Systems	3,800	Technology
4	Public Environmental Occupational Health	3,447	Life Sciences & Biomedicine
5	Economics	2,848	Social Sciences
6	Law	2,344	Social Sciences
7	Computer Science Theory Methods	2,296	Technology
8	Health Care Sciences Services	1,890	Life Sciences & Biomedicine
9	Medicine General Internal	1,853	Life Sciences & Biomedicine
10	Political Science	1,741	Social Sciences

Another interesting fact is a substantial number of articles that belong to one or more research areas. This is another indication that trust is not one authoritative category or research area but instead found across disciplines. For instance, 746 articles of WOS assigned to "Social Sciences - Other Topics" were also assigned to "Business & Economics". In Computer Science research area, 49.8% of 7,839 articles appear in more than one research area (Table 3).

Table 3 - Appearances of Computer Science (CS) articles in other Research Areas

	Total of Articles		Total of Appearances	
Appear only in CS Research Area	3,939	50.2%	3,939	29.5%
Appear in CS and 1 other Research Area	2,437	31.1%	4,874	36.5%
Appear in CS and 2 other Research Areas	1,359	17.3%	4,077	30.6%
Appear in CS and 3 other Research Areas	72	0.9%	288	2.2%
Appear in CS and 4 other Research Areas	32	0.4%	160	1.2%
	7,839		13,338	

# 2.2 Co-occurrence of keywords

After the preliminary analysis for hypothesis 1, we proceeded with the test for hypothesis 2: "Trust has different meanings across the disciplines". We assumed that the keywords of an article identify the meaning that trust carries on it, and that the articles of a specific discipline have similar meanings for trust. So that it was expected that the map of co-occurrence of keywords from articles of a specific discipline could identify the meanings of trust used in this discipline. Once these maps were different among the disciplines, we could validate the hypothesis 2.

We created the maps of keywords using a bibliometric tool, the VOSviewer version 1.6.5 (VOSVIEWER, 2016), which was developed by the Centre for Science and Technology Studies, Leiden University. This freely available software creates distance-based maps, which means that the distance between two items in the map reflects the strength of their relation. A stronger relation is represented with a smaller distance between two items. This feature can make it easy to identify clusters of related items (VAN ECK; WALTMAN, 2010), which are assigned by clustering techniques and

differentiated by colors. A direct relation between two items (co-occurrence) is represented by a line connecting both. Details about these techniques can be found in (WALTMAN; VAN ECK; NOYONS, 2010). Essentially, the graphics are generated over a network data already created, a network data imported from other software, or from bibliographic data (such as authors, references, citations, keywords, and organizations) extracted from scientific databases such as Web of Science.

VOSviewer offers a Network Visualization (Figure 4) in which the size of the label and the size of the circle are determined by the item's importance (frequency) among all items (VAN ECK; WALTMAN, 2010). It is also possible to set parameters to fine-tune this visualization, which are detailed in the user manual (VOSVIEWER, 2016).

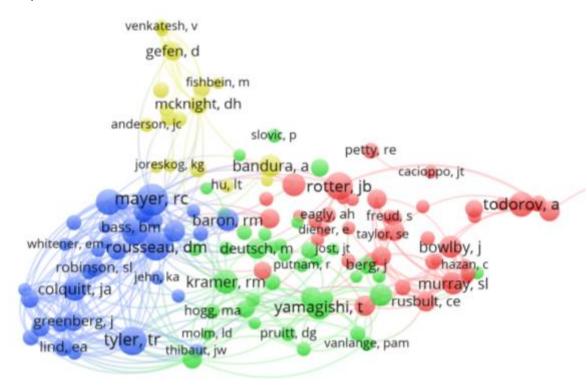


Figure 4 - Co-occurrence of authors in the Psychology research area

VOSviewer creates its maps in three steps: extracting the similarity matrix, calculating the coordinates of nodes, and then translating, rotating, and reflecting the map (VAN ECK; WALTMAN, 2010). The similarity matrix is obtained from a normalization of the co-occurrence matrix, which corrects the differences in the total number of occurrences or co-occurrences of items. Details about this normalization can be found in (VAN ECK; WALTMAN, 2009). The similarity matrix obtained in step one is an input to calculate the coordinates of all nodes using the "Visualization of

Similarities" (VOS) mapping technique, for which the higher the similarity between two objects, the higher the weight of their squared distance. The main goal of VOS is to minimize the weighted sum of squared Euclidean distances between all pairs of objects. A detailed description of the technique is presented in (VAN ECK; WALTMAN, 2007). This minimization does not have a unique solution, so that three transformations are applied at the end to ensure the same map from the same co-occurrence matrix.

# Co-occurrence maps

We extracted bibliographic data from Web of Science to create maps for four Research Areas: Psychology (Figure 5), Computer Science (Figure 6), Operation Research Management Science (Figure 29, Appendix D) and Sociology (Figure 30, Appendix D). A 'general' map with these four research areas was also created (Figure 7). Colors in this general map do not represent the research areas, like in other maps the colors represent clusters of keywords using techniques described in (WALTMAN; VAN ECK; NOYONS, 2010), that is, research areas are mixed.

We decided to get these areas to test our hypotheses of many meanings due to their large number of articles and their multidisciplinary. A more granular analysis of our hypothesis was done here using maps created from the five Computer Science categories: Information Sciences (Figure 31), Artificial Intelligence (Figure 32), Hardware (Figure 33), Software Engineering (Figure 34), and Theory and Methods (Figure 35). These maps are available in Appendix D.

See Appendix A for further details about bibliographic data extraction, Appendix C for details about maps settings on VOSviewer, and Appendix D for complementary maps.

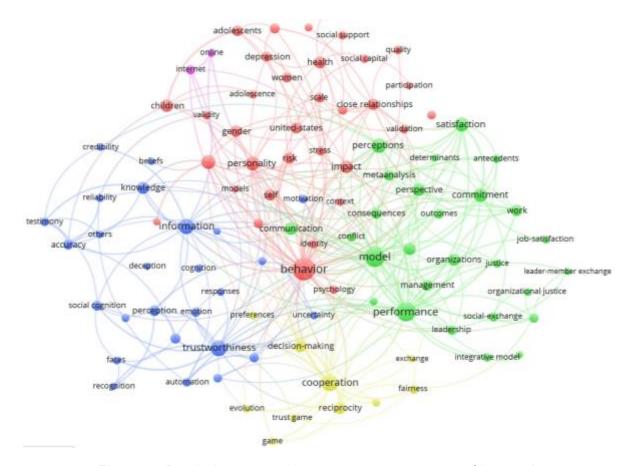


Figure 5 - Psychology research area co-occurence map of keywords

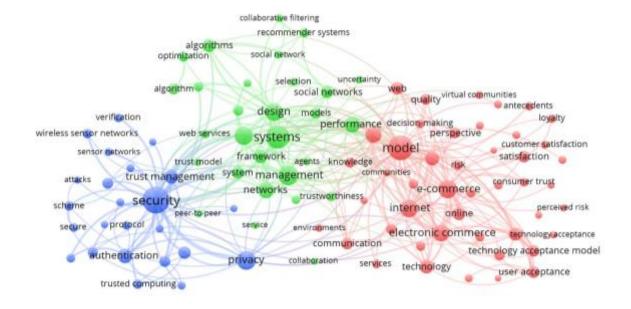


Figure 6 - Computer Science research area co-occurrence map of keywords

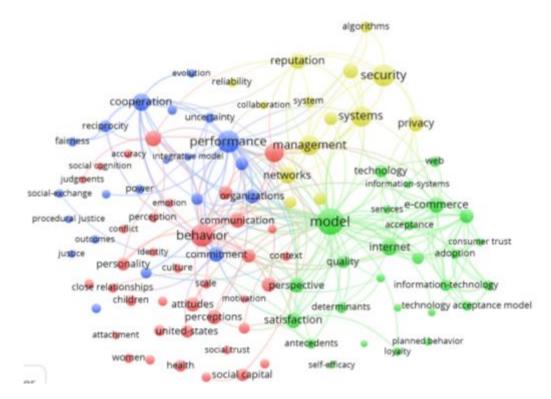


Figure 7 – Co-occurrence map of keywords from Computer Science, Psychology, Operation Research Management Science, and Sociology

# 2.3 Considerations about the literature mining

Distribution analysis showed that trust articles occur across a wide range of disciplines, and sometimes in more than one discipline the same time (Table 3, pg. 32). Considering the classification offered by the Web of Science, there is not a concentration of articles in one main area, Research Area, or WOS Category. For instance, Life Sciences & Biomedicine and Technology research areas have almost the same number of appearances, 27,000 and 26,950 respectively, and the difference between the first and the second in the rank is around 20% (Table 1, pg. 31). In an analysis of categories, the difference between different disciplines such as Computer Science Information Systems and Public Environmental Occupation Health is only 10% approximately (Table 2, pg. 31). Our literature analysis was able to confirm hypothesis 1 – "if trust is present in many situations of our daily lives, a wide range of disciplines must study it" - is true.

VOSviewer was very helpful to execute our co-occurrence analysis. Considering that different meanings of 'trust' could be identified by the keywords that occur with it, maps from Computer Science (Figure 6, pg. 35) and Psychology (Figure

5, pg. 35) research areas showed that trust meanings from these areas are significantly different. Computer Science meanings were mainly related to security, recommendation, management in social networks, and behaviors in e-commerce environment; whereas Psychology has meanings related to personality, cooperation, satisfaction, and trustworthiness. Even in a specific research area the categories may have different meanings. In Computer Science, Artificial Intelligence (Figure 32, Appendix D) has meanings like cooperation, communication, and similarity, whereas Software Engineering (Figure 34, Appendix D) has meanings such as cryptography, provenance, and reliability. We conclude from these different meanings across disciplines that our hypothesis 2 - "trust has different meanings across the disciplines" - is also true.

This analysis conducted from a respectable scientific database and method described above presented significant evidence about the suitcase perspective of trust.

# 3 Conceptual foundations of trust

Once the previous chapter showed the evidence that trust is a suitcase word, this chapter presents some conceptual foundations that support our model of trust under a suitcase perspective.

# 3.1 Many Definitions of Trust

Giving the many contexts in which trust appears in daily experiences, it is not surprising that trust is studied widely in applied and academic disciplines such as Business, Philosophy, Law, Psychology, Computer Science, Sociology, and Biology. One reflex of this multidisciplinary, associated whit the quantity of meanings, is that there is no consensus about what trust is.

Unfortunately for scholars (BAUER, 2015; CHO; CHAN; ADALI, 2015; JOSANG; ISMAIL; BOYD, 2007; MAYER; DAVIS; SCHOORMAN, 1995; MCKNIGHT; CHERVANY, 1996), there are dozens of definitions and models for trust, and some of them are considered elusive and ambiguous. See (BAUER, 2015) for a list of definitions. Furthermore, no one author, definition, or model is authoritative. Each area tends to view trust from its own perspective (MCKNIGHT; CHERVANY, 2001), focusing on sub-concepts to create its particularized version of trust (BAUER, 2015).

These definitions of trust are developed, in general, over empirical works that fit a determined type of research (MCKNIGHT; CHERVANY, 2001) or have a lack of concept-measurement consistency (BAUER, 2015). On the other hand, some initiatives created general models using as many definitions as possible, in attempt to escape from a narrow view. However, these models are highly conceptual in a way that are not workable.

Despite this plurality, some authors are largely cited as references across disciplines, such as (BAIER, 1986; BARBER, 1983; CASTELFRANCHI; FALCONE, 2000; GAMBETTA, 2000; HARDIN, 2002; LUHMANN, 1979; ROTTER, 1967). For instance, (MCKNIGHT; CHOUDHURY; KACMAR, 2002) and (MAYER; DAVIS; SCHOORMAN, 1995), originally from Management Sciences, are largely cited in Psychology Research Area, as observed in a co-citation analysis produced with the VOSviewer (Figure 4, pg. 33). (MAYER; DAVIS; SCHOORMAN, 1995) is the most

cited reference with 418 citations and (MCKNIGHT; CHOUDHURY; KACMAR, 2002) is the fifth with 279 citations from a universe of 177.794 citations in the area.

Models from (COFTA, 2007) and (MARSH, 1994) propose general views that encompass most part of other widely cited definitions, and although they are not suitable individually to be used with the suitcase perspective, together they enabled us to compose our new suitcase model. See item 3.9 for further details about these views.

## 3.2 Actors, Risk & Context

This item brings some ideas about actors, risk, and context related to trust.

### <u>Actors</u>

Actors lead a trust situation. The trustor is the actor that has an attitude of trust, mistrust, or distrust, toward another actor, the trustee. In its turn, the trustee "can be anyone who may perform a task, upon whom one might wish to rely" (O'HARA, 2012).

Trustor and trustee can be the same person when you trust yourself. Both of trustor or trustee might be humans, but they can be also machines, software, animals, roles (doctor, lawyer, teacher, football referee, etc.), institutions, and others. The use of non-humans in trust situations depends on the understanding of their intentions, capacities and incentives. For example, "A piece of software has no intention to work in anyone's interest, although it may be intended to work in the interests of those who have paid for it" (O'HARA, 2012). We decide not to turn on philosophical discussions about it, so that our suitcase's model of trust is open to any kind of actor.

Sometimes a third actor can be part of a trust situation. This third person is "someone with the credentials to issue a representation of behavior to which the trustee will conform" (O'HARA, 2012); However, the trustor has to give this authority to him and trust him. At this point, we establish a recursive trusting behavior. The presence of the third actor in trust situation is frequently associated in social networks and online services studies (GOLBECK, 2006; JOSANG; ISMAIL; BOYD, 2007).

### Risk

There is not a consensus in literature about the relationship between trust and risk, if risk is a requisite/antecedent to trust, is trust, or is an outcome of trust (MAYER; DAVIS; SCHOORMAN, 1995). However, for (LUHMANN, 1979) trust presupposes a situation of Risk. Without risk, trust does not exist, because others will meet our expectations. Without a possibility to choose, there is not a trust situation. The trustor can choose to accept the risk, or not. Once taking a behavior decision to trust, the trustor is willing to accept the risk (LUHMANN, 1979), and he is willing to risk some of his assets in the transaction (O'HARA, 2012).

### **Context**

Context are restrictions of the circumstance in which the trustee is claimed to be trustworthy. It is used to give importance to a trust situation and delimitate the horizon of our assessments. The restrictions might be a task, a particular role, or a specific people. Besides that, we can add environmental events, economic circumstances, among many other factors to specify the context. However, it has to be delimited with useful information.

Furthermore, a capacity to trust in some contexts would not imply a capacity to trust elsewhere (O'HARA, 2012). The trustee can be trustworthy in a context, but not in another. For example, we can trust someone drives safely in good weather conditions, but not in a context of pressure with rain, wind, and traffic.

# 3.3 Situation, Belief, Probability, and Behavior Decision

In a "Trust Situation" the trustor has to decide about the future. The trustor expects a specific behavior from the trustee, and there is always a risk involved that may take place that the trustee does not behave such as expected. It is not certain that this risk will occur because the action is in the future. The trustor acts by deciding to trust, mistrust, or distrust the trustee (see item 3.7 for definitions). This behavioral decision is based on the trustor's probability to take risks, the importance of situation and the involved risk (context-dependence), and the trustee's capacity and interest to

behave such as expected. Probabilities are involved in this behavior decision because the results are over a promised action, that can be realized, or not.

When our perception about a trustee's capacity and interest to behave as we expect are high, and our perception about the risks are low, our confidence to engage in a trust situation will be high. However, even if we are highly confident, not necessarily we produce a trusting behavior. We can trust that airplanes are safe, but produce a trusting behavior only when flying one.

Trust that drivers will respect the law and stop their cars before the crosswalk is an example of everyday trust situation. Ready to cross a street, pedestrians have to assess the probability that drivers have ability to operate the car, that drivers are not intending to deliberately hit them, that drivers are paying attention to the way, and so on. The risk of being hit by a car if the driver does not stop is very significant. We can collect some evidence about these expectations, looking for how they conduct their cars, if they are slowing down, if they are looking at the signal and us, and so on. However, we will not have all information to get the perfect assessment. Our decision will be based on the best judgement with the available data, composed with our capacity to take risks in general.

# 3.4 Three-part relation

Trust can be explained by a three-part relation (BAIER, 1986; HARDIN, 2002; LUHMANN, 1979). This relation is translated as "a trustor A that trusts, i.e. judges the trustworthiness of a trustee B with regard to some behavior X" (BAUER, 2015). A behavior in this case can also be interpreted as an action. (BAUER, 2015) named this relation as 'ABX', where A and B are actors and X the expected behavior.

# 3.5 Group and Cooperation

Trust is a major aspect of cooperation of any kind (MARSH, 1994). Agents working together implies Cooperation, when people "work and act together on a task, and share profits or benefits from doing so"; implies Collaboration, when "one works jointly with other on a task"; and implies Coordination, when "the act of causing parts

to function together or in a proper order". However, cooperation is not necessarily a result of trust, while trust is a result of cooperation-based mutual benefit of helping each other (GAMBETTA, 1988). Group formation is a deep and complex phenomenon, which will not be discussed in this material. However, we will consider that groups may be formed by cooperation behavior.

# 3.6 Types of Trust

Trust has many "types" (or "classes" or "sub-concepts"), such as Situational, Social, Interpersonal, Political, and General. Although they carry different meanings and have distinct limits, ABX relation can explain all of them. For instance, in case of Political Trust, which express our belief that the political class will behave such as expected in many contexts (O'HARA, 2012), B is set as the political class and A and X can vary according on the context. In Interpersonal Trust, which is an expectancy held by an individual or a group about the word or promise of another individual or group (ROTTER, 1967), A and B are individuals or groups for any expected behavior X. In general, these types have a simple variation of actors and expected behaviors.

Other types of trust are essentially related to non-rational components, results of traits of personality forged by family relationship, religion, or past remarkable experiences, such as moral, faith, reciprocity, and others. These types such as Basic Trust (MARSH, 1994), Disposition to Trust (MCKNIGHT; CHERVANY, 2001), and Propensity to Trust (MAYER; DAVIS; SCHOORMAN, 1995) can be explained also by an ABX relation. Nevertheless, in this case, A and B are the same actor and X is irrelevant.

In an attempt to study the variety and impact of types of trust in time, we performed an n-gram analysis with Google Books Ngram, which is an online search engine that charts frequencies of any set of comma-delimited search strings using a yearly count of n-grams found in sources printed between 1500 and 2008. Further details about the tool operation can be found in (GOOGLE, 2017). An n-gram is a contiguous sequence of n items such as syllables, letters, or words. For instance, 'suitcase word' is a 2-gram (or bigram). Figure 8 shows trends in seven bigrams from 1900 to the early 2000's.

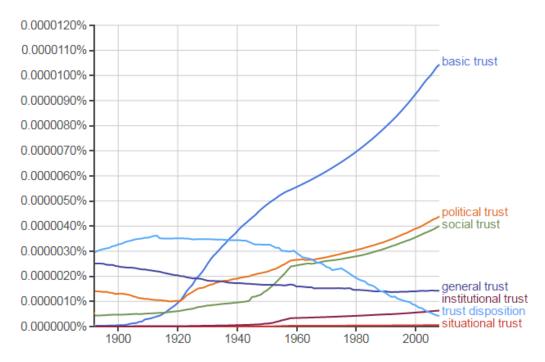


Figure 8 - Ngram distribution of types of trust

The x-axis shows the percentage of these bigrams over all the bigrams contained in Google sample of books written in English and published in the United States. Comparing terms among them, General trust, Political trust, and Trust disposition have been largely used since 18th century, although the use of the last one is in a decline since the 1940's. Considering that Trust disposition and Basic trust have similar use, we can assume that the first has been replaced by the second. The use of Political and Social trust started to rise after the 1920's, and increases after the 1960's, reflecting the economic and political issues of the period.

In general, we may consider that types of trust are reactions to temporal facts and reflect meanings of trust that are important at that time.

# 3.7 Trust dynamics and measurement

Distrust is not a complement for trust. If we cannot trust someone, it does not mean that we automatically distrust this person. We must have reasons to do it. Sometimes we do not have enough information to decide about the situation (Trust or Distrust). We can be in a state of complete ignorance, or we can have weak evidence to trust or distrust that cannot pass the "Limit of forgivability" (Undistrust) or the "Cooperation threshold" (Untrust).

These limit and threshold are change points in trust dynamics (Figure 9). We can treat the behavioral decision to trust as a binary decision in which trust is evaluated in either discrete or continuous scale. When trust surpasses the Cooperation threshold (trust) we can produce a cooperative behavior, whereas a value lower than a Limit of forgability (distrust) produces a non-cooperative behavior (CHO; CHAN; ADALI, 2015).

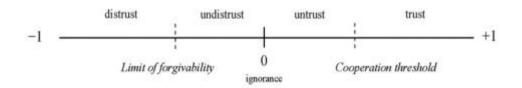


Figure 9 – The trust continuum, extracted from (CHO; CHAN; ADALI, 2015)

In a general way, we can define the states of trust as:

- Trust trust values are equal or greater than a Cooperation threshold;
- Untrust trust values are positive but not enough to reach the Cooperation threshold;
- Undistrust trust values are negative but greater than a Limit of forgivability;
- Distrust trust values are lower than the Limit of forgivability;

Daily experiences show that transitions among these states are not equal. In general, it takes time to trust someone, but it is easy to distrust it; and once in a distrust state, return to a trust state is harder than before.

Trust has "been broken, misplaced, abused, shaken, and violated. Occasionally it is repaired and rebuilt", but if we do not measure it, "we can ignore it, fail to cultivate it, and ultimately lose it" (THOM; HALL; PAWLSON, 2004).

### 3.8 Trustfulness and Trustworthiness

This item presents trustfulness and trustworthiness concepts.

### **Trustfulness**

Trustfulness is the trustor's general capacity to take risks (TULLBERG, 2007). This capacity is very similar to "disposition to trust" (MCKNIGHT; CHERVANY, 2001) and "propensity to trust" (MAYER; DAVIS; SCHOORMAN, 1995). Trustfulness concerns only the trustor and it is independent of trustee assessment. For example, one can have all reasons to trust in someone, but is not capable to take the associated risk for reasons that are related to himself. The opposite view can be also true; in certain cultures, a family member is always trusted, independent of assessment.

