Modeling human-like non-rationality in social agents

Kochanowicz, Jaroslaw Slawomir

2018

Kochanowicz, Jaroslaw Slawomir. (2018). Modeling human-like non-rationality in social agents. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73314>

<https://doi.org/10.32657/10356/73314>

Modeling Human-like Non-rationality in Social Agents

Jaroslav Slawomir Kochanowicz
School of Computer Science and Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree
of Doctor of Philosophy

2016

Abstract

Rational decision making is one deliberately aimed at the realization of well-defined goals in an environment, based on a formal representation of its state and logical deduction grounded in those representations. Conversely, non-rationality is characterized by a) informal or sub-symbolic processing instead of one based on formal representations; b) associative, habitual reactions rather than step by step deduction or optimization; c) lack of explicit goals or their link to decision making. Human species is the only one capable of consistent rationality, but the scale and gravity of human non-rationality are far greater than often suspected. Cognitive biases and other non-rational influences on key decisions occur disregarding explicit reasoning or goal realization, subconsciously and often in a certainty that they did not take place. Habitual or stereotypic deviations from rationality in human cognition are well documented as non-reducible to rational pursuit of the egoistic benefit or its occasional distortion with temporary emotional excitation. Non-rationality is also not limited to the episodic emotional deviation, but is a permanent 'design feature' of human cognition, making rationalistic 'homo economicus' approach incapable of tackling the great challenge of human behavior modeling. While social psychology has systematically addressed non-rationality of human cognition for decades, it is not true for computer science, as current paradigms do not sufficiently account for the non-rational character of humans. Despite its great relevance, non-rationality is not yet properly addressed in computational models of social agents.

Addressing the above problem, this thesis aims to study how various types of non-rationality can be addressed, described and modeled in the most relevant computational fields. The proposed models are implemented in the agent architectures and used in specific simulations to exemplify their ability to capture non-rationality within the personality, informal human cognition of culture and context social. To this end, thesis first

presents an interdisciplinary overview of how human-like non-rationality is addressed in Cognitive Architectures (CA), Multi-Agent Systems (MAS), Affective Computing (AF), Social Signal Processing(SSP), discussing inter-field trends and strengths and weaknesses of the specific disciplines and approaches. Based on this, elements of a paradigm of computational non-rationality modeling are proposed. Those include abandonment of the paradigmatic simplicity of MAS and closing the socio-cultural immergence-emergence loop by MAS-CAA merger. Also, an inclusion of the specific, implicit and internal culture-generating mechanisms and 'System 1' social context processing is postulated as a basis of personality and culture related non-rationality.

To realize those postulates, Social Context Cognition Model (SCCM), inspired among others by the Dual Process Theory, is proposed. SCCM is a connectionist model whose activation reflects agent's subconscious social associations induced by the situation. It captures non-conceptual, habitual and associative connotations and their consequences, rather than rule and reasoning based, calculative ones. It also provides 'System 1' modulation of behavior, non-appraisal affect, action, plan or norms feasibilities, usable both in entertainment and social simulations. Furthermore, a cognitive Dramaturgical Dissonance Model (DDM) is proposed. It is inspired by a Dramaturgical Metaphor and a theory of Cognitive Dissonance and used in a simulation of 'System 1' affect generation. Next, a novel Social Context based Personality Model (SCP) is proposed based on SCCM. It is a continuation and specification of the Cognitive-Affective Personality System theory and is an alternative to the dominating 'trait models' whose drawbacks as descriptive tools are multiplied when they are used as generating mechanisms. SCP is parametrized using sociological data and used in a series of simulations including multiple agent interactions. Model's properties are compared to 'trait models' in a formal validation, proving its superiority. Finally, the issue of parametrization of socio-cognitive models, like SCCM and SCP is addressed by a semi-automatic, crowd-sourcing based parametrization system. Its presentation consists of a data acquisition system outline, including the statistical methods used and a validation, showing usability of both the semi-automatic parametrization system and the parametrized socio-cognitive models as tools for acquiring and representing human-like non-rationality.

Acknowledgments

To my wife and daughter. Without you, there would be no point.
To my parents and teachers. Without you, there would be no way.

I would like to express my gratitude to my thesis advisors, Professor Tan Ah-Hwee and Professor Daniel Thalmann for their time, patience, support, and understanding during these past five years. I would also like to thank Brian Allen for his friendship, and advice during his time in Institute for Media Innovation. My gratitude also goes to Professor Nadia Thalmann and the Institute for Media Innovation for providing my scholarship and financial support.

I would also like to thank my friends, who helped me so much. Krystyna Milian and Keith Alden, for proofreading my unpolished papers and providing me with your great feedback and Jacek Kubiak, for fixing countless mistakes in my thesis. I especially thank my mom and dad for years of sacrifices and unconditional love and care. Above all, thank you, Ola. You are my best friend, soul-mate and my wife. We share everything. Your love, support, sense of humor and all the help kept me going.

Contents

Abstract	i
Acknowledgments	iii
List of Figures	x
List of Tables	xii
1 Introduction	1
1.1 Background and Motivation	1
1.2 Issues and Challenges	4
1.2.1 Theoretical challenges of 'non-rational' computing	4
1.2.1.1 Many fields, no paradigm. Externalist-internalist issue	4
1.2.1.2 'Homo-economicus'	7
1.2.1.3 Summary and the proposed contribution	9
1.2.2 Non-rationality and personality modeling	9
1.2.2.1 Issues of 'trait models'	10
1.2.2.2 Challenge of the socio-cognitive modeling	11
1.2.2.3 Summary and the proposed contribution	12
1.2.3 Non-rationality, norms and culture modeling	12
1.2.3.1 Collective and systemic culture representations	13
1.2.3.2 Norm emergence as a culture model. Challenge of sim- plistic intra-agent representations.	16
1.2.3.3 Summary. Micro-macro gap. Contributions, scope and granularity of work.	18
1.2.4 Non-rationality and social-context cognition	20
1.3 Research Objectives	21

1.4	Approach and methodology	22
1.5	Organization of Report	24
2	Literature study	26
2.1	Introduction, relevant concepts and definitions	26
2.1.1	Rationality, irrationality and non-rationality	26
2.1.2	'System 1' vs. 'System 2'	28
2.1.3	Cognitive vs. non-cognitive processes	29
2.1.4	Abandoning homo-economicus	30
2.1.4.1	Homo-economicus and traditional social sciences	31
2.1.4.2	Homo-economicus and computer sciences	31
2.1.5	Scope and structure of the work	32
2.2	Affective aspects of non-rationality	33
2.2.1	Affective computing	33
2.2.1.1	Legacy of traditional psychology	33
2.2.1.2	Reconstruction of emotion theories in computational models	34
2.2.2	Cognitive architectures	36
2.2.3	Cognitive-Affective Architectures and Artificial Humans	38
2.3	Personality modeling	39
2.3.1	Elements of the psychology of personality	39
2.3.1.1	Personality construct. Two traditions	39
2.3.1.2	Person-situation consensus. Interactionism and CAPS	40
2.3.2	Personality in computational models	41
2.3.2.1	'Trait models' in computational personality modeling	42
2.3.2.2	Non-trait models	43
2.4	Culture and norm modeling	45
2.4.1	Introduction. Culture in social sciences	45
2.4.1.1	Culture - the problematic concept	46
2.4.1.2	The micro-macro gap in the social sciences	46
2.4.2	What are the norms? Norms and culture.	47
2.4.3	Norm emergence theories	48

- 2.4.3.1 Theory of rational choice 49
 - 2.4.3.2 Theory of socialization 51
 - 2.4.3.3 Theory of social identity 51
 - 2.4.4 Culture in computer sciences 53
 - 2.4.4.1 Normative Multi-Agent Systems in social and culture simulation 53
 - 2.4.4.2 Two tendencies in normative Multi-Agent Systems . . . 54
 - 2.4.4.3 Multi-Agent Systems and Homo-economicus 56
 - 2.4.4.4 Computational models in cognitively rich agents 57
- 2.5 Social context cognition 61
 - 2.5.1 Psychology of social context cognition 61
 - 2.5.1.1 Constructive Perception Theory 62
 - 2.5.1.2 Dramaturgical Theory 62
 - 2.5.1.3 Cognitive dissonance theory in social context cognition . 64
 - 2.5.2 Social context in Social Signal Processing 64
 - 2.5.3 Social context in Cognitive Affective-Architectures 67
- 3 Social Context Cognition Model 70**
 - 3.1 Introduction 71
 - 3.1.1 Key design choices. Social prototypes as mediating units 73
 - 3.1.2 Connectionism, saturation, and 'depth' mechanism 77
 - 3.2 Proposed model 79
 - 3.2.1 General architecture 79
 - 3.2.2 'Depth' mechanism 80
 - 3.2.3 Details of data processing 81
 - 3.2.3.1 Feature extraction processing 82
 - 3.2.3.2 Feature classification 82
 - 3.2.3.3 Prototype association 85
 - 3.2.3.4 Prototype inhibition 88
 - 3.2.3.5 Energy release 89
 - 3.2.3.6 Output layer 90

3.3	'Depth' mechanism, correctness proof	91
3.4	Example setup and output presentation	94
3.4.1	SCCM creation and parametrization	94
3.4.2	Example character profile and scenario	95
3.4.2.1	Character	95
3.4.2.2	Scenarios	97
3.4.2.3	Believability, complexity and stereotypical characters	97
3.5	Believability validation	102
3.5.1	Introduction	102
3.5.2	Characters and simulation scenarios	103
3.5.3	SCCM and human context cognition output acquisition	104
3.5.4	Believability evaluation and analysis	104
3.6	Summary and conclusions	105
4	Dramaturgical Dissonance in Human-Like Non-Rationality Modeling	107
4.1	Introduction	109
4.1.1	Dramaturgical and dissonance theories	109
4.1.2	Structure of dramaturgical processing	110
4.2	Proposed model	110
4.2.1	Dramaturgical databases	110
4.2.2	Dramaturgical sub-processes	112
4.2.2.1	Situational expectancy	112
4.2.2.2	Historical expectancy	113
4.2.2.3	Expectation Influenced Role Perception	114
4.2.2.4	Current Role Attribution	114
4.2.2.5	Scene related processes	115
4.2.3	Dissonances and consonances	115
4.3	Dramaturgical-dissonance example	117
4.3.1	Perception in the scene	117
4.3.1.1	Dramaturgical databases	118
4.3.1.2	Agent History, Familiarity, and simulation specific parameters	118

4.3.2	Example simulation setup and results	119
4.3.2.1	Resulting dramaturgical situation	121
4.3.2.2	Resulting dissonances	122
4.4	Validation and comparison to SCCM	123
4.4.1	Introduction	123
4.4.2	Characters and simulation scenarios	124
4.4.3	DDM context cognition output acquisition	124
4.4.4	Believability evaluation and analysis	125
4.4.5	Conclusions from the statistical comparison	126
4.5	Summary and conclusions	127
5	SCCM based personality model	128
5.1	Introduction	129
5.1.1	Situation in personality modeling	130
5.1.2	Mediating units: prototypes, features and personality states . . .	131
5.2	Proposed model and its parametrization	132
5.2.1	Specification of SCCM	132
5.2.2	Psychological data based parametrization	133
5.2.2.1	Domain specific parameters	134
5.2.2.2	Simulation specific parameters	135
5.2.2.3	Agent and scenario generation	137
5.3	Result analysis	137
5.3.1	Subclass emergence	138
5.3.2	Objective role frequency and variability	139
5.3.3	Saturation analysis. Variability vs. scenario length.	140
5.3.4	Subjective popularity vs. objective victimization	141
5.4	Summary	143
6	Semi-automatic SCCM/SCP parametrization	144
6.1	Crowd-sourcing social context	145
6.2	SCCM/SCP crowd-sourcing	146
6.2.1	Application specific preparations	146

6.2.1.1	Sample selection and preprocessing	146
6.2.1.2	Initial feature selection	147
6.2.1.3	Subject selection	148
6.2.2	Data format	148
6.2.2.1	Input samples	148
6.2.2.2	Social context components	149
6.2.2.3	Social context component features	153
6.2.2.4	Question-answer set representation	154
6.2.3	Crowd-sourcing procedure	155
6.2.3.1	Context calibration	155
6.2.3.2	Component and feature collection	155
6.2.4	Data processing	157
6.2.4.1	Sample base and averaged values	157
6.2.4.2	Correlation analysis	159
6.2.4.3	Duplicate detection and merger	160
6.2.4.4	Feature evaluation	161
6.2.4.5	Model generation	163
6.3	Example application	164
6.3.1	Domain and samples	164
6.3.2	Validation and result analysis	165
6.4	Summary	166
7	Conclusions	167
7.1	Summary of contributions	167
7.2	Limitations of the presented models and future work	169
	References	175

List of Figures

1.1	Computational methodologies used in social human-like non-rationality modeling.	6
2.1	Example personality state, defined as current behavioral disposition, measured in different contexts for two individuals exhibiting equal personality trait	41
2.2	The high-level abstraction presented by CAPS. Details of the mediating units and its connections are left largely unspecified.	44
3.1	Social Context Cognition Model in an example agent architecture	72
3.2	Part of a connectionist model involved in the permanent saturation problem	75
3.3	Inhibition as an attempt to answer the saturation problem	76
3.4	Case of excitation for non normalized weights	77
3.5	The SCCM general structure	81
3.6	Ahmed's mediating units	96
3.7	Ahmed's prototype activation in the 'Friendly visit' scenario, t1	98
3.8	Ahmed's prototype activation in the 'Mosque' scenario, t1	98
3.9	Ahmed's prototype activation in the 'Friendly visit' scenario, t2	99
3.10	Ahmed's prototype activation in the 'Mosque' scenario, t2	99
3.11	Ahmed's prototype activation in the 'Friendly visit' scenario, t3	100
3.12	Ahmed's prototype activation in the 'Mosque' scenario, t3 (red font signifies a negative activation)	100
3.13	SCCM influence on Ahmed's Pleasure-Arousal-Dominance affective space during 'Friendly visit' scenario	101

3.14 SCCM influence on Ahmed’s Pleasure-Arousal-Dominance affective space
during ‘Mosque’ scenario 101

4.1 Details of the dramaturgical process 113

4.2 Selected scene-related activations 120

4.3 Selected total and actor specific role-related activations. 121

4.4 Example scene/role dissonances 123

5.1 The general structure of SCP 132

5.2 Selected prototype activations in SCCM’s of the simulated agents 139

6.1 Example images representing domain with marked objects and agents . . 156

6.2 SCP/SCCM crowdsourcing procedure 158

6.3 Snapshot of the role component control page 162

List of Tables

1.1	General glossary of terms	25
2.1	'System 1' vs. 'System 2' comparison	29
3.1	Symbol table	83
3.2	SCCN example input elements	95
4.1	Symbol table	111
4.2	Agent features linked to role attributions	118
4.3	Dramaturgical databases SS and RR	119
4.4	Features perceived by agent at given times	120
4.5	Agent historical attributions	122
5.1	Role influence on personality state expression [MJ97, Ble09]. 'O', 'C', 'E', 'A' and 'N' stand for openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism respectively.	136
5.2	Role features [MJ97]	136
5.3	Observable objects and their features [ØR01, FCG07]	136
5.4	Personality influence on action probabilities [Ble09]	137
5.5	Pure/mixed bullies and victims [MDBC06]	141
5.6	Number of victim-bullies depending on the scenario length.	141
5.7	Emergence of bully sub-classes.	142
6.1	Symbol table, crowd-sourced data	150
6.2	Symbol table, analysis-generated data	151

Chapter 1

Introduction

1.1 Background and Motivation

Describing and quantifying human behavior, including its social aspect, remains one of the greatest and most interdisciplinary scientific challenges. It dictates limits of our ability to formally and systematically model economical, cultural, political or religious phenomena in both macro and micro scale. Traditionally a challenge of social sciences, this problem has a growing relevance for computational disciplines including Cognitive Architectures (CA), Artificial Social Agents (ASA), Multi-Agent Systems (MAS), Affective Computing (AC), Social Signal Processing (SSP). However, individual and group behavior modeling poses a great difficulty for both traditional social sciences and modern computational models and remains a grand challenge, not only due to its complexity but also informal, fuzzy and non-rational nature of human cognition.

Rational decision making is one deliberately aimed at a realization of well-defined goals in an environment, based on a formal representation of its state and logical deduction grounded in those representations. Conversely, **non-rationality** is characterized by a) informal or sub-symbolic processing instead of one based on formal representations; b) associative, habitual reactions rather than step by step deduction or optimization; c) lack of explicit goals or their link to decision making. Human species is the only one capable of consistent rationality, but scale and gravity of human non-rationality are far greater

CHAPTER 1. BACKGROUND AND MOTIVATION

than it is often suspected or admitted. Image of eyes above coffee contribution box in the office inclines people to pay more honestly with no supervision [BNR06], without changing the situation from an egoistic-utilitarian or even moral perspective. Attribution of a positive feature like intelligence to a person primes us to attribute them with other positive qualities like honesty or generosity, regardless of any statistical correlation they may have, in a 'halo effect' example [Tho20]. Countless examples of this type occur not only disregarding any conceptual reasoning or rational goal realization but also subconsciously and often in a certainty that they could not take place 'in my case'. Habitual, gregarious, ritualistic, unconscious, stereotypic and erroneous deviations from rationality in human thinking and behavior are currently well documented [HNM05] as non-reducible to rational pursuit of the egoistic benefit or its occasional distortions with temporary emotional excitation, as often assumed. Non-rationality can no longer be perceived as a rare, affective abnormality in otherwise rational thinking, but as a systemic, permanent quality - 'a design feature' of human cognition.

Non-rational aspects of cognition do manifest in the individual behavior, but they are even more crucial in the group and social interactions and the emergence of socio-cultural phenomena. The interplay of human rationality and non-rationality in affect, personality or habits, result in the emergence of rituals, customs, stereotypes or prejudice, shaping social process and structure. Subconscious and sub-symbolic aspects of relations, identity, values, social roles or norms determine group character through the actions of agents perhaps even stronger than their formal and explicit, rational counterparts. Therefore the second main aspect of this work, i.e. the struggle to understand the cognitive aspect of social and cultural phenomena and its emergence in the micro and macro scales, is closely related to the problem of non-rationality.

Those challenges elude the rationalistic 'homo economicus' [JS44] approach and are problematic partly due to the fundamental issues with the paradigms popular in the

CHAPTER 1. BACKGROUND AND MOTIVATION

computational modeling, that do not sufficiently account for the non-rational, stereotypical, habitual and associative character of human behavior. While psychology has systematically addressed non-rationality of human cognition (including its non-emotional aspects) for decades [TK74], it is not the case for computer science. Especially its non-affective, socio-cognitive aspect is relatively underrepresented, despite its great relevance and normative-emotional power. Psychosocial phenomena like dynamic identification with subjectively defined groups, internalization and exchange of their operational stereotypes, roles, norms and patterns of context cognition are the ones we fail to address properly. Groups of agents acquiring and using such contents, relating to them emotionally, habitually and rationally in their behavior and personalities would provide new levels of psychological and sociological believability. At the same time, traditional social sciences remain fragmented and largely unsuccessful in the consolidation of micro and macro perspectives on culture [Mes11]. Some, like sociology, are described as still struggling to establish their proper scientific paradigm and most of their verbal models are not formal enough to specify computational models of the mentioned mechanisms [Bro08]. *An example of ultimate goal of such efforts in a micro scale* is a hypothetical Cognitive Architecture, solving a 'cultural Turing test' by observation and internalization of values, taboos, social contexts or stereotypes from a social environment. *A macro scale example* could be an emergence and evolution of artificial religious rituals within a Multi-Agent System, organized around historical events and psychologically critical topics (death, sex, war etc.) influencing goals and behavior of groups and individuals. Despite significant progress in both Cognitive Architectures and Multi-Agent Systems, simulation, theoretical models, or even a paradigm promising such solutions are far beyond our current reach. In most aspects, even formal theoretical criteria of success, or proper validation data do not exist.

1.2 Issues and Challenges

This section discusses the major issues and challenges of non-rationality modeling. Section 1.2.1 addresses the problem of developing a proper theoretical approach and paradigm of computational non-rationality modeling. Section 1.2.2 discusses the connection between personality modeling and non-rationality, while section 1.2.3 addresses problems related to culture, norms and their link to non-rationality. Section 1.2.4 presents social-context cognition as the key processes underlying various types of non-rationality.

1.2.1 Theoretical challenges of 'non-rational' computing

This section addresses theoretical issues and challenges of human-like non-rationality modeling, including:

- research fragmentation due to a greatly interdisciplinary nature,
- rationalistic inclinations of most solutions
- lack of a paradigm consistently promoting implementation of explicit models of non-rational cognitive mechanisms in agent-based social simulation,
- either exclusively externalistic or internalistic perspective of most approaches, breaking the social emergence-immersion loop.

1.2.1.1 Many fields, no paradigm. Externalist-internalist issue

Due to its depth and breadth, the social human behavior is addressed by many fields (see Figure 1.1), in various abstraction levels. For example, Social Signal Processing (SSP) is devoted to the analysis and generation of the observable human-like nonverbal signals. SSP has provided a considerable progress in 'machine analysis of relevant behavioural cues like blinks, smiles, crossed arms, laughter, and similar' [VPB09] but does not address

CHAPTER 1. ISSUES AND CHALLENGES

the underlying cognitive mechanisms. Cognitive Affective-Architectures (CAA) and Artificial Social Agents (ASA) take a less behavioristic approach and focus on the cognition process, including modeling of human-like memory, planning, affect elicitation, reason-affect relations, mood or personality [CTN07]. The paradigm of Multi-Agent Systems (MAS) has been shown to be effective in modeling of the group and social phenomena, with a multiplicity of relatively simple agents. Notably, there is no well-defined field investigating the challenging circular relationship between the group-cultural qualities and individual-psychological ones, including culture emergence.

Progress in human-like non-rationality modeling or even its preliminary survey are then problematic, due to an extremely interdisciplinary nature (see Figure 1.1), even excluding traditional psychology and sociology. The number and scope of the relevant fields lead to fragmentation of the existing solutions and inconsistent terminology. Additionally, while non-rationality greatly affects many computational fields addressing human behavior (like MAS, CAA, SSP and others), it is not the main object of research in any of them. Aspects of non-rationality are getting more attention in these fields and initial solutions begin to appear (for example in CAA [Sun13], ASA [DMP14] or MAS [Sun09]), but at the same time, implicitness of non-rationality and preoccupation with field-specific research goals leave it without a devoted domain. Therefore, despite the impressive development of the mentioned domains, the circular relation of individual non-rationality to the cultural emergence and social behavior is still insufficiently addressed.

Finally, progress in the non-rationality modeling is hindered by the great division between the externalist, objectivist approach to normative-cultural phenomena, represented by MAS, and the internalist, cognitive one presented by CAA. The theoretical and practical unification of those approaches are greatly needed. In the most general summary, MAS directly tackle the problem of norm emergence, but the depth of this pursuit is limited by the psychological complexity of the used agents. On the other hand, CAA

CHAPTER 1. ISSUES AND CHALLENGES

Figure 1.1: Computational methodologies used in social human-like non-rationality modeling. Clusters roughly reflect three main groups of approaches to culture modeling: agent methods (1), systemic approaches (2) and cognitive models (3). Cognitive models offer the most detailed model of culture on the individual, micro-level but address agents in isolation. Agent methods address macro-level cultural emergence from multiple agents, but their paradigms assume simplistic, often rule-based agents. Systemic approaches can connect micro and macro levels of social phenomena in a complex manner, but they are predesigned, static formulations of the observed phenomena and as such they do not include any emergence. Part of the methodological problem in culture modeling in general and in non-rationality modeling, in particular, is the gap between groups (1) and (3). This gap needs to be bridged with new approaches. Based on [ZMVH08]

CHAPTER 1. ISSUES AND CHALLENGES

provide complex, affectively and cognitively believable agents, but they are designed and used for investigation of the individual, and not social and cultural dimensions. The gap between those paradigms leads to either insufficient external socio-cultural emergence (in the case of CAA) or immergence of emerging patterns in the agent cognition (in the case of MAS). The problem lies outside of the complex-but-single individual or large-but-simple group modeling (see 3.2) and the challenge may be seen as an expansion of CAA with new habitual, socio-cultural mechanisms of human non-rationality and placing them in a social context. On the other hand, it is a development of MAS with those elements towards psychologically rich agents, enabling cultural emergence. It could be argued that this is a beginning stage of a domain forming process that may result in a 'non-rational computing' - domain using human-like non-rational agents in social simulation.

1.2.1.2 'Homo-economicus'

Humans act based on a mixture of rational thinking, emotion, but also genetic and socially obtained mechanisms, biases, and patterns of environment representation and decision making that circularly influence cultural emergence. Habitual and convention-related aspects may be seen as a third mode of 'reasoning', next to formal reasoning and affect, as they are neither rational nor emotional but instead influence and are influenced by those two decision-making modes. They include, for example, informal and associative aspects of social context cognition, habitual and emotion based self-penalization and innate social motivation structure (needs to socialize, to be moral or to be perceived as such etc.) and many other mechanisms. In the social sciences, they constitute social cognition and thus determine social dynamics and through it, the externally observable culture. In a context of the computational social simulation, they remain largely unexplored.

The above tendency, is a part of a broader rationalistic approach to human behavior, as decision-making models, from economy to MAS, often implicitly assume *'homo*

CHAPTER 1. ISSUES AND CHALLENGES

economicus' [JS44] i.e. perspective on humans as basically rational, self-interested and pursuing individual and autonomously selected goals as optimally as possible. To understand the situation in human non-rationality modeling, a certain historical perspective on this concept is needed. Over the last decades, social sciences have made a broad transition from the rationalistic perspective on humans in social sciences, to one consistently including non-rational aspects [Kah11]. While psychology traditionally emphasizes the influence of emotion on human behavior, the core of the new perspective is that the source of human non-rationality is also non-affective. The new message is: we are irrational even without emotional excitation. In computational simulations, the mainstream AI only began embracing similar notions only in the early 90's, when fields like Affective Computing (AC) or now influential models like OCC [OCC90] gained attention. Only recently elements of non-rationality became widely addressed in the computational science and until now mostly in its most traditional i.e. affective dimension. This is understandable due to the engagement in the traditional challenges of AI like planning, image recognition or language processing. Also, the background of the researchers and what Picard called emotion's 'stigma' in science [Pic95] played a part. Moreover, modeling of non-affective non-rationality suffers from the inherent inaccessibility of subconscious structures and limited data for validation of the models.

Explicit models of subconscious cultural mechanisms and their links to affective and rational processes are a vital missing element in human cognition modeling. Their application in conjunction with rational and affective aspects is crucial in increasing psychological soundness of artificial agents and the realism of norm and convention emergence. Addressing cultural emergence from the rationalistic, utilitarian perspective expanded with just affect may produce a collection of agents communicating emotional signals during the pursuit of goals, but complex socio-cultural emergence requires inherent tendency to maintain stereotypical group images, multiple identities, values or role patterns. Three

CHAPTER 1. ISSUES AND CHALLENGES

decades ago transition from cognitive to cognitive-affective models took place. Today, non-affective aspects of non-rationality must also be included.

1.2.1.3 Summary and the proposed contribution

Theoretical issues and challenges of human-like non-rationality modeling were discussed, including interdisciplinary nature and fragmentation of research, rationalistic inclinations and exclusively externalistic or internalistic approach and a lack of a paradigm dedicated to non-rational cognitive mechanism modeling in a context of the agent-based social simulation. This work proposes a small contribution to solution, by including a survey of research in several computational disciplines (including CAA, normative MAS, AC, SSP, ASA) and establishing terminology and a common conceptual frame, discussing implicit assumptions about non-rationality underlying those disciplines, their limitations and potential in contributing to non-rationality modeling (Chapter 2). Finally, possible directions of development are presented, with an outline of goals and needed mechanisms of non-rationality that are a base of specific models and implementations (Chapters 3-6).

1.2.2 Non-rationality and personality modeling

The concept of personality is an intuitive way to address individual-specific patterns in preferences, style, behavior and many idiosyncratic non-rational features. Personality is traditionally central to psychology and embodies the assumption about individual coherence and an identity existing as an object-like entity, invariable throughout contexts. Despite a century of intensive efforts to identify cross-situational consistencies, this assumption remains surprisingly unfruitful [MS95] and dispositional differences between individuals are smaller than changes within one individual between situations [Eps79]. This fact cannot be explained by the prominent 'trait models', like Five Factor Model (FFM)[MJ92]. An opposing tradition undermines the importance of the individual traits,

CHAPTER 1. ISSUES AND CHALLENGES

even reducing individuals to object manipulated by contexts. This century old 'most persistent cleavage in the social sciences' [Els89] reappears in the computational modeling.

Development of robotic and virtual artificial humans in the last three decades [ZMVH08] made personality modeling crucial, due to their use in training, advertising or entertainment. In this process, the 'trait personality' debate moved to the field of computing. Less obviously, also MAS are at the point where the quality of non-rational components, including personality, becomes a major limitation [KTT13, Kam13]. Most importantly, in many agent models, personality is de facto used to as an implicit model of the idiosyncratic features of non-rationality - a collective, simplified label for individual's stable behavioral patterns making personality relevant to non-rationality in social agents.

Next section addresses the issues and challenges of personality modeling from the perspective of non-rationality, including:

- generation of flat and predictable characters by dominating 'trait models'
- use of those descriptive tools, as prescriptive mechanisms of cognition
- simplistic trait vectors as a replacement of the complex socio-cognitive process
- vagueness of the alternative, socio-cognitive models in psychology
- uncertain status of identity's invariance in contextual approach
- challenge of the socio-cognitive model parametrization

1.2.2.1 Issues of 'trait models'

Domination of Five Factor Model (FFM) [MJ92], as a personality model in simulation, is understandable, due to its simplicity and easy parametrization. However, it comes at a price of limitations implied by any 'trait model'. The above discussed drawbacks of

CHAPTER 1. ISSUES AND CHALLENGES

such models, as descriptive tools, are multiplied when they are used as generating mechanisms. A fact that 'trait models' are descriptive, not prescriptive in terms of cognition is vastly ignored in the computational modeling. When used in behavior generation, those 'average attitude values' produce uniform, flat personalities undermining psychological believability. This is exactly due to the negation of contextual cognitive-affective changes. For example, 'agreeable' FFM-based agent is equally 'agreeable' in all contexts, with no structured, idiosyncratic variability, resulting in predictable, flat behavior and multi-faced characters (e.g. a fearful mother, brave when her children are threatened) are beyond this methodology. This flaw is not limited to FFM and extends to all 'trait models', like PEN [Eys91] or other, less popular ones.

1.2.2.2 Challenge of the socio-cognitive modeling

Cognitive-Affective Personality System (CAPS) theory [MS95] is a notable example of a recent attempt to overcome limitations of the 'trait models'. It redefines personality from a context-invariant feature set to an idiosyncratic process of context based disposition adaptations (i.e. not 'what one is in all contexts' but 'what one is in which context'). CAPS is introduced as a network of cognitive-affective 'mediating units'. It is empirically supported [MS95] and 'consistent with parallel findings showing significant amounts of variance attributable to Person X Situation interaction.' [CS99].

While promising, the contextual approach seemingly undermines the basic intuition that personality is an invariant characteristic. What remains stable, if personality is defined in terms of dynamic states? If the notion of personality does not refer to some invariance, what is its purpose? Also, similar models raise three issues from the practical computational perspective. To define personality in terms of responses to a specific context, one must define relevant differences in perceived contexts clearly. Such definition should ideally not be liberal, but reflect the psychologically relevant structure of the

CHAPTER 1. ISSUES AND CHALLENGES

situation. This is, however problematic, as 'there is no consensus regarding the most important "psychologically active features of situations"' [Ble09]. Secondly, if changes in the dispositions are due to activation network of hidden mediating units, what is their nature and how are they connected? Both issues are largely unsolved. Finally, a network of mediating units representing personality is complex and parameter-heavy. To validate and use them a method of parametrization (preferably as automatic as possible) must be proposed.

1.2.2.3 Summary and the proposed contribution

This section discussed issues and challenges of personality modeling from the perspective of non-rationality. Those included domination of the descriptive 'trait models' used as mechanisms prescribing cognition, leading to a replacement of the complex socio-cognitive processes, predictable characters, and vagueness of the alternative, socio-cognitive models, their issue of identity's invariance and the problem of parameterization. This thesis addresses those issues by proposing both a theoretical framework, models their implementation for social agents, based on the socio-cognitive approach.

1.2.3 Non-rationality, norms and culture modeling

This section discusses challenges of modeling culture in a context of human-like non-rationality. Culture is a broad concept. Its definition may vary from the totality of material and intellectual results of human activity (e.g. 'pyramids are an example of Egyptian culture') to a characterization of the mental tendencies and inclinations shared in a group (e.g. 'hierarchy and collectivism are a part of Asian culture'). This last perspective is close to our interest - culture as particularities (including non-rational ones) *shared* within a given population, similar to personality as describing the *individual* ones). Notably, this perspective subscribes to a cognitivist view i.e. it sees *culture as information*, as opposed to merely describing externally perceived behavior.

CHAPTER 1. ISSUES AND CHALLENGES

Societies generate cultural subgroups, traditions and conventions that are twofold non-rational and 'habitual': 1) they are based on human non-rational and habitual mechanisms, and 2) they form customs - non-rational 'habits within the society'. Discrimination, rituals, customs, prejudice or stereotypes are clear examples of such phenomena. However, everyday thinking is also characterized by such convention-driven elements that carry great normative and emotional power, but their informal and fuzzy nature makes capturing them a challenge. In general, traditional social sciences remain divided and largely unsuccessful in the consolidation of the micro and macro perspectives on culture [Mes11]. Some, like sociology, still struggle to establish their scientific paradigm [Bro08]. Understanding culture and its emergence is traditionally a challenge of social sciences, but its relevance for social simulation, MAS and CAA is growing. This section will briefly discuss 'systemic' and agent-based approaches to culture and their current shortcomings (see chapter 2 for more elaborate analysis).

1.2.3.1 Collective and systemic culture representations

The first group of culture models addresses it on a collective, systemic level, in an attempt to identify physical and abstract elements (resources, taxes, population, education etc.) and formalize their relations. At least two main subdomains can be identified at this level [ZMVH08].

The first methodology in this group is the formal Organizational Modeling emerging from a pragmatic attempt to capture the diversity of the organizational forms, the interaction between them and their underlying environmental conditions from a perspective of a well-defined pragmatic goal i.e. organization's survival and success under the limited resources and obstacles of various nature. Organizational problems are often subtle and avoid identification. Formal methods were developed to provide more rigor and reliability in the analysis. Formal Organizational Modeling addresses links between

CHAPTER 1. ISSUES AND CHALLENGES

organization's structure, behavior, and performance from the analytical and optimization perspective. The basic question (i.e. how to effectively and efficiently organize internal structure and activities to fit with the environment) is translated into narrower, problem specific subquestions about decentralization, formalization (of organization rules), links between hierarchy and information flow etc. under given goals and conditions. Traditional inductive-ethnographic (extracting structure from observed organizations) and deductive-optimization approaches (treating the structure as an optimization parameter in the mathematical model of the task, e.g. [PMP⁺02]) are currently supported and expanded by simulational approaches, that, although distinctly different, share features of both of them [LL01]. These methods have been successfully applied to address many general issues like resource security [HD12] or policy forming [GLR11], but also particular problems like the influence of paternalism on family firms [CNCM12]. Organization Models may provide formal verification of systemic properties, assuming constraints imposed by the models. They allow performance-related hypothesis formulation and validation as well as execution of specific scenarios on several levels of abstraction to model non-linear behavior and contain elements like formally defined department or position related activities and responsibilities and sometimes agent-based elements (discussed more in the next section)

The second approach is the System Dynamics Modeling dating to early 1960's [For97]. System dynamics modeling is a computer-based method of analysis and design support, addressing complex social, economic, ecological and other systems that include mutual interaction, feedback, and circular causality between its elements. This field has well-defined stages of problem identification, mapping, and modeling, designed to allowing gradually build confidence in the model outcome. [MMR13]. On the implementational level, system dynamics models usually use nonlinear, first-order differential/integral equations iteratively simulated in a discrete time intervals. Models are structure-oriented and

CHAPTER 1. ISSUES AND CHALLENGES

structural level is seen as a primary causal factor for the observed output behavior, aiming at the ability to see the dynamic patterns underlying specific (and possibly random) events. Similarly, individual decisions are seen as the outcome of the underlying system structure. [RA10]. System Dynamics Modeling has been successfully applied to numerous problems. “System dynamics has been applied to issues ranging from corporate strategy to the dynamics of diabetes, from the cold war arms race between the US and USSR to the combat between HIV and the human immune system. System dynamics can be applied to any dynamic system, with any temporal and spatial scale. In the world of business and public policy, system dynamics has been applied to industries from aircraft to zinc and issues from AIDS to welfare reform.’ System Dynamics Modeling has also been used by various political or military bodies to contribute to their understanding of military, diplomatic, economic, social and other characteristics of countries or regions to evaluate their strengths and weaknesses [RDW05]

Systemic models like the ones above often approach culture not as an effect of the interaction of individuals but as a collective entity. Behavior is often reduced to the structural-organizational level while soft, cognitive dimension of culture is neglected or represented collectively by cultural-analytic dimensions or features [AVA⁺09] (e.g. ‘collectivist vs. individualist’ dimension from Hofstede’s culture theory [Hof01]). Direct use of similar constructs forbids emergence from more basic features i.e. psychological contents of agents, their formulation or exchange. This remark applies to all such cultural metrics, even if their usefulness for group-analytic purposes is unquestionable. Analogically to ‘trait models’ in personality (1.2.2), models of culture require a modeling shift and specification of alternative, culture-forming mechanisms on the individual level, to enable emergence. [AVA⁺09, SPN⁺12]

CHAPTER 1. ISSUES AND CHALLENGES

1.2.3.2 Norm emergence as a culture model. Challenge of simplistic intra-agent representations.

The second group of models overcomes limitations of the collective, systemic representations using MAS and particularly Normative MAS. In MAS norms are on one hand used as pragmatic tools of distributed control, assuring properties of a designed system and on the other, as providing a bottom-up, individualistic explanation of the cultural phenomena. The term 'norm' is ambiguous in MAS, there is no standardization concerning its intra-agent representation [Neu12]. It may include simple constraints, obligations or rarely more complex mental objects, usually moral norms and values. In principle, even when the internal norm representations exist (and norms are not simply seen as an external, emerging behavior regularities) their cognitive complexity is low compared to the mechanisms postulated in psychology and sociology. This approach to culture modeling has been successful in domains from ethnological research to exploration of culture's influence on organizational performance or in solving engineering problems. For example, Kobti et. al propose cultural algorithms to analyze influence of various factors like kinship cooperation strategies [KRK04] or protein exchange [KR05] to simulate the early Anasazi settlements.

There have been several attempts to create paradigms of culture modeling using this approach. For example, Morris et. al. attempt to simulate 'cultural interactions as a multiagent system of high functioning agents that achieve an equilibrium of beliefs' [MRHU11]. In another attempt, Mc Breen et. al. use BDI framework to represent culture 'as a system of shared values, which are linked to the Hofstede dimensions of culture. This system is represented by so-called metanorms that define appropriate, culturally-varying, behaviour in different relational contexts. In this way culture affects the possibility of normative changes, in particular the acceptance of policies designed to issue new norms in a society'. [BTDH11]. In it's combination of MAS and collective

CHAPTER 1. ISSUES AND CHALLENGES

cultural component, this approach somewhat resembles the cultural algorithms. This separate cluster of algorithms contains a knowledge component i.e. 'belief space' that is a collective information container accessible to all the population members. If the belief space contains norms, cultural algorithms can be seen as an intermediate step between the normative MAS and collective models of culture. Cultural algorithms have been used outside of social or cultural simulations e.g. in large-scale multi-agent systems in industry [ROSR03, AA14] or even in classical algorithmic challenges, like the traveling salesman problem [KC09].

Despite those successes, at least since Epstein [DGE⁺00], there is a tendency in MAS, to identify culture with:

- (i) *norm emergence*, as behavioral pattern resulting from game-theoretic, utilitarian individual adaptation or a shared, normative knowledge pool on a population-level
- (ii) *external products of (i)* e.g. communication, trade and relation networks, production and consumption patterns, life span or housing distributions. [DGE⁺00]

An implicit assumption, that (i) and (ii) reasonably approximate culture, is based on a broad application of term 'norm' in psychology, i.e. as 'cultural product including values, customs, traditions etc.' [She36]. However, this assumption is unjustified, considering the narrow redefinition of norms within MAS. The fact that abstract conditional rules (that underly emergence in MAS) are not a good model of the innate habitual, affective and other complex normative mechanisms (that underly culture emergence according to social psychology), is often silently ignored. Cognitive over-simplicity, based on externalistic focus and a decision to create complexity using simple agents in MAS, has also a trivial, technical justification, as agents with complex cognitive components are hard to model, implement, parametrize and validate. Constraints and obligations used instead are easy but, as Neumann points out, 'it requires a lot to transform morality into a computational

CHAPTER 1. ISSUES AND CHALLENGES

approach. It is still an open research question how this can be realized' [Neu12]. This is even more true, in a broader case of culture. Attempts to address this issue exist but are relatively limited.

Rare MAS models including intra-agent cultural mechanisms [AVA⁺09, SPN⁺12] often use atomic culture representations (again, usually Hofstede's dimensions), simply moving them from the collective to subjective, intra-agent level. This is a progress, but modeling cognitive process underlying culture emergence with statistical, group-describing tools rather than proper psychological, generative mechanisms shares the limitations of the previously criticized 'trait vectors' in personality modeling. Instead, moving culture generation inside the agent architectures must not incorporate atomic vectors, but processes incorporating interplay of non-rational components like innate and developed social goals, identities, emotion and habits of social cognition etc.

1.2.3.3 Summary. Micro-macro gap. Contributions, scope and granularity of work.

In summary, despite their great usefulness, the mentioned methods have shortcomings from the perspective of culture modeling. Those include:

- absence of culture emergence in 'collective' culture models
- reduction of culture to simplistic norm emergence in MAS
- atomic cultural representations in agents, in the rare models that have any
- rare presence of roles, identities, norms and similar informal culture-related constructs in agent architectures
- general absence of associative, habitual and socio-cognitive culture forming mechanisms

CHAPTER 1. ISSUES AND CHALLENGES

- no well-defined links between socio-cognitive, rational and affective mechanisms

Issues with culture simulations using MAS are specific cases of a broader problem with the agent simplicity underlying MAS [HHC09, NH11, MLT09]. While it enables simulation of certain social phenomena, it is too far-reaching and insufficient to model culture emergence rooted in the human non-rationality. Next to the already mentioned transition from external, atomic and rationalistic representations of norms to their complex, internal and non-rational versions, a similar change must take place in concepts like role, group identity etc., that need to be redefined from their usual organizational context [Puj06, JADT12, Sav11] to a cognitive one. For example, 'role' is usually used in MAS to introduce objectively different agent functions or types within a group. Modelling it as a part of agent's cognition e.g. as a psychological 'stereotype' emerging from experience, observation and communication and formed differently for each individual (even within one culture or group), would allow to formally address problems of soft culture formation, exchange and internalization on a new level.

Both formal, objective, verbalized and informal, subjective or subconscious versions of those elements should have their place within an agent cognition as they have a different function, and mechanisms of change or influence and share little more than a name. Behavior does not only depend on objective, institutional roles (norms, group identities etc.) attributed in a formal organizational sense, but also emerge from the interplay of stereotypical roles subjectively associated with self and others, at a given moment, partly coherently or outright incoherently from their formal counterparts, as it is the case in actual human societies. This thesis contributes to those goals, by proposing such model of informal, non-rational dimension of cultural contents and its effect on agent actions and MAS leading to the emergence of external social patterns of behavior based on such model.

CHAPTER 1. ISSUES AND CHALLENGES

This problem is also directly related to the issue of the micro-macro gap (section 2.4.1.2). The above analysis points to a need for a human-like, non-rational, individualistic (micro), cognitive foundation of what discussed methods directly represent as a collective or simplistically emergent cultural object (macro). This thesis contributes to this objective by proposing a micro-scale formation of human-like non-rational, informal normative components that could become be a language of macro-level cultural internalization or exchange in a society of agents using it. It presents several approaches to such cognitive level formation and simulations using it off-line in small groups of agents while online learning and large-scale simulations are discussed as future work.

1.2.4 Non-rationality and social-context cognition

Human-like social context cognition is a common, underlying thread in the previously mentioned issues and it appears as a key mechanism in many aspects of human-like non-rationality. The need to transition from the 'trait' to 'contextual' personality model requires parametrization of the personality expressions with the specific contexts. Similarly social context cognition is a location of the informal, subconscious norms and other culture-forming elements. While formal values or norms are often valid in most contexts, informal normative dispositions are characterized by their *contextual dependence* and inconsistencies, implying a non-trivial management system based on a social context monitoring, rather than a static, uniform norm hierarchy. Exchange and synchronization of social context models and their normative consequences is a base of 'System 1' [Kah11] cultural transmission and immergence, qualitatively different from the formal norm propagation. Additionally, modulation of cognition and behavior with non-rational social context cognition may reproduce various cognitive biases ('framing' [Kah11] may be seen as plan feasibility modulation, stereotypical role attribution may implement 'halo effect' [Kah11] etc.), making modelling it even more relevant. Informal and associative

CHAPTER 1. RESEARCH OBJECTIVES

human-like social context cognition is a much-needed base, unifying personal and shared, transmitted aspects of non-rationality.

While human-like affect in social agents gained a considerable attention over the last 20 years, resulting in models of emotion and moods or their relation to planning, social context cognition is strongly underrepresented. Its human-specific structure and relation to other cognitive processes are responsible for many observable features of personality, culture, affect, and normative behavior. Social context cognition is, therefore, a key aspect of human-likeness that remains relatively unexplored and it remains uncertain how to model it. This thesis contributes to changing this situation, by surveying the works related to the context cognition, identification of the key components of this cognition and their relations. This theoretical contribution is followed by a technical one: selected elements are translated into a formal model, implemented, parametrized, used in specific MAS simulations and generated results are validated. More discussion about the existing works on social context and their relation to this thesis can be found in section 2.5.3.

1.3 Research Objectives

This research aims to develop a theory and implementation of computational models of human-like non-rationality that can subsequently be used in cognitive architectures of social agents in Multi-Agent Simulations. As part of the agent cognitive architecture, the proposed models will respond to a continual stream of sensory input from the environment and serve as components providing specific aspects of 'System 1' social context cognition including culture, personality or non-appraisal affect generation.

Towards this goal, a critical analysis of the current trends in Multi-Agent Simulations, Cognitive Affective Architectures, and other fields will be presented. Constructive suggestions will form a novel approach based on the cognitively rich agents with stereotypical, habitual and irrational human-like characteristics. Analysis and a new proposed

CHAPTER 1. APPROACH AND METHODOLOGY

research paradigm take a form of postulates of the artificial emergent culture including decomposition of the so-called 'vector representations' of personality and culture, a discussion of the CAA-MAS integration, in a framework of an emergence-immersion relation and a systematic specification of the postulated irrational cognitive elements.

Specific models will address the challenges of representing personality as a process, rather than a feature, modeling groups and subcultures in Multi-Agent Simulations, capturing their specific 'System 1' personality-culture systems. As the proposed socio-cognitive models are parameter-heavy, and must often rely on accurate real data, the methodology to address the challenging problem of parametrization of the provided models will also be addressed. It is based on a semi-automatic method, using crowd-sourcing. It will be described and used in a set of simulations tackling a real context cognition problem, used as a validation.

In the summary, the addressed research objectives include

- (i) Analysis of the state and current trends in computational human-like non-rationality modeling in relevant disciplines.
- (ii) Creation of a new approach aimed to overcome limitations of the current models.
- (iii) Development of the specific human-like non-rationality models of social context cognition, personality, culture and affect generation, based on this approach.
- (iv) Application and verification of those models in multi-agent simulations.
- (v) Development of method to automatize parametrization of the proposed model with real human data

1.4 Approach and methodology

This research addresses a problem of including human-like non-rationality in simulations in both theoretical and practical dimension. The former aspect includes a literature

CHAPTER 1. APPROACH AND METHODOLOGY

study, an overview of the situation within related computational fields, and an analysis of the selected theories from sociology, psychology, and even philosophy, to allow identification of the long and short term goals. Latter aspect will be based on this broad analysis and identification of key cognitive mechanisms. It consists of the development and application of specific models implementing those identified mechanisms and a method of their parametrization.

First, a connectionist Social Context Cognition Model and its extension modeling personality (a specification of psychological CAPS model) are proposed. The key problem of psychological socio-cognitive models like CAPS is an unlimited number of candidates for the relevant context elements and their cognitive encodings, and their selection is problematic due to the innate inaccessibility of the cognitive process. Regardless of the actual underlying mechanisms, humans not only have an innate ability to store concrete information about the individuals or situations but also abstract representations and their features. Such categorization can be functionally described as organized around 'prototypes' [dSNdSdL13]. Therefore, those prototypes will be encoded in a connectionist structure of the proposed models, not in the sense of reflecting actual biological mechanisms but the specific mentioned functional capability.

Secondly, a non-connectionist Dramaturgical Dissonance Model, a cognition model grounded in the Dramaturgical and Cognitive Dissonance theories is presented. Similarly to CAPS, Dramaturgical and Cognitive Dissonance theories require significant specification, to be usable in a computational cognitive model. As proposed connectionist and predicate models are used and validated as the architectural component of multiple interacting agents, the entire research is placed within the methodology of Agent Based Modeling, Multi-Agent Simulations and Cognitive Architecture modeling.

Finally, as proposed models are parameter-heavy and depend on data validity, parametrization is crucial and challenging when performed by hand (as in Chapter 3 and 4) or based

CHAPTER 1. ORGANIZATION OF REPORT

on the existing research data (as in Chapter 5). As no databases containing cognitive-habitual material similar to the one used in the presented models, it is impossible to test against models them and data collection support for testing and application is key. As a semi-automatic solution, a methodology of crowd-sourcing based parametrization of socio-cognitive models is presented in Chapter 6.

1.5 Organization of Report

The rest of this thesis has been organized as follows. Chapter 2 discusses works related to human-like non-rationality in Cognitive Architectures, Multi-Agent Simulations, Affective Computing and Social Signal Processing and includes components of the proposed paradigm, based on the potential and limitations of the analyzed existing work that provides a theoretical basis for later models. Chapter 3 introduces Social Context Cognition Model. Chapter 4 presents Dramaturgical Dissonance Model. Chapter 5 uses Social Context Cognition Model to implement process-like personality-culture component and its application in Multi-Agent Simulation. Chapter 6 presents semi-automatic parametrization method used to parametrize Social Context Cognition Model. Finally, Chapter 7 contains conclusions and future work.

CHAPTER 1. ORGANIZATION OF REPORT

Table 1.1: General glossary of terms

Acronym	Meaning
AC	Affective Computing
ASA	Artificial Social Agents
BDI	Belief Desire Intention Model
CA	Cognitive Architectures
CAA	Cognitive Affective-Architectures
CAPS	Cognitive-Affective Personality System
CDT	Cognitive Dissonance Theory
CPT	Constructive Perception Theory
DT	Dramaturgical Theory
FFM	Five Factor Model
MAS	Multi-Agent Systems
OCC	OCC Model
SCCM	Social Context Cognition Model
SIT	Social Identity Theory
SSP	Social Signal Processing
TOM	Theory Of Mind

Chapter 2

Literature study

This chapter provides a review and a survey of relevant work in four aspects of human-like non-rationality of interest: affective non-rationality, social context cognition, personality, and culture. It also includes the definitions of central concepts and provides a general historical perspective on the transition from 'homo economicus' [JS44] visible in social and computational theories.

2.1 Introduction, relevant concepts and definitions

In this section, most relevant terms are defined, to allow a precise presentation. It includes comments on the distinction between rationality, irrationality and non-rationality, a definition of 'System 1' and 'System 2' and 'cognitive' vs. 'non-cognitive' process dichotomy. Next, a general summary of the historical transition from the purely rationalistic models is presented. The section ends with the specification of the scope and structure of the following work.

2.1.1 Rationality, irrationality and non-rationality

Humans are not rational beings. Image of eyes above coffee contribution box in the office inclines people to pay more often without supervision [BNR06], without changing the situation from an egoistic-utilitarian or even moral perspective. Attribution of a

CHAPTER 2. INTRODUCTION, RELEVANT CONCEPTS AND DEFINITIONS

positive feature like intelligence to a person primes us to attribute them with other positive qualities like honesty or generosity, regardless of any statistical correlation they may have, in a 'halo effect' example [Tho20]. Countless examples of this type occur not only outside conceptual reasoning or rational goal realization but also subconsciously and often in a certainty that they do not and could not take place 'in my case'. Habitual, gregarious, ritualistic, unconscious, stereotypic and erroneous deviations from rationality in human thinking and behavior are currently well documented [HNM05] as non-reducible to rational pursuit of an egoistic benefit or its occasional distortion with temporary emotional excitation (as often assumed). Non-rationality can no longer be perceived as a rare affective abnormality in otherwise rational thinking, but as a systemic, permanent quality, 'a design feature' of human cognition.

In general, *rational* decision making is an analytic process that operates on a formal conceptual model of the situation and goals to be achieved in it. Potential alternative actions are represented similarly, and their consequences are calculated in a chain of reasoning that can be described, analyzed, and communicated. Evaluation of their expected outcomes from the position of goals allows optimal action selection.

Irrational decision is one that deviates from the optimal action, regardless of the nature of the decision-making process. *Non-rational* decision-making, on the other hand, is one not using explicit representations, formal reasoning, and sequential analysis. It is heuristic and associative and often fast. Also, such process cannot usually be communicated due to the aforementioned lack of formal and explicit representations (making the definition of non-rationality dependent on decision of what is considered formal/explicit in a given case).

Irrationality, therefore, deviates from rationality with regards to the outcome, while non-rationality with regards to the mode of operating, and one does not imply the other (although they are often correlated). Example: good heuristics may often provide a non-rational but correct (therefore, not irrational) solution, while an error in formal reasoning

CHAPTER 2. INTRODUCTION, RELEVANT CONCEPTS AND DEFINITIONS

provides an irrational but not non-rational answer. Although irrationality is occasionally mentioned, the primary interest of this thesis is human-like non-rationality, and the optimality of actions is secondary. Notably, the scope of this thesis includes human-like aspects whose optimality (and therefore irrationality) is problematic in evaluation (at least practically). Personality is a good example here - in general, it is hard to prove that some types of personality are more 'irrational' (i.e. less optimal) than others, yet personality certainly is a non-rational aspect of human-likeness.

2.1.2 'System 1' vs. 'System 2'

In the case of human decision makers, the relation between rationality and non-rationality is perhaps best addressed by the dual process theory [TK74]. Its foundations are as old as modern psychology and likely come from William James, who distinguished two different kinds of thinking: associative and 'true' reasoning. Over the decades, this concept has been developed and reformulated several times and has recently been popularized by Kahneman, who uses terms '*System 1*' (non-rational, associative intuition) and '*System 2*' (rational, conceptual reasoning, see Table 2.1 for details). Dozens of cognitive biases related to properties of 'System 1' are known [HNM05]. They include :

- 'Anchoring effect' - tendency to be influenced by irrelevant data. For example, when asked if a famous person was more than 100 years old when they died, people show a much greater tendency to overestimate the age at death than when the question referred to death at 35 [EG01]
- 'halo effect' - tendency to 'transmit' assessments of individuals from one area to another [Tho20]
- 'framing' - drawing different conclusions from the same information, depending on the presentation context [Kah11]

CHAPTER 2. INTRODUCTION, RELEVANT CONCEPTS AND DEFINITIONS

- 'pervasive optimistic bias' - tendency to overestimate benefits and underestimate costs e.g. an average expectation of a kitchen renovation cost was less than 19 thousand dollars, while the actual cost was almost 39 [KT77] thousand.
- 'substitution' - 'System 1' tendency to implicitly answer a simpler question in a place of a difficult one [KF04]
- anthropomorphism - tendency to characterize animals, objects, and abstract concepts as possessing human-like traits, emotions, intentions etc. [Mit96]

Table 2.1: 'System 1' vs. 'System 2' comparison

'System 1'	'System 2'
Unconscious	Conscious
Implicit	Explicit
Automatic	Controlled
Low Effort	High Effort
Rapid	Slow
Associative	Rule-Based
Contextualized	Abstract
Recognition, perception	Rules, comparisons, analysis
Parallel	Serial
Large Capacity	Small Capacity

2.1.3 Cognitive vs. non-cognitive processes

Another relevant distinction is the one between cognitive and non-cognitive processes, as 'non-cognitive' is often used where 'non-rational' could perhaps be a better alternative. For example, classification of affect theories into cognitive and non-cognitive ones (see 2.2.1.1) distinguishes the latter as involving 'only nonpropositional and possibly even nonconceptual mental representations' [RHD⁺13] and the former as requiring 'certain higher-order mental representations, in particular beliefs and desires' [RHD⁺13].

CHAPTER 2. INTRODUCTION, RELEVANT CONCEPTS AND DEFINITIONS

This is perhaps somewhat counterintuitive, as 'cognition' means simply *acquisition of knowledge* 'through thought, experience, and the sense' [Dic04] and could be realized by either 'System 1' or 'System 2', as long as the acquired conceptual associations are seen as a form of knowledge. Indeed, 'usefulness of this distinction has been questioned because of the vagueness and ambiguity of the term cognitive' [RHD⁺13]. Defining 'cognitive' process as acquiring some, possibly implicit, knowledge within the system (e.g. by drilling basketball throws) and 'non-cognitive' process as free of such acquisition (e.g. the same throws performed by a robot without learning capabilities) could perhaps help avoiding some confusion. The same time, rational vs. non-rational' or 'System 1 vs. System 2' distinction could be used to describe the mode of the process.

However, even assuming the initial definition, a clear difference between 'rational' and 'cognitive' processes must be observed. The latter definition only involves *some* 'higher order' representations or analysis while the former makes such demands on *all* steps of the process, and so e.g. cognitive emotion elicitation would still not be a rational process.