Trustfulness is essentially compounded by non-rational components, results of traits of personality forged by family relationship, religion, or past remarkable experiences, such as moral, faith, reciprocity, and others. (MCKNIGHT; CHERVANY, 2001) said that the "disposition to trust [such as the trustfulness] comes primarily from trait psychology, which says that actions are modeled by certain childhood-derived attributes".

Religion also affects trustfulness in many ways as a behavior guide. As examples, "for religion, in the absence of proof, believing is necessary", or "If you believe that a potential cooperator believes she will be punished if cheating on you, this is a reason for you to be trustful" (TULLBERG, 2007). Faith in humanity is a belief that others, without assessment, are usually competent, benevolent, honest/ethical, and predictable (MCKNIGHT; CHERVANY, 2001).

Some disappointments or good experiences can affect trustfulness, but it tends to be stable across situations (MAYER; DAVIS; SCHOORMAN, 1995) and people (MCKNIGHT; CHERVANY, 2001).

# **Trustworthiness**

Trustworthiness and trust are distinct concepts that should be viewed as interrelated concepts. They are independent because one can possess trustworthiness

even when there is no call for trust (BAUER, 2015). However, trustworthiness is an antecedent of trust (MAYER; DAVIS; SCHOORMAN, 1995), so that trustworthiness is essential to understand trust (O'HARA, 2012).

If someone is motivated to do what we trust him to do, then this person is trustworthy (HARDIN, 2002). In (O'HARA, 2012) words, trustworthiness is "the property that they will do what they say they will do. If they fail, then it will typically be for some reason outside their control", and "a person is trustworthy when she does what she says she will do."

(MCKNIGHT; CHERVANY, 2001) suggest that this property can be measured with 16 trust-related characteristics, organized in second-order conceptual categories such as predictability, integrity, benevolence, and competence. However, these characteristics, even wide, may limit the meanings of trust considered.

### 3.9 Models of Trust

This item present in details (MARSH, 1994) and (COFTA, 2007) models of trust, which were used as basis to model and simulate trust under a suitcase perspective.

#### Stephen Marsh's model of trust

Marsh's thesis (MARSH, 1994) is widely cited by computer science scholars, especially among those studying social networks, peer-to-peer systems, multi-agent systems, e-commerce management, and wireless sensor networks. His work develops a formalization and a computational model for trust, implements the model in a simulator called PlayGround and undertakes some experiments to demonstrate its feasibility.

The formalism (Table 4) is centered on the agent concept developed in Artificial Intelligence (AI) (MARSH, 1994). The key idea is to extend an AI agent's reasoning capabilities to reproduce trust behavior - that is to decide when and with whom to cooperate. The top-down methodology adopted by Marsh does not start with a particular definition of trust but instead, observes "the phenomenon and attempt to define a formalism which behaves in the same manner".

Table 4 - Summarized notation of Marsh's formalism

Name	Notation	Range	Description
Situations	α,β,	S	→ is a point in time for a specific agent
Agents	a, b, c,	A	→ concept developed in Artificial Intelligence
Societies	S <sub>1</sub> , S <sub>2</sub>	$S_x \in A$	→ agents grouped according to some metric
Utility	$U_x(\alpha)^t$	[-1,+1]	ightarrow utility of situation $lpha$ to agent x
Importance	$I_{x}\left(\alpha\right)^{t}$	[0,+1]	$\rightarrow$ importance of a situation $\alpha$ for agent x
Knowledge	$K_{x}(y)^{t}$	True / False	→ x met y at time t

To demonstrate the feasibility of his formalism, Marsh proposed an application in terms of a decision model for trust. There are three types of trust in this application: Basic, General, and Situational (Table 5). The Basic Trust (BT) derives from agent's past experiences, but does not have a connection with other agents, a specific situation, or the environment. Simply, "Alice trusts", and this can be applied to anything. However, it may increase or decrease in the time considering her new experiences of life. The General Trust (GT) denotes how much an agent trusts another in a complete perspective: "Alice trusts Bob". The GT represents a moment in the time, whereas the Estimated General Trust (EGT) considers all interactions between Alice and Bob in the time. Otherwise, the Situational Trust (ST) represents the trust in an agent in a specific situation: "Alice trusts Bob in this situation".

Table 5 - Types of trust in Marsh's formalism

Name	Notation	Range	Description	
Basic	$T_{x}^{t}$	[-1,+1)	→ basic disposition to trust someone or something	
General -	$T_x(y)^t$	[-1,+1)	$\rightarrow$ how much x trusted y at the time t	
	$\widehat{T_x(y)}$ )	[-1,+1)	→ how much x trusted y (estimated from all interactions)	
Situational	$T_x(y,\alpha)^t$	[-1,+1)	$\rightarrow$ how much x trusted y in situation $\alpha_x$ at the time t	

.

<sup>3</sup> Alice and Bob are widely used as personas in security problems.

To cooperate in a situation  $(\alpha)$ , Alice's (x) trust in Bob (y) must be greater than a threshold (eq. 1). This Situational Trust  $(T_x(y, \alpha), eq. 2)$  depends on the Importance  $(I_x(\alpha))$  and the Utility  $(U_x(\alpha))$  of the situation for Alice, that are respectively a rational economic and a subjective judgments of the situation, and on the Estimated General Trust (EGT or  $\widehat{T_x(y)}$ ) of Alice in Bob.

$$T_x(y,\alpha)^t > Coop\_Threshold_x(\alpha) \implies Will\_Cooperate(x,y,\alpha)$$
 (1)

$$T_x(y,\alpha)^t = U_x(\alpha) \times I_x(\alpha) \times \widehat{T_x(y)}$$
 (2)

The General Trust (GT or  $T_x(y)$ ) denotes how much an agent trusts another agent in general when the decision is being made without reference to a specific situation. For instance, Alice may say "I trust in Bob" based on an analysis of historic interactions with him.

The definition of the EGT depends on Alice's memory span and her Disposition to Trust. Alice can use the highest past value of GT in Bob in a similar situation (Optimistic disposition), the lowest value (Pessimistic), or a mean of values (Realistic, eq. 3) to define the EGT.

$$\widehat{T_{x}(y)} = \frac{1}{|A|} \sum_{\alpha \in A} T_{x}(y)$$
 (3)

A in eq. 3 represents the set of all situations similar to  $(\alpha)$  which Alice has experienced with Bob. If Bob is unknown in a similar situation, GT past values in any situation are used to define the EGT. When there is not a past value for GT, then Alice's Basic Trust (BT) is used.

The Cooperation Threshold (eq. 4) is an estimate of how much Perceived Competence and EGT mitigate the Perceived Risk, multiplied by the Importance of the situation to Alice. The Perceived Risk and the Perceived Competence of Alice in Bob depend on the knowledge between them from previous interactions. See (MARSH, 1994) for further details about formulas and concepts definitions.

Coop\_Threshold 
$$_{x}(\alpha) = \frac{\text{Perceived\_Risk}_{x}(\alpha)}{\text{PerceivedCompetence}_{x}(y,\alpha) + \widehat{T_{x}(y)}} \times I_{x}(\alpha)$$
 (4)

Despite its benefits, avoiding a particular definition for trust brings some important weaknesses to the formalism. As pointed by (MARSH, 1994), "the kind of trust we define (implicitly or explicitly) using the formalism, that is, the kind of behavior we produce from trusting agents, is indicative of only one subclass of trust (and there are many)". Furthermore, "however abstract the formalism may be, it still has to be founded on some aspect of trust, which is one of many types".

The "resultant behavior" produced by Marsh's formalism is a representation of trust, but does not do "a deep consideration of the aspects of a particular situation". For him, this is a strength of the formalism, since it is not limited by a specific situation, making it generalizable. Here, we partially agree with him. The formalism must be simple and generic and an attempt to consider all aspects of situations involving trust would be a disaster in terms of complexity and practicality. Thence, the "loss of specificity is, from this viewpoint, no bad thing", and can be accepted by most part of situations. However, the formalism must be also extensible to consider additional aspects of trust when necessary, once not all the meanings of trust are represented.

### Piotr Cofta's model of trust

(COFTA, 2007) aims to understand how confidence is built and how it is influenced by others. Confidence is presented as "the belief that the other part will behave in a way that is beneficial and desired for the first party", allowing people to engage in relationships by externalizing a set of actions (confident behavior). The trigger to the confident behavior can be defined by a limiting factor. When trustor's confidence value is higher than this threshold, the confident behavior can be produced.

Cofta's model for confidence is a self-referring sum of control and trust, driven and limited by complexity. Control and trust work together, and no one alone is an alternative. Figure 10 shows a graphic representation of how a confident behavior is produced in this context, which is also formalized in equations.

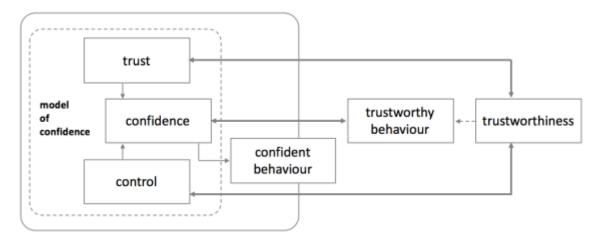


Figure 10 - Confidence behavior model, extracted from (COFTA, 2007)

On the assumption that confidence, and consequently the confident behavior, is produced by an assessment about trustee's trustworthy behavior (eq. 5), trust is defined as an estimate about trustee's trustworthiness (eq. 6), which is defined as the trustee's intentions toward the trustor in the transaction. The gap between these intentions and trustee's trustworthy behavior can be enforced by instruments of control (eq. 7).

$$conf_{trustor \rightarrow trustee} = est_{trustor}(twb_{trustee})$$
 (5)

$$trust_{trustor \to trustee} = est_{trustor}(tw_{trustee})$$
 (6)

$$ctrl_{trustor \rightarrow trustee} = est_{trustor}(twb_{trustee}) - est_{trustor}(tw_{trustee})$$
 (7)

#### where:

- est trustor the function to produce trustor's estimate of its argument
- *conf* confidence
- *twb* trustworthy behavior
- *tw* trustworthiness
- *ctrl* control

Considering all elements of those formulas in terms of subjective probability and numeric values, Cofta defines:

- Trust as an "estimate of a subjective probability of certain desired course of action in the absence of control".
- Control as "the difference between the estimate of the subjective probability of trustworthy behavior and trust."

One of the key points of Cofta's model is the complexity-horizon exchange. A trustor cannot include everyone and everything in his confidence behavior decision. The extent of the horizon for a transaction depends on several factors, such as physical and psychological predispositions, current cognitive capabilities, tiredness, etc., including trustor's ability to manage the number of entities and the relationships between them. The trustor has to eliminate irrelevant entities until he reaches an optimal complexity that provided a certainty for his decision.

Another key point is that trust and control require evidences (opinion, fact, signal), and the significance of each evidence is attenuated depending on the perceived honesty of the source. According to Cofta there are three classes of evidences for trust: continuity, competence, and motivation; and there are three classes for control: influence, knowledge, and reassurance. See (COFTA, 2007) for further details about these categories of evidence.

Recursion is also a key point for Cofta's model based in the "basic building block" for confidence (Figure 11). In this block, numbers from 1 to 5 are inputs to trust and control processing, and output 6 is the result of those inputs. The operators  $\oplus$  and  $\otimes$  have the intuitive meaning of addition and multiplication, respectively; however, "this does not imply that actual semantics of those operators or any numeric metric of confidence".

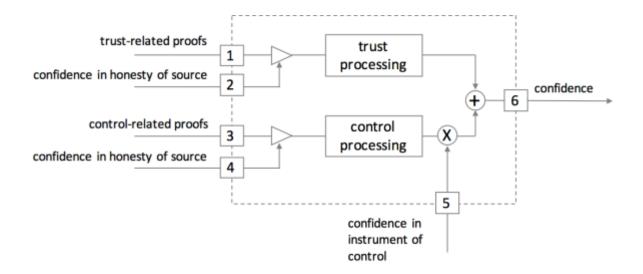


Figure 11 - Basic building block for confidence, extracted from (COFTA, 2007)

Trust and control processing depend on proofs (or evidences) that come from honest sources. The confidence in source honesty may be assessed with another block, i.e. entries 2 and 4 can be another building block for confidence. However, confidence in instruments are different from confidence in the trustee; while the trustee is expected to behave in a manner that is beneficial to trustor, instruments are only expected to be able to control the trustee. Figure 12 shows an example of recursion extracted from (COFTA, 2007). Alice must assess her confidence in three entities: Bob, Carol, and Dave. Carol is the main source of evidence about Bob, and Dave is an instrument to control him.

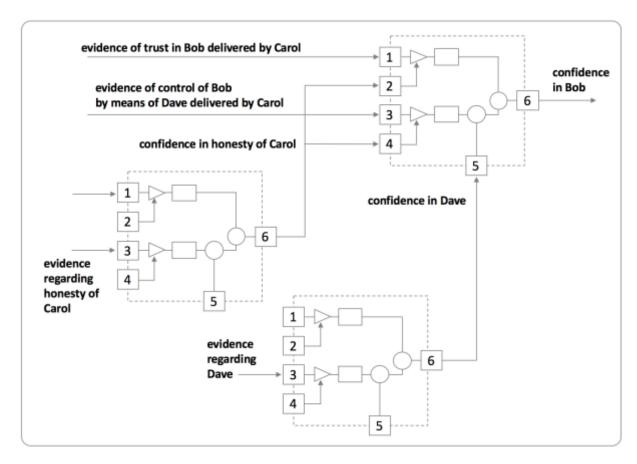


Figure 12 - Building blocks within a horizon of three entities, extracted from (COFTA, 2007)

The basic building block for confidence is part of the confidence behavior model presented in Figure 10. Table 6 presents a summary from the formalization proposed by Cofta, extracted from (COFTA, 2007). See pg. 36 in (COFTA, 2007) for further details about the formalism.

Table 6 - Summarized notation of Cofta's formalism

Name	Notation	Description
World	W	$W = \{A, E\}$
Entities	A	$A = \{ Alice, Bob, \dots, a_1, \dots x \}$
Entity	$a_1$	$a_1 = (name, cplx)$
Pieces of Evidences	Е	$E = \{e_1, \dots, e_n\}$
Evidence	$e_1$	$e_i = (a_k, a_l, subject, conf_{a_k \to a_l}^{subject}, cert_{a_k \to a_l}^{subject})$

#### where:

- $a_k$  entity that holds evidence
- $a_l$  entity that an evidence is about
- cplx complexity that is associated with dealing with such an entity
- *subject* label that is used to distinguish evidence
- $conf_{a_k \to a_l}^{subject}$  level of confidence in an evidence
- ullet  $cert_{a_k o a_l}^{subject}$  level of certainly about confidence in an evidence

Using the formalism, a trustor (Alice) confidence in a particular trustee (Bob) can be described as:

$$conf_{Alice \to Bob}^{subject} = trust_{Alice \to Bob}^{subject}(Et) \oplus \sum_{i=1}^{n} ctrl_{Alice \to Bob}^{subject}(Ec|y_i) \otimes conf_{Alice \to y_i}^{control}$$
(8)

#### where:

- *trust* (*Et*) level of trust from available evidence of trust *Et*
- $ctrl(Ec|y_i)$  level of control using instrument  $y_i$ ; from evidence Ec

$$Et = \{ et_x^{subject} \otimes conf_{Alice \to x}^{truth} \}$$
 (9)

$$Ec = \{ ec_x^{subject} \otimes conf_{Alice \to x}^{truth} \}$$
 (10)

#### where:

- $et_x^{subject}$  evidence related to trust in subject that is coming from x
- $ec_x^{subject}$  evidence related to control in subject that is coming from x
- $conf_{Alice \rightarrow x}^{truth}$  level of Alice's confidence the honesty of x
- $et_x^{subject} \in E_x^{subject}$  and  $ec_x^{subject} \in E_x^{subject}$

Our literature review across many disciplines of science showed us that trust is a widely discussed concept, with dozens of definitions and models, but no one of them is authoritative. However, the some authors are largely cited among disciplines, and have definitions and models that may support the suitcase perspective, such as (TULLBERG, 2007), (BAUER, 2015), (MARSH, 1994) and (COFTA, 2007). This chapter showed details about:

- the importance of the context to trust;
- how a trust situation and the types of trust may be decoded in a threepart relation;
- the importance of trustfulness and trustworthiness to trust decision;
- a description of the model and formalism of Cofta's and Marsh's works;

These concepts were essential to model trust under a suitcase perspective, as presented in the next chapter.

# 4 Modeling trust under a suitcase word perspective

In real life, suitcases may have many objects with different purposes. We take out these objects according to a specific need, and to do it, we must be capable to identify the objects and operate their functionalities. Model trust under a suitcase word perspective (i.e. a suitcase model) is an opportunity to go beyond a restrict trust model, and use a procedure to help people, scientists, governments, and companies to identify, decode, measure, and consequently, manage different meanings of trust according to each situation.

To achieve these goals, the suitcase model must go beyond (MARSH, 1994) and (COFTA, 2007) models and define what is expected as "General Trust" (Table 5, pg. 47) or "trust-related proofs" (Figure 11, pg. 52) for each situation. The new model must consider, as a guide, what is expected and what is evidenced in each situation, to measure trust between the trustor and the trustee. Without that, the measures may treat distinct situations in an equal and erroneous way.

Inspired by this analogy, we propose a framework with three components: 1) a suitcase model to handle different meanings of trust, 2) a procedure to define trust situations in terms of the suitcase model, and 3) a behavioral decision model of confidence to simulate cooperation (Figure 13). These components work together, since we need a reference (model) to define and extract trust values from real life situations, and then simulate cooperation or grouping.

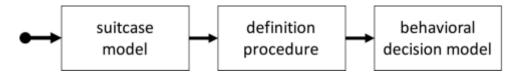


Figure 13 – Proposed framework to define trust under a suitcase perspective

Next items present these components and how they may work together.

### 4.1 Suitcase model

The suitcase model should be able to identify and structure in components as many meanings of trust as possible, making clear what are the expected behaviors in each situation.

Our proposal considers the conceptual foundations presented in previous chapter to combine them to support the suitcase perspective, building a more comprehensive model, which is one of the main contribution of this work.

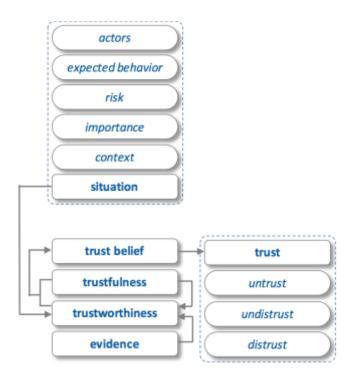


Figure 14 - Suitcase model components and their relations

Figure 14 shows the components of the new suitcase model. Broadly speaking, decision to trust (or distrust, untrust, undistrust) is a belief constructed from trustfulness and trustworthiness. Trustworthiness in its turn is a function of the evidence collected with influence of trustfulness. These components and their relation are described in the next items.

#### 4.1.1 Trust belief

Trust is a belief that the counterpart will behave as expected in a context.

#### 4.1.2 Trust and Confidence

Trust is not confidence. Trust works together with control to improve confidence

(COFTA, 2007), and consequently produces a confident behavior after a behavioral decision. A confident behavior allows engaging in a relationship, which may derive to a grouping phenomenon. We changed Figure 14 to include the relation between the suitcase model and the other components that produce an action of cooperation (Figure 15).

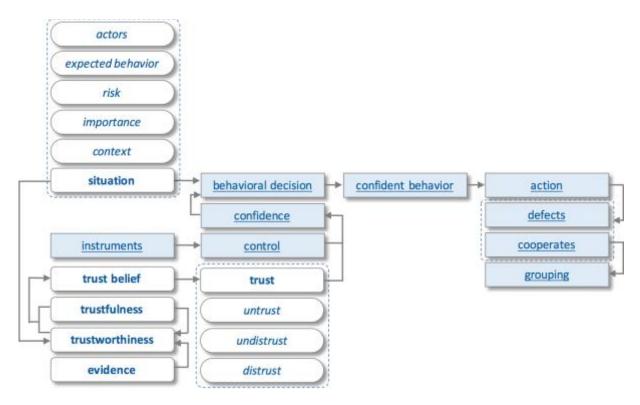


Figure 15 - Confident behavior components (blue) and their relations with the suitcase model of trust

The component 'confidence' in Figure 15 was originally described by (COFTA, 2007) using Figure 11 and eq. 8. Other components, i.e. the 'behavioral decision' and their subsequent components, are described in item 4.3.

#### 4.1.3 Trust situation with a behavioral decision of confidence

We expanded the original idea of three-part relation (ABX) from subsection 3.4 to include the control component (Figure 16). Since, control work together with trust to produce confidence, and consequently a cooperation action, it seems more workable

for trust simulation if all components that determine a behavioral decision of confidence are represented in the same situation. This new relation was named as ABX+C.

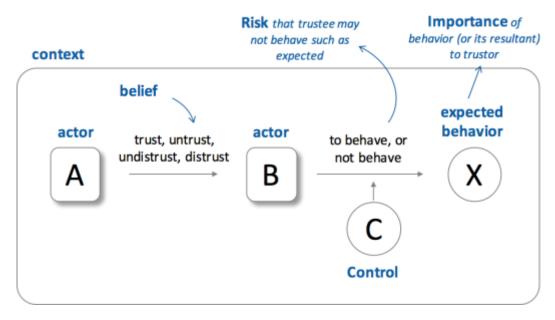


Figure 16 – ABX+C relation, an expanded three-part relation of trust

In this relation, actor A believes that actor B will behave such as X in a specific context, but there is a risk that B may not behave such as expected. A can be any type of actor (humans, societies, machines, animals, etc.) capable to trust another actor. The specific context, which is composed by known variables, such as political, economic, emotional, environmental, and others, can be simplified as 'favorable', 'neutral, or 'unfavorable' to reduce complexity of trustworthiness computation.

The importance of the expected behavior is also part of the situation. For instance, in situations where the importance is very low, we can accept higher risks. This risk does not have connections with trustee's capabilities and interests to behave such as expected. Risk in ABX+C situation is related to undefined and unexpected factors that are outside control, but can cause the trustee to fail.

Importance was also simplified in categorical values: favorable, neutral, or unfavorable for context, and high, medium, or low for importance. This was done to reduce the complexity of analysis, and minimally reduce the subjectivity of values. It is unpractical to decide how the many variables of the environment could affect trustworthiness and behavioral decision of confidence.