2.1.4 Abandoning homo-economicus

With the above basic definitions at hand, this section presents a brief historical summary of the observable transition from the exclusively rationalistic perspective on human behavior, to the one acknowledging human irrationality and non-rationality. Decision-making models in various fields (from economy to computational behavior modeling) assumed '*homo economicus*' [JS44] i.e. perspective on humans as basically rational, self-interested, and pursuing individual and autonomously selected goals as optimally as possible. To understand the situation in human non-rationality modeling, a historical perspective is needed. Two relevant trends are briefly presented in this section: one in social sciences and another one in computer modeling.

CHAPTER 2. INTRODUCTION, RELEVANT CONCEPTS AND DEFINITIONS

2.1.4.1 Homo-economicus and traditional social sciences

The first trend is a broad transition from the rationalistic perspective on humans in social sciences, to one including consistently non-rational aspects. While psychologists always emphasized the profound influence of powerful emotion on human behavior, the core of the new perspective is that the source of human non-rationality is also non-affective. The new message is: we are non-rational even without emotional excitation. As Kahneman puts it 'at any given time scholars in a particular field tend to share basic assumptions about their subject. Social scientists in the 1970s broadly accepted two ideas about human nature. First, people are generally rational, and their thinking is normally sound. Second, emotions such as fear, affection, and hatred explain most of the occasions on which people depart from rationality. Our article challenged both assumptions. We documented systematic errors in the thinking of normal people, and we traced these errors to the design of the machinery of cognition rather than to the corruption of thought by emotion' [Kah11]. This paradigm shift cannot be overestimated in any context of behavior modeling - a fact reflected by Kahneman's Nobel Prize in Economy.

2.1.4.2 Homo-economicus and computer sciences

The second trend is a gradual inclusion of human-like elements in Cognitive Architectures, Social Simulation, Multi-Agent Systems, etc. While its traces may be tracked to the early AI, mainstream AI embraced those notions only in the early 90's, when fields like Affective Computing or now-classical emotion models like OCC [OCC90] gained attention. Only relatively recently non-rationality became widely addressed in the computational science and until now mainly in its most traditional i.e. affective dimension. This delay is understandable due to the engagement in practical, 'rational' challenges (planning, image recognition, language processing, etc.), researcher's background, and what Picard called emotion's 'stigma' in science [Pic95]. Moreover, non-affective non-rationality modeling suffers from the inherent inaccessibility and a subtle, subconscious

nature of their underlying processes and structures and (in consequence) limited data for model validation. Following analysis of how most relevant domains and specific models include aspects of non-rationality, will provide an opportunity for concrete examples.

2.1.5 Scope and structure of the work

After defining key concepts and presenting the broadest background, this section will specify the scope of the thesis and the structure of the remaining part of this chapter.

The section 1.2 proposes a division of the broad problem of non-rationality into challenges related to emotional or affective non-rationality, social context cognition, personality and normative-cultural aspects. Naturally, non-rationality modeling also includes other challenges, notably human-like motivation structure, identity management and their dynamics, but they are not directly addressed.

As mentioned in section 1.2.1.1, there is no well-defined computational field focusing directly and comprehensively on issues related to human-like non-rationality. This makes the analysis problematic due to the fact that the mentioned aspects of non-nationality mentioned above find their way to various computational disciplines that use incompatible perspectives, terminologies, and methodologies. Some of those disciplines include Social Signal Processing, Cognitive Architectures, Affective Computing, Social Agents with Personality, and Normative Multi-Agent Systems in Computational Sociology.

The choice has been made to structure this analysis according to the structure of non-rationality, rather than divisions of the particular computational fields, as the former one more directly reflects the actual structure of the problem. As a consequence, the remaining part of the chapter is divided into sections addressing affective non-rationality (section 2.2), social context cognition (section 2.5), personality (section 2.4.1), and normative-cultural aspects (section 2.4), while each relevant computational domain is mentioned in one or more of those sections.

2.2 Affective aspects of non-rationality

This section addresses the traditionally most researched aspect of human-like non-rationality i.e. its emotional dimension, its conceptualization, and modeling, in computational simulations based on the traditional psychological theories.

2.2.1 Affective computing

This section briefly describes the situation and effort in the unification of various traditional emotion theories through formal specification on a level of computational model and implementation.

2.2.1.1 Legacy of traditional psychology

There are various ways in which countless theories of emotion can be divided. Perhaps the most basic distinction of emotion elicitation theories is into cognitive and non-cognitive ones (as presented e.g. by Reisenzein et. al. [RHD⁺13], also see 2.1.3). In psychology, the discussion about the scope of non-cognitive emotion is controversial, partly due to the obscurity of the term 'cognitive' (also see 2.1.3).

Cognitive theories of emotion see emotion as a consequence of agent's knowledge and beliefs about the situation while non-cognitive ones see it as a pre-cognitive process. The most relevant cognitive theory (and emotion theory in general) is the appraisal theory [Arn60, FKTS89], 'a predominant force among psychological perspectives on emotion and arguably the most fruitful source for those interested in the design of symbolic AI systems, as it emphasizes and explains the connection between emotion and cognition' [MGP10]. Appraisal theory argues that emotion is a result of judgment of the 'relationship between the events and an individual's beliefs, desires, and intentions' [SL90]. In some versions of the theory 'appraisals also trigger cognitive responses, often referred to as coping strategies - e.g., planning, procrastination or resignation feeding back into a continual

CHAPTER 2. AFFECTIVE ASPECTS OF NON-RATIONALITY

cycle of appraisal and re-appraisal' [RHD⁺13]. Next section presents how those theories have been used to create specific implementations.

Over the decades, psychology has developed countless emotion theories, and yet despite its achievements it remains far from any unified theory of human emotion. As Gratch et al. put it 'from the perspective of an applied affective computing researcher, psychology would ideally be at an advanced stage where the ultimate correct theory of emotion (UCTE) or even better, the ultimate unified theory of the mind (UUTM) had been attained. (...) Unfortunately, psychology has not yet arrived at this stage.' [MGP10]. Numerous, incompatible proposed theories are not only formulated verbally and informally rather than in quantitative models or computable languages and are often so vague that comparisons between them are problematic even for the experts. Selected theories of emotion were formalized into computational models (see next section), but an even greater number has not been formalized, is vague and does not give much hope for formalization and even less for unification or formal comparison with other theories. [MGP10].

2.2.1.2 Reconstruction of emotion theories in computational models

The domain of Affective Computing (**AC**) is a developing, interdisciplinary field attempting to introduce order to the above situation, using various modeling methods from isolated, partial models based on set-theory formalisms and predicate logic [BDK08] to agent-logics, BDI [Bra87, RG⁺95] and complete cognitive architectures [DMP14, Hud04]. The domain gained momentum after Picard's paper [Pic95], as a reaction to the tendency to neglect affect in computational models. AC encompasses the development of systems that can recognize, model, and generate human affects. It draws from computer science, psychology, and cognitive sciences. Besides the ability to understand, react to and produce affect, AC addresses a question how affect sensing and generating technology influences human-computer interactions and how it may improve them.

CHAPTER 2. AFFECTIVE ASPECTS OF NON-RATIONALITY

To achieve computational models of emotions, their initial reconstruction into formal but implementation-independent languages is useful. Its goal is to improve the original theories by greater formalization, enforcing clarification and disambiguation in preparation for implementation while avoiding complexities related to implementational details. The first language proposed for formalizing psychological emotion theories was the set theory [BDK08] due to its formalism, expressiveness, and close relation to computational theories. In this way, a general model of cognitive-appraisal theories was initially formalized [BDK08]. Later other formalisms were used, notably representation of emotion theories in agent logics that exchanged some expressiveness for easier specification. Especially Belief-Desire-Intention (**BDI**) model [RG⁺95] provided a convenient conceptual environment, as most emotion theories use beliefs and desires as their base. The fact that they are already provided by BDI has allowed an easy specification and on the other hand a link to rational decision making already specified with BDI. Therefore, many theories were realized using BDI e.g. emotions related to expectations [CL03], four basic emotions [Mey08], and most notably two formalizations [RHD⁺13, AHL09] of the 22 emotions of the OCC theory [OCC90] that became a dominant standard in many practical applications.

As mentioned in the previous section, the most relevant emotion theory (at least from the computational perspective) is the appraisal theory. Its implementations often include complex appraisal estimating algorithms based on rule sets, Markov processes [SMP10], reactive [RP05] and decision-theoretic plans [GM04] and complete, complex cognitive models [MLL09]. Many other computational implementations rooted in the theory of cognitive models have been proposed both with a quantitative [DL12, SDM09] and solely on quantitative [AHL09, LS11] characteristics of emotion. Some of them did not specify a quantitative connection to more basic elements of cognition (like goal/belief power)[SDM09] while others did. Some models (like EMA [GM04]) are perhaps not so

CHAPTER 2. AFFECTIVE ASPECTS OF NON-RATIONALITY

clearly linked to the formal theories of emotion (what Adam and Lorini refer to as 'ad hoc models' [AL14]), despite considerable their popularity in real life application. One of the main problems of such models is difficulty with comparisons due to different levels of granularity, methodology and implementational isolation. One attempt to overcome this issue is the application of Cognitive and Cognitive-Affective Architectures described in the next two sections.

Computational models like the ones above have found successful applications in theoretical human-computer interfaces, various artificial intelligence or companion systems and even theoretical psychology (as tools of simulational verification of theories). For example, Marsella and Gratch have used EMA to generate specific predictions about how human subjects will appraise and cope with emotional situations and argue that empirical tests of these predictions have implications for psychological appraisal theory (Gratch et al., 2009b, Marsella et al., 2009).

2.2.2 Cognitive architectures

As many researchers in AC have pointed out [Sun09, RHD⁺13], a complete Cognitive Architecture provides a significant advantage over the isolated models in both practical and theoretical terms. It does so as an environment embedding various theories and reasonably approximating aspects they do not address (by using other theories), rather than neglecting them. It also allows a closer comparison between the results of different approaches. This section analyses the domain of Cognitive Architectures, strongly motivated by similar sentiments in a context of general human cognition. Cognitive Architecture (**CA**) is a model of artificial, computational processes, aiming to resemble human cognitive abilities. It possesses a broad spectrum of intelligence, as opposed to being a solution to one particular narrow task. As Ng states 'cognitive architecture can be defined as a single system that is capable of producing all aspects of behavior,

CHAPTER 2. AFFECTIVE ASPECTS OF NON-RATIONALITY

while remaining constant across various domains and knowledge bases' [Ng09]. Sun puts it shortly: 'cognitive architecture is a domain-generic computational cognitive model' [Sun06].

When describing CAs, it is meaningful to distinguish between psychologically and engineering oriented ones. The latter are designed for purely pragmatic purposes, while the former aim at a relative structural and functional cognitive realism and human-likeness, and are designed as a research tool for cognitive scientists. 'Cognitive architecture provides a concrete framework for more detailed modeling of cognitive phenomena, through specifying essential structures, divisions of modules, relations between modules, and so on. Its function is to provide an essential framework to facilitate more detailed modeling and understanding of various components and processes of the mind. It forces one to think in terms of process, and in terms of detail. Instead of using vague, purely conceptual theories, cognitive architectures force theoreticians to think clearly. They are critical tools in the study of the mind' [Sun09]. For the engineering oriented architectures, psychological realism is a minor concern, and their design structure reflects problem space, rather than the human mind. Besides theoretical importance both psychological and engineering oriented CA also have a direct, practical and commercial use and have been used successfully, to solve complex tasks. In the context of this work, psychologically believable CA are more interesting. Examples of prominent CA implemented as symbolic, production based systems include SOAR [Lai12], Icarus [LC]. Epic [KM97], Clarion [Sun06] and ACT-R [And07] are examples of hybrid, neural-symbolic systems. Many detailed surveys exist [CTN07, DOP08, LLR09].

As mentioned in section 1.2, traditional AI (including CA) has been developed for a long time with a notable disregard to human-like non-rationality, including emotion. The most direct limitation of pure CA from the perspective of this thesis is the fact that most of them are focused on solving traditional challenges of AI like reasoning, planning,

language processing, etc. leading to initial marginalization of non-rationality even in its most visible, affective form. This remark is however not true for Cognitive-Affective Architectures, described below.

2.2.3 Cognitive-Affective Architectures and Artificial Humans

As the research in AC progressed, the general purpose CA were finally used as means of specification of affect theories, and over the last two decades, the shift from pure cognitive models to ones including affect can be observed. CA offered AC researchers freedom from irrelevant implementation efforts and decisions and a complete environment for the tested theories by providing the 'default', approximated answers. As a result, a subgroup of CA appeared: Cognitive-Affective Architectures (CAA) [ZMVH08]. CAA addresses not only specification of affect generation model, but also the relation between affect and rationality, allowing formal and more complete specification of emotion effect theories. CAA have been developed both as tools in academic research, human-machine interfaces [KASA08], entertainment [SS09], training or advertising [ZMVH08]. Usually, academic CAA are theoretically more sound, while non-academic CAA address a greater spectrum of features. Believable affective modeling of 'Artificial Humans' is a field in itself. Examples of CAA include: Soar-Emote [MIL04], that combined implementation of both physiological and emotion cognitive emotion theories in SOAR or appraisal theory implementation in a computational model EMA [MG06]. For a comprehensive overview of CAA, refer to [LSZ12, ZMVH08].

A subfield of AH with particularly developed intelligent affective capabilities are intelligent companion agents capable of longterm interaction and relationship forming. They are often developed for entertainment or achievement of particular therapeutic goals. Due to the nature of their interaction with humans such agents require advanced capabilities in terms of recognising and generating affect. Artificial companions usually rely on cognitive appraisal theories (see previous section) and continuously analyse the subjective

CHAPTER 2. PERSONALITY MODELING

significance of the events and selects coping mechanism, strategies and actions based on it. Paro is an example of such robotic agent (seal robot) used in a nursing home in attempt to improve well-being of the elderly [TTK05]. Also virtual characters have been used for similiar purposes, in order to improve health and longavity [VRB⁺12]. Recent and perhaps most advanced example is Adam's and Lorinis work using BDI model of emotion in companion architecture [AL14]. They present an emotion-reasoning agent performing coping strategies aimed at well-being improvement. Underlying PLEAID engine that we extended to implement Dastani and Lorinis BDI logic for graded emotions and coping strategies.

2.3 Personality modeling

Personality is an intuitive way to collectively and briefly label specific, idiosyncratic patterns in preferences, style, behavior, and other 'System 1' features of cognition. As such, personality modeling is central to this work. This section contains an overview of the work related to personality modeling, including theories within the psychology of personality and works in the computational synthesis of personality.

2.3.1 Elements of the psychology of personality

Selected aspects of the psychology of personality are described here by analyzing two personality conceptualization traditions present in psychology. Their relation and potential integration is an object of much effort and controversy in the field that translates strongly into issues of computational personality modeling.

2.3.1.1 Personality construct. Two traditions

Personality is traditionally a central concept in psychology. It embodied the assumption about individual behavioral coherence and consistency manifesting invariably throughout

CHAPTER 2. PERSONALITY MODELING

various contexts. A century long, intensive efforts in identification of features of this entity was surprisingly unfruitful [MS95], as differences in behavioral dispositions between individuals are comparable or smaller, than differences manifested by a given individual in various situations [Eps79] (Figure 2.1). This fact cannot be explained by the prominent 'trait models', for example, Five Feature Model (FFM)[MJ92], that emerged from this perspective. A parallel and well-established tradition also exist [Dur95, Gof59, Col04]. It undermines the importance of individual characteristics, in extreme cases reducing people to objects manipulated by social situations and contexts - the actual sources of social agency. This century old dichotomy, called 'the most persistent cleavage in the social sciences' [Els89] continues within the field of computational modeling.

2.3.1.2 Person-situation consensus. Interactionism and CAPS

Most notable criticism of the approach to personality, or even its very concept, came from Mischel, who's influential claims that 'with the possible exception of intelligence, highly generalised behavioral consistencies have not been demonstrated, and the concept of personality traits as broad dispositions is thus untenable' [Mis68] led many psychologists to question the very existence of personality. The resulting debate eventually led to the emergence of the interactionist perspective, which recognizes a complex interplay of personal and situational components in the behavioral disposition changes. Most personality psychologists currently attribute themselves to some version of this view and some researchers suggest that synthesis of the views has been achieved and that 'incorporating both personality processes and trait structures in research will move personality research forward' [FN08]. This implies, that attempts to eliminate contextual aspect from the definition of personality were finally abandoned as unsuccessful. Even redefinition of personality traits from concrete values to probability distributions proposed by Fleenon, while a significant improvement, still does not fully capture idiosyncratic variability in behavior dispositions (Figure 2.1). The contextual, state aspect seems unavoidable.

CHAPTER 2. PERSONALITY MODELING

Figure 2.1: Example personality state, defined as current behavioral disposition, measured in different contexts for two individuals exhibiting equal personality trait, when it is defined as an average behavioral disposition, *but possibly even as the statistical distribution of this disposition*. Context dependent, intrapersonal changes in behavioral dispositions are typically comparable in scale to inter-personal differences.

Cognitive-Affective Personality System (CAPS) theory [MS95] is a notable example of a recent attempt to conciliate this 'personality controversy'. It belongs to the interactionist tradition i.e. it redefines personality from a context-invariant feature set to an idiosyncratic process of context-based behavior disposition adaptations (i.e. not 'what you are in general?' but 'what you are in which context?'). CAPS is introduced as based on a network of cognitive-affective 'mediating units' and it is empirically supported [MS95] despite relatively limited research on intra-personal behavioral disposition changes. It is also 'consistent with parallel findings showing significant amounts of variance attributable to Person X Situation interaction.' [CS99].

2.3.2 Personality in computational models

As the development of virtual and robotic agents went through extreme growth in the last decades [ZMVH08], the 'personality debate' moved to the field of personality computing, which is still to fully embrace recent progress in the psychology of personality, including CAPS. In this section, work and trends in computational models of personality are dis-

CHAPTER 2. PERSONALITY MODELING

cussed, divided into an analysis of the domination of the 'trait models' in computational personality modeling and discussion about works using non-trait approach.

2.3.2.1 'Trait models' in computational personality modeling

A broad spectrum of problems addressed by personality computing is often classified into three categories [Pia13]: Automatic Personality Recognition, Automatic Personality Perception, and Automatic Personality Synthesis. Only the last one is addressed here. Personality synthesis, as entire personality computing, is dominated by 'trait models', especially FFM. Values representing the selected dimensions are often used as regulating parameters e.g. to define default mood [Geb05, Ksh02], character of the emotional dynamics [SFC⁺10], to modulate appraisal [AKG⁺00], or directly shape the style of the discourse [MW10], facial expression [MSM⁺12], gestures, and posture [NTB⁺11], and finally goals or preferences in reasoning.

Domination of FFM, as personality model in agent and social simulations, is understandable due to its simple form and natural parametrization. However, it comes at a price of severe limitations implied by any 'trait model'. The fact that 'trait models' are descriptive tools not meant as a prescription of cognitive mechanisms is regularly ignored in computational modeling. Unfortunately, drawbacks of those models as descriptive tools described in the previous section are multiplied when they are used in personality synthesis. As behavior generation mechanisms, those models produce predictive, flat personalities undermining character believability. This is precisely due to the absence of any contextual cognitive-affective changes regulating behavioral dispositions. It is linked to criticism of FFM, as capturing 'surface, easily noticeable aspects of personality and neglecting more private or **context-dependent ones**' [McA95]. In practice, 'agreeable' or 'extrovert' character using FFM is always equally 'agreeable' or 'extrovert' with no individualized contextual variability, resulting in a personality, that

CHAPTER 2. PERSONALITY MODELING

fails to surprise in a new context (like fearful mother, suddenly brave when her children are threatened etc.). This flaw is not limited to FFM - it extends to all 'trait models', like PEN [Eys91] and other, less established ones.

All this has obvious consequences for individual modeling but, less obviously, models of non-rational components (like personality) are becoming a bottleneck in multi-agent and agent-based social simulations [KTT13, Kam13].

2.3.2.2 Non-trait models

Socio-cognitive psychological models like CAPS (section 2.3.1.2) are promising, but raise two related issues, especially important in the computational context. First, to define personality in terms of patterns of reactions to contexts, model must distinguish between external situations that differ in their consequences for human context cognition. Such definition should ideally not be liberal - it must follow the psychologically relevant structure of the situation. This is problematic, however, as 'there is no consensus regarding the most important psychologically active features of situations' [Ble09]. Second, what is the nature of those hidden 'mediating units' and how are they linked, specifically? In psychological models, both aspects are usually addressed at a high level of abstraction, e.g. CAPS just lists five types of cognitive-affective units (encodings, expectancies/beliefs, affects, goals/values and competencies/self-regulatory plans) and presents the nature of their connection in very general terms (Figure 2.2).

Several computational models approaching personality synthesis within socio-cognitive approach exist. Sandercock's cognitive-affective framework [SPZ06] explicitly introduces context to personality manifestation represented by 'coping preferences' and weights in appraisal processes. Will [Mof97] is an emotion generation architecture, extended to include personality architecture, as emotional influences on the memory, aims, etc. lead to behaviors manifesting measurable personality traits. Faur [FCPM13] focuses on the

Figure 2.2: The high-level abstraction presented by CAPS. Details of the mediating units and its connections are left largely unspecified.

self-image as the key personality building process. The connectionist methodology has also been used to model personality related processes [PT05], with situation features as one of the inputs.

Most notable example, however, comes from Read et al. [RMB⁺10]. Authors explicitly link their work to CAPS, make specific assumptions about the mediating units, specify a personality process, and prove it manifests some of the characteristics of FFM. The model focuses on two goal/motivation systems (avoidance and approach) and the influence of the abstract situational features on the goal related mediating units. Its principal goal is to reconstruct the underlying dynamics of FFM. The discussed model captures some of the important goal-related situation dynamics but abstracts away a crucial part of social context cognition i.e. the process of establishing proposed high-level, goal-oriented situation features, based on the sensory input and intermediate context cognitions that are absent in the model. The link between the sensory input and de-

CHAPTER 2. CULTURE AND NORM MODELING

termination whether the situation facilitates specific goals is nontrivial and includes the implicit interplay of social roles, group stereotypes, and similar elements. By determining permanent, idiosyncratic patterns in decision-making and self-regulation, those elements contribute to personality, whether their source is private or cultural. Their subconscious activation modulates saliency of norms, goals or plans at any given time. For example, self-imposed limitations on socialization goals at a funeral or a party depend on the activation of such encodings. It's persistent, and individual uniqueness may lead to the characterization of one's personality as e.g. 'rude' or 'stiff' by others. This aspect is absent here and in many other models. This thesis aims to address it, with an intention of future integration with models focusing on motivation, self-image management, and other previously modeled aspects. Additionally, it will address a need to initially check the utility of the proposed models in simulations not only of individual agents but also in extensive multi-agent interactions to investigate its potential usefulness in such applications.

2.4 Culture and norm modeling

This section will present work related to the computational culture modeling from the perspective of non-rationality and selected sociological theories relevant from this point of view.

2.4.1 Introduction. Culture in social sciences

Culture is a very broad concept with substantially varying possible definitions. Culture may, for example, denote entirety of material and intellectual product of human activity (e.g. 'pyramids are an excellent example of Egyptian culture') or cognitive-behavioral features unique to a group (e.g. 'strong hierarchy and collectivism is a part of Asian

CHAPTER 2. CULTURE AND NORM MODELING

culture’). This last case is close to how culture is understood in this work. As attribution of personality was seen as an intuitive way to describe *individual* particularities of human behavior (section), the construct of culture is used to characterize particularities *shared* within a population. This section describes how culture, understood as above, is approached in social sciences and some of the related issues and challenges.

2.4.1.1 Culture - the problematic concept

Traditional social sciences remain torn because culture is central to them, and yet it defies their rigorous and fruitful research [Bro08]. As a result, many branches of social science surprisingly downplay (implicitly or explicitly) ‘culture’ as a term. While easy quantification of economic or structural factors leads to easily verifiable theories, the concept of culture became a conceptual ‘garbage can’ for poorly defined mechanisms. Its use, to explain any results by trivially labeling them as a ‘cultural influence’, made the term unpopular, as explaining nothing in reality. It has been argued that this inability to adequately address culture is a key factor keeping branches of social science ‘fractionated, each speaking their own, often mutually unintelligible languages and holding assumptions and theories that are mutually incompatible’ [Mes11].

2.4.1.2 The micro-macro gap in the social sciences

Ontologically, socio-cultural properties emerge from psychological ones. The inability of social sciences to capture this with sufficient formality results in a disparity between description levels. It leads to macro-scale models representing culture ‘as an object’ with underrepresented emergent, process-like aspect, leaving out accumulated micro-scale psychological knowledge. Fear of reductionist issues, i.e. reluctance to reduce sociology to psychology, plays a role here. The same time, the micro-macro gap must be primarily attributed to limitations of the traditional verbal models and technical difficulties of creating quantitative, computational ones, that surpass expertise of the social scientists.

CHAPTER 2. CULTURE AND NORM MODELING

A related problem is a question of behavioral vs. cognitive perspective on the culture that divides its researchers, as many of them approach the problem from the behavioristic standpoint, through a pattern of agent's actions, while others feel a necessity to involve internal beliefs and mental states into the definition. Despite the hope to reduce culture to externally measured regularities in observable behaviors, the computational, emergent, and agent-based models of culture require theories that see culture as information and information processing system.

2.4.2 What are the norms? Norms and culture.

Another fundamental concept, closely related to culture, is 'norm' - an equally ambiguous term. In related fields, it represents different ideas, which often have little in common. For an engineering oriented MAS designer, norms may be unbreakable laws of agent interactions, designed by hand and static, to ensure global system efficiency. For a social scientist, the meaning is very different i.e.: 'Like a grammar, a system of norms specifies what is acceptable and what is not in a society or group. And analogously to grammar, it is not the product of human design and planning. This view suggests that a study of the conditions under which norms come into being, as opposed to one stressing the functions fulfilled by social norms is important' [Bic06]. Similarly, term *norm emergence* is often used in computational simulations, without much consideration, while its meaning is a source of a heated debate within social sciences, visible in the research of the psychological and cognitive genesis of norms.

Even within MAS alone, implicit differences in how norms are understood can be profound and misleading. Many researchers see norms as flexible, decentralized methods forcing global system efficiency, defined explicitly and externally [ST, You04]. Such norms are designed and imposed top down, to eliminate unpredictable behavior and formally enforce certain global properties. On the other hand, in many social simulations, norms

CHAPTER 2. CULTURE AND NORM MODELING

are perceived as constructs, liberally created by agents, emerging from mechanisms of individual, local optimization and learning, serving agent's goals and not any particular global system efficiency. In such cases, conditions and speed of convergence to a uniform set of norms in a group are often investigated. However, a uniform convergence may not be desired in a third perspective, i.e. one concerned with realistic social simulations. When socio-psychological believability is the aim, agent 'subcultures' are expected to hold different norm sets, as norms are not always globally acceptable in the actual societies.

If not made explicit, those very different ways of understanding norms may be a source of confusion in a discussion about cognitive basis of normative behavior. Therefore, authors share Rauwolf's, opinion that 'In discussing social science's utilization of the term norm, our goal is not to reach a conclusion on a definition still debated or to get mired in a semantic argument. Rather, in articulating the breadth of the terminology, it may yield the knowledge requisite to deliberate upon the utility of passing aspects of the vernacular into the computational arena.' [RBDV13]. This work uses results from several fields with a substantially different perception on norms, yet unless stated otherwise, the third perspective on norms presented in the above paragraph is assumed.

2.4.3 Norm emergence theories

From a sociological perspective there are, in short, three basic stances on emergence of social norms: socialization [PSA⁺51], social identity [Taf81], and rational choice theory [Axe86, Rom55]. All make different, testable statements and neither explains the experimental evidence and thus explaining norm emergence completely and sufficiently [Bic06]. Those three perspectives are introduced briefly in this section, to give an often overlooked background for the cognitive mechanisms that need to be implemented in MAS-CAA.

CHAPTER 2. CULTURE AND NORM MODELING

2.4.3.1 Theory of rational choice

A model of norm emergence is considered as belonging to a rational choice theory, if it assumes, that norms emerge and are maintained purely as a result of compliance based on sanctions in a rational, benefit maximization process [CTN07]. 'If conformity to a norm attracts approval and transgression disapproval then conforming is rational' [TK59]. In this perspective, external sanctions and their rational expectations determine behavior. Norm internalization and maintenance mechanisms may be therefore reduced to a general learning of external feedback related to actions, rationally aimed at benefit optimization. In models based on this view, there are often no norm-specific internal representations, and norms are usually understood behaviorally: as patterns of behavior in a group, emerging from sanctioning and generic learning.

Rational choice theory faces multiple problems in explaining social phenomena, norm emergence, in particular, [Bic06, Sco71]. For example, many observable social norms do not involve sanctions [Bic06] and people have a tendency to conform to costly norms, even when sanctions are impossible, for example when its transgression cannot be detected. It has also been pointed out, that monitoring and sanctioning works well only in limited groups and repeated interactions, otherwise resulting anonymity renders it inefficient. Supporters of the rational choice theory have problems with explaining such results without resorting to a form of habitual internalization, with internal self-sanctioning system e.g. guilt [Sco71]. External sanctioning or penalization, as the only cause for norm maintenance, is also problematic, because for it to function, the agents observing transgression must consistently punish the transgressors. This may be costly and usually brings no direct benefit to the punishing agent. Furthermore, from the perspective of the penalized agent, it is not obvious when he is an object of a sanction and when actions of others are unrelated to his transgressions. All this suggests mechanisms beyond generic,

CHAPTER 2. CULTURE AND NORM MODELING

rational, egoistic optimization, acting both on the side of the penalized, and penalizing agents.

Another prediction of the rational theory that contradicts observations is the expectation that norms more favorable for nearly all agents should emerge faster, than the ones unfavorable for the general population. Dynamics of the norms related to e.g. discrimination or corruption is inconsistent with those expectations. Bicchieri et al. prove, that norm inefficiency is an insufficient condition for a norm to disappear [BR95]. In the case of corruption, they show, that in real populations norms tend to cycle back and forth between more honest and corrupted versions, instead of leaning towards honesty, as the rational choice theory would suggest. Similarly, it is not easy to explain how, in an environment dominated by an established norm, a new norm could emerge and dominate. For example, based on sanctioning, popular fashions and social consensuses with little functional consequence should be stable, as transgression is costly and disobedience brings no benefit. This is also the usual result of a typical MAS simulation based on such assumptions. After population behavior converges to one norm to a certain degree, local norms have little chance to spread, in most cases, and 'cascades of self-reinforcing support for a highly unpopular norm cannot occur in a fully connected social network.' [CWM05]. That is clearly not the case in the actual societies, where unpopular norms often dominate, in a relatively short period, with no specific requirements about connectivity. For example, new fashions may have no direct functional value and be at odds with dominating trends but dominate the group rapidly. This suggests more complexity and involvement of specific psychological mechanisms subjectively shaping and weighting networks of social connections, or otherwise allowing the agent to maintain penalized behavior under the external pressure from a social network. Finally, it is accepted, that humans do not predominately operate on a basis of explicit, formal, and calculative criteria, but intuition, emotionality, and habitual heuristics, that are susceptible to various

CHAPTER 2. CULTURE AND NORM MODELING

types of reasoning distortions and biases, leading to irrationality and suboptimal solutions [Bou05]. It is hard to present reasons, why normative behavior should be a special case.

2.4.3.2 Theory of socialization

Models belonging to the theory of socialization, also usually assume general agent's tendency to maximize egoistic utility. However, some form of self-imposed norm obedience is postulated, related to internalization in a socialization process that imprints norms in agent by defining needs and affective responses to various situations. This produces a socially shared value system, by shaping the individual personalities.

The resulting norm conformity is usually depicted as environment/penalization independent. This statement may intuitively seem too strong, and indeed, it produces experimentally rejected predictions. For example, it has been shown, that expectation of other's conformity influences one's conformity [BC10]. According to the socialization theory, norms, as part of self, should affect both behavior and beliefs directly. Behavior and normative beliefs should, therefore, be significantly correlated, and lack of this correlation is problematic. Quite shockingly, most people do not follow their own declared standards. As experiments show, the primary factor influencing behavior is not what a person thinks is right, but this person's beliefs about societal expectations [Bic06]. This points out a great distance between the declared norms and behavior. At the same time, most researchers agree, that some degree of social internalization is hard to reject.

2.4.3.3 Theory of social identity

Socialization theory implies that normative preferences, simply by becoming part of individual preferences and goals, render external conditions marginally important. Social identity theory contradicts this and addresses the limitations of socialization theory. Its

CHAPTER 2. CULTURE AND NORM MODELING

basic approach is a deconstruction of an idea of a monolithic self, into multiple, context-activated 'social-selves' i.e. 'that part of an individual's self-concept which derives from his knowledge of his membership of a social group' [Taf81]. This decomposition is either partial, leaving some element of 'true I' coupled with the changing, superficial social-selves, or it is complete. In the later case, no identity beyond a social construct exists. In both cases, supporters of the social identity theory state, that specific behavior tendencies are closely linked to the context-dependent social-selves. For that reason, normative behavior explanation has to be contextual. Taken to a full logical conclusion, this theory not only explains the weak correlation between the declared norms and behavior but also between the behavioral norms in various contexts.

Within certain social contexts, people may perceive themselves as members of a particular group and their role in it, which changes the structure of self, according to the perceived group and role character [Taf81]. Change in the self leads to change in the goals, values, the character of emotional responses, and other characteristics related to behavior. Therefore, habitual internalization of a specific norm is not the only cause of norm obedience. The desire to maintain social identities perceived as demanding specific norm obedience needs to be addressed. Indeed, experiments support the claim that collectively shared norms are far more consequential to the behavior than the personal ones [CRK90]. In the social identity theory, habitual internalization aspect may be non-existent, partial (refer to some norms) or full and parallel to identification. To a degree that a norm is not habitually internalized, but based on a flexible subscription to identity, tendency to obey it may be adopted quickly, as beliefs about the identity may change rapidly due to the context. Indeed, studies of fashions and fads show evidence that such perceptions may change very rapidly. While social identity theory relies on the structure of social context cognition, it does not define it per se. The problem of social context cognition modeling is addressed in section 2.5 and remaining part of the thesis contains a formal specification of social context cognition model.

2.4.4 Culture in computer sciences

This section addresses work in computational culture modeling. At least since Axelrod [DGE⁺00], computational models promise a unifying platform to describe the culture. This attempt to model behavioral, externalistic dimension of culture is parallel to an attempt to model internal, individual encoding of culture and its influence on the remaining cognitive processes and the resulting behaviour. However, despite the significant advances in agent modeling with emotion and personality, the challenge of modeling the influences of culture on a cognitive level have been considerably smaller. Agent architectures remain a platform of consolidation of micro and macro models of great and largely untapped potential. As Neuman puts it, 'cognitive architecture of the agents provides an objective reconstruction of the subjective dimension of norms. Potentially, this provides a bridge between positivistic and phenomenological theory traditions. On the other hand, the inclusion of norms into the framework of the principally individualistic approach of agent-based modeling provides an attempt to overcome the classical micro-macro divide.' [Neu12].

2.4.4.1 Normative Multi-Agent Systems in social and culture simulation

The above perspective is, however, far from realization. Even worse, the way in which culture is addressed in computational models is inherently limited. The addressed problem is most strongly anchored in Multi-Agent Systems (MAS), particularly in the normative MAS, that will be addressed first.

MAS are computational models of multiple interacting agents of varying complexity, autonomy, and intelligence, placed in an artificial environment. MAS have been successfully used in various domains to solve problems, which are difficult to model with a monolithic system, including problems in social simulation. Application of norms has been one of the important strategies of coordination, used in such systems for two

CHAPTER 2. CULTURE AND NORM MODELING

decades [WW95]. Normative MAS have been defined as, 'a multi-agent system organized by means of mechanisms to represent, communicate, distribute, detect, create, modify and enforce norms, and mechanisms to deliberate about norms and detect norm violation and fulfillment' [BVDTV08]. As Jing et al. observe: 'Many domains are characterized by agents that interact with each other in accordance with common rules or norms. In international trade, a trading network may include a variety of entities (e.g., software, organizations and people) that are largely autonomous, geographically distributed, and heterogeneous in terms of their operating environment, culture, social capital, and goals. At any given moment, most agents will be conditioned by different regulations and norms, originating from different institutional contexts'[45].

Since the early works, like the Epstein's famous article [EA96] promising culture simulation with MAS, various researchers have explored this paradigm. One important trend prominent in this tradition is the investigation of norms.

2.4.4.2 Two tendencies in normative Multi-Agent Systems

In MAS norms are on the one hand seen as pragmatic tools of distributed control assuring properties of a designed system and on the other as providing a bottom-up, individualistic explanation of social phenomena. The term 'norm' is ambiguous in MAS and intra-agent norm representation is not standardized [Neu12].

In some cases norm is established subjectively and individually, while in many other works objectivistic and collectivist notion of a norm can be found. Ullman-Margalit defines a norm as a 'prescribed guide for conduct or action which is generally complied with by the members of the society.' [UM77] while Savarimuthu, as (internal) expectations about (external) actions - 'norms are expectations of an agent about the behavior of other agents in the society.' [SC11]. As Bicchieri puts it 'some authors identify norms with observable, recurrent patterns of behavior. Others only focus on normative beliefs

CHAPTER 2. CULTURE AND NORM MODELING

and expectations. All find it difficult to explain the observed variance in norm-induced behavior, and each offers an explanation of conformity that is at best partial' [BM11]. Furthermore, some works focus on the emergent, and unplanned aspect of a norm, to make a clear distinction from designed laws or codes: 'it is not the product of human design and planning it is important in order to understand the differences between social norms and other types of injunction, such as hypothetical imperatives, moral codes or legal rules' [BM11]. At the same time, other researchers do not make such distinctions and include designed norms, as yet another type of a social norm. Other possibilities also exist. For example, Coleman defines norm using a notion of a 'right to control': 'norm concerning a specific action exists when the socially defined right to control the action is held not by the actor but by others' [CC94].

On the implementational level, most popular choice are *constraints* i.e. hard limitations on agent actions designed off-line or *obligations* i.e. explicit conditional prescriptions incorporated e.g. in BOID architectures [BvdT03]. Rarely more complex mental objects are mentioned, usually *moral norms* and *values* [BTDH11] that accumulate all relevant cognitive aspects ill-represented by 'if-then' rules. Even when a norm is not simply seen as an external behavior regularity (e.g. 'all agents use the right side of the road') the cognitive complexity of its representations is low compared to mechanisms postulated in sociology.

Despite those multiple possibilities, there is a dominating tendency in MAS, going back to Epstein [DGE⁺00] to identify culture with a) *norm emergence* as behavior pattern propagation, b) *external products of a)* e.g. trade and relation networks, production and consumption patterns, life span, and housing distributions. [DGE⁺00]. An implicit assumption that this reasonably approximates culture is based on a broad application of norms in psychology, as 'cultural product including values, customs, traditions etc.' [She36]. This is, however, unjustified considering the narrow redefinition of the term

CHAPTER 2. CULTURE AND NORM MODELING

in MAS. The fact that abstract conditional rules are not a good model of the innate habitual, affective, and other complex particularities of normative mechanisms discussed in social psychology is often silently ignored.

Norm emergence in MAS predominantly belongs to the 'economicus' tradition (see section 2.1.4), as described by Elster : 'One of the most persistent cleavages in the social sciences is the opposition between two lines of thought conveniently associated with Adam Smith and Emile Durkheim, between homo economicus and homo sociologicus. Of these, the former is supposed to be guided by instrumental rationality, while the behavior of the latter is dictated by social norms. The former is 'pulled' by the prospect of future rewards, whereas the latter is 'pushed' from behind by quasi-inertial forces. The former adapts to changing circumstances, always on the lookout for improvements. The latter is insensitive to circumstances, sticking to the prescribed behavior even if new and apparently better options become available. The former is easily caricatured as a self-contained, asocial atom, and the latter as the mindless plaything of social forces.' [Els89].

2.4.4.3 Multi-Agent Systems and Homo-economicus

This dominating, rationalistic trend and externalistic assumption to create complexity with simple agents promote socio-cognitive insufficiency. Next to the general rationalistic inclinations (see 2.1.4) this has also a trivial, technical justification: agents with complex cognitive components are hard to model, implement, parametrize, and validate. Constraints and obligations are easy to use, but 'it requires a lot to transform morality into a computational approach. It is still an open research question how this can be realized' [Neu12]. This is even more true in a broader case of culture.

Previously mentioned conflict of behavioristic and cognitive perspectives on culture in social sciences (section 2.4.1.2) is therefore reflected in the normative MAS, that may

CHAPTER 2. CULTURE AND NORM MODELING

roughly be divided into two major categories, as 'researchers in normative multi-agent systems have been influenced from two different perspectives: philosophy of law (prescriptive approach) and conventionalistic approach (emergence approach).' [SC11]. Similarly, Andrighetto et al. observe that 'domain of normative agent systems presents two main directions of research. The first focuses on intelligent agent architecture and capacity to decide on the grounds of norms and the associated incentive or sanction. Those studies have provided Normative Multiagent Systems with a formal analysis of norms, thus giving crucial insights to represent and reason upon norms. The second is focused on much simpler agents and on the emergence of regularities from agent societies' [ACCC10].

While useful in itself, this second and dominating group is limited in its potential to further substantially contribute to realistic social simulation, without embracing more complex normative cognition. Simulations of norm emergence based on this perspective were used in economical simulations [Kit93], investigation of simple behavioral patterns like the use a left or right side of the road [HK04]. In those and other cases, research on 'norm emergence' in fact means analysis of conditions required for a domination of selected pattern in the population of simplistic agents. That is a relatively easy task, yet its link to social phenomena is problematic, as decision making is trivialized and rationalized and environments simplified to single step interactions usually with instant clear feedback, etc. This mainstream 'game-theoretic' trend is important - it provides insights into mathematical properties of emergence in systems where simple agent actions adhere to the proposed rules. However, the degree to which underlying simplifying assumptions about decision making allow its application in the general emergence of the social phenomena is limited.

2.4.4.4 Computational models in cognitively rich agents

The rather new research on computational models cultural agents can be roughly divided into cognitive, model-based approaches and data-driven ones. Data-driven solutions ap-

CHAPTER 2. CULTURE AND NORM MODELING

ply general, machine learning solutions to learn human-labeled audio-visual datasets containing contents with culture-specific contents, in the attempt to generalize it. The typical approach is an analysis of differences in topics, gestures, tone, postures or communication choices between the targeted cultures, for example, German and Japanese cultural features in the case of CUBE-G project [RAB⁺07]. Due to the effort-heavy labeling process, solutions in this group have a solid empirical foundation, but a lack of underlying theoretical model leaves the question of generalization open and highly reliable on the sample selection. Poorly selected samples may distort cultural features and model's ability to generalize and even well-selected ones may provide the ability to transfer results to slightly altered contexts, also making comparisons challenging. Model-based approaches, on the other hand, are grounded in the existing cognitive models and use cultural parameters to alter their components (like goals or beliefs), according to the assumed theory. Examples include the extension of the FAtiMA cognitive model (see the last paragraph of section) with the representations of the Hofstede cultural dimension values for the culture of the character, cultural specific symbols, culturally specific goals and needs, and the rituals of the culture and the Tactical Language System [JBFW⁺04]. While such models, in general, have a higher consistency than the data-driven approach and are well suited for the synthetic applications, their application realistic simulation of the existing cultures may be problematic due to lack of empirical anchoring. Some other computational models of culture include negotiating agent NegoChat [RZSH⁺14] or PURB and PAL agents [ABG⁺17].

This is one approach to address culture. The same time, 'in the field of AI culture has most often been seen as a more or less explicit, internally homogeneous, externally distinctive, collective entity. However, there is data to suggest that culture is and is better treated in AI as identity in context the demonstration in a particular context by a given person of a set of practices that index to members of the same group, and members

CHAPTER 2. CULTURE AND NORM MODELING

of other groups, his/her cultural community membership.’ [?]. Therefore at least part of the solution to a problem of cognitive dimension of culture should be contextual behavior modulation, including contextual norm handling. As explained in the previous sections, most work on the emergence of norms is done from a rationalistic, game-theoretic perspective [SC11] and lacks explicit rich internal norm representations. ‘Usually, in the formal social scientific field, and more specifically in utility and (evolutionary) game theory, the spread of new norms and other cooperative behaviors is not explained in terms of internal representations’ [ACCC10].

Attempts to address problems resulting from this exist outside the MAS mainstream but are relatively limited. Breen [BTDH11] uses ‘metal norms’ to explicitly separate culture, and norms and to show how ‘culture affects the possibility of normative changes, in particular, the acceptance of policies’ [BTDH11]. Kochanowicz [KTT13, KTT15a] proposes a model for MAS uniting norms, culture, and personality on a level of social context cognition. Prominent architectures embracing rich, internal normative representation include: NoA [KN03], deliberative normative agents [CDJT00], and BDI-based architectures: EMIL-A [ACCP07], BOID [BDH⁺01] and NBDI architecture [dSNdSdL13]. Comprehensive surveys including cognitively-rich normative architectures exist [Neu10]. Of the above, EMIL-A is the most complete one, with implemented norm recognition, internalization, and selection processes. By comparison, all the other architectures assume that norms are automatically detected, or explicitly known to the agent, while, in general, norm detection problem is a very nontrivial one. Furthermore, BRIDGE, NoA, and BOID do not include reasoning about the norms.

EMIL-A will be briefly presented, as a state-of-the-art example. It includes a norm recognizer, classifying input in two modes: recognition of the previously known normative information and detection and classification of unknown normative input. In case the external normative input is new, norm internalization starts. It uses a normative frame,

CHAPTER 2. CULTURE AND NORM MODELING

with a dynamic schema used to recognize and categorize an external input, as normative or not. The recognition of a norm does not necessarily imply its full understanding or will to apply it. It only means that the new information has been classified as a norm. Known norms are kept in the normative board. Once enough information is gathered about the new norm, it is turned into a normative belief. This is done using a normative frame, by checking the agent's knowledge about the relevant frame variables. Normative belief becomes a candidate for norms, which might be used. The norms on the normative board are ordered by salience (the degree of norm activation) - and are considered in the usual BDI reasoning processes, as restrictions on the goals and intentions of the agent. In the case of a normative conflict, highest salience decides which norm will dominate. Norm adoption uses a 'why not' approach: a new norm has a slight preference to be adopted if the agent cannot find evidence that it creates conflicts.

Rare systems like EMIL-A are built with a correct assumption that lack of internal norm representation is a limiting factor in social simulation. Sole internal norm representation is nevertheless insufficient due to the entire spectrum of socio-cognitive mechanisms, of which norm acquisition and management is one important element. As normative cognition is related to and distributed among various subsystems including affect, personality, a structure of needs, group identity, the social context structure, etc., norm-managing agents like EMIL-A are the first, crucial versions of cognitive architectures developed for socio-cultural simulations.

Despite similar attempts, psychological mechanisms of culture emergence and internalization are highly non-trivial, and the solution to the posted problem exceeds traditional MAS expertise and require transitions from MAS far into the domain of cognitive architectures (CA). Deficiency in social cognition limits believability of groups, but also individuals and CA-MAS dependency goes both ways. It is true that 'Social simulation with multi-agent systems can benefit from incorporating cognitive architectures, as they

provide a realistic basis for modeling individual agents' [Sun09]. The same time, the benefit may be mutual: application of CA in MAS will also stimulate the development of the mechanisms relevant in the simulation of individuals, virtual characters, etc.

2.5 Social context cognition

'System 1' social context cognition has been identified as a key component of non-affective non-rationality (section 1.2.4). The question remains, how to model it? This section introduces several psychological theories related to context cognition and review of the state of work in relevant computational domains tackling this problem.

2.5.1 Psychology of social context cognition

Multiple psychological theories are attempting to explain the way in which humans alter their behavior through their perception of social context. We choose to describe three of them from the one proposing most general to most specific tips about the social context cognition. First, Constructive Perception Theory is presented as a principal foundation behind a contextual perception. Next, Dramaturgical Theory is described with its consequences for modeling of 'System 1' social cognition, more specifically elements of context structure. Finally, Dissonance Theory is presented, as a theory that may be used to model human-like some of the interferences and discrepancies between those elements of social context. Furthermore, section 2.4.3.3 discusses Social Identity Theory (SIT) as a norm-forming mechanism. From the perspective of this section, SIT may be seen as its continuation. Namely, a specification of how the interplay of components of social context cognition dynamically alters other functionalities including identity managements or normative aspects of cognition. These theories will later be fundamental in a formulation of the computational models of social context cognition.

CHAPTER 2. SOCIAL CONTEXT COGNITION

2.5.1.1 Constructive Perception Theory

Constructive Perception Theory (CPT) states that there are bottom-up and top-down processes used to construct understanding in human cognition. Its roots go as far as to the philosophy of Immanuel Kant suggesting that perception-understanding relationship is reciprocal i.e. both affected by and affecting our experience of the world. CPT implies a high-level, consolidated and integrating image of the situation constructed from based on all sensory information and previous knowledge. This consolidated, contextual level is constantly updated in bottom-up process and it is used to guide and inform perception in a top-down process. [SS16]. Such structure leads to critical information reconstruction (e.g. while adding an obstructed letter 'O' on a red road sign, when 'S', 'T' and 'P' are visible) and but also possible perception distortions or various cognitive biases (also see system 1). This also points to high-order thinking and learning in perception. During perception, hypotheses are formed and tested about percepts that are based on three things: sensory data, knowledge, and high-level cognitive processes. CPT points to the general parallel bottom-up and top-down nature of perception and a centralized, unified cognition level while the theories discussed in the next points provide hints what is the structure of this centralized level and its influence of perception.

2.5.1.2 Dramaturgical Theory

One of the core elements, implicitly assumed in the theories in section 2.2, is a constant human-like monitoring of the social context. However, its specific mechanisms and structure were not specifically addressed by those theories. One theory that does address them is the Dramaturgical Metaphor or Dramaturgical Theory [Gof59] of the social life. The Dramaturgical Theory (DT) had an extensive influence on psychology, sociology, and anthropology. According to its original, sociological formulation, all behavior is inherently social. It also confirms the assumptions of the Social Identity Theory (discussed

CHAPTER 2. SOCIAL CONTEXT COGNITION

in section 2.4.3.3), in that it maintains that identity is not a stable, independent entity but one continuously altered accordingly to specific social contexts. More precisely, DT claims that innate human cognition governing social behavior operates in terms of a stage, or dramaturgical performance analysis. While contents of the dramaturgical scripts are personally and culturally developed, its structure seems innate and fixed and composed of scenes, roles, and scripts of social life.

The concept of a 'social role' is central in DT, as a norm aggregating entity in culture. The role is not understood, like for example in the organizational modeling context, as an atom of a group functionality in an external, objective, institutional perspective (for example official role of the 'general' in the army or a 'father' in the family). In DT role is a continuously, inter-subjectively constructed normative unit, often informal (like a 'friend') and constantly updated.

Both role definitions and their subscriptions to agents are individual, subjective, and possibly conflicting with the views of other agents. They may also change frequently. Similarly, concepts of a 'scene', with available 'social scripts' refer to the informal perception of classes of social situations, with norms, goals, rituals, plans, or emotional settings, subjectively perceived as acceptable in this context. Innate dramaturgical cognition is constantly filled with individual content and regularly synchronized externally by social feedback, making it inter-subjective rather than purely subjective. Postulated 'role-playing' is used to construct important data and characteristics essential in various cognitive and affective mechanisms. Acceptable and expected actions, moods or goals are functions of a dramaturgical interpretation. Dramaturgical cognition may also cause irrational, human-specific biases i.e. degree of perceived 'behavior consistency' is more correlated with the ability to recognize the role the actor is manifesting, rather than with logical consistency of a goal pursuit. Other agent features could be similarly defined and extracted, based on the patterns of dramaturgical performance.

CHAPTER 2. SOCIAL CONTEXT COGNITION

2.5.1.3 Cognitive dissonance theory in social context cognition

While processing information related to context, e.g. in an implementation of DT, a multitude of social signals representing context interpretation are used. For example, those related to the general dramaturgical expectation about behavior in a given scene, others to historical experiences with the other actors, their perceived role feasibility, actual perception of their current behavior, etc. Those signals carry information in themselves, but the fact that they do, or do not contradict with each other, is also consequential. Tensions caused by contradictions between these signals are important indicators of abnormalities or irregularities in social behavior. For example, they may indicate uncertainty, requiring extra attention or information, or trigger a response of emotional or rational nature. In addressing this problem, the nomenclature of the Cognitive Dissonance Theory (CDT) [Fes62] is very convenient. According to CDT tensions between the concurring cognitions cause anxieties or other mental discomforts forcing the agent to minimize them through behavior or mental actions. Using this approach in dramaturgical cognition, consonances and dissonances relevant for context interpretation may be defined to modulate important variables like the affective state of agent or norms.

2.5.2 Social context in Social Signal Processing

There are several computational domains worth mentioning with relation to social context. One of them is the rapidly growing field of Social Signal Processing (SSP), interested in imitating a set of human 'ability to express and recognize social signals like turn taking, agreement, politeness, empathy, friendliness, conflict, etc.' [PCD⁺11]. The ultimate goal of SSP are computer systems 'capable of sensing agreement, inattention, or dispute, and capable of adapting and responding in real-time to these social signals in a polite, non-intrusive, or persuasive manner, are likely to be perceived as more natural, efficacious, and trustworthy' [PCD⁺11]. Three tasks are therefore at its core: analysis,

CHAPTER 2. SOCIAL CONTEXT COGNITION

modeling, and generation of social signals, understood as communicative and informative messages (conscious or not) related to social facts. These social facts are typically categorized as:

- social interaction - specific instance of social event signaled through nods manifesting agreement, blink manifesting intimacy, etc.
- social emotion - a subset of emotion related to others like envy or empathy, indicated by tone, gestures, etc.
- social evaluation - assessments of characteristics of a person or group based on certain standards.
- social attitudes - a broad category defined as a tendency to behave in a certain way towards an individual or a group.
- social relation - relation towards agent(s) or group(s) based on co-dependence of their goals e.g. dependency, competition, exploitation, etc. This concept notably includes social roles understood as behavior expectation patterns.

Signals related to the above social facts manifest through multi-modal cues like 'facial expressions, body postures, and gestures, vocal outbursts like laughter, etc.' which can be automatically analyzed by technologies of signal processing. This domain is at its early stage; it is still addressing its core research questions, and many problems have only been addressed recently. Works addressing non-verbal social interaction analysis in small groups have been presented by Gatica-Perez [GP09], Gunes surveys recognition of emotion [Gun10], and Kleinsmith affective bodily expression recognition [KBB13]. The works on evaluations and attitudes are more scarce. Examples include facial attractiveness estimation [KDL⁺08], automatic assessment of agreement [KDL⁺08]. In comparison, there

CHAPTER 2. SOCIAL CONTEXT COGNITION

is a great number of works related to social relations (roles in particular) e.g. recognition of roles in shows [SFV09] and meetings [PZN⁺08].

From the perspective of SSP, human-like non-rationality should not be seen as a separate category but rather one pervading all the mentioned social fact classes and cognitive processes causing them. For example, social evaluations and attitudes have been shown to be heavily influenced by various 'System 1' particularities, like the mentioned cognitive biases. Social relations often depend on non-rational, context and culture dependent patterns that have no justification but a habitually repeated cultural transmission. All of them are known to rely on stereotypes maintained to support mentioned habits or mental processes like self-identification and other human-specific social needs.

This relates to a potential weakness of SSP as an approach to human-like non-rational phenomena modeling. The very term 'Social Signal' ties attention to the *surface, specific and behavioral* aspect i.e. visible communication, while developments in cognitive science and psychology point to *underlying and general* characteristics and limitations of human thinking that do not solely influence social interaction but cognition in general. Traditional solutions often used in SSP like hidden Markov models, conditional random fields, or deep belief networks, are pragmatic and effective in specific tasks, but they do not take advantage of knowledge about human 'System 1', nor help to model it in a broader scope. We, therefore, agree with Pantic et. al. who pointed out that SSP would benefit from addressing some of its challenges, significantly signal contextuality, as 'one complex problem rather than a number of detached problems in human sensing, context sensing, and human behaviour understanding'[PCD⁺11]. Therefore, signal processing techniques effective at particular tasks (like gesture recognition) need to be placed in a context of general cognitive architectures (discussed in the next section) to provide a broad computational theory of human-like non-rationality.

2.5.3 Social context in Cognitive Affective-Architectures

The mentioned requirement, to place the specified theory in a concrete, 'default' cognitive environment is a typical problem faced in application or validation in various domains, including SSP. It was also one of the motivations to develop Cognitive Architectures (CA) and their use in affect modeling, as Cognitive-Affective Architectures (CAA) (section 2.2.3). Development of CAA from CA exemplifies an abandonment of rationalistic paradigm and inclusion of non-rational aspects, usually limited to emotion but sometimes (especially in non-academic Artificial Humans) also in a simple personality representations. However, even CAA are:

- mostly limited to affective non-rationality leaving out other 'System 1' particularities
- designed with a focus on individual and used in separation (with the exception of Clarion [Sun06])
- ignoring many mechanisms of social cognition crucial in sociological psychology

Therefore CAA (including Clarion) do not explicitly model 'System 1' mechanisms, particularly social and cultural ones like social context cognition, stereotypical agent perception, or human-like identity management. As a result, an interaction of multiple CAA does not manifest complex social structures, and other phenomena found in animals, let alone humans.

CA evolved into CAA by including affect but have yet to embrace the non-affective and socio-cognitive 'System 1' mechanisms. The complexity of the traditional CAA problems forced a focus on individual, affective models of influence on otherwise rational decision-making and little effort has been put into the development of CAA in social simulations. Paradoxically, this strong focus on individual may cripple social group

CHAPTER 2. SOCIAL CONTEXT COGNITION

simulation, as well as believability of the individual CAA. Idiosyncratic features manifest as temperament or affective dispositions manifesting in separation but also as normative and cultural patterns like interpretations of social contexts, taboos or subscriptions to group identities. Those elements are an important factor even in individual functioning but are easier to omit in lone agents than other features of affect and personality. 'Studies with synthetic characters have so far infrequently considered the link between behavior and culture, it may become an invisible background; directly encoded into the design' [AVA⁺09].