Control (C) is our contribution to the original ABX relation inspired by (COFTA,

2007). Instruments of control represented by C, such as law, influence, force, contracts, social rules, or hierarchy, can enforce B to behave such as expected (X). The inclusion of control inspires the name of our three-part relation version (ABX+C).

### 4.1.4 Trustworthiness computation

Trustor's belief about the future is defined by trustor's trustfulness and trustee's trustworthiness.

Trustfulness is the trustor's capacity to take risks, and it will be considered as a subjective probability in the scope of this work, i.e. trustfulness will be a probability derived from a personal judgement about the trustor's capacity to take risks in the future.

Trustworthiness is a probability about the trustee's capacity and interest to behave such as expected in the future. (COFTA, 2007) uses 'trust-related proofs' to compute this probability (see input 1 in Figure 11, pg. 52). Proofs are also related with control, see inputs 3, although in this work we consider control as a subjective probability.

The role of these proofs (or evidence)<sup>4</sup> is to give to the trustor a verification or confirmation that the trustee had capacity and interest to behave such as expected in the past.

It is not the role of an evidence, in our proposed computation, to give an opposite confirmation about the trustee capacity and interest, i.e. if we expect an evidence that Bob can drive a car, only the times when he did a successful road drive test will be collected as evidence. The other many times when he failed will be ignored to the computation. We adopted this strategy because the opposite confirmation can be part of an evidence: Bob passed the road test without fail.

Evidence may be about the expected behavior directly, or about other evidence that may lead to it. For instance, in a situation where Alice expects Bob to be a good student, Alice will remember past situations when Bob was her student, or she will expect evidence about his interest to learn, attention in class, and respect with the

<sup>4</sup> Evidence is "the available body of facts or information indicating whether a belief or proposition is true or valid". Proof, verification, attestation, and support for, are synonyms of evidence. Source: Oxford Dictionaries. Accessed on February 4<sup>th</sup> 2017 at <a href="https://en.oxforddictionaries.com/thesaurus/evidence">https://en.oxforddictionaries.com/thesaurus/evidence</a>.

teacher. This leading evidence are important for new situations, i.e., situations that have not be experienced by them as trustor and trustee.

In new situations or if Alice and Bob did not know each other, evidence can be collected by provided recommendation by others, or a reputation score in an ecommerce platform, for example. In these cases, the trustor must be confident to use the source, generating an iterated trust assessment. If the trustor distrusts the source, the evidence may lose impact in trustworthiness computation (See input 2 in Figure 17). In the scope of this work, we will not incorporate evidence collected from third-parties, recommending this as an evolution of the suitcase model. If it is not possible to collect an evidence from past situations, or from recommendations, then the trustor will count only on trustfulness to decide.

We proposed a formula to compute trustworthiness in terms of expected and collected evidence. Creating a list of expected evidence is a strategy to define the meaning of trust for the trustor in the specific situation, and to make trustworthiness assessment less subjective, and this is one of the main contributions of this work.

The trustor will assess how the list of collected evidence matches the list of expected evidence. The more complete is the match, the higher is the positive impact of the evidence on trustworthiness computation.

We modified the basic building block for confidence, Figure 11, to relate the evidence with trustworthiness, not to trust, to make clear the role of evidence in trustworthiness computation; and we added an input for expected evidence also (see input 1 and 7 in Figure 17). Control inputs and outputs were not changed.

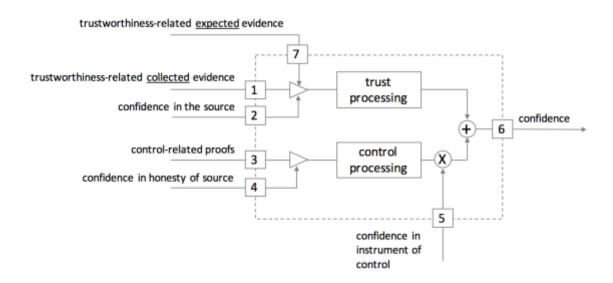


Figure 17 - New basic building block for confidence, modified from (COFTA, 2007)

Alice's assessment about Bob's trustworthiness depends on the list of expected evidence, the list of collected evidence, and the situation's components. See further details about the components in item 4.1.3. The situation's components have impact on the selection of expected evidence. For instance, in an unfavorable context of a highly important situation, the list of expected evidence tends to demand more from the trustee.

The trustor must limit the number of expected evidence to be collected, because it is impossible to verify everything, and the expected behavior must be considered as an expected evidence, because in cases where there is no leading evidence, the expected behavior will be the only one to be collected.

An evidence may be collected from similar or different situations in past. Figure 18 shows a summary of Alice's timeline of trust situations. The current situation (Situation 7) is preceded by other six trust situations. For instance, in June of 1995 (Situation 3), Alice expected that Charles could have capacity and interest to pilot a plane. She expected as evidence, in addition to the expected behavior, that he could have courage and quick reasoning. She collected as evidence that he had quick reasoning and courage, although after a behavioral decision, she decided to defect (D), as well as Charles (for unknown reasons). Table 7 shows a description of the notation used to cooperation between Alice and the trustees (Charles and Bob).

Table 7 - Cooperation notation in Alice's timeline of trust situations

Alice's action	Trustee's action	Notation
Cooperates	Cooperates	C/C
Cooperates	Defects	C/D
Defects	Defects	D/D
Defects	Cooperates	D/C

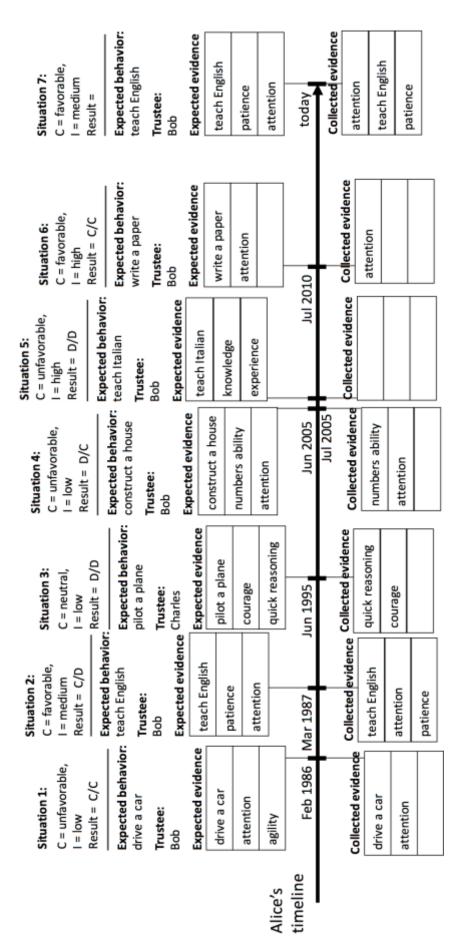


Figure 18 - Alice's timeline of trust situations

The conditions under which the evidence are collected has impact on trustworthiness computation. For instance, the distance of an evidence from present time has impact on trustworthiness computation, just like recurrence in time. For some people, evidence from long past situations are less valuable the ones from a recent past. In Figure 18, for example, 'attention' evidence from Situation 4 has less value than 'attention' from Situation 6. However, 'attention' is recurrent in time, and this may bring a positive impact to trustworthiness.

Evidence from different situations may lose importance in trustworthiness computation. For instance, 'attention' evidence from Situation 4 may have less value then 'teach English' and 'patience' evidence from Situation 2, once last situation is identical to the current one in terms of expected behavior and expected evidence (which reflects the context and the importance).

Another important factor from past situations is reciprocity. If the trustor distrusted the trustee in the last situation, any past evidence about the trustee may lose impact in current trustworthiness computation. The trustee may adopt voluntarily, or involuntarily, a 'Tit-for-Tat' strategy, when he will repeat the trustor's action in last situation. If the trustor did not cooperate in the last situation, the trustee will not cooperate in next situation too. See item 5.1 and (AXELROD, 1984) for further details about 'tit-for-tat' strategy. This is not the case in Figure 18, since Alice cooperated with Bob in Situation6.

To deal with these conditions, we included attenuators of the collected evidence in the basic building block of confidence: 1) distance from present time, 2) recurrence, 3) reciprocity, and 4) different situation. In other words, evidence from a distant past, evidence with low recurrence, defection in the last situation, and evidence from a different situation will have their importance diminished for trustworthiness computation (Figure 19).

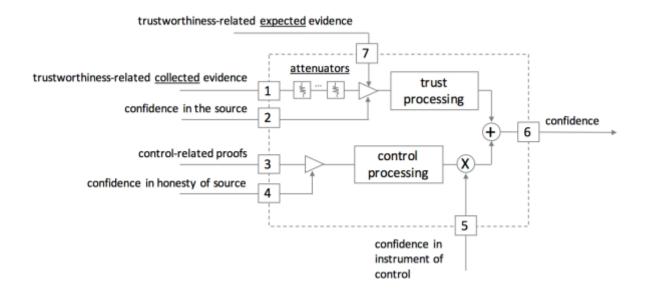


Figure 19 - Suitcase model version of the basic building block for confidence with attenuators, modified from (COFTA, 2007)

Trustor's memory span (m) has an important role in attenuation, because it limits the number of situation remembered for each trustee. Trustor will remember the last m situations with the specified trustee. In Figure 18 (pg. 63), for example, if Alice has a memory span of 4, she could remember four trust situations with Bob (Situation 6, Situation 5, Situation 4, Situation 2). The Situation 1 will not be remembered because the memory span, and Situation 3 will not be remembered because concerns to other trustee (Charles).

Attenuators are calibrated by trustor's trustfulness. For instance, if Alice in Figure 18 has a low trustfulness, she tends to ignore the 'attention' evidence from Situation 4, because it comes from a different situation, and the 'patience' evidence from Situation 2 will lose value for her, because it occurred a long time ago. However, a high trustfulness trustor would tend to ignore these conditions and consider both pieces of evidence in trustworthiness computation.

#### 4.1.5 Suitcase model formalization

Once we defined the components of the suitcase model (Figure 14, pg. 57) and how they are related with the components that produce the confident behavior (Figure 15, pg. 58), we present a formalization of our suitcase model (Table 8), which is based on (COFTA, 2007) formalism (Table 6, pg. 53).

Table 8 – Components of the suitcase model formalism

Name	Notation	Description
Actor	A	$A = \{Alice, Bob, \dots, a_1, \dots a_a\}$
Importance	I	$I = low \mid medium \mid high$
Context	С	$C = favorable \mid neutral \mid unfavorable$
Subject	subject	$subject_i = subject \ for \ wich \ Evidence \ is \ required;                                    $
Expected Behavior	Bex	Bex = subject of expected behavior
Situation	$s_j$	$s_j = is \ a \ situation \ composed \ by \ a \ Context,$ an Importance, and an Expected Behavior; $s_j = (C, I, Bex); \ SITUATION = \{s_1,, s_t\} \mid j = < t$
Piece of Evidence	$e_{i,j}$	$e_{i,j} = is \ a \ Piece \ of \ Evidence \ about \ a \ subject \ i$ $from \ a \ situation \ j;$ $e_{i,j} = \left(subject_i, s_j\right)$
Collected Evidence	Ec	
Collected Evidence Value	$\widehat{Ec}_{\iota}$	$\widehat{Ec}_i = is$ a numeric representation of a Collected Evidence to subject <sub>i</sub> ; It is an attenuated mean of all Piece of Evidence related to the subject <sub>i</sub> ;
Expected Evidence	Ex	$Ex = \{subject_1, \dots, subject_t\}$

Importance I Context С Collected Evidence Value Situation ,  $\widehat{Ec_i}$ SITUATION  $S_i$ Set of all Situations Expected Bex $e_{i,j}$  $Ec_i$ Ес Behavior All Collected Evidence Piece of Evidence Collected Evidence  $subject_i$ SUBJECT Ex Set of all Subjects All Expected Evidence

The Figure 20 shows a graphical summary of the formalism components.

Figure 20 - Graphical relation among the components of the suitcase model

The new suitcase model computes trustfulness as a subjective probability, whereas trustworthiness is a probability computed from expected and collected evidence. Therefore, trust is defined as a sum of these constructors:

$$trust_{Alice \rightarrow Bob}(Ex, Ec, s_j) = trustf_{Alice} + trustw_{Alice \rightarrow Bob}(Ex, Ec, s_j)$$
 (11)

where:

- $trustf_{Alice} = Alice's trustfulness subjective probability$
- $trustw_{Alice \rightarrow Bob}(Ex, Ec, s_j) = Alice's$  assessment on Bob's trustworthiness
- $s_i = current trust situation where Alice and Bob are actors$
- Ex = {subject<sub>1</sub>, ..., subject<sub>k</sub>} =
   set of k subjects related to Expected Evidence
- $Ec = \{Ec_1, ..., Ec_i\} = set \ of \ i \ Collected \ Evidence \ related \ to \ i \ subjects$
- $\circ$   $Ec_i = \{e_{i,1}, ..., e_{i,j}\} = all \ Pieces \ of \ Evidence \ collected \ related \ to \ subject_i$

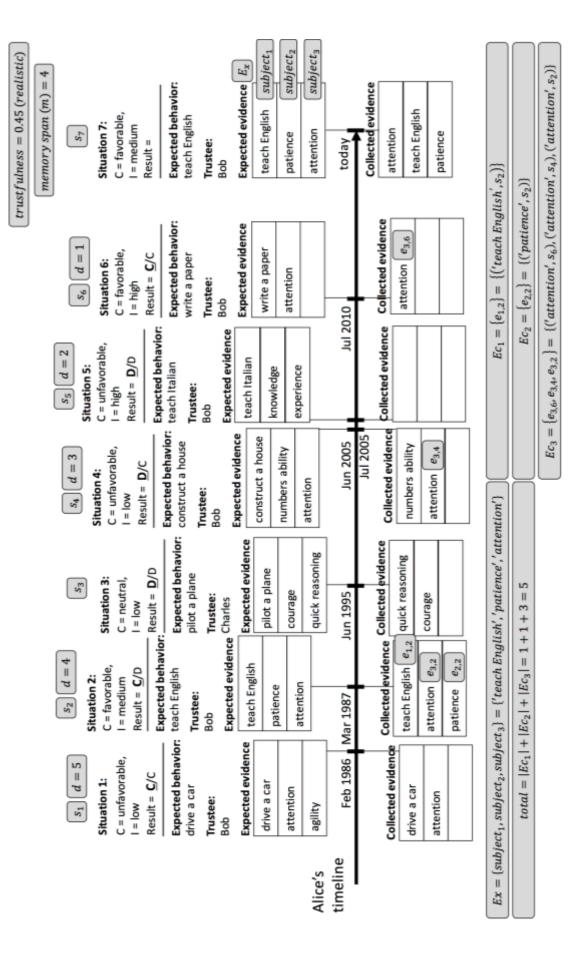


Figure 21 - Alice's timeline of trust situations marked with the suitcase model formalism

In the case of Figure 21 (pg. 68), considering that the current situation is  $s_j = s_7 = Situation$  7, the memory span is 4, and the trustfulness if 0.45 (realistic), these variables have the following values:

```
• Bex = 'teach \, English'

• C = 'favorable'

• I = 'medium'

• s_j = s_7 = (C, I, Bex) = ('favorable', 'medium', 'teach English')

• Ex = \{subject_1, subject_2, subject_3\} = \{'teach \, English', 'patience', 'attention'\}

• Ec = \{Ec_1, ..., Ec_i\} = \{Ec_1, Ec_2, Ec_3\}

• Ec_1 = \{e_{1,2}\} = \{('teach \, English', \, s_2)\}

• Ec_2 = \{e_{2,2}\} = \{('patience', \, s_2)\}

• Ec_3 = \{e_{3,6}, e_{3,4}, e_{3,2}\} = \{('attention', \, s_6), ('attention', \, s_4), ('attention', \, s_2)\}

• e_{3,6} = \{('attention', s_6)\}
```

The convex combination of trustfulness and trustworthiness (eq. 11) will result in a probability value between 0 and 1. This sum may be weighted, although in this work both values have the same weight (0.5). Trustfulness is also a value between 0 and 1, as well as the trustworthiness value.

 $\circ$   $e_{3,4} = \{('attention', s_4)\}$ 

 $\circ$   $e_{3,2} = \{('attention', s_2)\}$ 

Eq. 12 computes trustworthiness as a sum of all *Collected Evidence Values*  $(\widehat{Ec_i})$ , divided by the cardinality of the set of *Expected Evidence* (|Ex|). The division brings to computation the match between the expected and collected evidence lists.

The Collected Evidence Value  $(\widehat{Ec_i})$  is a numerical representation of the Collected Evidence  $(Ec_i)$ . Since the Pieces of Evidence  $(ec_{i,j})$  of an Evidence  $(Ec_i)$  were collected their values are '1'. It means that the trustor can rely on this Piece of Evidence to trust the trustee. Considering that the Collected Evidence Value  $(\widehat{Ec_i})$  is a mean of all Pieces of Evidence collected, the value of  $\widehat{Ec_i}$  is also '1' by default.

However, the *Piece of Evidence*, and consequently the *Evidence*, might be attenuated to reflect the conditions under which they were collected. With this purpose, we apply four attenuators:

- 1. a 'reciprocity' attenuator (*recip*) applied to the mean of all *Collected Evidence Values*, to reflect a defection from trustor in a previous situation, which must compromise any *Collected Evidence* (eq. 12);
- 2. a 'recurrence' attenuator (*recur*) applied to the mean of all *Pieces of Evidence*, to reflect a low recurrence of an *Evidence* (eq. 13);
- 3. a 'distance' attenuator (*dist*) applied to the *Piece of Evidence* to reflect how distant is the situation from the present (eq. 13);
- 4. 'different situation' attenuator (*diff\_sit*) applied to the *Piece of Evidence* that comes from a different situation than the current one (eq. 13);

The trustworthiness computation is done by:

$$trustw_{Alice \to Bob}(Ex, Ec, s_j) = recip_{Bob \to Alice} * \frac{\sum_{i=0}^{|Ec|} \widehat{Ec_i}}{|Ex|}$$
 (12)

where:

- $recip_{Bob \to Alice}$  reciprocity attenuator; if Alice cooperated in last  $situation, recip_{Bob \to Alice} = 1;$  if not, recip will follow Table 9
- |Ex| = cardinality of the Expected Evidence
- $\widehat{Ec_i} = Collected Evidence Value of a Collected Evidence (Ec_i)$

Collected Evidence of a subject i ( $Ec_i$ ) may contain n Pieces of Evidence ( $e_{i,j}$ ) obtained from n situations  $s_j \mid j = 1 \text{ to } n$ 

 $\bullet \quad \mathit{Ec}_i = \left\{e_{i,1}, \dots, e_{i,n}\right\} = \text{evidence related to } \mathit{subject}_i \text{ collected from } n \text{ situations}$ 

In the case of Figure 21 (pg. 68), these variables have the following values:

- $Ex = \{Ex_1, ..., Ex_i\} = \{subject_1, subject_2, subject_3\} \rightarrow |Ex| = 3$
- $Ec = \{Ec_1, ..., Ec_i\} = \{Ec_1, Ec_2, Ec_3\} \rightarrow |Ec| = 3$
- $Ec_1 = \{e_{1,2}\} = \{('teach\ English',\ s_2)\}$
- $Ec_2 = \{e_{2,2}\} = \{('patience', s_2)\}$
- $\bullet \quad Ec_3 = \left\{e_{3,6}, e_{3,4}, e_{3,2}\right\} = \left\{('attention', \ s_6), ('attention', \ s_4), ('attention', \ s_2)\right\}$
- $\circ$   $e_{3,6} = \{('attention', s_6)\}$
- $\circ$   $e_{3,4} = \{('attention', s_4)\}$
- $\circ$   $e_{3,2} = \{('attention', s_2)\}$

The Collected Evidence Value  $(\widehat{Ec}_i)$  of a Collected Evidence  $(Ec_i)$  is done by:

$$\widehat{Ec}_{i} = recur_{e_{i}} * \frac{\sum_{j=1}^{|Ec_{i}|} \left( e_{i,j} * \frac{diff\_sit e_{i,j}}{dist e_{i,j}} \right)}{|Ec_{i}|}$$
(13)

where:

- $e_{i,j} = 1 = \text{trustor can rely on this } Piece of Evidence to trust the trustee}$
- $|Ec_i|$  = cardinality of *Collected Evidence*
- $recur_{e_i} = recurrence$  attenuator; if the subject i appears in all past situation with Bob, then  $recur_{e_i} = 1$ , if not, recur will follow Table 9
- $diff\_sit_{e_{i,j}} = different$  situation attenuator for the  $Evidence\ e_{i,j}$ ; if the Piece of Evidence comes from an equal situation, then  $diff_{sit\ e_{i,j}} = 1$ , if not,  $diff_{sit}$  will follow Table 9
- $dist_{e_{i,j}} = distance$  attenuator for the Evidence  $e_{i,j}$ ; if the Piece of Evidence comes from the last situation, then  $dist_{e_{i,j}} = 1$ , if not,  $dist_{e_{i,j}}$  will follow Table 9

In the case of Figure 21 (pg. 68), these variables have the following values:

•  $subject_1 \rightarrow n = 1 \rightarrow |Ec_1| = 1$ 

- $e_{1.2} = 1$
- $subject_2 \rightarrow n = 2 \rightarrow |Ec_2| = 1$
- $e_{2.2} = 1$
- $subject_4 \rightarrow n = 3 \rightarrow |Ec_3| = 3$
- $e_{3,6} = e_{3,4} = e_{3,2} = 1$

Once we defined how the evidence are computed in trustworthiness, and how they are attenuated to reflect the conditions under which they were collected, the next item present our strategy to calibrate the attenuators.

#### **Attenuators calibration**

Attenuators in the suitcase model are calibrated by trustfulness value. Table 9 presents the calibration strategy adopted for attenuators using three categories of trustfulness values: optimistic, realistic, and pessimistic. Marsh proposed these categorical values for agents disposition to compute the Estimated General Trust (MARSH, 1994).

In the suitcase model, the optimistic agent is defined as an agent with high value of trustfulness, whereas the pessimistic has a low value. Realistic agents have a medium value of trustfulness. The range of values of these categories are subjective, such as the trustfulness value. The range must be set, by a subjective assessment, to reflect the distribution of optimistic, realistic, and pessimistic agents in the simulation.

For our analysis, we considered an equal distribution among the categories. Trustors with low values, from 0 to less than 0.33, tend to have a pessimistic assessment over an evidence, whereas high values of trustfulness, from more than 0.66 to 1 make them more optimistic to accept an evidence. The values out of these ranges reflect a realistic view of evidence.

		Range	of trustfulness	values
Attenuators	Notation	Pessimistic	Realistic	Optimistic
Distance from present	dist	$d^2$	d	1
Evidence recurrence	recur	$\frac{ Ec_n }{total^2}$	$\frac{ Ec_n }{total}$	1
Evidence from different situation	diff_sit	trustf	trustf	trustf
Reciprocity for defection in last situation	recip	trustf	trustf	trustf

Table 9 - Attenuators calibration by trustfulness range of values

The distance of an evidence from present (d) is defined as the number of past situations between the evidence situation and the present, including the referred situation. Those situations must concern only the trustor and the trustee under consideration. Trustors with a pessimistic trustfulness tend to consider this distance as much greater than it really is. On the other hand, high level of trustfulness overlaps this distance. Situation results (cooperation or defection) is irrelevant in this case.

For example, in Figure 21 (pg. 68), Alice is a Realistic agent, the attenuators have the following values:

- $dist_{e_{1,2}} = d = s_6 * 1 + s_5 * 1 + s_4 * 1 + s_3 * 0$  (Charles) +  $s_2 * 1 = 4$
- $dist_{e_{2,2}} = d = s_6 * 1 + s_5 * 1 + s_4 * 1 + s_3 * 0$  (Charles) +  $s_2 * 1 = 4$
- $dist_{e_{3.6}} = d = s_6 * 1 = 1$
- $dist_{e_{3,4}} = d = s_6 * 1 + s_5 * 1 + s_4 * 1 = 3$
- $dist_{e_{3,2}} = d = s_6 * 1 + s_5 * 1 + s_4 * 1 + s_3 * 0$  (Charles) +  $s_2 * 1 = 4$

Recurrence of an evidence has similar calibration to distance from present. It is defined as the number of appearance of an evidence, considering all piece of evidence remembered from the trustee (*total*), including those not collected. Situation results is also irrelevant in this case.

In the case of Figure 21 (pg. 68), the variable total will be computed as follows:

• 
$$m = 5 \rightarrow total = |Ec_1| + |Ec_2| + |Ec_3| = 1 + 1 + 3 = 5$$

• 
$$recur_{e_{1,2}} = \frac{|Ec_1|}{total} = 1/5 = 0.2$$

• 
$$recur_{e_{2,2}} = \frac{|Ec_2|}{total} = 1/5 = 0.2$$

• 
$$recur_{e_{3,6}} = \frac{|Ec_3|}{total} = 3/5 = 0.6$$

• 
$$recur_{e_{3,4}} = \frac{|Ec_3|}{total} = 3/5 = 0.6$$

• 
$$recur_{e_{3,2}} = \frac{|Ec_3|}{total} = 3/5 = 0.6$$

Although the 'attention' evidence appears three times (in Situation 2, Situation 4, and Situation 6), the recurrence will be computed as 2/3, because only Situation 6, Situation 5, and Situation 4 fit in memory, where 'attention' appears only two times. If memory span is 4, then Situation 2 is considered again, once Situation 3 is related to another trustee.

Calibration for evidence from different situations (*diff\_sit*) and reciprocity (*recip*) are done directly using the trustfulness value (*trustf*).

# Confidence computation

To simulate effectively the suitcase model, we also needed to formalize the other components that produce the cooperation action (Figure 15, pg. 58). For that we present the confidence computation extracted from Cofta (eq. 8), but modified to incorporate the suitcase model (eq. 11) by defining confidence as an convex combination of trust and control, and to define control as a multiplication of evidence from instruments of control and the confidence in these instruments, resulting in:

$$conf_{Alice \to Bob}^{subject} = trust_{Alice \to Bob}(Ex, Ec, s_j) + \sum_{i=1}^{n} crtl_{Alice \to Bob}^{subject}(Ect|y_i) * conf_{Alice \to y_i}^{control}$$
 (14)

We change the original notation for evidence of control to *Ect*. Confidence is a probability value between 0 and 1, and trust and control are weighted, although in this work both values have the same weight (0.5).

As mentioned before, although control can be modeled as a function of

expected and collected evidence, in this work we simplify control summation as a subjective probability, because trust is the focus of the new model.

Once we defined our suitcase model, we needed to create a procedure to decode real life trust situations in the components of our model.

# 4.2 Defining a trust situation

To manage trust under a suitcase word perspective, we need to define the meanings of trust in terms of the components of the suitcase model. We proposed a form (Table 10) to guide this definition.

Table 10 – Guide form to define trust in terms of the components of the suitcase model

Model components	Types of values
<u>Situation</u>	text
<u>Actors</u>	
• <u>Trustor</u>	
o name	string
o trustfulness	optimistic, realistic, or pessimistic
Trustee's name	string
Expected behavior	string
Expected evidence	
evidence 1	string
evidence 2	string
•	string
evidence n	string
Context	'favorable', 'neutral, or 'unfavorable'
Importance	'high', 'medium', or 'low'
Memory span	integer number

Table 11 shows an example of a trust definition from a hypothetic situation.

Table 11 – Example of definition from a trust situation.

Components	Types of Values
Situation	Alice believes that Bob will be a good teacher of mathematics. The expected behavior is highly important, because she has a final exam in two days, and she needs to learn fast. The context in unfavorable, because she did not attend the mathematics classes in the last 2 months
<u>Actors</u>	
• <u>Trustor</u>	
o name	Alice
o trustfulness	realistic
Trustee's name	Bob
Expected behavior	be a good teacher of Math
Expected evidence	
evidence 1	Good teacher
evidence 2	Knowledge in mathematics
evidence 3	Teaching method
evidence 4	Patience
Context	unfavorable
Importance	high
Memory span	10

# 4.3 A generalized model of behavioral decision

Our generalized model of behavioral decision is based on (MARSH, 1994), and its main goal is to reproduce the 'behavioral decision' component in Figure 15 (pg. 58), which might produce a confident behavior, and a cooperation action, consequently.

We started by adapting eq. 1 to carry the idea that the 'confident behavior' is produced by confidence (Figure 15, pg. 58):

$$conf_{Alice \rightarrow Bob}^{subject} > conf\_threshold_{Alice}(s_j) \Rightarrow confident\_behavior(Alice, Bob, s_j)$$
 (15)

As a consequence of this change, the cooperation threshold became a confidence threshold. In the original threshold (eq. 4), the perceived risk was mitigated

by the perceived competence and the estimated general trust (EGT) together; and the importance affected directly the risk.

Perceived competence and EGT might be considered as measures of trustee's capacity and interest to behave such as expected, respectively, so we replaced them by our confidence value, which has the same goal: estimate the trustee's capacity and interest to behave such as expect. This change also brings control as an additional option to mitigate the risk (eq. 16).

$$conf_{threshold}_{Alice}(s_j) = \frac{Risk(s_j)}{conf_{Alice \to Bob}^{subject}} * \frac{I(s_j) + C(s_j)}{2}$$
(16)

Considering eq. 15 and eq. 16, the trustor will produce a confident behavior if trustor's confidence in Bob to behave such as expected is greater than the risk multiplied by the importance and mitigated by the confidence in Bob.

At the end, the eq. 15 can be rewritten as:

$$\left(conf_{Alice \to Bob}^{subject}\right)^{2} > Risk(s_{j}) * \frac{I(s_{j}) + C(s_{j})}{2} \Rightarrow confident\_behavior(Alice, Bob, s_{j})$$
(17)

# 4.4 Considerations about modeling trust under a suitcase perspective

Getting evidence that trust is a suitcase word was the first step to model trust under this perspective. After that, we guided our review under this perspective to collect conceptual foundations, definitions and models of trust from scientific literature.

That knowledge supported our model, the suitcase model, which is a result of trustor's trustfulness and trustee's trustworthiness (TULLBERG, 2007). Considering that trust works together with control to improve confidence (COFTA, 2007), and that a confident behavior allows engaging in a relationship, we also proposed a generalized model of behavioral decision of confidence. This decision model was based on (MARSH, 1994) implementation. This complementary model was required to make our suitcase model workable, since trust alone could not produce a confident behavior to

promote cooperation.

We also proposed a less subjective method to measure trustworthiness in terms of expected and collected evidence. Creating a list of expected evidence was a strategy to identify which meanings of trust are being considered to computation. It would be difficult to find evidence for trustworthiness if we do not know what is important for the trustor in a situation. Besides that, a finite list of terms may reduce the complexity of analysis, once we have to measure only the most important evidence for each situation (important in trustor's eyes).

The list of expected evidence might be considered an over simplification of the trustworthiness assessment, since we, as humans, are able to remember and process a more elaborated list of evidence. However, our proposed list does not have a limit, and its size can be adapted to each situation. We intended to keep the model workable allowing it to be used in trust management by nonscientists users.

To bring the suitcase model even closer to real-life, attenuators were proposed to capture the conditions in which the evidence was collected. It is possible to refine this idea by identifying and including other kinds of attenuators; however, this might increase the complexity of trustworthiness assessment, and might reduce its workability, since it could increase the information to be retrieved. Evidence and instruments of control were considered as subjective probabilities in this work. Although it is also possible to measure control in terms of expected and collected evidence of control, in this work, we decided to focus in trust.

Our model of trust under a suitcase perspective was designed to be more comprehensive and workable as possible. We proposed also a procedure to define a real trust situation in the components of this suitcase model, and we presented a behavioral decision model of confidence, which completes our framework to manage trust, making it possible to create a simulator for the suitcase model.

# 5 Simulating the suitcase model

This chapter presents a simulation of our suitcase model built as a modified version of the PlayGround simulator (MARSH, 1994). We start this chapter with a presentation of the Iterated Prisoner's Dilemma, which is a central concept for the original PlayGround, and then we present a replica of the PlayGround developed for Internet version. Finally, we show how the suitcase model was incorporated to this replica to create the PlayGround 2.0.

# 5.1 The Iterated Prisoner's Dilemma

The Prisoner's Dilemma (AXELROD, 1980) is a well-known two-players game studied by social sciences, business, computer science, and other fields of science. The game takes place in a prison, where two prisoners will be charged in 5 years of reclusion. At a certain moment, both receive an offer to testify against the other prisoner in return of a reward. The offer is done individually, not being allowed the communication between them. Both players (prisoners) have to decide either to cooperate with one each other (remaining in silent) or to defect (incriminating the other player).

Table 12 gives the canonical matrix that represents the payoff considering both decisions, and also gives the original scoring adopted in (AXELROD, 1980). If both players decide to cooperate, they receive a Reward (R) with 3 years of payoff and go to prison for 2 years. If one player defects whereas the other player cooperates, the defector will set free, receiving a Temptation (T) payoff of 5 years, and the other player will serve 5 years in prison (with a Sucker's payoff of zero). Finally, if both players defect, both will serve 4 years in prison, receiving a Punishment (P) payoff of 1 year.

Table 12 - Payoff Matrix of Prisoner's Dilemma, extracted from (AXELROD, 1980)

	Cooperate	Defect
Cooperate	R, R	S, T
Defect	T, S	P, P

	Cooperate	Defect
Cooperate	3, 3	0, 5
Defect	5, 0	1, 1

The condition T > R > P > S must be respected to assure the incentive of both sides to be selfish (individual rationality), and to assure the higher payoff to both sides for mutual cooperation over mutual defection (group rationality) (AXELROD, 1980).

If two players play the Prisoner's Dilemma more than once in succession, they can remember their play, and they aiming to maximize their total payoffs changing strategies, then we have an Iterated Prisoner's Dilemma (IPD) game (AXELROD, 1980). In IPD game, the basic Prisoner's Dilemma is a game stage in a repeated game. Maximizing the total payoff is the central point of the IPD.

Many strategies have been proposed to achieve this goal. For instance, using the 'Tit for Tat', one of the most successful strategy (AXELROD, 1984), a player will first cooperate, and after that he will always replicate the opponent's previous action. If the opponent had been cooperative, the agent is cooperative; if not, the agent will not cooperate until a cooperation action from the opponent. The 'Alternator' player strategy will alternate between cooperating and defecting ignoring the counterpart actions. Tournaments were create to determine the most successful strategy (AXELROD, 1980, 1984).

In Axelrod's Tournament, players propose their own strategy and the winner player (prisoner) achieves the lesser year of imprisonment at the end. Libraries are available to run and develop these tournaments in modern languages such as Python (KNIGHT, 2015).

An Axelrod's Tournament environment (Figure 22) is composed by Players, who Play against each other in a series of Turns driven by a Strategy. The strategy is a set of instructions that produce an Action (Cooperation (C) or Defection (D)) for each turn. The possible outcomes for each turn are: Mutual cooperation (C, C), Defection (C, D) or (D, C), and Mutual defection (D, D).

A Match is a consecutive number of turns between two specific players, and a Win is attributed to the player who has the higher total score at the end of a match. All matches between all players, ignoring the order of assignment, is a Round Robin. A Tournament is a predetermined number of Round Robins.

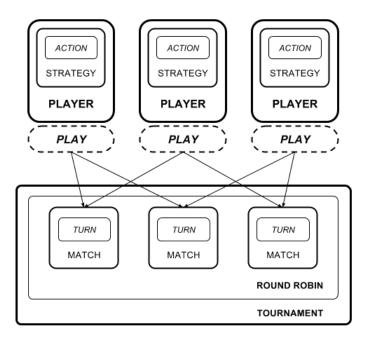


Figure 22 - Axelrod's Tournament environment

# 5.2 Replicating the original PlayGround

In this section, we present the replication of the original playground. We started by describing the original implementation, then we listed some assumptions about operation details, and finally discussed the replication results.

# 5.2.1 The original PlayGround

(MARSH, 1994) developed two implementations of his formalism. The first considered trust as a strategy in an Axelrod's Tournament and proved the feasibility of the formalism. The second, called the PlayGround, implemented the formalism more completely, by introducing some additional features to the tournament concept, aiming to create an environment closer to the real world. The PlayGround was a graphical application (Figure 23) developed in HyperCard that simulated interactions among agents.

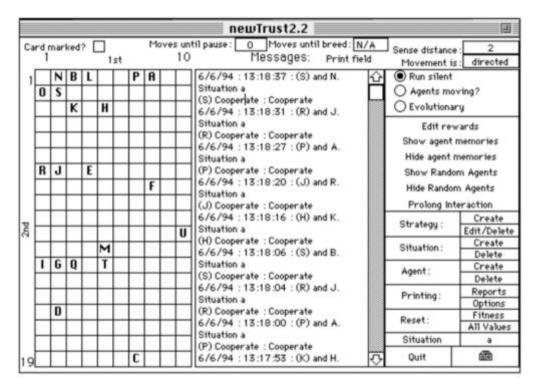


Figure 23 - A screenshot of the PlayGround, extracted from (MARSH, 1994)

Agents moved around a cellular grid. An interaction occurred when an agent tried to move to a place that was already occupied. If occupied, agents played a round of the Iterated Prisoner's Dilemmas game in which they 'Cooperate' or 'Defect'. The payoff matrix varied according to the situation assigned to the interaction, which was randomly assigned, or defined as the same for all interactions. A generic payoff (T=>5, R=>3, P=>1, and S=>0) was used for most experiments. At the end of all movements, the agent with highest score won.

The follow items are important details extracted from (MARSH, 1994) about PlayGround operation which were considered essential to the replication.

## **Cycle of operation**

The code bellow extracted from (MARSH, 1994) represents the full cycle of the PlayGround:

```
repeat
for each agent do
begin
move
if another agent 'met' then
interact
adjust trust values if necessary
save results
end if
end
until maximum number of moves reached
```

## **Trusting agent**

Like in the tournament, agents have strategies (Random, Cooperator, Defector, Tit for Tat, and others) to decide to 'Cooperate' or to 'Defect'. Marsh introduced a new strategy, which implements the proposed formalism for trust. These 'Trusting' agents consider 'trust' to act and may use it to move. Their memory is limited, so that only "x" interactions for each opponent can be considered to decide the actions.

## **Disposition to Trust**

As described on item 3.9, the Estimated General Trust depends on trustor's Disposition to Trust as follow (Table 13):

Disposition to trust	How Estimated General Trust is calculated
Optimistic	Use the maximum value of all General Trust past values
Realistic	Use the average of General Trust values in similar situations (eq. 3)
Pessimistic	Use the minimum value of all General Trust past values

The historic values assessment is limited by trustor memory.

#### **Directed movement**

The direction of movement for all agents in the field is chosen randomly among North, South, West, and East, except for the trusting agent, which is able to choose a movement toward a trustworthy agent. This directed movement was designed to reflect human behavior, in which an individual usually has access to a circle of trustworthy people allowing agents to form groups, and also to reflect external factors that make impossible the desired movement. A group is a cluster of neighbor agents in the grid.

This choice is limited by a configurable 'range of vision', equal for all directions, for example, two positions in each direction. However, although the trusting agent can choose the direction, not necessarily he can take it. Depending on his set up, he can: (1) choose his movement, (2) receive a random movement, or (3) influence the movement. This last option is a mix of the previous two, which uses a weighting of probabilities to determine the influenced direction: 50% for desired direction, 10% to opposite direction, and 20% for the others. Which means that he may move to an undesired direction. In both cases, random or influenced movements, an interaction can occur between the same agents on next round.

#### **Trust adjustments**

Other feature in which the PlayGround approximated real life was the adjustment of trust values after an interaction. These adjustments apply to Basic Trust, General Trust, and Perceived Competence. After each interaction, adjustments are done in trust values for both trusting agents. In general, Marsh proposes standard variations in which trust increase if other cooperates and decreases if not. For instance, after an interaction between Alice and Bob, from an Alice perspective as a trustor, if Alice cooperates and Bob defects, Alice's Basic Trust will decrease 1%, her General Trust on Bob will decrease 10%, and her Perceived Competence about Bob will decrease 1%. Table 14 show details about all trust values adjustments after an interaction.

Table 14 - Trust values adjustments after an interaction between Alice and Bob

After an interaction between Alice and Bob where	Alice's Basic Trust	Alice's General Trust on Bob	Alice's Perceived Competence about Bob
Both Cooperate	+1%	+10%	+1%
Alice Cooperates and Bob Defects	-1%	-10%	-1%
Alice Defects and Bob Cooperates	+5%	+1%	+1%
Both Defect	-5%	-10%	-1%

#### **Experiments**

The first experiment run by Marsh wanted to test if trust could be educated. Divided in two parts, the first one forced two Realistic agents to play a IPD. They had an unbound memory and they could not move around the grid. The initial values of A were extremely low and set up to distrust B (defect); whereas B was inclined to cooperate. After one hundred interactions, agent A seemed to move slowly to a cooperative behavior, whereas B always cooperated. Considering that A was a Realistic agent with unbound memory, all past interactions were composing the mean for its Estimated General Trust in B, so that if Basic trust values were increased due to B's cooperation, the mean of all values pushed A to a slow movement toward cooperation.

In part II of this experiment, A became C, an Optimistic agent, and B became D, with the exactly same characteristic and trust values. Memory of both agents remains unbounded. After only a dozen interactions, C chose to cooperate with D. Some assumptions were done by Marsh, among which the difference of learning was done by the way the Realistic and Optimistic agents compute the Estimated General Trust. It seems that the mean, in the first case, held a slow learning, whereas a selection of highest past value showed a faster result. Tables with detailed outputs extracted from the original work can be found in Appendix E.

In fact, Experiment A was essential and enough to support our replication. Experiments B and C realized tests to determine memory span effects in the decision. Sequential experiments tested the influence of other strategies, movement style,

evolution, and number of interactions, especially in group formation. For instance, experiment F Part II tested the grouping behavior among 21 agents (6 Random, 5 Optimistic, 5 Pessimistic, and 5 Realistic) in the situation a, with directed movement and range of vision 2, after 250 interactions. See further details about these experiments in (MARSH, 1994).

# 5.2.2 Assumptions about original implementation

Implementation information was collect in (MARSH, 1994). However, it was not enough to a successful replication of original simulator. Listed below are some important factors about the implementation that were assumed by us:

- Memory Span: If an agent has a memory span of 4, and in his life this
  agent had 21 interactions equally distributed with 3 different agents, the
  trustor will remember at maximum 12 interactions.
- Grid definition: The field is 'flat', which means that the boundaries are
  well defined. An agent cannot move to East if he is on the eastern most
  position of the field. He will remain in the same position. However, we
  decide to allow configuring a wrapped field.
- <u>Directed movement 1</u>: when deciding a directed movement, the trusting agent will consider only the first agent that he sees on a direction. If the agent sees two neighbors in the North, he will consider the closest one to decide his movement, even if the other is more trustworthy.
- <u>Directed Movement 2</u>: The trusting agent will consider only the first (and closest) agent that he sees in a direction to choose a movement. A simulated interaction using the generic situation is done with existent neighbors in four directions. The greatest difference between the Situational Trust and the Cooperation Threshold (ST-CT), even negative values, will define the chosen movement. In case of absence of neighbors, the movement will be random.

- Interaction: When two trusting agents interact, both play the roles of trustor and trustee for the exactly same situation, because both need trust to decide the action, so that in a unique interaction there are two plays ("A trust B" and "B trust A") to the same situation. Trust values are adjusted for trustors and trustees after each play.
- Trust adjustments: For a Realistic agent, before calculating Situational Trust, a mean of past General Trust (GT) is saved as a new GT value in memory to use in future calculations. After each interaction, the adjusted value from the mean will be saved as another new GT value. This is not true for Optimistic and Pessimistic agents, for whom only the GT originated from adjustments are saved as new values.
- <u>Values</u>: Although they had formulae, or suggestions of calculations, Importance, Utility, Risk, Costs, and Perceived Competence values did not have their origin detailed in (MARSH, 1994). In addition, it was only possible to get initial values for Alice and Bob. Other agents did not their initial values known. Therefore, we decided to use random values between Alice and Bob's lowest and highest values to initialize other agents. We adopted the same strategy for Basic Trust. General Trust started with null values. Zero and one (1) values cannot be accepted for Competence, Control and Perceived Competence values.

# 5.2.3 Replication results

The original PlayGround replica for Internet was implemented in PHP and it is available online at (CARTOLANO, 2016). The first goal of the new simulator was replicate 1) Experiment A Part II, 2) Experiment F Part I, and 3) Experiment F part II. Output tables extracted from (MARSH, 1994) were essential during development to propose and test the assumptions described above. Without these outputs, we could not replicate the original simulator.