In summary, pure CA had started to include affective mechanisms by forming CAA, but still have to embrace the socio-cultural ones, on the larger scale. At the same time, indirect influences of DT are visible in some of the existing architectures. Other, less numerous models include concepts of roles [CSJ⁺10, DMHvdB12, PI01, Rom55] or social context [KCCS05, FJ13, JADT12]. Selective inclusion of the 'dramaturgical' elements shows changing research focus and at the same time most models are partial in their use of context elements, from the perspective of 'System 1' realization of DT. Perhaps the most complete model is an extension of the FAtiMA model [AVA⁺09] analyzed here to give an example. FAtiMA's authors propose a formal XML describing cultural 'rituals', 'symbols', and positions in Hofstede's cultural space [Hof01]. 'Rituals' are cultural plans used by FAtiMA planner or to generate appraisal. 'Symbols' translate perception to meaning e.g. 'wave-hand' to 'greeting'. The prescribed 'dictionary' allows cultural translations to the structure of meaning (e.g. 'wave-hand' to 'greeting' link) but the structure itself is not influenced and hard-coded (e.g. the very existence of 'greeting'). This way only the surface of the cultural differences is touched, where in fact they are far deeper and introduce very concepts like 'greeting' or their socio-affective connotations. Additionally, cultural modulation is limited to five dimensions with their predefined consequences e.g. 'in a collectivist culture, events good for others are appraised more

CHAPTER 2. SOCIAL CONTEXT COGNITION

positively'. 'Culture/personality vectors', are another problem, and the fundamental difficulties concerning them will be addressed later (sections 2.3.2 and 2.4.1). Finally, models including purely predicate-based socio-cultural cognition, like FatiMA, may not be ideal for 'System 1' with connectionist models being more suitable.

Chapter 3

Social Context Cognition Model

Social situations constitute a reality sui generis and therefore need a warrant analysis in their own right

Erving Goffman

This chapter presents a Social Context Cognition Model (SCCM), as published in [KTT15b, KTT13]. SCCM is a connectionist model. The network activations it establishes reflect agent's informal, 'System 1' [Kah11] social associations induced by a perceived situation. It captures non-conceptual, habitual and associative connotations and their behavioral consequences, rather than rule and reasoning based, calculative ones. As an implementation of the social dimension of 'System 1', it may simulate its various effects on cognition and behavior discovered in humans [Kah11]. Those include generation of non-appraisal affect, adjusting the action, plan or norm feasibilities, or influencing personality state expressions, depending on needs of the particular simulation and agent architecture in which it is embedded (section 3.1). SCCM is usable in simulations oriented both on entertainment (e.g. games) and scientific, social simulation, and may be parameterized based on the sociological data, by hand or by semi-automatic methods.

This chapter presents the use of SCCM in both manually created, entertainment-oriented characters and those generated with sociological data or in semi-automatic parametrization, usable in social simulation. The model is validated with the use of

CHAPTER 3. INTRODUCTION

objective psychological and sociological data and the subjective believability tests. In this chapter subjective believability tests of hand-made SCCM in entertainment-oriented characters is presented. Our results confirm that SCCM is promising in this kind of application. Statistical comparison of believability with human data shows that 'System 1' context cognition produced by handmade SCCM may be of similar quality to hand-prepared samples, especially if the second round of adaptation of the prototypes statistically shown as 'unbelievable' is performed.

The remaining part of this chapter is organized as follows. Section 3.1 provides a background to the social context cognition in the frame of non-rationality and presents a general overview of the SCCM and justification of the design choices underlying it and addresses the saturation problem, appearing in the recurrent connectionist models like SCCM. Section 3.2 is a detailed presentation of the data processing occurring in SCCM and contains a formal correctness proof of the used 'depth mechanism', regarding solving the saturation problem. Section 3.3 contains a formal correctness proof of the used 'depth mechanism', while section 3.4 includes a case study of a character and scenario creation and the example generated outputs. Section 3.5 describes the believability validation procedure and its results. The chapter ends with section 3.6 that contains summary and conclusions.

3.1 Introduction

Chapter 2 discussed aspects of non-rationality modeling underrepresented in computational models. Consequences of this underrepresentation were analyzed and, specific directions suggested. Among others, an informal, 'System 1' [Kah11] social context cognition model was identified as a major missing element underlying many of them: uniting culture, personality, motivation, normative behavior and various other aspects of human non-rationality (Figure 3.1).

CHAPTER 3. INTRODUCTION

Figure 3.1: Social Context Cognition Model in an example agent architecture

While social context cognition influences many aspects of human-likeness, it remains uncertain how to model it. This section addresses the general methodological questions and design decisions before the actual model is proposed. The candidates for model components (i.e. mediating units) establishing social cognition are identified, and the rationale behind the particular chosen social prototypes is presented. The selection of the connectionist methodology is justified, and the nature of the connections between the mediating units is discussed. Finally, the problem of saturation encountered by the recurrent connectionist models like the one presented is discussed.

3.1.1 Key design choices. Social prototypes as mediating units

The issue of how to contextualize human behavior has been a longstanding dilemma [Rob07] and the key problem is the potentially unlimited number of candidates for the relevant encodings of context cognition. Verbal, socio-cognitive psychological models raise two related issues. First, context model must distinguish between external situations that differ in their consequences for human context cognition. Such definition should ideally not be liberal - it must follow the psychologically relevant structure of the situation. However, this is problematic, as 'there is no consensus regarding the most important psychologically active features of situations' [Ble09]. Second, what is the nature of the hidden 'mediating units' and how are they linked, precisely? In psychological models, both aspects are usually addressed at a high level of abstraction but in a computational context-cognition model, those components must be specified.

Two observations are important. First, human cognition is inherently social, to a degree that in terms of social simulation 'context' may be approximated by 'social context'. Second, human 'natural categories' are organized around 'prototypes' [NC84]. Also, social cognition is shaped by prototypical encodings [CMS82] i.e. humans not only store concrete information about the individuals or situations, but also abstract representations and their features. Following prototype-related mediating units were selected:

Role prototypes. Social roles introduce predictability in human relations, by unifying behavioral expectations of the ones attributed to them. As relevant predictors of the manifested behaviors 'social role with its immanent demands is a reasonable concept for contextualizing personality' [Ble09]. Notably, also beliefs about the roles of others predict behavior [ØR01] and attitudes towards others are based on roles attributed to them [Olw94]. In summary, the influence of the primed social role on the behavior shows stable regularities [MS95], and roles are the prime candidates for mediating units in a model contextualizing behavior [Rob07].

CHAPTER 3. INTRODUCTION

Group prototypes. Perceived group identity is also a well-established modulator of behavioral dispositions. People identifying with a group behave differently and ‘saliency of the group affects the perception of the environment’ [CRR07]. Moreover, people primed to perceive different identities, change differently. For example, self-identifying as a ‘general football fan’ versus ‘fan of the team X’ may change behavior in a dramatically different way [LPER05]. Similarly, identification with religious, professional, subculture, criminal or other groups also causes specific behavioral dispositions.

General situation prototypes. A general situation prototype is the broadest context representation used. It is understood as a collective ‘meaningful environment encountered by an individual at a given time’ [CM98]. Situations understood in this broad way, are good behavior disposition predictors [WM88]. Also, prototypical situation categorizations are highly shared: ‘naive perceivers agree about person-situation matches, sharing knowledge of the most prototypic behaviors and personality types associated with different types of situations. Research suggests that such knowledge about social situations might prove useful for the perceiver as an actor to plan and regulate behavior.’ [CMS82].

Prototype features. As roles, groups and situations have been well studied in terms of their underlying features, numerous candidates have been identified. This suggests an intermediate processing step, between the sensory and the prototype unit activation. There is some consensus that the main dimensions leading to social role activations are power and warmth [Ble09]. In the case of such situations, valence and authority were noted as relevant [GZKB14]. For groups, perceived competence and amiability are probable key dimensions [CFK⁺09]. Additionally, certain prototypes are directly activated by specific sensory data like symbols, uniforms, or behavior.

Connections between the above prototypes must also be considered. The active role, group and situation prototypes influence manifestation of social context cognition on

CHAPTER 3. INTRODUCTION

Figure 3.2: Part of a connectionist model involved in the **permanent saturation** problem. Positively associated units are strongly activated, circular self-excitation may remain after the external context relaxation. At $T=1$, relevant feature detection stimulates units and proper maximum activation in the given context is reached ($T=5$). Input relaxes, but units are correlated strongly and their excitation exchange is sufficient to indefinitely maintain high activation ($T=10$, $T=15$).

a behavioral level, as model's 'outputs'. However, they also reinforce each other. For example, situation cognitions alter probabilities of role assumptions [ØR01, GZKB14] and perceived role feasibility was influenced by 'lesson' or 'break' situations. To reflect that, SCCM includes recurrent connections between the prototypes. Finally, activation of social prototypes has also relatively stable consequences, other cognitive consequences like goals or norm saliency changes and SCCM may also be used for their modulation.

CHAPTER 3. INTRODUCTION

Figure 3.3: **Inhibition** as an attempt to answer the saturation problem. Two strongly connected unit groups (A and B) form a network. Inhibition allows proper activation (I), but due to the great self-excitation large group (A) enters permanent saturation, while small (B) behaves properly (II). If decay or attenuation are increased to stop A's saturation (IV), B is unexcitable in contexts to which it should respond to (III). As the amount of the self-excitation is proportionate to the number of strongly correlated units, inhibition parameters not only need to be tuned to each network but may potentially have to be infinitely large.

CHAPTER 3. INTRODUCTION

Figure 3.4: Weights are not normalized and units vary in their excitability based on the data. In this example, role R1 is difficult to excite while role R2 may be easily activated by several sources. If normalization is used to deal with the problems of saturation, R2 becomes very hard to excite and activation patterns contained in the data are reversed and corrupted.

3.1.2 Connectionism, saturation, and 'depth' mechanism

To reflect the associative and subsymbolic nature of the prototypical influences, SCCM is designed as a connectionist model, while circular connections between prototypes make it also recurrent. Any connectionist model with recurrent, positive feedback faces issues of stability and responsiveness. SCCM is no exception (Figure 3.2). When positively associated units are stimulated strongly together, their circular, self-perpetuation excitation may cause permanent or chaotic excitation. The most popular method that may decrease the probability of permanent saturation is activation inhibition (by decay over time or attenuation of the propagated signals) that may decrease the probability of permanent saturation. A classical example [Gro82] uses on-center off-surround inhibition

$$\frac{d}{dt}x_i = -Ax_i + (B - x_i)I_i - x_i \sum_{k \neq i} I_k \quad (\text{Eq. 3.1})$$

where A is the decay parameter, B is the maximum excitation, I_k is the input k and x_i is excitation of unit i . Simple models with an even number of recurrent connections

CHAPTER 3. INTRODUCTION

may, therefore, eliminate this problem by selecting a good inhibition value. This is, for example, the case of Read's model [RMB⁺10] that only includes two recurrent connections for each goal node. However, in a recurrent network with an arbitrary number of unnormalized weights, it may be impossible. To avoid permanent saturation of multiple and strongly correlated units inhibition must be high, to counter their great excitation. However, this inhibits less connected units with correct activations (Figure 3.3). Both situations lead to continuously misleading context detection and permanent system insensitivity. In a general case, when an arbitrary number of recurrent connections for each unit is allowed (like in SCCM), this danger cannot be eliminated (but minimized at best), when no middle ground in the inhibition value exists. Moreover, even when such hand-tuning is achievable, any weight change (e.g. partial randomization of SCCM) negates it.

While introducing weight normalization makes inhibition selection easier by removing the inequality in recurrent connections, it is also problematic. First, SCCM may be based on psychological data that would be distorted in the process. Second, in SCCM, each unit may potentially be fully activated by several sources, including observed features, detected groups, roles, situations, and depending on the data, sums of excitatory weights may vary, by as much as an order of magnitude. Therefore, normalization makes units strongly connected to multiple sources unexcitable (Figure 3.3) and weakly connected ones easily excitable, perverting the activation patterns.

To address the above problems, a new method of dealing with this issue is introduced in SCCM. In general, stability is a complex system-dynamic issue, but a specific assumption about the unit activation function solves instability and permanent saturation problem. This is done based on the tracing of the signal's distance from its perceptual source. As postulated in the grounded cognition theory [Bar08], all internal representations, including social prototypes, are anchored in the sensory input. As activation

CHAPTER 3. PROPOSED MODEL

spreads from perception to abstract social cognitions, it is possible to measure how much of the unit's activation comes from perceptual grounding and how much from other abstract units. Portions of activation may then be assigned a 'depth' (distance traveled from the sensory source). It may be used e.g. to inhibit propagation of signals too far from their perceptual 'anchor'. This allowed fast and correct context detections without strong activation suppression, permanent saturation or chaotic instability from outdated, 'abstract' signals. The details of the 'depth' mechanism are presented in the next section.

3.2 Proposed model

This section specifies the Social Context Cognition Model (SCCM). It includes the general architecture description, the detailed data processing and the formal correctness proof of the presented saturation problem solution.

3.2.1 General architecture

SCCM consists of mediating units implemented as neurons, organized in clusters, grouped in four layers (Figure 3.5).

Input layer (I). Observed objects, locations, a behavior of agents, etc. lead to social prototype activations, depending on the simulation. Perception of such elements is represented by this layer.

Feature layer (X). More abstract (and simulation independent) context features are recognized in this layer. They are grouped in three sets, containing: social roles (X_r), groups (X_g) and situations (X_s). X_r consists of power and warmth (general role indicators) and role-specific symbols (e.g. age, formal position, etc.). Likewise, X_g and X_s contain general group indicators (competence, amiability), situation indicators (valence, authority presence), specific group-symbols (uniform, flag, etc.) and situation-symbols (time, place, objects, etc.).

CHAPTER 3. PROPOSED MODEL

Prototype classification layer (Y). Actual detection of social prototypes is represented in this key layer. It consists of three clusters of mediating units: role classification of agents ($\mathbf{Yr}$), total situation classification ($\mathbf{Ys}$), and social group classification ($\mathbf{Yg}$). Activations in this layer represent agent's hypothesis that given prototype describes the context.

Output layer (Z). Social context effect on the cognition and the behavioral disposition is represented in the output layer. Each unit corresponds to a distinct behavioral entity like personality state manifestation, plan, or norm feasibility (see Figure 3.1).

The connections between the units represent associative influences (Figure 2). Observation inputs use feature activating connections ($\mathbf{Ir}$, $\mathbf{Ig}$, $\mathbf{Is}$) to excite the feature layer, that in turn activate fitting prototypes, via feature to prototype connections ($\mathbf{Wr}$, $\mathbf{Wg}$, $\mathbf{Ws}$). Associations among the role, situation, and group prototypes (e.g. 'lesson' situation may have an excitatory influence on a 'teacher' role but inhibitory on a 'clown') use inter-prototype class connections ($\mathbf{Vrg}$, $\mathbf{Vgr}$, $\mathbf{Vrs}$, $\mathbf{Vsr}$, $\mathbf{Vsg}$, $\mathbf{Vgs}$). Finally, the activated prototypes introduce related personality state via disposition changing connections ($\mathbf{Pr}$, $\mathbf{Pg}$, $\mathbf{Ps}$). Each of the N potentially observed agents corresponds to a separate instance of role ($\mathbf{Yr}$) and role feature ($\mathbf{Xr}$) clusters, while group and situation features and prototypes describe total context uniquely. This causes the top-down and bottom-up flow of circular, global-local exchange between the situation/group ($\mathbf{Ys/Yg}$) and role ($\mathbf{Yr}$) perceptions known in some cognition models [MR81].

3.2.2 'Depth' mechanism

In a general case, saturation problem addressed in section 3.1.2 is a complex system-dynamics issue. However, specific domain allows SCCM to use a limiting unit activation method, eliminating instability and permanent saturation based on the tracing of the signal's distance from the perceptual grounding. As postulated by the grounded cognition [Bar08], all internal representations including social prototypes, are anchored in

CHAPTER 3. PROPOSED MODEL

Figure 3.5: The SCCM general structure

perceptual data. As the activation spreads from perception to social units, it is, therefore, possible to measure the amount of the unit's activation from its perceptual grounding (as opposed to excitation by other units). Activation portions may then be assigned a 'depth' (distance it traveled from the perceptual source) used to modulate them e.g. to stop propagating those too far from the perceptual 'anchor'. This allows to combine fast and correct context detection without strong activation suppression (see above) while eliminating permanent saturation and chaotic instability by limiting propagation of the signals that are too outdated and 'abstract' (i.e. far from the concrete perceptual source). This is also feasible since 'System 1' non-rationality modeled here operates using direct, associative correlation rather than prolonged deduction chains. Details of the used mechanism are presented in the next section.

3.2.3 Details of data processing

Units in the model differ in their processing, depending on the cluster, but all activation and weight values belong to $[-1, 1]$. Units in a given cluster are referred to with a lower index (e.g. Xg_i represents the activation of the i -th social group prototype). Roles and role features are additionally indexed with agent number (e.g. $Xr_{n,i}$ and $Vr_{n,i}$ are i -th role feature and role prototype (respectively) of the n -th perceived agent). Connections are indexed with source and destination indexes (e.g. $Ws_{i,j}$ is a connection between the

CHAPTER 3. PROPOSED MODEL

i-th situation feature and the j-th situation prototype). An additional index in roles and role feature connections indicates an agent number (e.g. $Wr_{n,i,j}$ refers to a connection between i-th role feature and j-th role prototype, for the n-th agent) (see Table 1).

Prototype potentials consist of D 'depths'. The perceptual inputs (i.e. feature classification) are included in zeroth potential 'depth', and each propagation jump causes the signal to be stored one 'depth' higher. The d-th 'depth' ($0 \leq d < D$) of the potential is referred to as $Yr_{n,j}^d$, Yg_j^d and Ys_j^d , similarly potential changes due to association, inhibition, and energy release have 'depths' e.g. $\Delta_a Yg_n^d$. As mentioned, when a portion of the potential spreads to other units, it increases 'depth', and 'depth' D potentials are no longer propagated. Value D then effectively sets a limit to influence for a given portion of perception. This simple solution effectively allows building coherent situation interpretations through association, while eliminating the boundless, exponential and circular self-perpetuating effect without tedious and ultimately ineffective parameter tuning (see 4.4).

3.2.3.1 Feature extraction processing

The input layer is a point of contact with the external world, and its structure depends completely on the specific simulation environment. Similarly, while a list of relevant prototypes or features may be selected relatively stable for a given problem (e.g. 'defender' or 'victim' as roles in a bullying problem), their specific activation method relies on the simulation frame. Therefore, unlike with the prototype and personality state layers, it is impossible to prescribe them in a general case, but the next section shows how this may be done for a particular domain.

3.2.3.2 Feature classification

Prototype layer is a central processing element in SCCM. It is the most complex part of SCCM and consists of the feature classification (using Wg , Ws , and Wr), the associative

CHAPTER 3. PROPOSED MODEL

Table 3.1: Symbol table

Mediating units	
I_j	j-th input element
Xs_j	j-th feature of current situation
Xg_j	j-th feature of current group
$Xr_{n,j}$	j-th feature of n-th perceived agent
Ys_j^l	l-th level of j-th social situation detection
Yg_j^l	l-th level of j-th social group detection
$Yr_{n,j}^l$	l-th level of j-th social role attribution to n-th agent
Z_j	j-th personality state dimension
Links to prototype features	
$Ir_{n,i,j}$	i-th input's link to n-th agent's j-th feature
$Is_{i,j}$	i-th input's link to j-th situation feature
$Ig_{i,j}$	i-th input's link to j-th group feature
Links to prototypes	
$Ws_{i,j}$	i-th situation feature's link to j-th situation
$Wg_{i,j}$	i-th group feature's link to j-th group
$Wr_{n,i,j}$	i-th role feature's link to j-th role for n-th agent
Prototype associations	
$Vsg_{i,j}$	i-th situation's link to j-th group
$Vgs_{i,j}$	i-th group's link to j-th situation
$Vsr_{n,i,j}$	i-th situation's link to n-th agent's j-th role
$Vrs_{n,i,j}$	n-th agent's i-th role link to j-th situation
$Vgr_{n,i,j}$	i-th group's link to n-th agent's j-th role
$Vrg_{n,i,j}$	n-th agent's i-th role link to j-th group
Feature based prototype detection parameters	
$Hs_{i,j}$	i-th dimension of j-th situation prototype's center
$Ds_{i,j}$	i-th dimension of j-th situation prototype's deviation
$Hg_{i,j}$	i-th dimension of j-th group prototype's center
$Dg_{i,j}$	i-th dimension of j-th group prototype's deviation
$Hr_{n,i,j}$	n-th agent's i-th dimension of j-th role's center
$Dr_{n,i,j}$	n-th agent's i-th dimension of j-th role's deviation
Simulation specific step sizes	
s_c, s_a	simulation specific step sizes for classification (s_c),
s_i, s_r	association (s_a), inhibition (s_i), release (s_r) and
s_p	personality change(s_p) phases

CHAPTER 3. PROPOSED MODEL

unit excitation from other clusters (using V_{rg} , V_{gr} , V_{rs} , V_{sr} , V_{sg} , V_{gs}), the competitive inhibition within the own cluster and the energy decay.

In the feature classification phase, prototype units act similarly to radial basis neurons. The Gaussian function

$$\rho(\|\mathbf{x}\|) = \exp(-\|\mathbf{x} - \mathbf{c}\|), \quad (\text{Eq. 3.2})$$

that is often used with the Euclidean norm, creates a symmetrical Gaussian 'bell' curve. Instead, SCCM units are not only parametrized by the Gaussian curve center (**cent**), but also by it's spread over the input dimensions (**dist**). This allows defining prototypes not only in terms of feature medians but also their standard deviations, shaping the Gaussian curve accordingly. Resulting multidimensional Gaussian function takes the following general form:

$$\rho(x) = \exp\left(-\sum_{i=1}^n \frac{(x_i - \text{cent}_i)^2}{2 * \text{dist}_i^2}\right). \quad (\text{Eq. 3.3})$$

This gives additional flexibility and directly corresponds to the typical available psychological data (see section 5.2.2.2 and tables 5.1, 5.2 for concrete, literature based examples). Based on this general form, *feature classification based role activation change* is given by

$$\Delta_c Y_{r_{n,j}}(t) = \exp\left(-\sum_{i=1}^{|X_r|} \frac{(X_{r_{n,i}}(t-1) - H_{r_{n,i,j}})^2}{2D_{r_{n,i,j}}^2}\right), \quad (\text{Eq. 3.4})$$

feature classification based situation activation change is equal to

$$\Delta_c Y_{s_j}(t) = \exp\left(-\sum_{i=1}^{|X_s|} \frac{(X_{s_i}(t-1) - H_{s_{i,j}})^2}{2D_{s_{i,j}}^2}\right), \quad (\text{Eq. 3.5})$$

and finally, *feature classification based group activation change* is equal to

$$\Delta_c Y_{g_j}(t) = \exp\left(-\sum_{i=1}^{|X_g|} \frac{(X_{g_i}(t-1) - H_{g_{i,j}})^2}{2D_{g_{i,j}}^2}\right), \quad (\text{Eq. 3.6})$$

CHAPTER 3. PROPOSED MODEL

where $\mathbf{Hr}_{n,i,j}$, $\mathbf{Hs}_{i,j}$, $\mathbf{Hg}_{i,j}$, $\mathbf{Dr}_{n,i,j}$, $\mathbf{Ds}_{i,j}$ and $\mathbf{Dg}_{i,j}$ are the respective Gaussian centers and standard deviations in each of the feature space dimensions based on the psychological data. This change is incorporated into the unit activations:

$$Yr_{n,j}(t) = (1 - s_c) * Yr_{n,j}(t - 1) + s_c * \Delta_c Yr_{n,j}(t), \quad (\text{Eq. 3.7})$$

$$Ys_j(t) = (1 - s_c) * Ys_j(t - 1) + s_c * \Delta_c Ys_j(t), \quad (\text{Eq. 3.8})$$

$$Yg_j(t) = (1 - s_c) * Yg_j(t - 1) + s_c * \Delta_c Yg_j(t), \quad (\text{Eq. 3.9})$$

where $s_c \in (0, 1)$ is a classification based change step size. If the 'depth' mechanism is used, the feature classification component instead constitutes the input (zeroth) 'depth' of activations:

$$Yr_{n,j}^0(t) = (1 - s_c) * Yr_{n,j}^0(t - 1) + s_c * \Delta_c Yr_{n,j}(t), \quad (\text{Eq. 3.10})$$

$$Ys_j^0(t) = (1 - s_c) * Ys_j^0(t - 1) + s_c * \Delta_c Ys_j(t), \quad (\text{Eq. 3.11})$$

$$Yg_j^0(t) = (1 - s_c) * Yg_j^0(t - 1) + s_c * \Delta_c Yg_j(t). \quad (\text{Eq. 3.12})$$

3.2.3.3 Prototype association

Prototype association and inhibition form a recurrent part of the SCCM network. While they are activated by the feature classification, prototype units also influence each other, based on their connections i.e. units excite those positively associated with them and inhibit negatively associated ones. This reflects the conditional probabilities

CHAPTER 3. PROPOSED MODEL

encoded in the social context perception (for example, some social groups or situations are perceived as less probable when certain roles are present).