Our result for Experiment A Part II was the same as Marsh's; the optimistic agent C started to cooperate at the 13<sup>th</sup> interaction. However, we found unexpected

output values for agent C. Adjustment for Basic and General Trust values appears to be reversed until agent C starts to cooperate as discussed in (CARTOLANO; SARAIVA; STEVENSON, 2016). Table 15 shows how this inversion produced a different output from what was expected according to Table 14 or (MARSH, 1994). We named columns from A to F to facilitate identification.

Table 15 - Difference between Marsh's and expected output for agent C

	Expected	output values	for agent C	Marsh's o	utput values f	or agent C
	Α	В	С	D	E	F
#	Basic	General	Coop. Thresh	Basic	General	Coop. Thresh
13	0.143668	0.090146	0.572641	0.090147	0.143666	0.572918
12	0.136827	0.089253	0.600133	0.089253	0.136827	0.600438
11	0.130311	0.088369	0.628908	0.088369	0.130311	0.629233
10	0.124106	0.087494	0.659009	0.087494	0.124106	0.659361
9	0.118196	0.086628	0.690498	0.086628	0.118196	0.690885
8	0.112568	0.085770	0.723436	0.085770	0.112568	0.723855
7	0.107208	0.084921	0.757881	0.084921	0.107208	0.758334
6	0.102103	0.084080	0.793899	0.084080	0.102103	0.794395
5	0.097241	0.083248	0.831561	0.083248	0.097241	0.832097
4	0.092610	0.082424	0.870933	0.082424	0.092610	0.871504
3	0.088200	0.081608	0.912075	0.081608	0.088200	0.912702
2	0.084000	0.080800	0.955076	0.080800	0.084000	0.955752
1	0.080000	0.080000	1.000000	0.080000	0.080000	1

Considering that agent C defected and agent D cooperated until interaction number 12 (included), it was expected that agent C's Basic Trust should be increased by +5% and General Trust by 1% after each interaction until both cooperate. In this case it was expected that Basic Trust A1(0.080000) would be A2(0.080000 \* 1.05 = 0.084000), General Trust B1(0.080000) would be B2(0.080000 \* 1.01 = 0.080800), and Cooperation Threshold would be C2(0.955076) according to the equations on item 3.9.

However, Marsh's output shows that Basic Trust D1(0.080000) was adjusted +1% resulting in D2(0.080000 \* 1.01 = 0.080800) and the General Trust E1(0.080000) was adjusted by +5% resulting in E2(0.080000 \* 1.05 = 0.084000). This produced an inverted result for General and Basic Trust which directly impacted the other columns Situational Trust and Cooperation Threshold.

We considered this inversion in the original output unexpected because there are no reasons for that described in (MARSH, 1994), and because after cooperation (13° interaction) adjustments became in line with expected. Despite this, both Marsh's original and replication output produced a Situational Trust value greater than Cooperation Threshold after 12° interaction.

Table 16 shows our output for agent C in Experiment A Part II replication extracted from (CARTOLANO; SARAIVA; STEVENSON, 2016). Table 25 and Table 26 in Appendix E show the output extracted for agent C in (MARSH, 1994), whereas Table 27 and Table 28 show output for agent D.

We obtained the exact same values of the original output for agent D.

Table 16 - Agent C partial results of Experiment A Part II replication

#	Basic	General	Competence	Situation	Threshold	Action
13	0.143668	0.090146	0.007212	0.620646	0.572641	С
12	0.136827	0.089253	0.007141	0.591093	0.600133	D
11	0.130311	0.088369	0.007070	0.562944	0.628908	D
10	0.124106	0.087494	0.007000	0.536138	0.659009	D
9	0.118196	0.086628	0.006931	0.510607	0.690498	D
8	0.112568	0.085770	0.006862	0.486294	0.723436	D
7	0.107208	0.084921	0.006794	0.463139	0.757881	D
6	0.102103	0.084080	0.006727	0.441085	0.793899	D
5	0.097241	0.083248	0.006660	0.420081	0.831561	D
4	0.092610	0.082424	0.006594	0.400075	0.870933	D
3	0.088200	0.081608	0.006529	0.381024	0.912075	D
2	0.084000	0.080800	0.006464	0.362880	0.955076	D
1	0.080000	0.080000	0.006400	0.345600	1.000000	D

In a further effort to validate Marsh's simulations, we reran his experiment F Part I on the speed and efficiency in group formation. The grid was populated with 6 Random, red, and 5 Optimistic, green, Realistic, yellow, and Pessimistic, gray agents. Agents were configured to move directed toward trustworthy agents using a range of vision of 2 cells. Initial random values were generated for Competence, Importance, Risk, and Basic Trust, considering the range of values from Experiment A part II. Because the initial conditions are generated randomly, and the direct movement has a random component, it is not possible to duplicate Marsh's results exactly. However, after being assigned randomly on the grid at the beginning of the experiment (Figure 24), agents were also grouped after 250 interactions (Figure 25) like in original experiment.

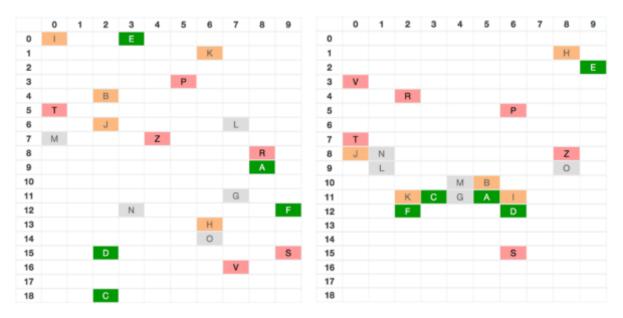


Figure 24 - Initial agent's distribution in experiment F part I replication

Figure 25 - Final agent's distribution in experiment F part I replication

Table 17 presents the top and bottom five agents that held highest and lowest scores after 250 and 10,000 movements. Our results are different from Marsh's results, because none of our Random agents appears among first agents. Instead, Random agents are among the last in rank.

Table 17 - Top and bottom agents after 250 and 10,000 movements

		After 250	Movemen	ts	Α	fter 10,000	) Moveme	nts
	Ag	gent	Score	# Int	Ag	gent	Score	# Int
1	Р	Maria	456	143	R	Bob	13,340	5,579
2	R	Hardin	408	144	0	Charles	13,243	5,669
3	R	Kyle	354	123	0	Etienne	13,220	5,507
4	R	lan	339	134	Р	Neo	12,848	5,211
5	0	Fred	321	134	Р	Oddin	12,567	5,129
17	0	Charles	214	127	RD	Silva	6,650	2,822
18	RD	Silva	201	65	RD	Rafa	6,242	2,777
19	RD	Peter	199	81	RD	Peter	6,218	2,788
20	RD	Tom	171	84	RD	Tom	5,938	2,767
21	RD	Zoom	139	60	RD	Zoom	5,669	2,667

This result is related to the number of interactions that agents have during their movements. An agent may need one or more movements to interact whit other agents. Trusting agents were more efficient in terms of interactions per movements, once they tend to direct their movement to more trustworthy agents.

After replicating the experiments, we proposed a new test for the original model by analyzing successful cooperation in time. Initial random values were saved from Experience F Part I and used in five rounds. Optimistic, Realistic, Pessimistic, and Random values in Figure 26 are means of the values of their agents in the five rounds. When simulations are done until 10,000 movements, clear differences among the strategies emerge. Optimistic and realistic agents have a mean of 71% and 60% of successful cooperation among all interactions and they increase successful cooperation consistently, while the pessimistic and random agents have lower means and show decreasing percentage of successful cooperation.

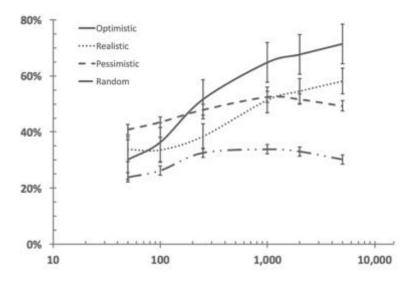


Figure 26 - Percentage of successful cooperation in time for trusting agents

# 5.3 PlayGround 2.0 - the suitcase version

We present in this item the PlayGround 2.0 (Figure 27), which was created using the Internet replica presented before with minimal changes. We implemented the suitcase model on it as a new strategy to players, the 'Suitcase Trustor'. The agents that adopt this strategy are the 'suitcase agents'.

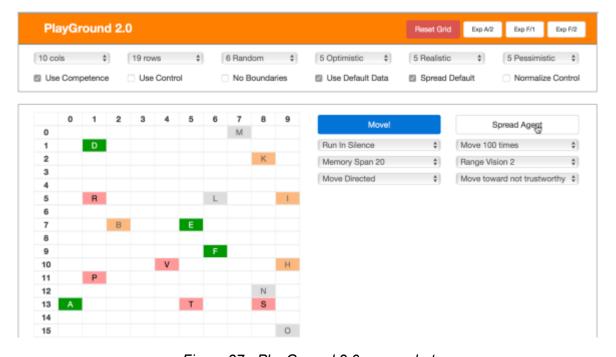


Figure 27 - PlayGround 2.0 screenshot

We did not change the original 'Trustor' strategy from (MARSH, 1994); so that the trustor agent and the suitcase agent could play together in the field, in companion with the 'Cooperator', 'Defector', and 'Random' agents. As well in the original simulator, the winner strategy in the PlayGround 2.0 achieves the highest score at the end of all movements.

## **General features of version 2.0**

The concept of agents on the field, moving around, and playing when meet each other, is the same of the Internet replica, and the cycle of operation, described in item 5.2.1, is followed systematically.

The suitcase agent has the same 'field' features of its predecessor. It has a range of vision in the field and a memory span, and may decide or influence its movements according to a trust assessment of neighbors.

An important evolution is the random assignment of situations during the cycle of operation. Until reach their maximum number of moves, both trusting agents will be confronted with different situations, but only the suitcase agents will decide according to the different meanings of trust presented.

#### Main changes

The main difference between both trusting agents are related to the trust decision and the behavioral decision that leads to cooperation.

Although the results are possibly the same ('cooperates' or 'defects'), the trustor agent decides to cooperate using trust (eq. 1), whereas the suitcase agent uses confidence (eq. 15). This is an important difference, since confidence is a wider concept that incorporates control to enforce the expected behavior (eq. 14).

Furthermore, the suitcase agent considers trust as result of the trustor's trustfulness and the trustee's trustworthiness values (eq. 11). By its turn, the trustfulness is computed in terms of expected collected evidence for each situation (eq.12), and this allow the trustor have a different understood about the many meanings of trust.

## **Simulator configuration**

Users must previously decide the area of the grid and the number of agents in the field. After create the grid and spread the agents on it, users can configure the number of movements in each cycle, and if these movements are directed, random, or partially random. The rank of agent by score is updated after each cycle.

In terms of initial configuration of the suitcase agent, it is required to define their types, and their importance, context, and their list of expected evidence for each situation. When choosing a type among optimistic, realistic, or pessimistic suitcase agent, a random value is assigned to trustfulness considering a pre-defined range: pessimistic, from 0 to less than 0.33, realistic, from more than 0.33 to 0.66, and optimistic, from more than 0.66 to 1. These ranges were chose assuming an equal distribution of types in the population (see item 4.1.5).

The Importance and Context categories are translated to values in the simulator. The high level of Importance and the unfavorable context are translated to 0.75. The medium level and the neutral context are translated to 0.5, and finally, the low level and the favorable context are 0.25.

The differences of initial configurations from previous simulator required a minimal change on the interface.

## <u>List of expected evidence</u>

The list of expected evidence is central for the operation of the 'Suitcase Trustor' *strategy*, because it is used as a filter to collect evidence from past situations. Before starts an interaction, the suitcase *agent* recovers in *memory*, i.e., remembers, all past situations with the trustee, limited by the memory span, then their pieces of evidence remembered are filtered using the list of expected evidence. When a piece of evidence is collected, some additional information about the situation (expected behavior, importance, context, and trustor's action) are collected too.

Figure 28 presents a relation between these entities in the 'Suitcase trustor' strategy.

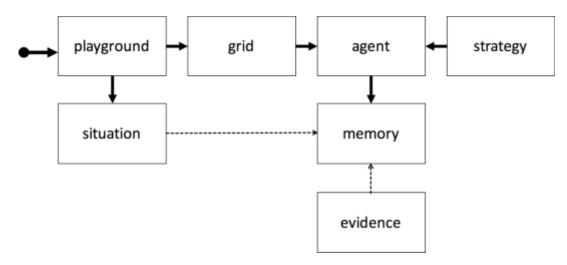


Figure 28 – Relation between entities in the 'Suitcase Trustor' strategy of the PlayGround 2.0.

#### **Attenuation**

As a final main aspect of the 'Suitcase Trustor' we defined the strategy to attenuate the collected evidence. By considering that the role of these attenuators is diminish the importance of a piece of evidence if necessary, to reflect the conditions in which it was collected, we defined the value of a piece of evidence as '1' by default. This means that under the best condition the trustor can rely on the piece to trust the trustee. However, in different conditions the value will be attenuated.

## 5.4 Considerations about the model simulator

The main goal of this chapter was to present how our simulator was developed and some its details. We decided to use the original PlayGround as a basis to our simulator, so that we could make use of a known structure to run our model.

We had to replicate the original tool in an Internet version, because the code was no longer accessible, which demanded from us some time to understand the deeper details of the PlayGround, including make assumptions about some operational details, once the information was not find in the original text, and finding some discrepancy in the results of the original work. However, this replication effort was essential to understand the model and adapt it to the suitcase model, and to modify the original simulator to create the PlayGround v2.0.

# 6 Case study in Citizen Science

The main goal of this case study was to test the hypothesis that trust can be simulated under a suitcase perspective with potential to handle different meanings in Citizen Science projects (item 1.4, pg. 24). We did not want to compare strategies or to analyze actors grouping aspects in this work. A successful simulation would use the procedure to define real-life trust situations in terms of the suitcase model (item 4.2), and plot the agents in the field, being able to trust using what is the most important to them in each one of the situations.

# 6.1 Scenario of Citizen Science in Biodiversity

A target for real-world study in Citizen Science in the Biodiversity domain is eBird, a very successful citizen science project managed by the Cornell Lab of Ornithology (E-BIRD, 2016); eBird's goal is to engage people, mostly non-scientists, to register and share occurrences of birds from across the globe. In 2016, e-Bird was producing 5-10 million observations of bird species per month.

Two real world decisions involving eBird are of interest: 1) Should the eBird team accept the data collected by an eBird participant as valid? 2) Should a scientist use a data set published by eBird team? To answer these questions one would want to have a way to check the data quality and to know if the participant and the eBird team are trustworthy.

e-Bird address issues of data quality by using machine learning to quantify and control observer variability (KELLING et al., 2015a) and to correct both spatial and temporal bias in data (KELLING et al., 2015b). These processes are supported by a review system that is based on local experts. New initiatives to enhance data quality using the similarity of observer profiles are being developed.

Ultimately, however, people managing eBird must trust the observers, and people using eBird data must trust the eBird staff to provide "good" data.

# 6.2 Defining the Citizen Science situations

Our case study recreated hypothetical trust situations from the eBird project in the PlayGround 2.0. We tried to place in the field a sample of situations with potential to have different meanings in the eyes of trustor, so that the procedure could demonstrate how the suitcase model may be capable to define many meanings of trust and make them workable to be simulated.

We proposed for this case study three actors (Table 18): the eBird, a scientist, and a participant ('a citizen scientist'). The eBird represents the institutions that run the eBird Project. The scientist is a professional of science that has interest in citizen science data, and the participant is a citizen with experience in bird watching.

Range of Memory Trustfulness Name Actor Vision Span Participant 1 P1 Realistic 10 2 Scientist 1 S1 Optimistic 5 2 eBird Е Realistic 10 3

Table 18 - Actors of the citizen science case study simulation

The actors (or agents) would interact under nine trust situations:

- **1.** eBird believes that participants can report data accurately.
- 2. eBird believes that participants are committed to the project.
- **3.** eBird believes that participants may work in groups.
- **4.** Scientist believes that eBird have good data.
- **5.** Scientist believes that eBird has a professional reviewer team.
- **6.** Participant believes that eBird will protect its sensitive data.
- **7.** Participant believes that eBird does not have economic ambitions.
- **8.** Participant believes that eBird is a serious project.

We decided to define some combinations among actors and situations. This means that the scientist would define two situations, the participant would define three situations, and finally, eBird would define three situations.

These combinations could expose different reactions, because each actor might have a unique understanding about trust in each situation. For instance, the participant may expect from eBird a large number of participants as an evidence that it is a serious project, or it may be satisfied with a respectable institution behind the project. The participant may also have different meanings for different situations, i.e., the expected evidence for 'is a serious project' may be different for 'do not have economic ambitions'.

Sit#	Trustor	Trustee	Expected behavior	Context	Importance	Expected Evidence 1	Expected evidence 2	Expected evidence 3
-	eBird	Participants	Participants can report data accurately	unfavorable	high	can report data accurately	have a good data quality score	can observe the native species
7	eBird	Participants	can report data accurately	neutral	wol	can report data accurately	ı	ı
က	eBird	Participants	are committed to the project	unfavorable	medium	are committed to the project	respond to all requests to observe	share space and time with others
4	eBird	Participants	are committed to the project	favorable	high	are committed to the project	have a high frequency contribution	observe the native species
2	eBird	Participants	may work in groups	unfavorable	wol	may work in groups	share space and time with others	observe with inexperienced users
9	eBird	Participants	may work in groups	neutral	high	may work in groups	accepted observations are > 70%	1
7	S	eBird	have good data	neutral	high	have good data	has a quality assurance process	has provenance data
80	S	eBird	has a professional reviewer team	neutral	medium	has a professional reviewer team	response time	number of reviews per month
6	Ξ	eBird	will protect my sensitive data	unfavorable	wol	will protect my sensitive data	has a respectable institution behind	is a serious project
10	Ξ	eBird	will protect my sensitive data	neutral	high	will protect my sensitive data	ı	ı
F	E	eBird	do not have economic ambitions	neutral	wol	do not have economic ambitions	ı	ı
12	E	eBird	do not have economic ambitions	unfavorable	medium	do not have economic ambitions	open access to all data	publish its balance sheet
13	E	eBird	is a serious project	favorable	medium	is a serious project	has a large number of participants	ı
14	Σ	eBird	is a serious project	favorable	medium	is a serious project	has a respectable institution behind	ı

Table 19 - Map of situations and expected evidence for the actors of the citizen science case study simulation

## Applying the definition procedure

The simulation started by applying the procedure to define the situations in terms of the suitcase model components (item 4.2, pg. 75). Table 19 lists the 14 trust situations defined in this case study. Each situation is described with its trustor, trustee, expected behavior, context, importance, and until three expected evidence.

We decided to duplicate the situations involving eBird and the P1 actors, to create situations with the same expected behavior, but with different importance and context. For instance, the eBird expects that the participants can report data accurately in Situations 1 and 2; however, in the first situation, the context is 'unfavorable' and the importance is 'high', whereas in the second the context is 'neutral' and the importance is 'low'. This difference reflects, in the list of expected evidence, which one is more demanding in terms of expected evidence for the more important and unfavorable situation.

Some expected evidence was also shared among situations, to make it possible that a piece of evidence would be collected from different situations in the future. For instance, eBird could collect expected evidence that a participant 'share space and time with others' in two types of situations 'are committed to the project' and may work in groups'.

#### Considerations about the definition procedure

Table 19 is an evidence that we can use the procedure to define different trust situations, with different meanings, using the suitcase model.

Possibly, other trust managers could select another list of expected evidence for all the situations in this simulation, or make an interview with the actors to find the real expected evidence, or replace them by quality measures (completeness, integrity, consistency, etc.), or even subjective values.

However, the possibility to have infinite arranges of expected evidence is a consequence of the suitcase perspective, which is one of the main contributions of this work. This possibility to use what is the most important for the trustor, in his eyes, or in the eyes of the trust manager, to make a trust decision is an important evolution.

Cofat's model (COFTA, 2007) goes closer to this, but in a subjective way. Since

the proofs of trust are not expected, it is not possible to know when a proof must be considered, and how this must be done. In the other hand, Marsh's model (MARSH, 1994) uses subjective measures for the capacity (Perceived Competence) and the Interest (Estimated General Trust) to behave such as expected. In these subjective values, the proofs of trust are less clear than Cofta's.

Once, we successfully used the procedure to define real-life trust situations, we present in the next item the agent's reaction according to each situation.

# 6.3 Reaction under different situations

Using the suitcase model, it is obvious that different situations, with distinct expectations, must generate different reactions translated into numbers. However, this is an evolution compared with the original PlayGround, in which the situation has impact only after the decision is made to cooperate or not, when the payoff matrix is selected.

We present in this item a comparison of the reactions of Participant 1 (P1) to demonstrate how the expectations and evidence can be used to consider what is most important in each situation. Those reactions considered the agent's timeline of trust situations described in Table 20.

P1's timeline of trust situations Sit# Trustee Piece of Evidence Collected 1 Piece of Evidence Collected 2 Piece of Evidence Collected 3 Result d 14 eBird is a serious project C/C 1 13 eBird is a serious project D/D 2 12 eBird open access to all data publish its balance sheet D/D 3 11 eBird do not have economic ambitions C/D 10 eBird will protect my sensitive data has a respectable institution behind is a serious project D/D 5 eBird will protect my sensitive data has a respectable institution behind is a serious project C/C 6

Table 20 – Participant P1's timeline of trust situations

We compared the values of a virtual 15<sup>th</sup> interaction in the cases of P1 playing again a Situation 9 (S9), or a Situation 14 (S14) with eBird. See Table 19, pg. 99, for further information about these situations.

P1's reactions were evaluated in terms of the trustworthiness values, since this

is the component of trust sensitive to the situation. The behavioral decision was not analyzed since it is a decision that happens after the trustor is impacted by the situation. Participant 1 is a Realistic agent and has a memory span of 10 (Table 18, pg. 97).

The following items present the two scenarios: 1- considering a new Situation 9, and 2- considering a new Situation 14, described using the suitcase model formalism (Table 8, pg.66).

## Scenario 1 – P1 and eBird playing a new S9 in 15th interaction

In this scenario, the trustor P1 expects the eBird to protect his sensible data.