If the depth algorithm is not used, the entire unit activation is influenced by the whole activation of all associated units. *Association based role activation change* equals

$$\Delta_a Y_{r_{n,j}}(t) = \frac{\sum_{i=1}^{|Y_g|} (V_{gr_{n,i,j}} * Y_{g_i}(t-1)) + \sum_{i=1}^{|Y_s|} (V_{sr_{n,i,j}} * Y_{s_i}(t-1))}{\ln Y_{r_{n,j}}}, \quad (\text{Eq. 3.13})$$

association based scene activation change is equal to

$$\Delta_a Y_{s_j}(t) = \frac{\sum_{i=1}^{|Y_g|} (V_{gs_{i,j}} * Y_{g_i}(t-1)) + \sum_{n=1}^N \sum_{i=1}^{|Y_{r_n}|} (V_{rs_{n,i,j}} * Y_{r_{n,i}}(t-1))}{\ln Y_{s_j}}, \quad (\text{Eq. 3.14})$$

and *association based group activation change* is equal to

$$\Delta_a Y_{g_j}(t) = \frac{\sum_{i=1}^{|Y_s|} (V_{sg_{i,j}} * Y_{s_i}(t-1)) + \sum_{n=1}^N \sum_{i=1}^{|Y_{r_n}|} (V_{rg_{n,i,j}} * Y_{r_{n,i}}(t-1))}{\ln Y_{g_j}} \quad (\text{Eq. 3.15})$$

The incoming connection sums are defined as:

$$\ln Y_{r_{n,j}} = \sum_{i=1}^{|Y_g|} |V_{gr_{n,i,j}}| + \sum_{i=1}^{|Y_s|} |V_{sr_{n,i,j}}|. \quad (\text{Eq. 3.16})$$

$$\ln Y_{s_j} = \sum_{i=1}^{|Y_g|} |V_{gs_{i,j}}| + \sum_{n=1}^N \sum_{i=1}^{|Y_{r_n}|} |V_{rs_{n,i,j}}|, \quad (\text{Eq. 3.17})$$

$$\ln Y_{g_j} = \sum_{i=1}^{|Y_s|} |V_{sg_{i,j}}| + \sum_{n=1}^N \sum_{i=1}^{|Y_{r_n}|} |V_{rg_{n,i,j}}|. \quad (\text{Eq. 3.18})$$

CHAPTER 3. PROPOSED MODEL

Based on this, activation values are updated before the next phase:

$$Y_{r_{n,j}}(t) = (1 - s_a) * Y_{r_{n,j}}(t - 1) + s_a * \Delta_a Y_{r_{n,j}}(t), \quad (\text{Eq. 3.19})$$

$$Y_{s_j}(t) = (1 - s_a) * Y_{s_j}(t - 1) + s_a * \Delta_a Y_{s_j}(t), \quad (\text{Eq. 3.20})$$

$$Y_{g_j}(t) = (1 - s_a) * Y_{g_j}(t - 1) + s_a * \Delta_a Y_{g_j}(t), \quad (\text{Eq. 3.21})$$

where $s_a \in (0, 1)$ is a simulation specific, association based, change step size. If 'depth' mechanisms are used, each activation 'depth' for each of the prototype types is calculated separately:

$$\Delta_a Y_{r_{n,j}}^d(t) = \frac{\sum_{i=1}^{|Y_g|} (V_{gr_{n,i,j}} * Y_{g_i}^{d-1}(t - 1))}{\ln Y_{r_{n,j}}} + \frac{\sum_{i=1}^{|Y_s|} (V_{sr_{n,i,j}} * Y_{s_i}^{d-1}(t - 1))}{\ln Y_{r_{n,j}}}. \quad (\text{Eq. 3.22})$$

$$\Delta_a Y_{s_j}^d(t) = \frac{\sum_{i=1}^{|Y_g|} (V_{gs_{i,j}} * Y_{g_i}^{d-1}(t - 1))}{\ln Y_{s_j}} + \frac{\sum_{n=1}^N \sum_{i=1}^{|Y_r|} (V_{rs_{n,i,j}} * Y_{r_{n,i}}^{d-1}(t - 1))}{\ln Y_{s_j}}, \quad (\text{Eq. 3.23})$$

$$\Delta_a Y_{g_j}^d(t) = \frac{\sum_{i=1}^{|Y_s|} (V_{sg_{i,j}} * Y_{s_i}^{d-1}(t - 1))}{\ln Y_{g_j}} + \frac{\sum_{n=1}^N \sum_{i=1}^{|Y_r|} (V_{rg_{n,i,j}} * Y_{r_{n,i}}^{d-1}(t - 1))}{\ln Y_{g_j}}. \quad (\text{Eq. 3.24})$$

where $0 < d < D$. D-th activation 'depth' is only influenced by the d-1 'depth' of the associated neighbor, as the activation changes are computed as follows:

CHAPTER 3. PROPOSED MODEL

$$Yr_{n,j}^d(t) = (1 - s_a) * Yr_{n,j}^d(t-1) + s_a * \Delta_a Yr_{n,j}^d(t), \quad (\text{Eq. 3.25})$$

$$Ys_j^d(t) = (1 - s_a) * Ys_j^d(t-1) + s_a * \Delta_a Ys_j^d(t), \quad (\text{Eq. 3.26})$$

$$Yg_j^d(t) = (1 - s_a) * Yg_j^d(t-1) + s_a * \Delta_a Yg_j^d(t), \quad (\text{Eq. 3.27})$$

3.2.3.4 Prototype inhibition

Similarly to the association phase described above, in-class inhibition without 'depth', mechanism activation changes due to inhibition are given by

$$\Delta_i Yr_{n,j}(t) = \sum_{i=1}^{|Yr|} (V_{irr_{n,i,j}} * Yr_{n,i}(t-1)), \quad (\text{Eq. 3.28})$$

$$\Delta_i Ys_{n,j}(t) = \sum_{i=1}^{|Ys|} (V_{ss_{n,i,j}} * Ys_{n,i}(t-1)), \quad (\text{Eq. 3.29})$$

$$\Delta_i Yg_{n,j}(t) = \sum_{i=1}^{|Yg|} (V_{gg_{n,i,j}} * Yg_{n,i}(t-1)) \quad (\text{Eq. 3.30})$$

and the updated activation values equal

$$Yr_{n,j}(t) = (1 - s_i) * Yr_{n,j} + s_i * \Delta_i Yr_{n,j}(t), \quad (\text{Eq. 3.31})$$

$$Ys_j(t) = (1 - s_i) * Ys_j + s_i * \Delta_i Ys_j(t), \quad (\text{Eq. 3.32})$$

$$Yg_j(t) = (1 - s_i) * Yg_j + s_i * \Delta_i Yg_j(t). \quad (\text{Eq. 3.33})$$

CHAPTER 3. PROPOSED MODEL

where $s_i \in (0, 1)$ is a simulation specific, inhibition based, change step size. In a version with 'depth' mechanism, activation changes are equal to

$$\Delta_i Y_{r,n,j}^d(t) = \sum_{i=1}^{|Y_r|} (V_{rr_{n,i,j}} * Y_{r,n,i}^{d-1}(t-1)), \quad (\text{Eq. 3.34})$$

$$\Delta_i Y_{s,j}^d(t) = \sum_{i=1}^{|Y_s|} (V_{ss_{i,j}} * Y_{s,i}^{d-1}(t-1)), \quad (\text{Eq. 3.35})$$

$$\Delta_i Y_{g,j}^d(t) = \sum_{i=1}^{|Y_g|} (V_{gg_{i,j}} * Y_{g,i}^{d-1}(t-1)) \quad (\text{Eq. 3.36})$$

and the updated activations are

$$Y_{r,n,j}^d(t) = (1 - s_i) * Y_{r,n,j} + s_i * \Delta_i Y_{r,n,j}^d(t), \quad (\text{Eq. 3.37})$$

$$Y_{s,j}^d(t) = (1 - s_i) * Y_{s,j} + s_i * \Delta_i Y_{s,j}^d(t), \quad (\text{Eq. 3.38})$$

$$Y_{g,j}^d(t) = (1 - s_i) * Y_{g,j} + s_i * \Delta_i Y_{g,j}^d(t). \quad (\text{Eq. 3.39})$$

3.2.3.5 Energy release

The last component of the prototype layer processing is the energy decay process, ensuring a 'cool down' of the prototypes after the activating conditions pass. Without the 'depth' mechanism energy release is equal to

$$Y_{r,n,j}(t) = (1 - s_r) * Y_{r,n,j}(t-1), \quad (\text{Eq. 3.40})$$

$$Y_{s',j}(t) = (1 - s_r) * Y_{s,j}(t-1), \quad (\text{Eq. 3.41})$$

CHAPTER 3. PROPOSED MODEL

$$Yg'_j(t) = (1 - s_r) * \max(0, Yg_j(t - 1)), \quad (\text{Eq. 3.42})$$

where $s_r \in (0, 1)$ is the energy release parameter. When the 'depth' mechanism is used, the formulas take the following form:

$$Yr_{n,j}^d(t) = (1 - s_r) * Yr_{n,j}^d(t), \quad (\text{Eq. 3.43})$$

$$Ys_j^d(t) = (1 - s_r) * Ys_j^d(t), \quad (\text{Eq. 3.44})$$

$$Yg_j^d(t) = (1 - s_r) * Yg_j^d(t), \quad (\text{Eq. 3.45})$$

where $0 \leq d < D$.

3.2.3.6 Output layer

The effect social context cognition established in the prototype layer has on the remaining parts of the agent, is represented by the simulation-specific output layer. It consists of N independent units ($Z_i, i \leq N$), corresponding to distinct needed cognitive and behavioral entities (Chapter 5 proposes an example using five personality states, to contextualize behavior). Those output units are activated by the prototype layer. Change in their activation is calculated as

$$\Delta Z_j(t) = \frac{\sum_{i=1}^{|Ys|} Ps_{i,j} * Ys_i(t)}{\text{InP}_j} + \frac{\sum_{i=1}^{|Yg|} Pg_{i,j} * Yg_i(t)}{\text{InP}_j} + \frac{\sum_{n=1}^N \sum_{i=1}^{|Yr|} Pr_{n,i,j} * Yr_{n,i}(t)}{\text{InP}_j}, \quad (\text{Eq. 3.46})$$

CHAPTER 3. 'DEPTH' MECHANISM, CORRECTNESS PROOF

where

$$\text{InP}_j = \sum_{i=1}^{|Ys|} |\text{Ps}_{i,j}| + \sum_{i=1}^{|Yg|} |\text{Pg}_{i,j}| + \sum_{n=1}^N \sum_{i=1}^{|Yr|} |\text{Pr}_{n,i,j}|.$$

If the 'depth' mechanism is used, all prototype activation 'depths' must first be integrated into a complete activation value, that will influence the output layer:

$$Yr_{i,n}(t) = \sum_{d=0}^{D-1} Yr_{i,n}^d(t) \quad (\text{Eq. 3.47})$$

$$Ys_i(t) = \sum_{d=0}^{D-1} Ys_i^d(t) \quad (\text{Eq. 3.48})$$

$$Yg_i(t) = \sum_{d=0}^{D-1} Yg_i^d(t) \quad (\text{Eq. 3.49})$$

Whether or not the 'depth' mechanism is used, output units are updated according to

$$Z_j(t) = (1 - s_p) * Z_j(t-1) + s_p * \Delta Z_j(t), \quad (\text{Eq. 3.50})$$

where $s_p \in (0, 1)$ is the output layer change step size.

3.3 'Depth' mechanism, correctness proof

Let G be any prototype group i.e. $G = Yg \cup Yr \cup Ys$. For simplicity, the activation 'depth' d of unit $g \in G$, at the time t , is denoted as $g^d(t)$ and its total activation as $g(t)$ ($g^d(t)$ and $g(t)$ are respectively equal to $Yr_{n,j}^d(t)$ and $Yr_{n,j}(t)$ (if $g \in Yr$), $Yg_j^d(t)$ and $Yg_j(t)$ (if $g \in Yg$) or $Ys_j^d(t)$ and $Ys_j(t)$ (if $g \in Ys$), for certain j , n irrelevant for the proof).

CHAPTER 3. 'DEPTH' MECHANISM, CORRECTNESS PROOF

Let max. potential in G , at the time t , be

$$G(t) = \text{MAX}_{g \in G}(|g(t)|) \quad (\text{Eq. 3.51})$$

and max. input to G (from feature classification):

$$G_{in}(t) = \text{MAX}_{g \in G}(g^0(t)). \quad (\text{Eq. 3.52})$$

Let $G^d \subset G$ denote units with non-decaying d -th 'depth' activation i.e.

$$G^d = (g \in G: g^d(t) \not\rightarrow 0, \text{ when } t \rightarrow \infty) \quad (\text{Eq. 3.53})$$

With those definitions, we now formalize Lemma 1.

Lemma 1

For a unit group G , any positive s_c, s_a, s_i parameters solve the saturation problem i.e.

$\forall_G s_c > 0, s_a > 0, s_i > 0 \Rightarrow G$ has no saturation problem, i.e. $G(t) \rightarrow 0$, if $G_{in}(t) \rightarrow 0$.

Let opposite be true: for any $s_c > 0, s_a > 0, s_i > 0$, there exists G , with a decaying external stimulation, containing non-decaying unit activations, i.e

$$\exists_G: G_{in}(t) \rightarrow 0, G(t) \not\rightarrow 0. \quad (\text{Eq. 3.54})$$

If so, at least one non-empty G^d set exists, i.e

$$G(t) \not\rightarrow 0 \Rightarrow \exists_d: G^d \neq \emptyset \quad (\text{Eq. 3.55})$$

Let d_0 be the lowest d satisfying the above i.e.

$$G^{d_0} \neq \emptyset \text{ and } G^d \neq \emptyset \Rightarrow d_0 \leq d \quad (\text{Eq. 3.56})$$

Let $g \in G^{d_0} \Rightarrow g(t)^{d_0} \not\rightarrow 0$. Two options exist:

a) $d_0 = 0$, i.e. 'depth' is stimulated externally, by perception, then,

CHAPTER 3. 'DEPTH' MECHANISM, CORRECTNESS PROOF

$$g^0(t) \not\rightarrow 0. \quad (\text{Eq. 3.57})$$

Based on (4), (6) or (8), $g^0(t)$ is updated with:

$$g^0(t) = (1 - s_c) * g^0(t - 1) + s_c * \Delta_c g(t), \quad (\text{Eq. 3.58})$$

but $G_{in}(t) \rightarrow 0$, so $\Delta_c g(t) \rightarrow 0$, contradicting $g^0(t) \not\rightarrow 0$.

b) $d_0 > 0$, i.e. 'depth' is stimulated by other unit's 'depth' d_0-1 . Then $g^d(t)$ update is based on (11) and (23), (13) and (24) or (15) and (25). The association update is equal to:

$$g^{d_0}(t) = (1-s_a) * g^{d_0}(t - 1) + s_a * \Delta_a g^{d_0}(t), \quad (\text{Eq. 3.59})$$

while the inhibition update takes the following form:

$$g^{d_0}(t) = (1-s_i) * g^{d_0}(t - 1) + s_i * \Delta_i g^{d_0}(t). \quad (\text{Eq. 3.60})$$

$\Delta_a g(t)$ is given by (12), (14) or (16) and is equivalent to:

$$\Delta_a g^{d_0}(t) = \frac{\sum_{i=1}^N (x_i * g_i^{d_0-1}(t - 1))}{X}, \quad (\text{Eq. 3.61})$$

where N is a number of units connected to g , g_i are their activation values and x_i are connection weights. The same formula holds for $\Delta_i g(t)$, given by (20), (21) or (22). But $d_0 - 1 < d_0 \Rightarrow g_i^{d_0-1} \rightarrow 0$, so $\Delta_a g^{d_0}(t) \rightarrow 0$ and $\Delta_i g^{d_0}(t) \rightarrow 0$, and therefore $g^{d_0}(t) \rightarrow 0$, contradicting $g^{d_0}(t) \not\rightarrow 0$ and proving the lemma.

3.4 Example setup and output presentation

At least three methods to generate data for the proposed model are possible: a) using research-based, psychological data, b) using automatic model parametrization, and c) using manual parametrization. While the first two methods, suitable for social simulations and similar applications demanding objective accuracy and validation, are addressed in other parts of this work, this chapter uses the manual method, feasible for entertainment (e.g. game characters) and other applications focusing on subjective believability rather than objective accuracy.

To depict fundamental advantage of the presented model over typical 'vector feature' personalities, a choice to implement a multisided character was made. It requires dramatic yet believable and understandable social context interpretation changes, underlying culture-personality process, producing a credible, consistent and at the same time surprising character, in spite of its general simplicity. Creation of an artificial, believable character is a creative process. Following sections describe how it may take a more systematic shape in the case of SCCM, starting with an initial character profile.

3.4.1 SCCM creation and parametrization

As a part of the model evaluation, several characters and scenarios allowing them to manifest multi-sided context cognition and personalities were prepared. The selection of relevant perception elements (Table 3.2) was relatively easy but establishing prototypes and their connections was more challenging. Each character included a selection of 30 unique prototype units. Initial network was created from intuitive relations like 'being a part of a religious group makes the role of a devotee far more probable' or 'military uniform and American nationality activate archetype of oppressor for the perceived agent', and translated into connections with intuitive initial values. At this point, a precise 'personality tuning' started. In the prepared environment, changes in connection weights

CHAPTER 3. EXAMPLE SETUP AND OUTPUT PRESENTATION

Table 3.2: SCCN example input elements

Situation element	Used input perception units
ethnicity	Arab, American, European
religion	Muslim, Catholic, Jew, unknown
clothes	eastern, western, US uniform
weaponry	unarmed, gun, knife, uncertain
attitude	hostile-friendly, serious-joking
position	dominant vs. weaker
relation history	none, giver, taker
location	'mine'-'their', home, mosque
relevant objects	dangerous, sacred, precious
relevant groups	friends, devotees, soldiers
relevant people	family member, imam, boss

resulted in real-time generation of behavioral outcomes, showing both global SCCM state (e.g. Figures 3.7 - 3.12) and specific HU activation history, greatly supporting creative artist in the design process. Even so, the process required time and interpretation of the consequences of weight changes for output activations.

3.4.2 Example character profile and scenario

To achieve a goal of multisided character creation, cultural, religious, professional, and private aspects was envisioned for each of the stereotypic characters.

3.4.2.1 Character

The example character presented here, was an agent who is a sociable, benevolent, introvert with friends, somewhat serious and stiff in a professional environment and when his religious values or honor were perceived as threatened, he became an aggressive individual. This way initial sketch of 'Ahmed the friendly jihadist' was created. Character profile implied religious, private and professional roles, situations or groups (Figure 3.6) to dynamically adjust his mood (Figure 3.13) based on the cultural specification and the current context.

CHAPTER 3. EXAMPLE SETUP AND OUTPUT PRESENTATION

Figure 3.6: Ahmed’s mediating units

CHAPTER 3. EXAMPLE SETUP AND OUTPUT PRESENTATION

3.4.2.2 Scenarios

After the process of the character creation, several scenarios have been developed. Two examples are briefly presented here. In those examples, similar events led to extremely different SCCM reaction due to the reaction of Ahmeds context cognition to subtle differences in cultural context settings. In both cases, Ahmed encounters an American, who is first joking and friendly but later his jokes are perceived as unfriendly. The cultural context leads to dramatically different CSSM activation.

Scenario 'Friendly visit': Ahmed is in his house with friends. As seen in Figure 3.7, he promptly took a role of a host/friend and recognized a group style as 'family'. The situation is private, relaxed (Figure 3.13). The important values are hospitality, respect for life, and entertainment. American appeared wearing western civilian clothes and due to his outlook, friendly behavior, and SCCM state Ahmed attributed him a friends role in T2 (Figure 3.9). As seen in T3 offensive joke did not manage to trigger negative HUs. Ahmed's was less friendly, but the situation remained calm (Figure 3.11).

Scenario 'Mosque': Located in a mosque with an Imam, a group of devotees and sacred objects in T1 Ahmed took a role of a Muslim in a sacred situation. Dominant values were religious ones (Figure 3.8). Americans behavior did not change, but he was wearing army uniform and was attributed soldiers role. Due to sacred context and soldiers presence situation became serious in T2 (Figure 3.10). After Americans unfriendly joke Ahmed attributed him a role of an oppressor, infidel, and enemy. Religious values, honor, and norms of punishment for crime dominated, while respect for life and norm of not harming subsided (Figure 3.12) and affect shifted from stress to anger between T2 and T3 (Figure 3.14).

3.4.2.3 Believability, complexity and stereotypical characters

The character is on one hand stereotypical in order to make presentation clear and compelling and on the other hand, complex enough to cause the emergence of multifaceted

CHAPTER 3. EXAMPLE SETUP AND OUTPUT PRESENTATION

Figure 3.7: Ahmed's prototype activation in the 'Friendly visit' scenario, t1

Figure 3.8: Ahmed's prototype activation in the 'Mosque' scenario, t1

CHAPTER 3. EXAMPLE SETUP AND OUTPUT PRESENTATION

Figure 3.9: Ahmed's prototype activation in the 'Friendly visit' scenario, t2

Figure 3.10: Ahmed's prototype activation in the 'Mosque' scenario, t2

CHAPTER 3. EXAMPLE SETUP AND OUTPUT PRESENTATION

Figure 3.11: Ahmed's prototype activation in the 'Friendly visit' scenario, t3

Figure 3.12: Ahmed's prototype activation in the 'Mosque' scenario, t3 (red font signifies a negative activation)

CHAPTER 3. EXAMPLE SETUP AND OUTPUT PRESENTATION

Figure 3.13: SCCM influence on Ahmed's Pleasure-Arousal-Dominance affective space during 'Friendly visit' scenario (compare below).

Figure 3.14: SCCM influence on Ahmed's Pleasure-Arousal-Dominance affective space during 'Mosque' scenario. A and U are moments of American's appearance and becoming unfriendly respectively. Emotional changes are believable and correspond to situational context and SCCM activation

CHAPTER 3. BELIEVABILITY VALIDATION

(context-dependent) and unexpected dynamics. Despite somewhat simplified and caricatural character, observed the relation between religious identity potentially leading to extremist behaviors and introvert or peaceful and friendly identities at work or home, and their observed smooth, contextual transitions and patterns not far from the ones known from the police reports and psychopathological literature. For example, Mikulin-
 cer et al. report that 'avoidant people generally had problems with nurturance (being cold, introverted, or competitive), and anxious people had problems with emotionality (e.g., being overly expressive). These problems seem to underlie insecure individuals self-reported loneliness and social isolation (e.g., 40) and their relatively low relationship satisfaction, more frequent relationship breakups, and more frequent conflicts and violence' [MS12]. Therefore, in this case, believability and stereotypical character of the example were not at odds. Additionally, Chapter 5 contains less stereotypical, more nuanced and data-based parametrization.

3.5 Believability validation

As previously explained, this chapter provides subjective validation of SCCM's believability. It is presented in this section. It contains an introduction and an overview of the validation procedure, description of character and scenario creation, human and SCCM based data acquisition, subjective believability measurement and, finally, its statistical analysis with conclusions.

3.5.1 Introduction

As a part of the validation procedure, artificial SCCM-based agents were created and placed in a series of narratives coded, as simulation scenarios to generate prototype activation changes similar to those presented in section 3.4. Human subjects were presented with psychological profiles of the created characters and scenario descriptions and asked

CHAPTER 3. BELIEVABILITY VALIDATION

to provide the most believable prototype activation for a given character and scenario. Finally, activation outputs obtained from humans and SCCM were randomly presented to human subjects asked to evaluate their believability for a given character profile and the scenarios. This subjective evaluation was then statistically analyzed for differences in believability in the human and SCCM output.

It has been assumed that the CSSM prototypes were well selected and model complexity is sufficient to reflect the expected changes contextual changes. The procedure consisted of the following steps:

- preparation of the character profiles and scenario descriptions (see section 3.4.2 for a specific example)
- hand preparation of the character profiles and scenario descriptions to subjects (see section 3.4.2.1 and 3.4.2.1 for a specific example)
- hand-generation of CSSM model based on the profiles (as presented in Figure 3.6)
- presentation of the character profiles and scenario descriptions to the human subjects and collection of the believable prototype activations from them (similar to figures 3.7-3.12)
- generation of similar activations from the CSSM-based character models
- statistical analysis

The above points are described in more detail in the remaining part of the section.

3.5.2 Characters and simulation scenarios

Five multi-faced characters were hand-designed according to the principles exemplified above. Besides the 'Ahmed, the friendly jihadist', the characters included 'Sam, the

CHAPTER 3. BELIEVABILITY VALIDATION

arrogant soldier', 'Joe, the scarred macho', 'Betsy, the brave mother' and 'John, the insecure achiever'. Each character was captured in a short 'psychological profile' depicting main behavioral and cognitive characteristics in different contexts. Profiles were used to guide the SCCM design for each character, to obtain the human version of the data representing each character's behavior (see below) and in the testing believability of the results. Each created SCCM shared the input format and the number of the prototype units (30) but varied in the specific choice of prototypes and their connections. Moreover, ten narratives were encoded, as simulation scenarios exposing agents to input that was an object of context cognition, similarly to the 'mosque' example described above.

3.5.3 SCCM and human context cognition output acquisition

The above setup resulted in 30 (prototypes) x 5 (agents) x 10 (scenarios) potential required prototype activations, to be obtained from SCCM-based characters and humans. SCCM simulations were performed for all the possible sample combinations. Human data collection included a presentation of randomly selected combinations to twenty subjects with 200 samples gathered in total. Subjects were presented with the character profiles and scenario details and asked to provide what they thought was the most believable activation pattern in the presented case (e.g. given the presented scenario, provide the most believable degree to which John perceived the context as a 'conflict').

3.5.4 Believability evaluation and analysis

After the acquisition of SCCM and human samples, the evaluation process was performed. New group of subjects was presented with agent profiles and scenarios descriptions and asked to quantitatively evaluate the believability of the activation outputs obtained before, by responding to questions like 'from 0 to 5, how believably do you think does this represent Betty's perception of the situation as a 'friendly conversation' in the described

CHAPTER 3. SUMMARY AND CONCLUSIONS

scenario?'. Both SCCM and human generated outputs were measured for their subjective believability.

The obtained quantitative and independent believability data about human (μ_1) and SCCM (μ_2) outputs was subject to two-sample analysis. The initial analysis of human (N= 200, M=4.01, SD=1.30) and SCCM (N = 1500, M=3.79, SD=1.72) showed that pooled methods of analysis are applicable, and their standard deviation was comparable ($1.72/1.30 = 1.32 \leq 2$). The difference in believability evaluation between μ_1 and μ_2 was investigated based on the gathered samples with the null hypothesis that there is a difference between the two groups in the terms of subjective context cognition believability. The mean difference between the samples was 0.18, t-value was 1.7435 (DF = 1698) and standard error of difference was 0.126. The actual difference between SCCM and humans was in -0.4679 to 0.0279 range with 95% confidence. P-value was 0.0814 suggesting that although the believability median was a higher for the human subjects, the hypothesis about the higher believability of human data could be only confirmed with low statistical significance.

Moreover, when the analysis was conducted for the prototypes separately, it was clear that their believability was uneven (N=30, SD = 1.12). The prototypes could be roughly divided into two clusters: the major group of 'believable' prototypes and a minority of the 'unbelievable' ones. As a part of the test and after a brief analysis of the low believability scenario-prototype cases, the connections of the 'unbelievable' prototypes were readopted and the second round of validation took place. For the improved SCCMs the new believability (N= 1500, M=3.96, SD=1.47) was almost equal to the human data.

3.6 Summary and conclusions

In this chapter, Social Context Cognition Model was proposed, as an approach to informal, 'System 1' social cognition modeling capturing idiosyncratic and cultural agent

CHAPTER 3. SUMMARY AND CONCLUSIONS

particularities. The presented model captures subtle, non-linear relations between context components and various shifts in cognition. Connectionists nature and temporal dependencies encoded in the model allow realistic effects like subsymbolic priming with subtle context components causing measurable and believable changes in behavior. Concerning the granularity addressed by the model, the assumed sensory input must be preprocessed into the (expert provided or hand-picked) elements directly meaningful for the modeled context. Example implementation and character creation process were discussed, to show, how such an approach may be used to create agents with multifaceted behavioral and cognitive features through a context sensitive network of social interpretations causing cognitive-affective shifts in response to perceived situations. Moreover, subjective believability tests of hand-made SCCM in entertainment-oriented characters was presented with statistical analysis. Results confirm that model is promising, and statistical comparison of believability with human data shows that 'System 1' context cognition produced by SCCM may potentially be of similar quality to hand-prepared samples. SCCM may be applied to existing agent architectures by linking it to their cognitive and affective elements (Figure 3.1). SCCM may provide psychologically richer and multi-sided characters, realizing complex cultural and individual character profiles and specifications.

Chapter 4

Dramaturgical Dissonance in Human-Like Non-Rationality Modeling

All the world's a stage, And all the men and women
merely players; They have their exits and their
entrances, And one man in his time plays many parts,

As You Like It , William Shakespeare

In the last chapter, we have presented Social Context Cognition Model (SCCM) - a connectionist model capturing agent's informal social associations induced by a perceived situation. SCCM models a part of 'System 1' non-rationality by implementing social context cognition for multi-sided characters responding to changing contexts. Its believability has been experimentally confirmed. However, SCCM does not address all aspects of non-rationality related to social context cognition and some of its features may show weaknesses in certain applications. As a connectionist model, SCCM provides a 'black box' context interpretation and it does not include a formal, step by step processing, for example, of the historical, personalized data about the habitual reactions and expectations to agents. For the applications that rely not only on the final context cognition but also on its partial results and their relations, a non-connectionist, step-by-step context processing is more useful.

CHAPTER 3. SUMMARY AND CONCLUSIONS

This chapter proposes a Dramaturgical Dissonance Model (DDM), as published in [KTT12, KTT15b]. DDM is a component, usable in social situation modeling, inspired by the Dramaturgical [BE05, Gof59] and Cognitive Dissonance [Fes62] theories, that provide a psychological frame for such inferences. While Dramaturgical Theory analyses social interaction structure, Cognitive Dissonance Theory addresses cognitive adaptations and emotional reactions to incoherences in partial beliefs implying different features of the same cognition object, in this case - social context cognition. Model is used to generate example behaviors depicting capability to produce believable behaviors based on the relevant psychological theories. Example scenarios and simulations are provided, to show how habitual, social role and scene patterns could be used to generate emotional responses or plans in a believable way. Case scenarios resulting in a flexible tool providing dramaturgical interpretation of dynamic, social situations, cognitive dissonances within them and their behavioral consequences. Additionally, dramaturgical context interpretation is used as the basis of social affect generation. DDM is easily usable in reasoning, planning or affect generation. Its believability is tested and a comparison between the DDM and SCCM is provided including an analysis of their relative strengths and weaknesses and practical application suggestions.

As DDM is a solution to 'System 1' context cognition problem, it is an alternative to SCCM. DDM shows its advantage in applications where a) dissonances, consonances and other formal relations between partial context cognitions are relevant b) processing personalized to known agents is relevant and substantial individualized data is available c) subjective believability, rather than scientific accuracy is the target, as the non-typical data required by DDM may be harder to acquire than it is a case of SCCM. Conversely, SCCM is better in cases where a 'black box' solution providing social context cognition is sufficient. It also shows greater promise in scientific simulations, as the data required for parametrization and objective validation are more standard, easier to obtain and may even be obtained automatically, as shown in Chapter 6.

CHAPTER 4. INTRODUCTION

The remaining part of this chapter is organized as follows. Section 4.1 links Dramaturgical Theory and Cognitive Dissonance with the problem at hand. Section 4.2 presents the details of the proposed Dramaturgical Dissonance Model. Section 4.3 contains a dramaturgical-dissonance example and the example generated outputs while section 4.4 contains the validation results and comparison with the Social Context Cognition Model presented in the previous chapter. Finally, section 4.5 contains summary and conclusions.

4.1 Introduction

This section addresses the general motivation and design decisions before the Dramaturgical Dissonance Model model is presented. It briefly introduces dramaturgical and dissonance theories and describes the structure of the dramaturgical analysis implemented in the model.

4.1.1 Dramaturgical and dissonance theories

Social experience is contextual according to the Dramaturgical Theory. While its specific content is cultural, its structure is innate. It consists of social roles, scenes etc. representing the inter-subjective state of a 'social game' and the current situational expectations or actions feasibilities. A Dramaturgical Dissonance Model (DDM) is proposed, extending classical agent model with those features. The main components of DDM are 'Role attribution' and 'Scene recognition' processes, continuously interpreting the current dramaturgical situation. The first one assigns plausible roles to agents, while the second holistically interprets the entire scene, using sensory information, defined cultural databases and past subjective data on the tendencies of known agents. This state of the current dramaturgical situation may be used, for example, to generate emotions and moods, alter adequacies of goals or plans etc.

CHAPTER 4. PROPOSED MODEL

4.1.2 Structure of dramaturgical processing

As explained, human behavior and decision making are modulated in the social context cognition - a process modeled in the previous chapter. However, according to the Cognitive Dissonance Theory, relations between partial results in cognition (e.g. inconsistent in partial data) lead to dramatic cognitive effects. In order to model this, various potentially conflicting components must be formally defined in the cognitive process structure for consonances and dissonances to be defined between them. Most basic and habitual level of context cognition is the *uniform role expectancy* for all agents. In the second stage, the uniform, general stereotypical data are personalized, as the habitual reaction is modulated with expectations towards specific familiar agents, based on the past experiences with them in a *historical role expectancy*. Depending on the perceived level of familiarity, personal role-playing tendencies and feasibility are incorporated into the overall dramaturgical expectation. Finally, the established *expectation influences perception*. Mental acts aim to decrease cognitive dissonances (in this case, expectation vs. observation) and therefore are known to alter perception. Those gradual steps of establishing social context are formalized in DDM and will be described in the next section.

4.2 Proposed model

This section specifies the Dramaturgical Dissonance Model (DDM). It includes the Dramaturgical databases, dramaturgical processes and consonances and dissonances between them.

4.2.1 Dramaturgical databases

We will now describe the data used in the model. Dramaturgical correlations between roles, scenes and perception (RS, SR, SS, RR, RPER, SPER) constitute the first database set, second is the current dramaturgical interpretation (CA and CS) and last is the agent

CHAPTER 4. PROPOSED MODEL

Table 4.1: Symbol table

A, R, S, E, G	Known Agents, Roles, Situations, Emotional states, Goals
$WS \in ST$	Current (WS) and possible (ST) world states
SPER, RPER	Perception influence on roles/scenes SPER: $\mathcal{P}(ST) \times S \rightarrow [-1, 1]$ RPER: $\mathcal{P}(ST) \times A \times R \rightarrow [-1, 1]$
SR, RS, RR, SS	Role/scene correlations. RS: $R \times S \rightarrow [-1, 1]$ SS: $S \times S \rightarrow [-1, 1]$, RR: $R \times R \rightarrow [-1, 1]$
CA, CS	Role (CA) and situation (CS) attribution $CA_t: A \times R \rightarrow [-1, 1]$, $CS_t: S \rightarrow [-1, 1]$
F	Agent familiarity in a context. F: $A \times S \rightarrow [0, 1]$
HA	Past role attributions. $HA_t: A \times R \times S \rightarrow [-1, 1]$
AF	Agent-role feasibility. AF: $A \times R \rightarrow [-1, 1]$.
r_f, s_f	Role/scene filters $\in [1, \infty)$
SX	Situational role expectancy. SX: $R \times S \rightarrow [-1, 1]$
TX	Total role expectancy. TX: $R \rightarrow [-1, 1]$
THX	Agent's total historical attribution expectation. THX: $A \times R \rightarrow [-1, 1]$
IRP, ISP	Influenced role/scene perception. IRP: $A \times R \rightarrow [-1, 1]$, ISP: $S \rightarrow [-1, 1]$
MaxA	Maximum role attribution. MaxA: $R \rightarrow [0, 1]$
DSE	Scene expectation/ DSE: $S \rightarrow [-1, 1]$
RC, RD	Role dissonance and consonance RC/RD: $(a, r) \rightarrow [-1, 1]$
SC, SD	Scene dissonance and consonance SC/SD: $(s) \rightarrow [-1, 1]$

history (HA, F) serving as an operational memory of agent dramaturgical tendencies and capabilities. For the sake of simplicity above databases are assumed to be static in the simulation. Relevant symbols are defined in Table 4.1.

The observation of agent features in current world state (WS_t) is linked to its perceived role feasibility. Dramaturgical databases provide rooting of dramaturgy in perception by RPER and SPER. Function values represent levels of correspondence between perception and a role/scene attribution. Example: If 'agent 1' has a uniform and holds a gun $RPER(WS_t, a1, 'policemen')$ may have a high value. Beside relation to perception, databases store relations between dramaturgical elements. Some roles and scenes are perceived as 'natural' together in a given culture, while others not. Also, two scenes or roles may be perceived as positively correlated, neutral or excluding each other. Role

CHAPTER 4. PROPOSED MODEL

to scene implications of feasibility are stored in RS (and opposite in SR), scene to scene implications in SS and role to role is RR. Example: 'teacher' role and a 'lesson' situation are correlated positively. In many cultures, a role of a father and a husband are positively while the professor and boyfriend negatively correlated.

Above elements provide a shape of a social reality used to generate the current dramaturgical interpretation. Its main part consists of current role attributions (CA) and scene recognitions (CS). Agent also maintains operational dramaturgical memory. The perceived familiarity with an agent in a given situation type is stored in F and actual past role attributions in a given scene in HA. Finally, due to features, status etc. agents are perceived as more or less fit to certain roles. This information is stored in AF. Example: in Western culture, a girl of twelve will not be perceived as fit to the role of a 'wife'.

4.2.2 Dramaturgical sub-processes

We will now describe stages of the dramaturgical processes. It consists of two parallel pipelines - one for role attribution and one for scene processing (Figure 4.1). To realize role attribution a general or uniform role expectancy is calculated. Then, experience-based, historical component, specific for the observed agent is integrated into it and expectation influenced perception takes place. Finally, a new agent to role attribution is produced. Scenes are analyzed analogously, with the exception of the agent-specific step that does not apply.

4.2.2.1 Situational expectancy

The role attribution starts with the calculation of role expectancy for a scene detection:

$$SX_{t+1}(r, s) = \underline{CS_t(s)}^{sf} * SR(r, s), \quad (\text{Eq. 4.1})$$

CHAPTER 4. PROPOSED MODEL

Figure 4.1: Details of the dramaturgical process

where s_f is a simulation specific role sensitivity filter and $\underline{X}$ equals X , if $X > 0$ and 0 otherwise. Next, total dramaturgical role expectancy (TX) is calculated by integrating above expectations from all scene detections in the current context:

$$TX_t(r) = \frac{\sum_{s \in S} SX_t(r, s)}{\sum_{s \in S} \|SR(r, s)\|} \quad (\text{Eq. 4.2})$$

4.2.2.2 Historical expectancy

The next step is an individualization of the dramaturgical expectations, based on the past experiences with a given agent. Depending on the perceived level of familiarity, personal role-playing tendencies and feasibility are incorporated into the overall dramaturgical expectation. Example: if agent knows that certain 'teacher' is contradicting typical dramaturgical expectations (e.g. behaves informally during lessons), a role of a 'friend' may be incorporated in expectations of his behavior. X is experience based expectation of agent's role-playing in a scene:

CHAPTER 4. PROPOSED MODEL

$$X_t(a, r, s) = \underline{CS_t(s)}^{s_f} * (HA_t(a, r, s) + AF(a, r)) \quad (\text{Eq. 4.3})$$

THX integrates all scene influences into a historical expectancy of agent-role attribution:

$$THX_t(a, r) = \frac{\sum_{s \in S} F(a, s) * X_t(a, r, s)}{\sum_{s \in S} \|SR(r, s)\|} + \frac{\sum_{s \in S} (1 - F(a, s)) * SX_t(r, s)}{\sum_{s \in S} \|SR(r, s)\|} \quad (\text{Eq. 4.4})$$

Two elements of the formula represent the agent-specific and the uniform expectation components. The smaller subjective agent familiarity (F) in a given context, the less individualized and more stereotypical the expectation.

4.2.2.3 Expectation Influenced Role Perception

In the next step THX is used in forming that expectation influenced role perception (IRP). Mental acts tend to decrease cognitive dissonances (e.g. expectation vs. observation) here by altering perception interpretation. Role-playing classification is influenced by the expectations of the observer to the degree determined by the perception flexibility (p_f), representing individual tendency to adjust perceptions towards expectations:

$$IRP_t(a, r) = (1 - p_f) * RPER(WS_t, a, r) + p_f * THX_t(a, r) \quad (\text{Eq. 4.5})$$

4.2.2.4 Current Role Attribution

Finally, current role attribution CA_t is established through stabilization of the IRP signal, depending on the dramaturgical stability parameter (ds):

$$CA_{t+1}(a, r) = ds * CA_t(a, r) + (1 - ds) * IRP_{t+1}(a, r) \quad (\text{Eq. 4.6})$$

CHAPTER 4. PROPOSED MODEL

4.2.2.5 Scene related processes

In parallel to the role attribution, scene recognition takes place. The first step is calculation of the given role's maximum attribution among current agents:

$$\text{MAXA}_t(r) = \text{MAX}_{a \in A}(\text{CA}_t(a, r)^{r_f}) \quad (\text{Eq. 4.7})$$

It is used in the calculation of the dramaturgical scene expectation:

$$\text{DSE}_t(s) = \frac{\sum_{r \in R} (\text{MAXA}(r)_t * \text{RS}(r, s))}{\sum_{r \in R} \|\text{RS}(r, s)\|} \quad (\text{Eq. 4.8})$$

This dramaturgical feasibility of the scenes is then used to calculate expectation biased scene perception, analogically to how THX is used to calculate IRP:

$$\text{ISP}_t(s) = (1 - p_s) * \text{PerS}(\text{WS}_t, s) + p_s * \text{DSE}_{t+1}(s) \quad (\text{Eq. 4.9})$$

Finally, scene recognition (analogical to CA) is established:

$$\text{CS}_{t+1}(s) = (1 - s_s) * \text{ISP}_{t+1}(s) + s_s * \text{CS}_t(s) \quad (\text{Eq. 4.10})$$

4.2.3 Dissonances and consonances

As mentioned, the above dramaturgical classification signals do not only provide information in themselves. Their dissonances and consonances have consequences for agent's emotional or rational states and actions following them. To express that, role and scene coattributions are defined as:

$$\overline{\text{RR}}_t(a, r, r') = \underline{\text{CA}_t(a, r)^{r_f}} * \underline{\text{CA}_t(a, r')^{r_f}} \quad (\text{Eq. 4.11})$$

CHAPTER 4. PROPOSED MODEL

$$\overline{RS}_t(a, r, s') = \underline{CA}_t(a, r)^{rf} * \underline{CS}_t(s)^{sf} \quad (\text{Eq. 4.12})$$

$$\overline{SS}_t(s, s') = \underline{CS}_t(s)^{sf} * \underline{CS}_t(s')^{sf} \quad (\text{Eq. 4.13})$$

Agent-role attribution dissonance (RD) and consonance (RC) with scenes and roles can be calculated as:

$$RD_t(a, r) = - \sum_{s \in S} \underline{SR}(r, s) * \overline{RS}_t(a, r, s) - \sum_{r' \in R} \underline{RR}(r, r') * \overline{RR}_t(a, r, r') \quad (\text{Eq. 4.14})$$

$$RC_t(a, r) = \sum_{r' \in R} \underline{RR}(r, r') * \overline{RR}_t(a, r, r') + \sum_{s \in S} \underline{SR}(r, s) * \overline{RS}_t(a, r, s) \quad (\text{Eq. 4.15})$$

where the first component of dissonance/consonance refers to the relation with scenes and second with roles. Analogically, scene dissonance (SD) and consonance (SC) are equal to:

$$SD_t(s) = - \sum_{s' \in S} \underline{SS}(s, s') * \overline{SS}_t(s, s') - \sum_{r \in R} \underline{SR}(r, s) * \overline{RS}_t(r, s) \quad (\text{Eq. 4.16})$$

$$SC_t(s) = \sum_{s' \in S} \underline{SS}(s, s') * \overline{SS}_t(s, s') + \sum_{r \in R} \underline{SR}(r, s) * \overline{RS}_t(r, s) \quad (\text{Eq. 4.17})$$

Above values determine if a given dramaturgical element fits well into the rest of current scene. Example: situation classified as 'party', creates dissonance with a perceived serious, 'office-like' role and a consonance with 'relaxed' roles.

Another dissonance occurs between role attribution and dramaturgical expectations:

$$CD(a, r) = \|\underline{CA}_t(a, r) - \underline{TX}_t(r)\| \quad (\text{Eq. 4.18})$$

CHAPTER 4. DRAMATURGICAL-DISSONANCE EXAMPLE

Our final example is a dissonance between role attribution and historical expectations towards given observed agent:

$$HD(a, r) = \|CA_t(a, r) - \frac{\sum_{s \in S} CS_t(s)^{sf} * HA(a, r, s)}{\sum_{s' \in S} CS_t(s')^{sf}}\| \quad (\text{Eq. 4.19})$$

Above functions answer questions: 'To what degree agent violates my general expectations about behavior?' and 'To what degree agent behaves like before?' respectively. Other dramaturgical dissonances and consonances with direct consequences for cognition could be proposed.

4.3 Dramaturgical-dissonance example

To showcase the proposed model, implementation was created with an agent using the DDM to analyze the social situation in a virtual location. The agent was provided with the perception of features of the present agents and objects depending on their relative positions, light etc. In the presented example, the observer perceives three other agents and creates interpretations of the social scene as it unfolds and the perceived features and behaviors change. Changing perception fluidly influences observer's understanding of the occurring social situation and the role of each agent in it. Such approach was assumed to present the model in several distinct contexts using minimal space at the same time validating temporal integrity and continuity of DM's social classification.

4.3.1 Perception in the scene

To provide a perception with a degree of realism and make social interpretation challenging, agent features with dramaturgical consequences (actions, behavior style, uniform, objects held, identity etc.) are only detected at certain distances, also depending on the light level at the object positions. In the presented scenario each perceived agent shares

Table 4.2: Agent features linked to role attributions

	police	gangster	victim	normal	joker
uniform	0.8	-0.5	-0.5	-1	0
casual	0.2	0.1	0.5	0.5	0.3
bandana	-0.8	1	-0.2	-0.4	0
aggression	0.2	1	-1	-0.8	0
laughter	-0.8	-1	-1	0.2	1
confidence	1	1	-1	0.2	0
fear	-0.5	-0.8	1	-0.2	-0.4
gun	1	1	-1	-1	0
knife	-1	1	-1	-1	0
escape	-1	0	1	-0.2	0

a personal history with the observing agent, who subjectively attributes role-playing capability and tendencies based on it. Upon entering the scene observed agent remains anonymous until recognition.

4.3.1.1 Dramaturgical databases

Various observable features of agents and scene were linked to the dramaturgical interpretation of roles and scenes constituting the RPER and SPER functions. As an example, a subset of agent features influencing the value of RPER is presented in Table 4.2. A simple dramaturgical structure used by the perceiving agent in this example consists of five types of social scenes: 'casual walk', 'police intervention', 'gang fight', 'robbery' and a 'funny situation (joke)'. Relevant social roles are 'citizen', 'policemen', 'victim', 'gangster' and a 'joker'. Some dramaturgical correlations (SS, RR) can be found in Table 4.3.

4.3.1.2 Agent History, Familiarity, and simulation specific parameters

Agent familiarity (F) for all agents in our example is very high: 1 for 'casual walk' and 'joke' and 0.7 for other situations, representing observer's strong confidence in his expectations of the behavior of others, even in the rather unusual situations. The specific

CHAPTER 4. DRAMATURGICAL-DISSONANCE EXAMPLE

Table 4.3: Dramaturgical databases SS and RR

SS	intervention	robbery	fight	walk	joke
intervention	0	-0.2	-0.4	-1	-1
robbery	-0.2	0	-0.2	-1	-1
fight	-0.4	-0.2	0	-1	-1
walk	-1	-1	-1	0	0
joke	-1	-1	-1	0	0
RR	policemen	gangster	victim	normal	joker
policemen	0	-1	-0.6	0.2	-0.8
gangster	-1	0	-1	-1	-1
victim	-0.6	-1	0	0.1	-0.8
normal	0.2	-1	0.1	0	0
joker	-0.8	-1	-0.8	0	0

history of the agent historical attributions can be found in Table 4.4 and represents observer’s perception of all agents as generally normal people with agents 1-2 somewhat of ‘jokers’ and agent 3, as a typical ‘victim’ of jokes. General simulation parameters were chosen to fit the assumptions of the example. Since we ignore the problem of perception continuity and assume uninterrupted role-playing perception d_s was assumed 1. To show an influence of mild dramaturgical expectation bias on social perception p_f was assumed 0.3; since dramaturgical roles and scenes are sharply defined s_f and r_f had a value of 1.

4.3.2 Example simulation setup and results

In the presented example, the observer is placed in a partly dark alley, walking towards three initially unidentified agents. In the initial phase ($t=1$ to $t=4$, see Table 4.3, 4.4) observer is not able to perceive any detailed features of agents 1 and 2 but recognizes agent 3 located in a brighter part of the alley as his friend, moving away from the scene possibly running. As the observer moves closer ($t = 5$ to $t = 8$) he perceives behavior of agent 1 as aggressive and his outfit as ‘gangster-like’. Moving even closer ($t = 9$ to $t = 13$) observer notices that both agent 1 and 2 carry weapons (knife and a gun respectively). Agent 2 wears a casual outfit of a normal person but behaves very confidently (‘policemen-like’).

CHAPTER 4. DRAMATURGICAL-DISSONANCE EXAMPLE

Table 4.4: Features perceived by agent at given times

Perception	Value	Start	End	Agent
recognition	1	1	20	3
escaping	1	1	13	3
scarred	0.5	5	13	3
aggressive	1	5	13	1
bandanna	1	5	20	1
holds knife	1	9	20	1
holds gun	1	9	20	2
casual	1	9	20	2
confident	1	9	13	2
recognition	1	14	20	1,2
laughing	1	14	20	1,2

Figure 4.2: Selected scene-related activations

Finally, (t=14 to t=20) agent is close enough to detect that two men are in fact his friends, who start laughing.

Presented example scenario provides perceptions that at various stages suggest different social scenes and roles to the human observer. We analyze how DM responded to them by producing social classifications of varying scenes using dramaturgical knowledge.

In this section, we discuss resulting social context recognition generated by an agent using DM. Figures 4.2 and 4.3 provide generated dramaturgical data that form stable

CHAPTER 4. DRAMATURGICAL-DISSONANCE EXAMPLE

Figure 4.3: Selected total and actor specific role-related activations.

social context classifications easily understandable and intuitively translatable into believable human expectations that could occur in similar settings. We will now analyze them while going through several key dramaturgical points that have emerged.

4.3.2.1 Resulting dramaturgical situation

Initially, agent establishes a stable classification of a rather normal, 'casual' situation (Figure 4.2), despite behavior of Agent 3 (Figure 4.3 and Table 4.5) running from the scene, whose attribution of a 'normal person' role is equaled by his 'victim' classification - a fact somewhat unexpected in the context of the THX values (causing also mild 'victim' role dissonance in Agent 3, see Figure 4.4). At time $t = 5$ observation of the aggressive behavior and gangster-like uniform of Agent 1 leads to a dramatic shift in his role attribution as a probable 'gangster'. This attribution influences global scene detection suggesting a likely 'robbery' and both uniform and agent specific expectations lead to the increased 'victim' attributions in other agents. Another turning point is

CHAPTER 4. DRAMATURGICAL-DISSONANCE EXAMPLE

Table 4.5: Agent historical attributions

Agent 1-2	policemen	gangster	victim	normal	joker
intervention	-1	-1	0.5	1	-1
robbery	-1	-1	0.5	1	-1
gang fight	-1	-1	0.5	1	-1
casual walk	-1	-1	-1	1	0.5
joke	-1	-1	-1	0.3	1
Agent 3	policemen	gangster	victim	normal	joker
intervention	-1	-1	0.5	1	-1
robbery	-1	-1	0.5	1	-1
gang fight	-1	-1	0.5	1	-1
casual walk	-1	-1	-1	1	-0.4
joke	-1	-1	0.5	1	-0.1

t=9, where Agents 1 and 2 are finally perceived clearly enough to notice that they carry weapons. Role attributions change instantly: Agent 1 is perceived as a 'gangster', now with full conviction, however, two strong possibilities for Agent 2 exist ('gangster' and a 'policemen') and create great uncertainty (Agent 2 as 'policeman', Figure 4.4). Consequently, scene detection shifts, as the observer is now certain that some unusual, dangerous scene is taking place, but due to uncertainty in the role attributions cannot decide if it is a 'police intervention', 'gang fight' or (less probably) a 'robbery'. This context understanding remains stable until t = 15 when light change allows detection of identities of Agents 1 and 2 and as their individual history is included in dramaturgical interpretation, THX is personalized and together with the perceived laughter changes the situation classification into a 'joke' and agents into 'jokers'.

4.3.2.2 Resulting dissonances

Due to the space constraints, we have selected some representative dissonances to present the type of useful information they may provide in the social context cognition. Figure 4.4 includes three dissonances related to actor-role attributions (RD), one for each observed actor. Additionally, scene recognition (SD) dissonance related to the 'robbery' is

CHAPTER 4. VALIDATION AND COMPARISON TO SCCM

Figure 4.4: Example scene/role dissonances

presented. It is easy to notice that attribution dissonances appear in points where also human observer could feel that perceived role or scene is contradicting other aspects of the social context. Initial weird behavior of Agent 3 who seems to run away like a 'victim' in a casual situation, slight doubt about if Agent 1 could be 'policemen' instead of a 'gangster' or a dominating uncertainty about the Agent 2 being a 'policemen' or 'gangster' are all understandable and potentially consequential for agent's cognition, affect, and decision-making. As each of the dissonances is caused by specific roles and scenes, dissonance discomfort could be used to e.g. to create a goal to find out if Agent 2 is, in fact, a 'policeman' or a 'gangster' etc.

4.4 Validation and comparison to SCCM

This section contains validation of DDM's subjective believability and its comparison to results of SCCM. The section is divided into an introduction to the procedure, description of character and scenario creation, human and SCCM based output acquisition, subjective believability measurement with its statistical analysis, comparison with SCCM and conclusions about their potential application in different simulation types.

4.4.1 Introduction

The procedure of believability validation is analogous to the one used for SCCM in the previous chapter). As a part of it, five artificial agents (analogous to those used in SCCM validation) were created and placed in a series of simulation scenarios to generate

CHAPTER 4. VALIDATION AND COMPARISON TO SCCM

DDM-based social context cognitions. Example of such scenario execution is presented in section 3.4. As was the case with SCCM, human subjects were presented with profiles of the created characters and scenario descriptions and asked to provide the most believable prototype activation for a given character and scenario. This time, however, personal history of the agent relations was also included in the presentation. Next, activation outputs obtained from humans and SCCM were randomly presented to subjects who were asked to evaluate their believability in the provided character-relations-scenario sets. This subjective evaluation was then statistically analyzed for differences in believability between the human, DDM and SCCM data.

4.4.2 Characters and simulation scenarios

The same five agent profiles, depicting main behavioral and cognitive characteristics in different contexts, that were used with the SCCM were hand-implemented in DDM and coded in the dramaturgical databases. The created characters differed in the internal structure and as before, ten narratives were encoded as simulation scenarios exposing agents to input that was an object of context cognition.

4.4.3 DDM context cognition output acquisition

In the previous chapter, the validation setup resulted in 30 (prototypes) x 5 (agents) x 10 (scenarios) potential required prototype activations, to be obtained from SCCM-based characters and humans. As DDM includes role and situation but not group prototypes, the number of samples related to prototypes was smaller. However, to validate the effect of dissonances and consonances on believability, affective states influenced by them were included in the generated and acquired samples. social context cognition in SCCM and DDM was linked to three affective states (pleasure, arousal, and dominance) that in the case of DDM were also influenced by relevant dissonances/consonances. The total

CHAPTER 4. VALIDATION AND COMPARISON TO SCCM

number of generated, potentially gathered and analyzed samples was 23 (20 prototypes + 3 affective states) $\times$ 5 (agents) $\times$ 10 (scenarios-relations setups).

As in previous validation of SCCM, DDM simulations were performed for all the possible sample combinations, together with the missing 3 $\times$ 5 $\times$ 10 outputs for the SCCM affective state. As the scenarios were expanded with the agent relationship data, human data collection was repeated. Similarly to SCCM validation, it included a presentation of randomly selected combinations to twenty subjects with 200 samples gathered in total. Subjects were presented with the character profiles and scenario details (this time expanded with relationship history) and were asked to provide what they thought was the most believable activation pattern in the presented case.

4.4.4 Believability evaluation and analysis

After the acquisition of the samples from DDM, SCCM, and humans, evaluation was performed. Subjects were presented with the agent profiles and scenarios/relation descriptions and asked to quantitatively evaluate the believability of obtained outputs by responding to questions like 'from 0 to 5, how believably do you think does this represent this agent's arousal activation in the described scenario?'. DDM, SCCM, and human output was measured for subjective believability. The analysis method was similar to the one performed for SCCM, but three independent believability data sets were compared: human (μ_1), SCCM (μ_2) and DDM (μ_3) outputs. To provide a fair comparison with SCCM, the original, unimproved version of SCCM was used for all the agents (see the last paragraph of section 3.5.4).

Again, the comparison of human (N= 200, M=4.09, SD=1.38), DDM (N = 1150, M=3.82, SD=1.67) and SCCM (N = 1150, M=3.89, SD=1.78) standard deviations showed that pooled methods of analysis are applicable. The mean difference between the human and DDM samples was 0.27, t-value was 2.1615 (DF = 1348) and standard

CHAPTER 4. VALIDATION AND COMPARISON TO SCCM

error of difference was 0.125 and the actual difference between DMM and humans is estimated to be the in 0.0246 to 0.5154 range with 95% confidence. P-value was 0.0308 confirming human superiority with an average statistical significance. The difference between the DDM and SCCM samples was not statistically: relevant t-value was 1.3894 (DF = 2208), standard error of difference was 0.072 and P-value was 0.1648.

More interesting insights, however, are a result of an analysis of the specific sample types. Despite similar means, DDM and SCCM differed greatly in their believability in the prototype and affect related samples. Comparison of the human (N= 171, M=4.08, SD=1.36), DDM (N = 1000, M=3.80, SD=1.65) and SCCM (N = 1000, M=3.92, SD=1.76) prototype samples and their affective human (N= 29, M=4.16, SD=1.01), DDM (N = 150, M=3.97, SD=1.24) and SCCM (N = 150, M=3.71, SD=1.55) counterparts shows that while SCCM performed better in the case of pure social context cognition (and on average), DDM obtained better results in the affect believability.

This is due to the fact that the emotional state is sensitive not only to social context but also the relation between its partially processed components i.e. cognitive consonances and dissonances. This is a very relevant result, as it translates to the quality of any cognitive-affective human function that relays on those relations. On the other hand, the superiority of SCCM may be attributed to its more complete social context mapping (that also includes groups) and perhaps its more human-like connectionist nature.

4.4.5 Conclusions from the statistical comparison

In terms of subjective believability compared to human data, both models performed reasonably in several scenarios and for several characters. More detailed believability analysis showed that while SCCM seems to be better for applications requiring only 'black box solution' to social context cognition interpretation without relying on relations of its internal components, DMM is better for simulations where dissonances/consonances or

CHAPTER 4. SUMMARY AND CONCLUSIONS

other relations between the partial context components are relevant and where personalized, historical data about agent relations is relevant. Finally, SCCM is better suited to research-based, scientific simulations as data it requires in parametrization is much easier to obtain from psychological and sociological research than highly specific data required by DDM. Next chapter presents such scientific data based approach to SCCM parametrization in its application to personality modeling.

4.5 Summary and conclusions

In this chapter, Dramaturgical Dissonance Model was presented as a solution to the problem of modeling informal, 'System 1' social context cognition underlying idiosyncratic and cultural agent particularities. DDM is an alternative to SCCM's (presented in the previous chapter), and both models may be embedded and used in the existing agent architectures and may provide psychologically richer and multi-sided characters, realizing complex cultural and individual character profiles and specifications. Example implementation and character creation process were discussed and subjective believability comparison with both human and SCCM data was presented.

Statistical analysis of the results not only confirms that both models model are promising in the generation of believable social context modulated human-like particularities, potentially of quality similar to human samples, but also provides specific insights into strengths and weaknesses of each model that translate into suggestions about their potential in specific simulation settings. Analysis suggests that SCCM is overly better for applications requiring 'black box solution' to social context cognition interpretation, while DMM may be better suited for the cases where personalized, historical data about agent relations and relations between the partial context components are relevant.

Chapter 5

SCCM based personality model

Why, my man is created from the outside, that is, he is inauthentic in essence - he is always not - himself, because he is determined by form, which is born between people. His 'I', therefore, is marked for him in that 'interhumanity'. An eternal actor, but a natural one, because his artificiality is inborn.

Witold Gombrowicz

This chapter presents Social Context based Personality (SCP) model - a direct specialization of Social Context Cognition Model (SCCM) to the problem of personality modeling, as published in [KTT15a, KTT16b]. Personality is understood in this thesis, as a set of idiosyncratic behavioral dispositions, which were conceptualized in the literature as both traits and states[MS95, Hua09]. Those two aspects reflect accordingly context-pervading (traits) and context-dependent (states) individual and unique particularities of behavior. Reconciliation of those two aspects is problematic and psychology of personality 'has struggled with the task of conceptualizing the interplay between the stable and dynamic properties of personality systems' [Per90]. Applied to agent personality modeling, SCCM can be seen as an attempt to incorporate both of those aspects rather than to reduce personality to the static feature vector, as it is the case in popular 'trait models'. SCP aims to overcome weaknesses of the existing 'trait' models, by proposing a connectionist model grounded in the personality theories largely ignored in the computational personality modeling.

CHAPTER 5. INTRODUCTION

The remaining part of this chapter is organized as follows. Section 5.1 provides a background to the social context cognition in the frame of non-rationality and presents a general overview of the SCP in the framework of the non-rationality modeling. Section 5.2 explains the SCP's specification and parametrization process based on the actual psychological data, including a case character and scenario creation in section 5.2.2.3. Section 5.3 describes the believability validation procedure and its results. The chapter ends with section that contains the summary and conclusions.