- Bex = 'will protect my sensible data'
- trustfulness = Realistic = 0.47 (random value between from 0.33 until 0.66)
- C =' unfavorable'
- I = 'low'
- $s_i = s_{15} = (C, I, Bex) = ('unfavorable', 'low', 'will protect my sensible data')$
- $Ex = \{subject_1, subject_2, subject_3\} =$  $\{'will\ protect\ my\ sensible\ data', 'has\ respectable\ instituion\ behind', 'is\ a\ serious\ project'\} =$ |Ex| = 3
- $recip_{eBird \rightarrow P1} = P1's$  action in last situation  $s_{14}$  was 'Cooperation' = 1
- total = all Pieces of Evidence remembered = 11
- $Ec = \{Ec_1, ..., Ec_i\} = \{Ec_1, Ec_2, Ec_3\} = |Ec| = 3$

#### subject<sub>1</sub> = will protect my sensible data

$$\circ \quad Ec_1 = \{e_{1,9}, e_{1,10}\} = \{(subject_1, s_9), (subject_1, s_{10})\}\$$

$$\circ |Ec_1| = 2$$

o 
$$recur_{e_1} = \frac{|Ec_1|}{total} = \frac{2}{11} = 0.18$$

#### $e_{1.9} = 1$

$$\circ s_{15} = s_9 \rightarrow diff\_sit_{e_{19}} = 1$$

o 
$$dist_{1.9} = 6$$

#### $e_{1.10} = 1$

$$\circ \quad s_{15} \neq \ s_{10} \rightarrow \ diff\_sit_{e_{1,10}} = trustf = 0.47$$

o 
$$dist_{1.10} = 5$$

$$\circ \widehat{Ec_1} = 0.181 * \frac{\left(1 * \frac{1}{6}\right) + \left(1 * \frac{0.47}{5}\right)}{2} = 0.181 * \frac{(0.167 + 0.094)}{2} = 0.024$$

## subject<sub>2</sub> = has respectable instituion behind

o 
$$Ec_2 = \{e_{2,9}, e_{2,10}\} = \{(subject_2, s_9), (subject_2, s_{10})\}$$

$$\circ$$
  $|Ec_2|=2$ 

$$\circ$$
  $recur_{e_2} = \frac{|Ec_2|}{total} = \frac{2}{11} = 0.18$ 

#### $e_{2.9} = 1$

$$\circ \quad s_{15} = s_9 \rightarrow diff\_sit_{e_{1,9}} = 1$$

o 
$$dist_{1.9} = 6$$

## $e_{2.10} = 1$

$$\circ \quad s_{15} \neq s_{10} \rightarrow diff\_sit_{e_{2.10}} = trustf = 0.47$$

o 
$$dist_{2.10} = 5$$

$$\circ \quad \widehat{Ec_2} = 0.181 * \frac{\left(1 * \frac{1}{6}\right) + \left(1 * \frac{0.47}{5}\right)}{2} = 0.181 * \frac{(0.167 + 0.094)}{2} = 0.024$$

## $subject_3 = is \ a \ serious \ project$

$$\circ \quad Ec_3 = \left\{e_{3,9}, e_{3,10}, e_{3,13}, e_{3,14}\right\} =$$

$$\{(subject_3, s_9), (subject_3, s_{10}), (subject_3, s_{13}), (subject_3, s_{14})\}$$

$$\circ |Ec_3| = 4$$

$$\circ recur_{e_3} = \frac{|Ec_3|}{total} = \frac{4}{11} = 0.36$$

#### $e_{30} = 1$

$$\circ \quad s_{15} = s_9 \rightarrow diff\_sit_{e_{3.9}} = 1$$

o 
$$dist_{3.9} = 6$$

#### $e_{3.10} = 1$

$$o s_{15} \neq s_{10} \rightarrow diff\_sit_{e_{3,10}} = trustf = 0.47$$

o 
$$dist_{3.10} = 5$$

#### $e_{3.13} = 1$

$$\circ \quad s_{15} \neq s_{13} \rightarrow diff\_sit_{e_{3,13}} = trustf = 0.47$$

o 
$$dist_{3,13} = 2$$

#### $e_{3.14} = 1$

$$\circ \quad s_{15} \neq s_{14} \rightarrow diff\_sit_{e_{3,14}} = trustf = 0.47$$

o 
$$dist_{3.14} = 1$$

$$\circ \quad \widehat{Ec_3} = 0.36 * \frac{\left(1 * \frac{1}{6}\right) + \left(1 * \frac{0.47}{5}\right) + \left(1 * \frac{0.47}{2}\right) + \left(1 * \frac{0.47}{1}\right)}{4} = 0.36 * \frac{0.167 + 0.094 + 0.235 + 0.47}{4} = 0.087$$

$$\circ \quad trustw_{P1 \rightarrow eBird} = \ recip_{eBird \rightarrow P1} * \frac{\sum_{i=0}^{3}\widehat{Ec}_{i}}{3} = 1 * \frac{0.024 + 0.024 + 0.087}{3} = \frac{0.135}{3} = 0.045$$

# Scenario 2 - P1 and eBird playing a new S14 in 15th interaction

In this scenario, the trustor P1 expects the eBird to protect his sensible data.

- Bex = 'is a serious project'
- trustfulness = Realistic = 0.47 (random value between from 0.33 until 0.66)
- C = 'favorable'
- I = 'medium'
- $s_i = s_{15} = (C, I, Bex) = ('favorable', 'medium', 'is a serious project')$
- Ex = {subject<sub>1</sub>, subject<sub>2</sub>} =
   {'is a serious project', 'has respectable instituion behind'} = |Ex| = 2
- $recip_{eBird \rightarrow P1} = P1's$  action in last situation  $s_{14}$  was 'Cooperation' = 1
- total = all Pieces of Evidence remembered = 11
- $Ec = \{Ec_1, ..., Ec_i\} = \{Ec_1, Ec_2\} = |Ec| = 2$

#### $subject_1 = is a serious project$

- $c_1 = \left\{e_{1,9}, e_{1,10}, e_{1,13}, e_{1,14}\right\} = \left\{(subject_1, s_9), (subject_1, s_{10}), (subject_1, s_{13}), (subject_1, s_{14})\right\}$
- $\circ$   $|Ec_1| = 4$
- $\circ$   $recur_{e_1} = \frac{|Ec_1|}{total} = \frac{4}{11} = 0.36$

#### $e_{1.9} = 1$

o 
$$s_{15} \neq s_9 \rightarrow diff\_sit_{e_{19}} = trustf = 0.47$$

o 
$$dist_{1.9} = 6$$

## $e_{1,10} = 1$

$$\circ \quad s_{15} \neq s_{10} \rightarrow diff\_sit_{e_{1,10}} = trustf = 0.47$$

$$\circ \quad dist_{1.10} = 5$$

#### $e_{1.13} = 1$

$$\circ \quad s_{15} = s_{13} \rightarrow diff\_sit_{e_{1,13}} = trustf = 1$$

o 
$$dist_{1.13} = 2$$

#### $e_{1.14} = 1$

$$\circ \quad s_{15} = s_{14} \rightarrow diff\_sit_{e_{114}} = trustf = 1$$

$$\begin{array}{l} \circ \quad dist_{1.14} = 1 \\ \circ \quad \widehat{Ec_1} \ = 0.36 * \frac{\left(1 * \frac{0.47}{6}\right) + \left(1 * \frac{0.47}{5}\right) + \left(1 * \frac{1}{2}\right) + \left(1 * \frac{1}{1}\right)}{4} = 0.36 * \frac{0.078 + 0.094 + 0.5 + 1}{4} = 0.150 \\ \underline{subject_2} = has \ respectable \ instituion \ behind \end{array}$$

$$\circ \quad Ec_2 = \left\{e_{2,9}, e_{2,10}\right\} = \left\{(subject_2, s_9 \;)\;, (subject_2, s_{10})\right\}$$

$$\circ |Ec_2| = 2$$

$$\circ$$
  $recur_{e_2} = \frac{|Ec_2|}{total} = \frac{2}{11} = 0.18$ 

## $e_{2.9} = 1$

$$\circ \quad s_{15} \neq s_9 \rightarrow diff\_sit_{e_{29}} = trustf = 0.47$$

o 
$$dist_{2.9} = 6$$

#### $e_{1.10} = 1$

$$\circ \quad s_{15} \neq s_{10} \rightarrow diff\_sit_{e_{2.10}} = trustf = 0.47$$

o 
$$dist_{2.10} = 5$$

$$\circ \quad \widehat{Ec_2} = 0.18 * \frac{\left(1 * \frac{0.47}{6}\right) + \left(1 * \frac{0.47}{5}\right)}{2} = 0.18 * \frac{0.078 + 0.094}{2} = 0.016$$

o 
$$trustw_{P1 \to eBird} = recip_{eBird \to P1} * \frac{\sum_{i=0}^{2} \widehat{Ec_{i}}}{2} = 1 * \frac{0.150 + 0.016}{2} = \frac{0.166}{2} = 0.083$$

In the scenario where the situation was S9, the trustfulness value was 0.045, and in the other scenario, where the situation was S14, the value was 0.083. These results are discussed in the next item.

# 6.4 Considerations about the case study

Once our main goal was to verify if the definition procedure could decode different trust situations, i.e., if the model could support different meanings of trust, and if the expectations and the evidence could distinguish what was most important for the trustor in each situation, we considered our simulation a success. In this work, we did not want to compare the impact of the expectations between themselves, but we wanted to demonstrate the fact that it is possible to consider them on the decision, making the decision less subjective than in other models such as those from (MARSH, 1994) and (COFTA, 2007).

Table 19 (pg. 99) was an evidence that different trust situations, with different meanings, may be decoded using the definition procedure and the suitcase model. The suitcase model has shown to be comprehensive and workable at same time, since it can handle different trust situations while considering what is relevant to each trustor.

As a first observation of the agent's reactions, the trustfulness values showed that P1 agent reacted differently to Situation 9 (0.045) and Situation 14 (0.083), independently of a comparison between values. This is the most important to us, since we wanted to discuss how to compute trustworthiness giving different importance to the expectations and the evidences.

In a second observation, it is possible to verify that the situation with more of the same situations in the past, more recurrence of *Pieces of Evidence*, and more *Pieces of Evidence* from a recent past, achieved the highest value of trustfulness.

For instance, in scenario 2,  $e_{1,13}$  and  $e_{1,14}$  had a positive impact in  $\widehat{Ec_1}$  computation, once they were *Pieces of Evidence* collected from the same situation  $(s_{15}=s_{14}=s_{13})$ , and from a recent past  $(dist_{1.14}=1,dist_{1.13}=2)$ . The importance of these attenuators was verified in the same scenario, once the *Pieces of Evidence* of the *Evidence Ec*<sub>2</sub> were collected in opposite conditions, i.e. from a distant past and from different situations, resulting in a value  $(\widehat{Ec_2}=0.016)$  that is nine times lower than the previous *Evidence*  $(\widehat{Ec_2}=0.150)$ . The difference of quantity of *Pieces of Evidence* do not affect this conclusion, once they compose a mean in the *Collected Evidence Value* computation.

These conclusions can also be applied to scenario 1. For instance, the *Evidence*  $(Ec_3)$  which has pieces from more recent situations  $(e_{1,14},e_{1,13})$ , has a greater value than other *Evidence*  $(\widehat{Ec_1}=0.024,\widehat{Ec_2}=0.024,\widehat{Ec_3}=0.087)$ . Furthermore, the *Pieces of Evidence* from the same situations have more positive impact than others from different situations (for example,  $diff_-sit_{e_{1,9}}=1$  and  $diff_-sit_{e_{2,10}}=0.47$ ).

The reciprocity attenuators were not tested in this simulation, because the last interaction of the trustor was not a 'defection'. However, their impact can be compared with the impact of the recurrence attenuator. In both scenarios, it was verified that the *Evidence* with more recurrence had a positive impact in *Collected Evidence Value*. For instance, in scenario 1, the recurrence of *Evidence*  $Ec_3$  ( $recur_{e_3} = 0.36$ ) had a more positive impact than the recurrence of *Evidence*  $Ec_1$  and  $Ec_2$  ( $recur_{e_1} = recur_{e_2} = 0.18$ ).

We verified in this case study that the list of Expected Evidence, and the

evidence attenuation, allows an objective computation of the different reactions of the trustor to each specific situation, what may confirm our hypothesis "that trust can be simulated under a suitcase perspective with potential to handle different meanings".

# 7 Discussion

The main objective of this research was to model trust under a suitcase word perspective and apply it to a case study. We divided this objective in two hypotheses to be verified: 1) Trust is a suitcase word and 2) Trust can be modeled and simulated under a suitcase word perspective with potential to handle different meanings (item 1.4, pg. 24).

## Verifying the 'suitcase word' hypothesis

Aiming to verify the first hypothesis, we presented in Chapter 2 an analysis that collected evidence in scientific literature about the suitcase perspective of trust.

We verified that a wide range of disciplines has studied it, and that no one of them has an authoritative definition or model for trust. The articles retrieved were published in roughly equal percentage across Social Sciences (35%), Life Science & Medicine (29%), and Technology (29%) domains. Furthermore, almost half of the 55,249 articles belong to more than one research area, including different domains.

As a complementary analysis, we constructed maps of co-occurrence keywords from different research areas, assuming that the articles keywords articles could distinguish the many meanings of trust. Maps show different clusters of related keywords, which was inferred as an evidence that they represented one or more meanings.

The evidence that trust occurs in different areas of science, provided by the distribution analysis of articles, and that trust has different meanings in these areas, provided by the co-occurrence maps of keywords, were enough to confirm our hypothesis that trust is a suitcase word, under (MINSKY, 2006) definition of the term.

## **Learning with the PlayGround replica**

To learn more details about (MARSH, 1994), we decided to replicate the PlayGround in an Internet version. This was essential to understand his formalism in a practical way. However, the implementation information was not enough to allow an exact replication, so that we made some assumptions about details.

The replication results were as expected. However, we found a discrepancy between the implementation instructions and the output values provided by the author (MARSH, 1994). Unfortunately, we did not find reasons to invert the adjustment of agent C General and Basic trust values in Experiment A Part II. This fact did not affect our simulator, once we were able to replicate the error exactly as shown in the original output. At the end, we did not consider this inversion in our simulator.

The changes in the Internet replica were minimal to run the suitcase model and construct the PlayGround 2.0. We created a new strategy under the suitcase perspective, the 'Suitcase Trustor', so that we could have both trustor agents in the field.

#### Verifying the 'simulating' hypothesis

Having the PlayGround 2.0, we verified the hypothesis that trust could be simulated under a suitcase perspective, i.e. with potential to handle many meanings of trust. We proposed a case study in the domain of Citizen Science in Biodiversity, where trust as a direct impact on the production and use of biodiversity data.

The main goal of this case study was not to compare strategies or to analyze actors grouping, but to verify if suitcase agents could react differently according to each situation, i.e., if actors could have a different understanding about the meanings of the word 'trust'.

We promoted a simulation with an institution (eBird), a scientist, and one participant ('a citizen scientist'), which were confronted with different situations of trust. Furthermore, the same actor could play in similar situations, which different conditions of importance and context, so that we could capture the influence of these components in the decision.

As expected, different values for trustworthiness were presented due to the list of expected evidence, and its attenuation. This list considered what is important for trustworthiness in the eyes of the trustor, or in the eyes of the trust manager. The reactions observed confirmed our hypothesis "that trust can be simulated under a suitcase perspective with potential to handle different meanings".

## 8 Final remarks

This chapter presents the conclusions and proposed future steps of this research.

## 8.1 Conclusions

From our work we can affirm that trust is a suitcase word, under Minsky definition of the term (MINSKY, 2006), and this can change the way we identify, model, and simulate it.

Based on this assumption and other conceptual foundations (BAUER, 2015; COFTA, 2007; MARSH, 1994; TULLBERG, 2007) we have been able to create our model of trust under a suitcase perspective, which confirms our hypothesis(item 1.4, pg. 24). We also found that it is possible to define trust situations, making them less subjective and more workable to measure. This allows users and trust managers to decode a real-life situation in components of a model.

The trust-as-a-suitcase-word model was defined as a result of trustor's trustfulness and trustee's trustworthiness values (TULLBERG, 2007). Trustworthiness, in its turn, was defined as a function of the expected and collected evidence. Defining a list of expected evidence reduces the complexity of analysis, once only what is the most important for the trustor will be considered.

The attenuators proposed can be used to relativize the impact of evidence from a distant past, an evidence with low recurrence, a trustor's defection in last situation, and an evidence from different situation. They help capture the conditions in which the evidence was collected.

With the PlayGround 2.0 applied in a case study in the domain of Citizen Science in Biodiversity, we verified that the suitcase agents could react differently according to each situation. As expected, the trustor's values for trustworthiness changed according to the importance of each situation.

This is possible because the suitcase model is a less subjective method to measure trustworthiness as compared to (MARSH, 1994) and (COFTA, 2007) models, once their models do not define what is "General Trust" (Table 5, pg. 47) or "trust-

related proofs" (Figure 11, pg. 52) for each specific situation.

Marsh did not want to specify the meaning of trust in each situation, and this approach makes the measurement more subjective, requiring additional aspects to make a decision. In the other hand, Cofta goes closer to define what is important for each situation, but in a subjective way. Since his proofs of trust are not expected, it is not possible to know when a proof must be considered, and what to do with it in terms of importance. As cited before, each area or model tends to view trust from its own perspective (MCKNIGHT; CHERVANY, 2001). The suitcase model wants to delegate to its users the task of defining the perspectives, i.e., the expectations.

They can use a list of expected evidence as a strategy to identify which meanings of trust are being considered in the situation. It would be difficult to find evidence for trustworthiness if we do not know what is important for the trustor in a situation. Besides that, a finite list of terms may reduce the complexity of analysis, once we have to measure only the most important evidence for each situation (important in trustor's eyes). These expectations are evidence that confirm our hypothesis "that trust can be simulated under a suitcase perspective with potential to handle different meanings".

### 8.2 Future work

Despite all the positive results, we anticipate a number of areas that offer opportunities for research and improvement of the model.

Considering the hypothesis that trust is a suitcase word, it would be beneficial to analyze in which scope and depth trust has many meanings. Although the Web of Science (WOS) material did support our conclusion that trust has different meanings, it would be interesting to expand our analysis to more research areas in WOS or other respectable databases, such as PubMed or Scopus. More data could explicit in which level and extension trust has many meanings across science.

Furthermore, there is also an opportunity for adjustment of the trustworthiness values. Due to the attenuation logic, they tend to be always lower, and this might generate very small values that lose significance in face of other values, such as the trustfulness.

The list of expected evidence in trustworthiness might be also weighted, to give

a hierarchy to what is important to the trustor. The sum of trustworthiness and trustfulness in trust computation could be also weighted.

In this work, we did not define trustfulness as a function of expected and collected evidence, but this is something interesting to do, once it can reduce its subjectivity. Context, importance, and risk must also be developed in this way to reduce subjectivity.

Our research achieved success in confirm our hypotheses. However, this was an introductory work, focused in the application of the suitcase word perspective. More than a new model, this work aimed to bring a new approach to treat trust, which may also serve as inspiration to other works.

## **Bibliography**

ALABRI, A.; HUNTER, J. Enhancing the quality and trust of citizen science data. **Proceedings - 2010 6th IEEE International Conference on e-Science, eScience 2010**, p. 81–88, 2010.

AXELROD, R. Effective Choice in the Prisoner's Dilemma. **The Journal of Conflict Resolution**, v. 24, n. 1, p. 3–25, 1980.

AXELROD, R. The Evolution of Cooperation. New York: Basic Books, 1984.

BAIER, A. Trust and Antitrust. **Ethics**, v. 96, n. 2, p. 231–260, 1986.

BARBER, B. **The Logic and Limits of Trust**. 1st editio ed. [s.l.] Rutgers University Press, 1983.

BAUER, P. C. Three Essays on the Concept of Trust and its Foundations. [s.l.] University of Bern, 2015.

BISBY, F. A. The quiet revolution: biodiversity informatics and the internet. **Science** (New York, N.Y.), v. 289, n. 5488, p. 2309–2312, 2000.

CARTOLANO, E. **PlayGround 2.0**. Disponível em: <www.trust.cartolano.com.br>. Acesso em: 14 ago. 2016.

CARTOLANO, E.; SARAIVA, A. M.; STEVENSON, R. D. PlayGround 2.0: Simulating behavior decisions with trust and control computations. **Proceedings - 2016 14th Annual Conference on Privacy, Security and Trust (PST)**, Auckland, 2016, pp. 400-406.

CASTELFRANCHI, C.; FALCONE, R. Trust and control: A dialectic link. **Applied Artificial Intelligence**, v. 14, n. 8, p. 799–823, 2000.

CHO, J.-H.; CHAN, K.; ADALI, S. A Survey on Trust Modeling. **ACM Computing Surveys**, v. 48, n. 2, p. 1–40, 2015.

COFTA, P. Trust, Complexity and Control: Confidence in a Convergent World. Chichester: John Wiley & Sons Ltd, 2007.

E-BIRD. e-Bird. Disponível em: <www.ebird.org>. Acesso em: 29 ago. 2016.

FUKUYAMA, F. **Trust: The Social Virtues and The Creation of Prosperity**. New York: Free Press, 1996.

GALLOWAY, A. W. E.; TUDOR, M. T.; HAEGEN, W. M. W. The Reliability of Citizen Science: A Case Study of Oregon White Oak Stand Surveys. **Wildlife Society Bulletin**, v. 34, n. 5, p. 1425–1429, 2006.

GAMBETTA, D. Can we trust. **Department of Sociology, University of Oxford**, n. May 2002, p. 213–237, 1988.

GAMBETTA, D. Can We Trust Trust? 2000.

GOLBECK, J. Generating predictive movie recommendations from trust in social networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v. 3986 LNCS, p. 93–104, 2006.

GOOGLE. Google Books Ngram, 2017.

HARDIN, R. Trust and Trustworthiness. New York: Russell Sage Foundation, 2002.

INATURALIST. **iNaturalist**. Disponível em: <a href="http://www.inaturalist.org">http://www.inaturalist.org</a>. Acesso em: 29 ago. 2016.

ISPOT. **iSpot**. Disponível em: <a href="http://www.ispotnature.org/">http://www.ispotnature.org/</a>>. Acesso em: 29 ago. 2016.

JOSANG, A.; ISMAIL, R.; BOYD, C. A survey of trust and reputation systems for online service provision. **Decision Support Systems**, v. 43, n. 2, p. 618–644, 2007.

KELLING, S. et al. Taking a "Big Data" approach to data quality in a citizen science project. **Ambio**, v. 44, p. 601–611, 2015a.

KELLING, S. et al. Can observation skills of citizen scientists be estimated using species accumulation curves? **PLoS ONE**, v. 10, n. 10, p. 1–20, 2015b.

KNIGHT, V. Axelrod Python library.

LUHMANN, N. Trust and Power. Chichester: John Wiley & Sons Ltd, 1979.

MARSH, S. P. Formalising Trust as a Computational Concept. [s.l.] University of Stirling, 1994.

MAYER, R. C.; DAVIS, J. H.; SCHOORMAN, F. D. An Integrative Model of Organizational Trust. **The Academy of Management Review**, v. 20, n. 3, p. 709–734, 1995.

MCKNIGHT, D. H.; CHERVANY, N. L. The Meanings of Trust. Minneapolis: [s.n.].

MCKNIGHT, D. H.; CHERVANY, N. L. What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology. **International Journal of Electronic Commerce**, v. 6, n. 2, p. 35–59, 2001.

MCKNIGHT, D. H.; CHOUDHURY, V.; KACMAR, C. Developing and validating trust measures for e-commerce: An integrative typology. **Information Systems Research**, v. 13, n. 3, p. 334–359, 2002.

MINSKY, M. The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. [s.l.] Simon & Schuster, 2006.

O'HARA, K. **A General Definition of Trust**. Southampton: [s.n.]. Disponível em: <a href="http://eprints.soton.ac.uk/273193/">http://eprints.soton.ac.uk/273193/</a>>.

REEF. REEF. Disponível em: <www.reef.org>. Acesso em: 29 ago. 2016.

ROTTER, J. B. A new scale for the measurement of interpersonal trust. **Journal of Personality**, v. 35, n. Mh 11455, p. 651–665, 1967.

SILVERTOWN, J. A new dawn for citizen science. **Trends in ecology & evolution**, v. 24, n. 9, p. 467–71, set. 2009.

SOURCE, J. P. C. Citizen Science: Can Volunteers Do Real Research? **BioScience**, v. 58, n. 3, p. 192–197, 2008.

THOM, D. H.; HALL, M. A.; PAWLSON, L. G. Measuring patients' trust in physicians when assessing quality of care. **Health Affairs**, v. 23, n. 4, p. 124–132, 2004.

TULLBERG, J. Trust-The importance of trustfulness versus trustworthiness. **Journal of Socio-Economics**, v. 37, n. 5, p. 2059–2071, 2007.

VAN ECK, N. J.; WALTMAN, L. 2007 VOS: A new method for visualizing similarities between objects. **Advances in Data Analysis**, n. April 2006, p. 299–306, 2007.

VAN ECK, N. J.; WALTMAN, L. 2009 How to Normalize Cooccurrence Data? An Analysis of SomeWell-Known Similarity Measures. **Journal of the American Society for Information Science and Technology**, v. 60, n. 8, p. 1635–1651, 2009.

VAN ECK, N. J.; WALTMAN, L. 2010a Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, v. 84, n. 2, p. 523–538, 2010.

VOSVIEWER. VOSviewer.

WALTMAN, L.; VAN ECK, N. J.; NOYONS, E. C. M. 2010b A unified approach to mapping and clustering of bibliometric networks. **Journal of Informetrics**, v. 4, n. 4, p. 629–635, 2010.

WIGGINS, A. et al. **Mechanisms for data quality and validation in citizen science**. Proceedings - 7th IEEE International Conference on e-Science Workshops, eScienceW 2011. **Anais**...2011

WIGGINS, A.; CROWSTON, K. From conservation to crowdsourcing: A typology of citizen science. **Proceedings of the Annual Hawaii International Conference on System Sciences**, p. 1–10, 2011.

WIKIPEDIA. **Wikipedia: Trust, but verify**. Disponível em: <a href="https://en.wikipedia.org/wiki/Trust">https://en.wikipedia.org/wiki/Trust</a>, but verify>. Acesso em: 14 ago. 2016.



# A. Extraction of articles records from Web of Science

Search on the Thompson Reuters' Web of Science was focused on the database "Web of Science Core Collection". The words "trust" or "trustworthy" or "trustworthiness" were filled in the field "Topic", that covers the title, abstract, author keywords, and Keywords Plus, which are additional keywords suggested by Web of Science from significant frequently occurring words in the titles of articles. For distribution analysis we used the option "Analyze Results" ranked by "Research Areas" and "Web of Science Categories", with minimum record count (threshold) of 1, and displayed by 500 at a time. These data were downloaded with the "Save Analysis Data to File" tool. To support the co-occurrence analysis, we downloaded the records by using the tool "Save to Others File Formats" and the options "Full Record and Cited References" and "Other Reference Software", which agree with the VOSViewer's requirements. Due to a platform restriction, records were downloaded only by 500 per time. We proceeded a general search without filters, a filtered search using the four Research areas, and another filtered search with Computer Science categories. The following are examples of search descriptions extracted from Web of Science:

- <u>General</u>: You searched for: TOPIC: (trust) OR TOPIC: (trustworthy) OR TOPIC: (trustworthiness) Refined by: DOCUMENT TYPES: ( ARTICLE ) AND LANGUAGES: (ENGLISH) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI. Results: 55,249 (from Web of Science Core Collection)
- <u>Computer Science</u>: You searched for: **TOPIC**: (trust) OR **TOPIC**: (trustworthy) OR **TOPIC**: (trustworthiness) **Refined by: DOCUMENT TYPES**: (ARTICLE) AND **LANGUAGES**: (ENGLISH) AND **RESEARCH AREAS**: (**COMPUTER SCIENCE**) **Timespan**: All years. **Indexes**: SCI-EXPANDED, SSCI, A&HCI, ESCI. **Results**: **7,839** (from Web of Science Core Collection)
- <u>Psychology:</u> You searched for: **TOPIC:** (trust) OR **TOPIC:** (trustworthiness) OR **TOPIC:** (trustworthy) **Refined by: DOCUMENT TYPES:** ( ARTICLE ) AND

LANGUAGES: (ENGLISH) AND RESEARCH AREAS: (PSYCHOLOGY) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI. Results: 5,207 (from Web of Science Core Collection)

- <u>Sociology</u>: You searched for: TOPIC: (trust) *OR* TOPIC: (trustworthiness) *OR* TOPIC: (trustworthy) Refined by: DOCUMENT TYPES: ( ARTICLE ) AND LANGUAGES: ( ENGLISH ) AND RESEARCH AREAS: ( SOCIOLOGY ) Timespan: All years. Indexes: SCI- EXPANDED, SSCI, A&HCI, ESCI. Results: 1,451 (from Web of Science Core Collection)
- Operation Research Management Science: You searched for: TOPIC: (trust) OR TOPIC: (trustworthy) OR TOPIC: (trustworthiness) Refined by: LANGUAGES: (ENGLISH) AND DOCUMENT TYPES: (ARTICLE) AND RESEARCH AREAS: (OPERATIONS RESEARCH MANAGEMENT SCIENCE) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI. Results: 1,196 (from Web of Science Core Collection)

# B. Hierarchy relation among Web of Science categories

Table 21 presents the hierarchy relation of the terms used by the Web of Science to assign articles to Research areas and Categories of the science. The Research area terms were explicit assigned to main areas 5, whereas the categories were only listed without a direct relation to main areas or Research Areas 6. We made relations according to the name of terms using a method with three verifications. 1) In cases of exact name match of terms for Research area and category, such as Dance, Allergy, and more 119 others (total of 121, or 48%), we set a relation between both. 2) A total of 67 terms (27%), in which the Research area was part of the name of the category, such as "Literature" Research Area and "Literature American" category, were directly related. 3) The other 64 terms (25%) such as "Mathematics" Research area and "Logic" categories were related according to our experience and to colleague's recommendations.

Main Areas (MA):

- AH Arts & Humanities
- LB Life Sciences & Biomedicine
- PS Physical Sciences
- SS Social Sciences
- TC Technology

App – Appearances

<sup>5</sup> Accessed on the January 12, 2017.

Table 21 - Hierarchy relation between categories from Web of Science

#	MA	Research Area (Su)	Web Of Science Categories (Wc)	# App
1	АН	Architecture	Architecture	64
2	AH	Art	Art	90
3	AH	Arts & Humanities Other Topics	Folklore	5
4	AH	Arts & Humanities Other Topics	Humanities Multidisciplinary	159
5	AH	Asian Studies	Asian Studies	44
6	AH	Classics	Classics	10
7	AH	Dance	Dance	13
8	AH	Film Radio Television	Film Radio Television	33
9	AH	History	History	392
10	AH	History	Medieval Renaissance Studies	13
11	AH	History & Philosophy Of Science	History Of Social Sciences	167
12	AH	History & Philosophy Of Science	History Philosophy Of Science	422
13	AH	Literature	Literary Reviews	17
14	AH	Literature	Literary Theory Criticism	5
15	AH	Literature	Literature	68
16	АН	Literature	Literature African Australian Canadian	6
17	AH	Literature	Literature American	5
18	AH	Literature	Literature British Isles	10
19	АН	Literature	Literature German Dutch Scandinavian	1
20	AH	Literature	Literature Romance	4
21	AH	Literature	Literature Slavic	1
22	AH	Literature	Poetry	5
23	AH	Music	Music	46
24	AH	Philosophy	Philosophy	422
25	AH	Religion	Religion	272
26	AH	Theater	Theater	14
27	LS	Agriculture	Agricultural Economics Policy	148
28	LS	Agriculture	Agricultural Engineering	8
29	LS	Agriculture	Agriculture Dairy Animal Science	50
30	LS	Agriculture	Agriculture Multidisciplinary	166
31	LS	Agriculture	Agronomy	61
32	LS	Agriculture	Horticulture	24
33	LS	Agriculture	Soil Science	21
34	LS	Allergy	Allergy	13
35	LS	Anatomy Morphology	Anatomy Morphology	10
36	LS	Anesthesiology	Anesthesiology	60
37	LS	Anthropology	Anthropology	269
38	LS	Behavioral Sciences	Behavioral Sciences	268
39	LS	Biochemistry & Molecular Biology	Biochemical Research Methods	115
40	LS	Biochemistry & Molecular Biology	Biochemistry Molecular Biology	140
41	LS	Biodiversity Conservation	Biodiversity Conservation	209

42	LS	Biophysics	Biophysics	36
43	LS	Biotechnology Applied	Biotechnology Applied	179
+3	LO	Microbiology	Microbiology	173
44	LS	Cardiovascular System & Cardiology	Cardiac Cardiovascular Systems	161
45	LS	Cardiovascular System & Cardiology	Peripheral Vascular Disease	65
46	LS	Cell Biology	Cell Biology	61
47	LS	Cell Biology	Cell Tissue Engineering	10
48	LS	Critical Care Medicine	Critical Care Medicine	135
49	LS	Dentistry Oral Surgery Medicine	Dentistry Oral Surgery Medicine	181
50	LS	Departamental Biology	Biology	143
51	LS	Dermatology	Dermatology	80
52	LS	Developmental Biology	Developmental Biology	11
53	LS	Emergency Medicine	Emergency Medicine	153
54	LS	Endocrinology Metabolism	Endocrinology Metabolism	175
55	LS	Entomology	Entomology	17
56	LS	Environmental Sciences & Ecology	Ecology	457
57	LS	Environmental Sciences & Ecology	Environmental Sciences	985
58	LS	Environmental Sciences & Ecology	Environmental Studies	1,215
59	LS	Environmental Sciences & Ecology	Green Sustainable Science Technology	179
60	LS	Environmental Sciences & Ecology	Ornithology	51
61	LS	Evolutionary Biology	Evolutionary Biology	59
62	LS	Fisheries	Fisheries	38
63	LS	Food Science Technology	Food Science Technology	342
64	LS	Forestry	Forestry	145
65	LS	Gastroenterology Hepatology	Gastroenterology Hepatology	104
66	LS	General & Internal Medicine	Andrology	1
67	LS	General & Internal Medicine	Medicine General Internal	1,853
68	LS	General & Internal Medicine	Medicine Legal	97
69	LS	General & Internal Medicine	Medicine Research Experimental	296
70	LS	Genetics Heredity	Genetics Heredity	357
71	LS	Geriatrics & Gerontology	Geriatrics Gerontology	173
72	LS	Geriatrics & Gerontology	Gerontology	206
73	LS	Health Care Sciences & Services	Health Care Sciences Services	1,890
74	LS	Health Care Sciences & Services	Health Policy Services	1,287
75	LS	Health Care Sciences & Services	Primary Health Care	264
76	LS	Hematology	Hematology	70
77	LS	Immunology	Immunology	235
78	LS	Infectious Diseases	Infectious Diseases	370
79	LS	Integrative Complementary Medicine	Integrative Complementary Medicine	77
80	LS	Marine Freshwater Biology	Marine Freshwater Biology	56
81	LS	Medical Ethics	Medical Ethics	332

82	LS	Medical Informatics	Medical Informatics	452
83	LS	Medical Laboratory Technology	Medical Laboratory Technology	41
84	LS	Microbiology	Microbiology	106
85	LS	Mycology	Mycology	2
86	LS	Neurosciences & Neurology	Clinical Neurology	288
87	LS	Neurosciences & Neurology	Neuroimaging	23
88	LS	Neurosciences & Neurology	Neurosciences	385
89	LS	Nursing	Nursing	1,662
90	LS	Nutrition Dietetics	Nutrition Dietetics	281
91	LS	Obstetrics Gynecology	Obstetrics Gynecology	385
92	LS	Oncology	Oncology	469
93	LS	Ophthalmology	Ophthalmology	121
94	LS	Orthopedics	Orthopedics	136
95	LS	Ortorhinolaryngology	Audiology Speech Language Pathology	31
96	LS	Otorhinolaryngology	Otorhinolaryngology	95
97	LS	Paleontology	Paleontology	15
98	LS	Parasitology	Parasitology	41
99	LS	Pathology	Pathology	74
100	LS	Pediatrics	Pediatrics	411
101	LS	Pharmacology Pharmacy	Pharmacology Pharmacy	387
102	LS	Physiology	Physiology	39
103	LS	Plant Sciences	Plant Sciences	57
104	LS	Psychiatry	Psychiatry	1,029
105	LS	Public Environmental Occupational Health	Public Environmental Occupational Health	3,447
106	LS	Radiology Nuclear Medicine Medical Imaging	Radiology Nuclear Medicine Medical Imaging	188
107	LS	Rehabilitation	Rehabilitation	420
108	LS	Reproductive Biology	Reproductive Biology	63
109	LS	Respiratory System	Respiratory System	198
110	LS	Rheumatology	Rheumatology	96
111	LS	Sport Sciences	Hospitality Leisure Sport Tourism	407
112	LS	Sport Sciences	Sport Sciences	191
113	LS	Substance Abuse	Substance Abuse	176
114	LS	Surgery	Surgery	584
115	LS	Toxicology	Toxicology	64
116	LS	Transplantation	Transplantation	74
117	LS	Tropical Medicine	Tropical Medicine	68
118	LS	Urology & Nephrology	Urology Nephrology	117
119	LS	Veterinary Sciences	Veterinary Sciences	165
120	LS	Virology	Virology	14
121	LS	Zoology	Zoology	90
122	PS	Astronomy Astrophysics	Astronomy Astrophysics	121
123	PS	Chemistry	Chemistry Analytical	108
124	PS	Chemistry	Chemistry Applied	34
125	PS	Chemistry	Chemistry Inorganic Nuclear	13

126	PS	Chemistry	Chemistry Medicinal	21
127	PS	Chemistry	Chemistry Multidisciplinary	74
128	PS	Chemistry	Chemistry Organic	7
129	PS	Chemistry	Chemistry Physical	102
130	PS	Crystallography	Crystallography	6
131	PS	Electrochemistry	Electrochemistry	58
132	PS	Geochemistry Geophysics	Geochemistry Geophysics	63
133	PS	Geology	Geology	11
134	PS	Geology	Geosciences Multidisciplinary	170
135	PS	Mathematics	Logic	34
136	PS	Mathematics	Mathematical Computational Biology	142
137	PS	Mathematics Mathematics		117
138	PS	Mathematics	Mathematics Applied	1,057
139	PS	Mathematics	Mathematics Interdisciplinary Applications	399
140	PS	Mathematics	Statistics Probability	213
141	PS	Meteorology Atmospheric Sciences	Meteorology Atmospheric Sciences	149
142	PS	Mineralogy	Mineralogy	8
143	PS	Mining Mineral Processing	Mining Mineral Processing	18
144	PS	Oceanography	Oceanography	69
145	PS	Optics	Optics	122
146	PS	Physics	Physics Applied	63
147	PS	Physics	Physics Atomic Molecular Chemical	114
148	PS	Physics	Physics Condensed Matter	31
149	PS	Physics	Physics Fluids Plasmas	24
150	PS	Physics	Physics Mathematical	103
151	PS	Physics	Physics Multidisciplinary	138
152	PS	Physics	Physics Nuclear	32
153	PS	Physics	Physics Particles Fields	91
154	PS	Polymer Science	Polymer Science	17
155	PS	Thermodynamics	Thermodynamics	49
156	PS	Water Resources	Limnology	22
157	PS	Water Resources	Water Resources	250
158	SS	Archaeology	Archaeology	78
159	SS	Area Studies	Area Studies	328
160	SS	Business & Economics	Business	4,236
161	SS	Business & Economics	Business Finance	1,269
162	SS	Business & Economics	Economics	2,848
163	SS	Communication	Communication	1,112
164	SS	Criminology Penology  Criminology Penology		362 38
165	SS			
166	SS	Demography	Demography	96
167	SS	Education & Educational Research	Education Educational Research	1,098

168	ss	Education & Educational Research	Education Scientific Disciplines	383
169	SS	Education & Educational Research	Education Special	74
170	SS	Ethnic Studies	Ethnic Studies	108
171	SS	Family Studies	Family Studies	396
172	SS	Geography	Geography	446
173	SS	Geography	Geography Physical	75
174	SS	Government & Law	Law	2,344
175	SS	International Relations	International Relations	539
176	SS	Linguistics	Language Linguistics	83
177	SS	Linguistics	Linguistics	137
178	SS	Psychology	Psychology	446
179	SS	Psychology	Psychology Applied	1,127
180	SS	Psychology	Psychology Biological	83
181	SS	Psychology	Psychology Clinical	589
182	SS	Psychology	Psychology Developmental	495
183	SS	Psychology	Psychology Educational	195
184	SS	Psychology	Psychology Experimental	606
185	SS	Psychology	Psychology Mathematical	26
186	SS	Psychology	Psychology Multidisciplinary	1,371
187	SS	Psychology	Psychology Psychoanalysis	68
188	SS	Psychology	Psychology Social	1,224
189	SS	Public Administration	Public Administration	880
190	SS	Social Issues	Ethics	936
191	SS	Social Issues	Social Issues	597
192	SS	Social Sciences Other Topics	Political Science	1,741
193	SS	Social Sciences Other Topics	Social Sciences Biomedical	1,101
194	SS	Social Sciences Other Topics	Social Sciences Interdisciplinary	1,307
195	SS	Social Sciences Other Topics	Social Sciences Mathematical Methods	316
196	SS	Social Work	Social Work	556
197	SS	Sociology	Sociology	1,440
198	SS	Urban Studies	Planning Development	686
199	SS	Urban Studies	Urban Studies	479
200	SS	Women's Studies	Women S Studies	158
201	TC	Acoustics	Acoustics	26
202	TC	Automation Control Systems	Automation Control Systems	223
203	TC	Computer Science	Computer Science Artificial Intelligence	1,275
204	TC	Computer Science	Computer Science Cybernetics	339
205	тс	Computer Science	Computer Science Hardware Architecture	844
206	TC	Computer Science	Computer Science Information Systems	3,800
207	TC	Computer Science	Computer Science Interdisciplinary Applications	951
208	TC	Computer Science	Computer Science Software Engineering	1,626

209	тс	Computer Science	Computer Science Theory Methods	2,296
210	TC	Construction Building Technology	Construction Building Technology	89
211	TC	Energy Fuels	Energy Fuels	217
212	TC	Engineering	Engineering Aerospace	56
213	TC	Engineering	Engineering Biomedical	82
214	TC	Engineering	Engineering Chemical	141
215	TC	Engineering	Engineering Civil	320
216	TC	Engineering	Engineering Electrical Electronic	1,607
217	TC	Engineering	Engineering Environmental	189
218	TC	Engineering	Engineering Geological	19
219	TC	Engineering	Engineering Industrial	664
220	TC	Engineering	Engineering Manufacturing	224
221	TC	Engineering	Engineering Marine	8
222	TC	Engineering	Engineering Mechanical	122
223	TC	Engineering	Engineering Multidisciplinary	381
224	TC	Engineering	Engineering Ocean	10
225	TC	Engineering	Engineering Petroleum	28
226	TC	Imaging Science Photographic Technology	Imaging Science Photographic Technology	43
227	TC	Information Science Library Science Information Science Library Science		1,531
228	TC	Instruments Instrumentation	Instruments Instrumentation	134
229	TC	Materials Science	Materials Science Biomaterials	2
230	TC	Materials Science	Materials Science Ceramics	4
231	TC	Materials Science	Materials Science Characterization Testing	17
232	TC	Materials Science	Materials Science Coatings Films	8
233	TC	Materials Science	Materials Science Composites	9
234	TC	Materials Science	Materials Science Multidisciplinary	93
235	TC	Materials Science	Materials Science Paper Wood	18
236	TC	Materials Science	Materials Science Textiles	8
237	TC	Mechanics	Mechanics	128
238	TC	Metallurgy Metallurgical Engineering	Metallurgy Metallurgical Engineering	49
239	TC	Microscopy	Microscopy	3
240	TC	Nuclear Science Technology	Nuclear Science Technology	101
241	TC	Operations Research & Management Science	Ergonomics	390
242	TC	Operations Research & Management Science	Industrial Relations Labor	279
243	TC	Operations Research & Management Science	Management	4,594
244	TC	Operations Research & Management Science	Operations Research Management Science	1,173
245	TC	Remote Sensing	Remote Sensing	33
246	TC	Robotics	Robotics	30
247	TC	Science & Technologies Other Topics	Multidisciplinary Sciences	779

248	тс	Science & Technologies Other Topics	Nanoscience Nanotechnology	31
249	TC	Spectroscopy	Spectroscopy	31
250	TC	Telecommunications	Telecommunications	1,613
251	TC	Transportation	Transportation	164
252	TC	Transportation	Transportation Science Technology	148
		146 research areas	252 categories	92,765

## C. VOSviewer configuration

The maps of co-occurrence keywords were created with the option "Create a map based on bibliographic data" of VOSviewer, which runs over files exported from Web of Science, Scopus, PubMed, and RIS. The settings used to create the maps where described in Table 22. The number of articles selected in these setting is described on Table 23 and Table 24.

Table 22 - VOSviewer configuration for WOS co-occurrence maps

Parameter	Value/Option
Mapping approach	Map based on bibliographic data
Type of analysis	Co-occurrence
Unit of analysis	All keywords
Counting method	Fractional counting
Minimum number of occurrences of a keyword	15
Max number of keywords to be selected	100 (manual limitation)
Visualization - Scale	1.0
Visualization - Weights	Occurrences
Visualization – Labels – Size variation	0.50
Visualization – Max. lenght	30
Visualization – Lines – Size variation	300

Table 23 - Numbers from VOSviewer configuration

Research Areas	Docs	Keywords	Threshold
Computer Science	7.839	18.655	448
Operation & Management Science	1.181	5.085	94
Psychology	5.182	14.125	539
Sociology	1.451	5.156	113

Table 24 - Distribution of Documents, Keywords, and Threshold between Web of Science Categories

	Co-Occurrence (keywords)		
Web of Science Categories	Docs	Keywords	Threshold
Computer Science Information Science	3.854	10.458	238
Artificial Intelligence	1.295	4.359	49
Hardware	857	2.512	22
Software Engineering	1.640	4.845	60
Theory Methods	2.310	4.745	62

# D. Maps of co-occurrence

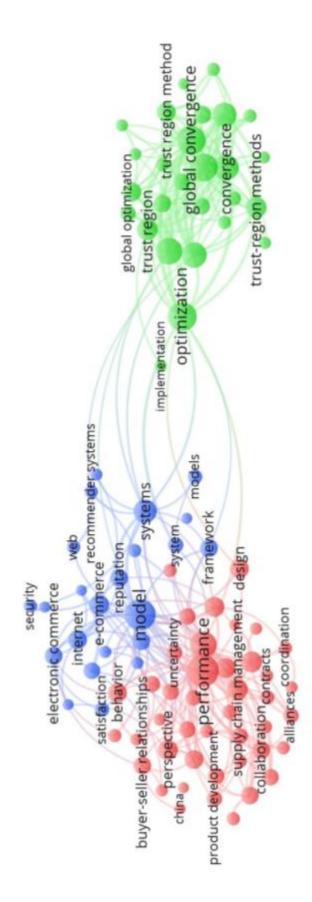


Figure 29 - Co-occurrence map for Operations & Management Science Research Area in Web of Science

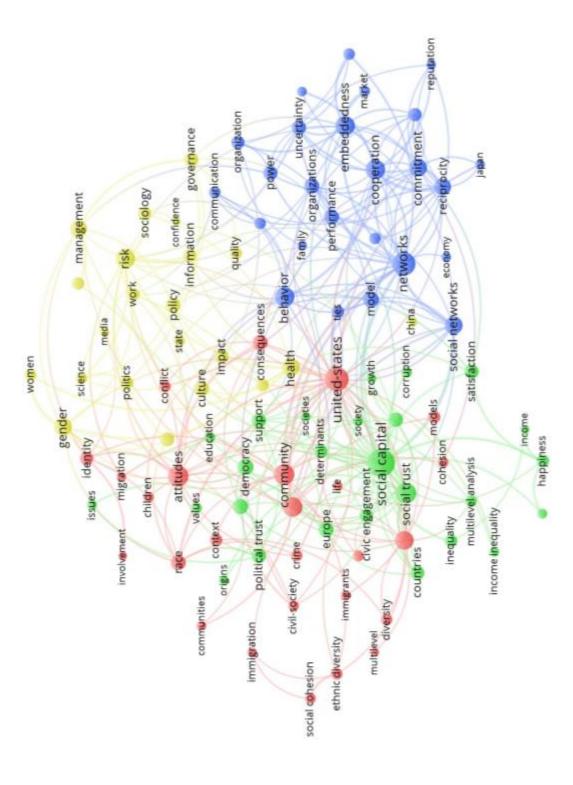


Figure 30 - Co-occurrence map for Sociology Research Area in Web of Science

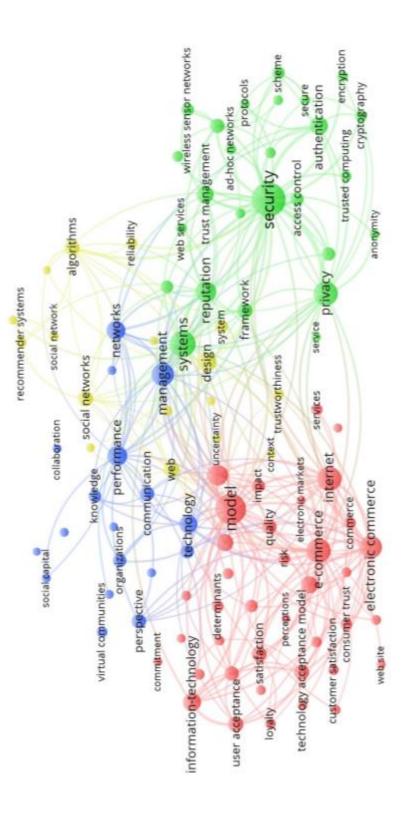


Figure 31 - Co-occurrence map for Computer Science Information Systems Category in Web of Science

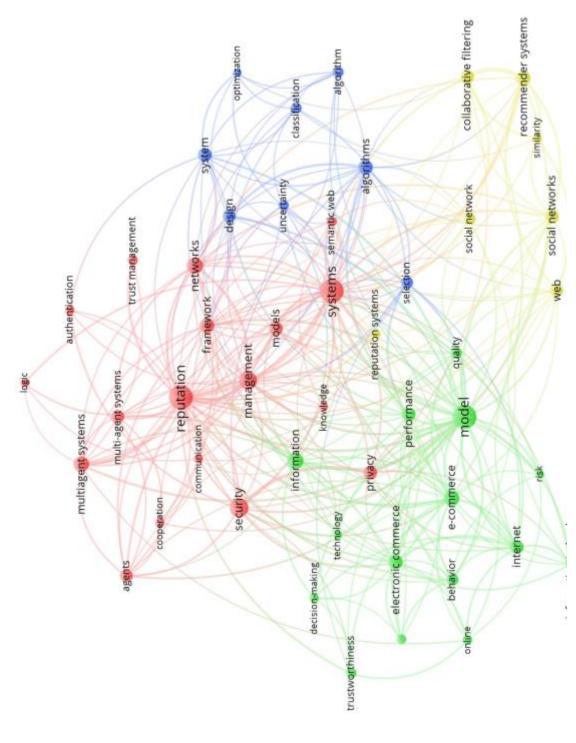


Figure 32 - Co-occurrence map for Artificial Intelligence Category in Web of Science

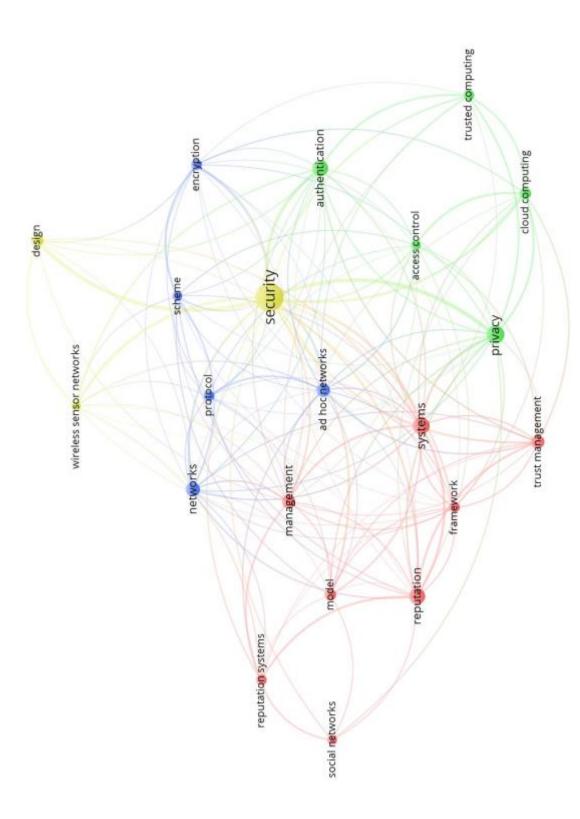


Figure 33 - Co-occurrence map for Hardware Category in Web of Science

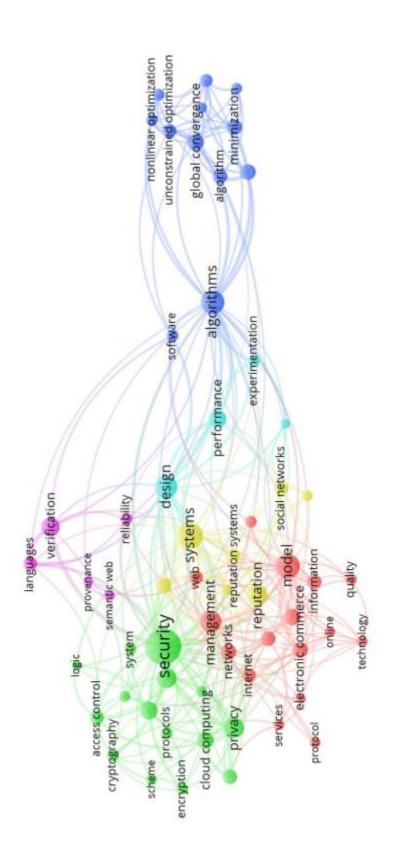


Figure 34 - Co-occurrence map for Software Engineering Category in Web of Science

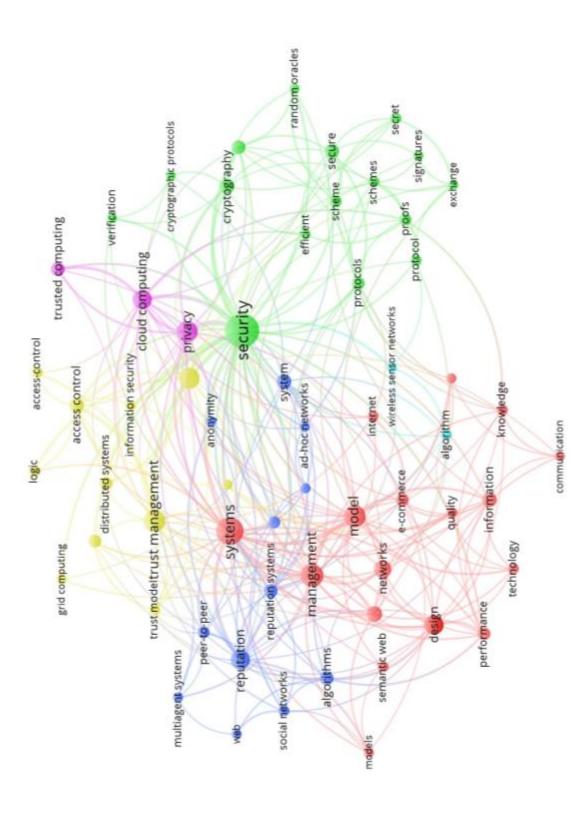


Figure 35 - Co-occurrence map for Theory Methods Category in Web of Science

## E. Results of Marsh's PlayGround

The following tables were extracted from (MARSH, 1994). They were essential during replication of the original PlayGround (Item 5.2), once they are the only evidence about its internal calculations. All tables are samples of outputs from experiment A part 2. Table 25 and Table 26 are results from agent C, whereas Table 27 and Table 28 are results from agent D.

### Agent C (Optimistic) results of Experiment A Part II

Table 25 - Agent C partial results of Experiment A Part II, extracted from (Marsh 1994)

Iteration	Other	Basic	Genl.	Other	Agent	Other	Agent	Other
no.	Name	trust	trust.	Comp	choice	choice	payoff	payoff
14	D	0.091957	0.173836	0.007284	Cooperate	Cooperate	3	3
	D	0.091047	0.158033	0.007212	Cooperate	Cooperate	3	3
13	D	0.091047	0.158033	0.007212	Cooperate	Cooperate	3	3
	D	0.090146	0.143666	0.007141	Cooperate	Cooperate	3	3
12	D	0.090146	0.143666	0.007141	Defect	Cooperate	5	0
	D	0.089253	0.136825	0.00707	Defect	Cooperate	5	0
11	D	0.089253	0.136825	0.00707	Defect	Cooperate	5	0
	D	0.088369	0.13031	0.007	Defect	Cooperate	5	0
10	D	0.088369	0.13031	0.007	Defect	Cooperate	5	0
	D	0.087494	0.124105	0.006931	Defect	Cooperate	5	0
9	D	0.087494	0.124105	0.006931	Defect	Cooperate	5	0
	D	0.086628	0.118195	0.006862	Defect	Cooperate	5	0
8	D	0.086628	0.118195	0.006862	Defect	Cooperate	5	0
	D	0.08577	0.112567	0.006794	Defect	Cooperate	5	0
7	D	0.08577	0.112567	0.006794	Defect	Cooperate	5	0
	D	0.084921	0.107207	0.006727	Defect	Cooperate	5	0
6	D	0.084921	0.107207	0.006727	Defect	Cooperate	5	0
	D	0.08408	0.102102	0.00666	Defect	Cooperate	5	0
5	D	0.08408	0.102102	0.00666	Defect	Cooperate	5	0
	D	0.083248	0.09724	0.006594	Defect	Cooperate	5	0
4	D	0.083248	0.09724	0.006594	Defect	Cooperate	5	0
	D	0.082424	0.09261	0.006529	Defect	Cooperate	5	0
3	D	0.082424	0.09261	0.006529	Defect	Cooperate	5	0
	D	0.081608	0.0882	0.006464	Defect	Cooperate	5	0
2	D	0.081608	0.0882	0.006464	Defect	Cooperate	5	0
	D	0.0808	0.084	0.0064	Defect	Cooperate	5	0
1	D	0.0808	0.084	0.006464	Defect	Cooperate	5	0
Start	D	0.08	0.08	0.0064	Defect	Cooperate	5	0

Table 26 - Agent C final results of Experiment A Part II, extracted from (Marsh 1994)

Iteration	Other	Costs	Ben-	Uti-	Risk	Other	Coop.	Sit.	Impor-
no.	Name		efits	lity		Comp.	Thresh	Trust	tance
14	D	0.1	0.6	6	0.12	0.007284	0.52286	0.682703	0.72
	D	0.1	0.6	6	0.12	0.007212	0.52286	0.682703	0.72
13	D	0.1	0.6	6	0.12	0.007212	0.572918	0.620637	0.72
	D	0.1	0.6	6	0.12	0.007141	0.572918	0.620637	0.72
12	D	0.1	0.6	6	0.12	0.007141	0.600438	0.591084	0.72
	D	0.1	0.6	6	0.12	0.00707	0.600438	0.591084	0.72
11	D	0.1	0.6	6	0.12	0.00707	0.629233	0.562939	0.72
	D	0.1	0.6	6	0.12	0.007	0.629233	0.562939	0.72
10	D	0.1	0.6	6	0.12	0.007	0.659361	0.536134	0.72
	D	0.1	0.6	6	0.12	0.006931	0.659361	0.536134	0.72
9	D	0.1	0.6	6	0.12	0.006931	0.690885	0.510602	0.72
	D	0.1	0.6	6	0.12	0.006862	0.690885	0.510602	0.72
8	D	0.1	0.6	6	0.12	0.006862	0.723855	0.486289	0.72
	D	0.1	0.6	6	0.12	0.006794	0.723855	0.486289	0.72
7	D	0.1	0.6	6	0.12	0.006794	0.758334	0.463134	0.72
	D	0.1	0.6	6	0.12	0.006727	0.758334	0.463134	0.72
6	D	0.1	0.6	6	0.12	0.006727	0.794395	0.441081	0.72
	D	0.1	0.6	6	0.12	0.00666	0.794395	0.441081	0.72
5	D	0.1	0.6	6	0.12	0.00666	0.832097	0.420077	0.72
	D	0.1	0.6	6	0.12	0.006594	0.832097	0.420077	0.72
4	D	0.1	0.6	6	0.12	0.006594	0.871504	0.400075	0.72
	D	0.1	0.6	6	0.12	0.006529	0.871504	0.400075	0.72
3	D	0.1	0.6	6	0.12	0.006529	0.912702	0.381024	0.72
	D	0.1	0.6	6	0.12	0.006464	0.912702	0.381024	0.72
2	D	0.1	0.6	6	0.12	0.006464	0.955752	0.36288	0.72
	D	0.1	0.6	6	0.12	0.0064	0.955752	0.36288	0.72
1	D	0.1	0.6	6	0.12	0.006464	1	0.3456	0.72
Start	D	0.1	0.6	6	0.12	0.0064	1	0.3456	0.72

## Agent D (Realistic) results of Experiment A Part II

Table 27 - Agent D partial results of Experiment A Part II, extracted from (Marsh 1994)

Iteration	Other	Basic	Genl.	Other	Agent	Other	Agent	Other
no.	Name	trust	trust.	Comp	choice	choice	payoff	payoff
14	С	0.777611	0.811524	0.668746	Cooperate	Cooperate	3	3
	C	0.769912	0.737749	0.662125	Cooperate	Cooperate	3	3
13	C	0.769912	0.808414	0.662125	Cooperate	Cooperate	3	3
	C	0.762289	0.734922	0.655569	Cooperate	Cooperate	3	3
12	C	0.762289	0.664197	0.655569	Cooperate	Defect	0	5
	C	0.769989	0.737997	0.662191	Cooperate	Defect	0	5
11	C	0.769989	0.66723	0.662191	Cooperate	Defect	0	5
	C	0.777767	0.741367	0.66888	Cooperate	Defect	0	5
10	C	0.777767	0.670583	0.66888	Cooperate	Defect	0	5
	C	0.785623	0.745092	0.675636	Cooperate	Defect	0	5
9	C	0.785623	0.674329	0.675636	Cooperate	Defect	0	5
	C	0.793559	0.749255	0.682461	Cooperate	Defect	0	5
8	C	0.793559	0.67857	0.682461	Cooperate	Defect	0	5
	C	0.801575	0.753967	0.689355	Cooperate	Defect	0	5
7	C	0.801575	0.683452	0.689355	Cooperate	Defect	0	5
	C	0.809672	0.759391	0.696318	Cooperate	Defect	0	5
6	C	0.809672	0.689196	0.696318	Cooperate	Defect	0	5
	C	0.817851	0.765773	0.703352	Cooperate	Defect	0	5
5	C	0.817851	0.696157	0.703352	Cooperate	Defect	0	5
	C	0.826112	0.773508	0.710457	Cooperate	Defect	0	5
4	C	0.826112	0.704969	0.710457	Cooperate	Defect	0	5
	C	0.834457	0.783299	0.717633	Cooperate	Defect	0	5
3	C	0.834457	0.716918	0.717633	Cooperate	Defect	0	5
	C	0.842886	0.796575	0.724882	Cooperate	Defect	0	5
2	C	0.842886	0.7353	0.724882	Cooperate	Defect	0	5
	C	0.8514	0.817	0.732204	Cooperate	Defect	0	5
1	C	0.8514	0.774	0.732204	Cooperate	Defect	0	5
Start	C	0.86	0.86	0.7396	Cooperate	Defect	0	5

Table 28 - Agent D final results of Experiment A Part II, extracted from (Marsh 1994)

Iteration	Other	Costs	Ben-	Uti-	Risk	Other	Coop.	Sit.	Impor-
no.	Name		efits	lity		Comp.	Thresh	Trust	tance
14	С	0.1	0.6	6	0.026667	0.668746	0.003048	0.708239	0.16
	C	0.1	0.6	6	0.026667	0.662125	0.003048	0.708239	0.16
13	C	0.1	0.6	6	0.026667	0.662125	0.003068	0.705525	0.16
	C	0.1	0.6	6	0.026667	0.655569	0.003068	0.705525	0.16
12	C	0.1	0.6	6	0.026667	0.655569	0.003047	0.708477	0.16
	C	0.1	0.6	6	0.026667	0.662191	0.003047	0.708477	0.16
11	C	0.1	0.6	6	0.026667	0.662191	0.003025	0.711712	0.16
	C	0.1	0.6	6	0.026667	0.66888	0.003025	0.711712	0.16
10	C	0.1	0.6	6	0.026667	0.66888	0.003003	0.715288	0.16
	C	0.1	0.6	6	0.026667	0.675636	0.003003	0.715288	0.16
9	C	0.1	0.6	6	0.026667	0.675636	0.00298	0.719285	0.16
	C	0.1	0.6	6	0.026667	0.682461	0.00298	0.719285	0.16
8	C	0.1	0.6	6	0.026667	0.682461	0.002956	0.723808	0.16
	C	0.1	0.6	6	0.026667	0.689355	0.002956	0.723808	0.16
7	C	0.1	0.6	6	0.026667	0.689355	0.002931	0.729015	0.16
	C	0.1	0.6	6	0.026667	0.696318	0.002931	0.729015	0.16
6	C	0.1	0.6	6	0.026667	0.696318	0.002904	0.735142	0.16
	C	0.1	0.6	6	0.026667	0.703352	0.002904	0.735142	0.16
5	C	0.1	0.6	6	0.026667	0.703352	0.002875	0.742568	0.16
	C	0.1	0.6	6	0.026667	0.710457	0.002875	0.742568	0.16
4	C	0.1	0.6	6	0.026667	0.710457	0.002843	0.751967	0.16
	C	0.1	0.6	6	0.026667	0.717633	0.002843	0.751967	0.16
3	C	0.1	0.6	6	0.026667	0.717633	0.002804	0.764712	0.16
	C	0.1	0.6	6	0.026667	0.724882	0.002804	0.764712	0.16
2	C	0.1	0.6	6	0.026667	0.724882	0.002754	0.78432	0.16
	C	0.1	0.6	6	0.026667	0.732204	0.002754	0.78432	0.16
1	C	0.1	0.6	6	0.026667	0.732204	0.002667	0.8256	0.16
Start	C	0.1	0.6	6	0.026667	0.7396	0.002667	0.8256	0.16