5.1 Introduction

This section provides a short presentation of the problems related to personality modeling. The concept of personality is an intuitive way to address individual-specific patterns in preferences, style, behavior and other idiosyncratic features of non-rationality. Therefore, its modeling is one of the topics central to this thesis. As mentioned in Chapter 1, social context cognition has strong ties to the problem of personality modeling and many aspects of non-rationality that personality denotes. After presenting formal models capturing social context cognition i.e. SCCM (chapter 3) and DDM (chapter 4), it is now possible to address the problem of personality modeling and some of its related challenges.

Relevant processing details and SCP are described, and processes of its parametrization involving selection of the key mediating unit classes and their connections is presented. While chapters 3 and 4 addressed hand-made, entertainment-oriented applications and their subjective believability, this chapter, and chapters 6 present use of SCCM in more serious, scientific social simulations based on sound psychological and sociological data in their parameterization and validation. This brings new challenges in the case of parameter-heavy models like SCP/SCCM. This chapter addresses them for a real modeling problem in school bullying domain modeling. SCP is parametrized with existing psychological data and the simulation results are compared with specific quantitative

observed psychological and sociological regularities. Results show that SCCM/SCP is able to recreate various regularities observed in the real data and is promising in both entertainment and social simulation oriented applications.

5.1.1 Situation in personality modeling

Personality is traditionally a central concept in psychology. It embodies the assumption about individual behavioral coherence and consistency manifesting invariably throughout various contexts. A century long, intensive efforts in the identification of features of this entity was surprisingly unfruitful [MS95], as differences in behavioral dispositions between individuals are comparable or smaller than differences manifested by a given individual in various situations [Eps79]. This fact cannot be explained by the prominent 'trait models', like Five Feature Model (FFM)[MJ92] that emerged from this perspective. A parallel and well-established 'contextual' tradition also exists [Dur95, Gof59, Col04]. It undermines the importance of individual characteristics, in extreme cases reducing individuals to objects manipulated by social situations and contexts - the true sources of social agency. This century old dichotomy, called 'the most persistent cleavage in the social sciences' [Els89] continues within the field of computational modeling.

Cognitive-Affective Personality System (CAPS) theory [MS95] is a notable example of a recent attempt to conciliate this 'personality controversy'. It belongs to the interactionist tradition i.e. it redefines personality from a context-invariant feature set to an idiosyncratic process of context-based behavior disposition adaptations (i.e. not 'what you are in all contexts' but 'what you are in which context'). CAPS is introduced as based on a network of cognitive-affective 'mediating units' and it is empirically supported [MS95] despite relatively limited research on intra-personal behavioral disposition changes. It is also 'consistent with parallel findings showing significant amounts of variance attributable to Person X Situation interaction.' [CS99].

CHAPTER 5. INTRODUCTION

While promising, CAPS raises two related issues. To define personality in terms of context responses, one must define different contexts clearly. Such definition should ideally not be liberal - it must follow the psychologically relevant structure of the situation. This is, however, problematic, as 'there is no consensus regarding the most important "psychologically active features of situations"' [Ble09]. Secondly, if changes in dispositions are due to the network of hidden mediating units, what is their nature and how are they connected? Despite listing five broad types of cognitive-affective units (encodings, expectancies/beliefs, affects, goals/values and competencies/self-regulatory plans), both issues are left unsolved by CAPS, as it addresses them at a high abstraction level (Figure 2.2).

5.1.2 Mediating units: prototypes, features and personality states

Specific context detection leads to an individual and context-specific behavioral disposition changes in the personality state. Personality state requires representation in SCP, and FFM is a good candidate for several reasons. Strong criticism of FFM as a trait model is not a postulate to eliminate it entirely. Instead, it may represent current, rather than average or otherwise context independent behavioral disposition. This possibility has been noticed before. Bleidorn suggests that as individuals can be described by FFM's values on average, so can their current behavioral disposition, and more generally 'personality states are considered as trait-content manifestations in short-term, continuous and concrete ways of acting, feeling and thinking that could be described in the same way as traits' [Ble09]. FFM's five values are also an ideal final, 'output' layer of the mediating units (Figure 5.1), for two practical reasons. First, any model providing dynamism in FFM's values may instantly replace the traditional FFM in any architecture using it without other architectural changes. Considering the popularity of FFM in agent modeling, this is a great benefit. Second, FFM is well investigated in terms of correlation

Figure 5.1: The general structure of SCP

with other psychologically relevant factors, including other potential mediating units influencing personality state represented as FFM’s values, or behaviors that could result from them, making parametrization of the model significantly easier.

After defining the representation of the personality state (i.e. five FFM values used as states), other context-cognition modulating components must be discussed. The key problem is the potentially unlimited number of candidates for the relevant encodings of context cognition and ‘the issue of how to contextualize personality psychology constructively has been a longstanding dilemma’ [Rob07]. Fortunately, the issue has been addressed by SCCM for a more general case and much of the research in the psychology of personality suggests, that social role, situation and group prototypes assumed in SCCM are also the ones relevant also in personality state expressions.

5.2 Proposed model and its parametrization

This section describes the specification of SCCM into SCP and a process of its parametrization for applications requiring scientific data about human context cognition.

5.2.1 Specification of SCCM

SCP is a specification of SCCM, achieved by replacing the general output layer (see 3.2.3) with a layer implementing personality state expression changes. In SCP, the personality

CHAPTER 5. PROPOSED MODEL AND ITS PARAMETRIZATION

state induced by social context is conceptualized as five values of the FFM i.e. openness, conscientiousness, extraversion, agreeableness, and neuroticism (see 5.1.2). Each of them is encoded as an activation of a unit in the personality state layer (represented by Z_i , $i \in \{1, 2, 3, 4, 5\}$). As mentioned in 5.1.2 there is a consensus that particular context components (like roles) influence personality expressions, but the details of this influence are unknown. In the presented model, it is hypothesized that each component has either positive or negative influence on each of the dimensions of personality. This is captured by their change in each time-step:

$$\begin{aligned} \Delta Z_j(t) = & \frac{\sum_{i=1}^{|Ys|} P_{s_{i,j}} * Y_{s_i}(t)}{\text{InP}_j} + \\ & \frac{\sum_{i=1}^{|Yg|} P_{g_{i,j}} * Y_{g_i}(t)}{\text{InP}_j} + \\ & \frac{\sum_{n=1}^N \sum_{i=1}^{|Yr|} P_{r_{n,i,j}} * Y_{r_{n,i}}(t)}{\text{InP}_j}, \end{aligned} \quad (\text{Eq. 5.1})$$

where

$$\text{InP}_j = \sum_{i=1}^{|Ys|} |P_{s_{i,j}}| + \sum_{i=1}^{|Yg|} |P_{g_{i,j}}| + \sum_{n=1}^N \sum_{i=1}^{|Yr|} |P_{r_{n,i,j}}|.$$

Each FFM value is updated according to:

$$Z_j(t) = (1 - s_p) * Z_j(t-1) + s_p * \Delta Z_j(t), \quad (\text{Eq. 5.2})$$

with s_p , as the personality change phase step size. More details about SCCM/SCP processing can be found in section 3.2.3.

5.2.2 Psychological data based parametrization

This section describes the use of SCP in a specific domain. For easier parametrization and stricter validation, a domain with a relatively well researched social context perceptions

CHAPTER 5. PROPOSED MODEL AND ITS PARAMETRIZATION

and their consequences for the personality states was required. The problem of school bullying fulfilled these criteria. First, the basic assumed constants are presented, followed by parameters related to input, feature and prototype layers. Finally, the degree of data completeness is discussed.

Scientifically sound SCCM/SCP parametrization includes determination of: a) role, situation, and group prototypes; b) associations between them; c) their link to personality states; d) prototype features; e) their link to prototype activations; f) identification of the input elements; g) their link to feature activation; h) SCP's influence on agent architecture using it, e.g. by connecting unit activations with agent action or plan probabilities. Solutions to a)-e) are domain specific i.e. are shared by simulations within a given domain, while f)-h) are simulation specific and will vary within domain with time scale, detail level, focus, etc.

All above stages should ideally be based on solid quantitative data as much as possible. At the same time, investigation of contextual influence on personality state expressions is a relatively new and unexplored perspective, data in most domains is scarce and incomplete at best. Entertainment-oriented projects using SCP for rich character generation may focus on subjective believability tests rather than confrontation with actual data (see Chapter 3). In more serious social simulations based on SCCM/SCP obtaining such data may be a part of the modeling effort or a reasonable compromise between existing data, own research and common sense may be required. Chapter 6 deals with semi-automatic parametrization that eliminates much of this effort. Example parametrization presented here relies on the existing scientific research where possible (around 70% of parameters) and own questionnaires investigating correlations between social prototypes.

5.2.2.1 Domain specific parameters

Domain parametrization consists mostly of the specification of the prototypes relevant for the domain, their influence of prototypes on the personality states, their relations

CHAPTER 5. PROPOSED MODEL AND ITS PARAMETRIZATION

and (partly) features activating them (Tables 5.1 and 5.2 for some examples). While such data may not be easy to obtain, it is reusable i.e. other simulations with the same domain are likely to share it. Role prototypes used, were roles traditionally pointed to by the most researchers (Bully, Defender, Outsider, and Victim [GOD06]) plus two non-bullying examples of the affiliation and power (Friend, Student) [Rob07]. Lesson and the unstructured Breaks [GZKB14] served as situation prototypes and Students, Friends and Gang-like prototypes were perceived groups.

5.2.2.2 Simulation specific parameters

The main part of the simulation specific parametrization includes setting the links between external object perceptions and the prototype features and the influence of the SCP on agent's actions. Agents perceived world and object features like time, location, presence of authority (the teacher) and other agents etc. Processed features of the objects and agents were linked with the relevant role prototype features. Key objects and features were selected according to the literature suggesting elements relevant to contextualization in this domain [ØR01] and universal warmth-competence dimensions [FCG07]. For example, the degree of aggressive and friendly actions perceived in agent's behavior was used to determine his warmth, cumulative warmth of all agents determined warmth as group feature. Current personality state was linked to the probability of the engagement in the actions and interactions with the perceived agents. Those consisted of decision to study or not, display of friendly or hostile affects, verbal or physical attack etc. and their probability changes were based on the research of personality state influence on the aggressive actions [TE05] (see Table 5.4 for personality-action relations). Simulation-specific parametrization was finalized by setting the number of perceived agents attributed social roles to five, assuming uniform role influence for all the agents, i.e. $Wr_n = Wr_m$, $Pr_n = Pr_m$, $Vgr_n = Vgr_m$, $Vrs_n = Vsr_m$, for $5 \geq m, n > 1$, and prototype association symmetry i.e. $Vgr = Vrg$, $Vgs = Vsg$, $Vrs = Vsr$.

CHAPTER 5. PROPOSED MODEL AND ITS PARAMETRIZATION

Table 5.1: Role influence on personality state expression [MJ97, Ble09]. 'O', 'C', 'E', 'A' and 'N' stand for openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism respectively.

	O	C	E	A	N
Friend	-0.18	0	0.52	0.2	-0.1
Student	-0.09	0.38	-0.24	0.12	0.08
Bully	0.15	-0.21	0.31	-0.19	0.22
Victim	-0.07	-0.35	-0.23	-0.26	0.31
Defender	0.11	0.36	0.29	0.42	-0.36
Outsider	-0.13	0.41	-0.27	0.04	-0.10

Table 5.2: Role features [MJ97]

Parameter	Bully	Victim	Defender	Outsider
Schol. comp. (mean)	2.3	2.8	3.9	3.2
Schol. comp. (deviation)	0.65	0.66	0.51	0.65
Social acc.(mean)	2.5	2.9	4.3	3.3
Social acc. (deviation)	0.8	0.65	0.52	0.65
Appearance (mean)	2.10	2.5	3.1	2.50
Appearance (deviation)	0.72	0.68	0.65	0.89
Conduct (mean)	2.20	3.19	4.2	3.50
Conduct (deviation)	0.34	0.33	0.38	0.48

Table 5.3: Observable objects and their features [ØR01, FCG07]

World object states	
Teacher	behavior supervision
Task	assignment
Location	in/outside classroom
Time	class/break time
Agent states	
Outfit	gang, student
Task	focus, ability
Physical features	appeal
Behavior to agents	aggressive, friendly

Table 5.4: Personality influence on action probabilities [Ble09]

	O	C	E	A	N
Physical attack	-0.12	-0.26	-0.19	-0.34	0.08
Verbally attack	0.04	-0.05	0.15	-0.30	0.20
Hostility	-0.33	-0.16	-0.30	-0.24	0.33
Protection	0.12	0.26	0.19	0.34	- 0.1
Task following	0.0	0.45	0.0	0.21	-0.11

5.2.2.3 Agent and scenario generation

The steps described so far address the parametrization of social context cognition consequences in an 'average' individual in a given simulated group. In a multi-agent simulation, it requires personalization for each agent, either manually or by partial randomization. In this case, 600 'student' agents were created automatically based on the psychological data wherever possible. For example, each personality state influence on each action probability [TE05] (Table 5.4) has standard deviation used in the randomization.

Additionally, several semi-random scenarios were created, including breaks, lesson, and transitions. Scenarios were relatively long and had multiple context changes, to compare the performance of SCCM/SCP with and without the depth mechanism (see 3.2, 3.2.3), in terms of resistance to saturation, as long scenarios and multiple context changes are more prone to saturation and insensitivity.

5.3 Result analysis

To initially test the presented model, the data generated in multiple simulations similar to the ones presented in the previous section needs to be compared to the statistical data on bullying incidents. 600 'student' agents were automatically created by introducing a random variation to the generic SCCM and placed in the semi-random scenarios. This section analyzes the obtained results, comparing SCCM's behavior and internal states with the available data.

5.3.1 Subclass emergence

As shown in Figure 5.2, agents using SCCM/SCP may spontaneously generate behavior subclasses well known in the bullying literature. Researchers investigating school bullying recognize several typical bullying roles and interaction patterns. Surprisingly, many of those patterns emerge from the random micro-differences in the agent's SCCM/SCP in the simulations. One example is a bully instigator-follower dynamics (Figure 5.2), a behavior pattern known in the literature.

In this case, Agent1 (A1) became a victim of Agents 2 and 3 (A2/A3). A2's neutral perceptions of A1 changes, as A2 sees A1 as a Victim ($T=1$), mostly due to A2's subjective perception of low Appearance, as a strong cause for victimization. This perception activates A2's Bullying situation prototype, which in turn activates Bully self-role prototype of A2 ($T=3$). This causes lowered Conscientiousness (Table 5.1). When it decreases sufficiently, A2 verbally attacks A1 ($T=4$). As A3's Appearance-Victim link is weaker, he sees A1 as a Victim only after A2's verbal attack, triggering the Bullying-Bully activation sequence ($T=7$). A3's Bully-Conscientiousness link is stronger than A2's. Even lower Conscientiousness leads A3 to a physical attack ($T=9$). Lesson start, influences bullies differently ($T=10$). As A2's Lesson prototype is sensitive, he stops the attack, while A3 fails to recognize the new context and stops the attack only when the teacher arrives ($T=20$). As a result, the instigator (A2) withdraws before the physical attacks and before the teacher's arrival, when only the follower (A3) is still bullying.

Similar familiar, persistent behavioral role-like patterns emerged spontaneously. Agents (like A2 or A3, shown in Figure 5.2) maintained individual styles consistent through all simulations, while showing multi-faced, variable behaviors e.g. bullying in specific contexts and protective behaviors in others. Statistical aspect of this phenomena was investigated using the example of the bully instigators and bully followers. SCCM/SCP-bullies were categorized into two groups: those who only joined bullying (followers) and

CHAPTER 5. RESULT ANALYSIS

Figure 5.2: a) (Left) Selected prototype activations in SCCM’s of the simulated agents in the analyzed example (see section 5.3.1 for detail analysis) b) (Right) This scenario is also an example of SCP/SCCM’s resistance to saturation, when ‘depth’ mechanism is used. While the initial context cognition is almost identical with and without the mechanism, strong detections cause saturation that is permanent without the ‘depth’ mechanism. With ‘depth’ mechanism, SCCM resists saturation without any global parameter changes (see section 3.3).

those also instigating bullying activities. This was compared with research [SLB⁺96] reporting human data (Table 5.7). Even though the total percentage of bullies reported by this research was somewhat higher, the relation between the followers and instigators was maintained. As previously, SCCM/SCP produced stable results regardless of the scenario length used.

5.3.2 Objective role frequency and variability

The most basic statistical test for SCCM/SCP is a bullying frequency test. There is a well-known trend in the bullying statistics: victims constitute between 7% and 15% of youth (with most research reporting 10-13% [ØR01, BU92, MDBC06]) and bullies between 6% and 12%. Numbers are higher for younger and difficult youth, but proportions are maintained. This data corresponds well to the SCP’s results: **15.1% victims and 12.8% bullies.**

Such results are, however, obtainable with a well calibrated ‘personality trait’ model.

CHAPTER 5. RESULT ANALYSIS

The actual challenge is to obtain human (low to medium) role stability, shown by in the longitudinal and short-term research [SLL98, SKB⁺05, CHK⁺05]. For example, aggressive victims or victim-bullies switching roles contextually, constitute around 33% of those involved in bullying [MDBC06] i.e. 7% of total population. SCP contextual variance patterns were much closer to the human data than those of 'trait models'.

As experiments confirm (Table 5.5), even a well calibrated FFM model, used in a simulation otherwise identical to the one described in section 6, is outperformed by SCCM in addressing this phenomenon. Normal distribution of the personality traits, calibrated to the generate the desired victimization level (15%) over the identical number of characters and scenarios was used (note that SCCM does not need such calibration). The 'trait model' generated predictably uniform bullying behavior: only 1.1% of the total population manifested mixed behaviors, as victim-bullies (Table 5.5). In the case of SCCM it was 7.6% of the population, showing SCCM's superior ability to generate mixed, contextualized behaviors coherent with the psychological data.

5.3.3 Saturation analysis. Variability vs. scenario length.

For longer scenarios with more complex social context changes, saturation problem becomes an issue, as discussed in section 3.1.2 and Figure 5.2. While SCCM with the 'depth mechanism' perform almost identically in the short scenarios, with no or single social context transitions, the lengthy scenarios show the superiority of the SCCM with the 'depth mechanism'. Table 5.6 illustrates results of the scenario sets with multiple scenarios merged to gradually increase the length. In general, as N scenarios of length L each (set A), are combined into N/M scenarios lasting $M \cdot T$ (set B) the obtained results differ more and more as M grows. The transition from the previous part of the scenario, instead of starting from a neutral state introduces expected differences. However, as transitioned states are close to neutral on average, the statistical difference should be minor and should not influence the role variability. That is, the spectrum of the prototype

CHAPTER 5. RESULT ANALYSIS

Table 5.5: Pure/mixed bullies and victims [MDBC06] .

Group	SCCM	FFM	Humans ^[MDBC06]
Pure bullies	31 (5.1%)	72 (12%)	36.5 (6.1%)
Pure victims	45 (7.5%)	83 (13.8%)	47.5 (7.9%)
Victim-bullies	46 (7.6%)	7 (1.2%)	42.7 (7.1%)
Population	600	600	600

Table 5.6: Number of victim-bullies depending on the scenario length.

Scenario length	'depth' mechanism	no 'depth' mechanism	FFM
50	46 (7.6%)	46 (7.6%)	7 (1.2%)
100	48 (8%)	37 (6.1%)	10 (1.6%)
200	43 (7.1%)	31 (5.2%)	8 (1.3%)
500	46 (7.6%)	27 (4.5%)	9 (1.5%)
1000	44 (7.3%)	25 (4.1%)	7 (1.1%)

activations presented by a given SCCM should be theoretically similar in set A and B. Without the 'depth' mechanism (Table 5.6), that is not the case due to the saturation. After each major social context change, there is a risk, that highly activated part of the network will maintain a self-perpetuating activation and the later stimulation will not cause a proper social context change. As M grows, variability drops, signaling saturation issues. This is not the case, however, if 'depth' mechanism is used (Table 5.6). It allows equally good performance, regardless of the scenario length, confirming the stability of SCCM with the 'depth' mechanism. [MJ97, Ble09]

5.3.4 Subjective popularity vs. objective victimization

While the previous sections analyzed the external behavior of SCCM/SCP-based agents, this one investigates their internal states and subjective prototype (role) attributions (absent in the 'trait models'). Socio-psychological phenomena may be investigated using both objective and reported subjective perspective. Also, analysis of the relation between those two types of data may be insightful. Models using subjective social prototype representations (like SCCM) may address such questions - a feat impossible for 'trait

CHAPTER 5. RESULT ANALYSIS

Table 5.7: Emergence of bully sub-classes.

Group	SCCM(50)	SCCM(200)	Humans
Follower bullies	31 (5.2%; 40.2%)	31 (5.1%; 39.2%)	40.8 (6.8%; 45.4%)
Instigator bullies	46 (7.6%; 59.8%)	48 (8.0%; 60.7%)	49.2 (8.2%; 54.6%)
Total bullies	77 (12.8%; 100%)	79 (13.1%; 100%)	90 (15%; 100%)

models’. Because of this advantage, SCCM may be used to reconstruct specific findings from the literature about the relations between the subjective social perceptions and the objective behavior patterns like ‘correlation between victimization and peer rejection ($r = .57$) is stronger than that between victimization and peer acceptance ($r = -.36$)’ [CB98]

Specifically, the relation between the subjective popularity/acceptance and the objective victimization was investigated. Popularity is a proven predictor of victimization. For the purpose of this example, peer acceptance and rejection were estimated by other agent’s attributions of ‘friendly’ and ‘unfriendly’ roles. Length and the degree of attribution of a Friend role to the agent determined agent’s acceptance level. Similarly, the rejection was based on a Victim role attribution. The global correlation between the objective victimization (amount of received aggression) and rejection and acceptance was 0.73 and -0.22 respectively, showing a similar pattern to the reported human behavior.

The relation between popularity and attributed roles was also assessed. The collective popularity ratios in humans for victims, bullies, and outsiders were -0.39, -0.21 and 0.25 respectively, as ‘not-involved children have been found to be more popular than bullies who, in turn, were found to be more popular than victims’ [BU92]. Another research, states that only 1% of subjects lacked a friend and all of them were bullied [ØR01]. Out of all agents, 0.3% was not nominated as a Friend by any agent and 5.1% never received ‘friendly interaction’. Regardless of the assumed definition of ‘having a friend’, they were all bullied.

5.4 Summary

In this chapter, SCP - a new personality model has been presented based on SCCM, addressing relevant limitations of the existing personality models. It has been implemented and parametrized for a specific domain. The generated results were tested against the existing data proving model's ability to generate context-dependent behavior, and model's properties were proven superior to that of 'trait personality models'. Also, SCP simulated many relations between subjective and 'objective elements of bullying reflecting those found in research on actual people - the purpose for which, unmediated 'trait models' cannot be used at all. Moreover, simulations using SCP further verified 'depth mechanism's' resistance to stagnation in highly correlated and activated network parts, preventing 'saturation problem', making it usable in agents exposed to complex and changing social contexts. This solution could be useful in other connectionist models with circular connections. While further validation, application to new domains and application of learning mechanisms to automatic model parametrization are still required, SCCM/SCP was proven promising in the context of both social simulation and individual character generation, from serious scientific applications to entertainment.

Chapter 6

Semi-automatic SCCM/SCP parametrization

This chapter addresses the problem of automatic parameterization of SCCM and SCP that were presented in previous chapters of the thesis, as published in [KTT16a]. Both psychology [MS95, CS99] and developments in the behavior simulation [KTT13, Neu08] suggest that current personality models ignore and oversimplify underlying social context cognition that modulates behavior in humans. Human behavior is inherently contextual. Disposition changes due to context differences have a similar scale to the inter-personal differences. Also, people with equal personality traits, in fact, react very differently to the same social contexts. It is increasingly clear that personality as a static, trait pervading all contexts must be redefined, as a pattern of responses to specific contexts [KTT15a]. Popular personality models [MJ92] ignore this and produce static agents e.g. equally 'open' or 'friendly' in all contexts and do not manifest contextual variability as in a case of an introvert person who opens up in a domestic context etc. While socio-contextual models like SCP, presented in the previous parts of this thesis are a promising solution to this problem, their parametrization is a crucial unsolved issue. Due to its nature, socio-cognitive data is inaccessible, subconscious, informal and sensitive to subtle framing changes. It is both qualitative (as it dictates discreet social categories) and quantitative (associations between them). Obtaining it is a significant challenge, as it

CHAPTER 6. CROWD-SOURCING SOCIAL CONTEXT

is often incomplete or non-existent in psychological literature. This chapter addresses the issue of automatic parameterization of SCCM and SCP, whose parametrization from psychological data is costly and inapplicable in many cases. It contains the summary of the relevant research including social context and crowd-sourcing literature and the outline of the SCP. It presents a new method of its crowd-sourcing based parametrization, including the statistical analysis used in the implemented system and system validation. Semi-automatic SCP generation may be used in human-like behavior modulation or personality models but also in multi-agent culture simulations and other applications.

6.1 Crowd-sourcing social context

As mentioned, SCP manual parametrization is an unsolved problem, limiting in application. The solution proposed here is a parameter extraction from the crowd-sourced social context data. Crowd-sourcing is a model of outsourcing work to fluid groups with little specific expertise or management structures. Recent years brought the great growth of crowd-sourcing applications, some also in cognitive-affective computing. The most popular use of crowd-sourcing in the affective computing is the ground truth label collection, somewhat similar to the qualitative part of the proposed system.

One example is VidL, a distributed video-labeling tool specifically designed for labeling affective data [EP09]. Another case is a 'Guess What?' [ROR11] game crowd-sourcing affective video data labeling of social situations. In many such cases 'there is no objective ground-truth' [MM14] and the agreement between the subjects is low, reflecting the subjective, culture-variant and framing sensitive nature of the problem. While in most of the applications this is problematic, SCP is designed to target such interpretation differences. Also, unlike other systems, SCP requires not only probing of labels but also their quantitative relations. In this sense, there is no real parallel system to compare against.

6.2 SCCM/SCP crowd-sourcing

In this section, a semi-automatic parametrization procedure of SCCM and SCP is described. SCP and SCCM are connectionist models using mediating units organized in clusters, grouped in four layers (input, feature, prototype and personality state layer) linked with associations with processing taking place in four main phases: feature classification, prototype association, energy release and output state change phase.

As shown in the previous chapters, SCP parametrization may be successfully obtained manually i.e. either from psychological or hand-made data. This is, however, tedious, due to a number of parameters and problematic, as the sociological data may be distributed among heterogeneous and incomplete sources or may be unavailable altogether for a given context, forcing developers to conduct such research themselves. For those reasons semi-automatic model parametrization described below is of key importance.

To use SCP or SCCM in a selected domain, related social context cognition in a represented human group must be obtained. Due to the nature of the models, two parameter classes are required: discrete (nodes i.e. mediating units of each type) and continuous (node connection parameters i.e. unit associations). This section describes application specific sample selection and preprocessing, initial feature set selection, presents data format, crowd-sourcing procedure and processing of the obtained data to obtain SCP.

6.2.1 Application specific preparations

Before automatic procedure takes place, domain and application specific preparations must be made, including sample selection and preprocessing, subject selection and choice of initial features.

6.2.1.1 Sample selection and preprocessing

For the crowd-sourcing to take place domain must be represented as a sensory data sample set to be presented to the users. Generation or selection of images, videos,

CHAPTER 6. SCCM/SCP CROWD-SOURCING

simulations and/or text properly covering domain is a first step of the parametrization. For example, if the model is to represent behavior modulation in the context of family context, recreation, and professional sport, presented scenes must represent the spectrum from personal to professional activities with varying levels of professionalism, styles etc. Sample granularity and scope must reflect the initial domain concept otherwise obtained data may not fully address the original problem.

To calculate the relationship between the presented sensory samples and the social context cognition elicited in the users chosen samples require brief preprocessing i.e. labeling of the key object and agent features present in each sample. Selected key objects, agents, and their features must be relevant for the investigated domain context cognition (again issue of granularity and scope is crucial). It is also worth considering that it imposes possible perceptual input in the agents generated in the model generation process. Those objects and agent features may be obtained either by hand, automatically or by crowd-sourcing (in questions like: 'what situation is this? what objects and agent features make you think so?'). In the case of simulation based sample generation, labeling may often be generated trivially with the samples.

6.2.1.2 Initial feature selection

Component features are an intermediate level between perceptual input and the social context cognition components. The situation, role or group features underlying component activation are more fuzzy and subconscious than the components themselves but are more application invariant. In most cases, role and group features include engagement, domination, and positive-negative spectrum, while situation features private-public and professionalism spectrum. Often best approximation requires prolonged domain investigation by the experts and may be equivocal. From the computational perspective, the best features maintain the most component distinguishing ability while reducing the

CHAPTER 6. SCCM/SCP CROWD-SOURCING

input dimension and increasing generalization. The proposed system uses a predefined candidate set, that is narrowed based on those qualities (see section 6.2.4.4).

6.2.1.3 Subject selection

When the samples and their descriptions are ready, subjects for crowdsourcing must be selected. There are several decisions to be made concerning the strictness of selection and supervision of the data gathering. Depending on the nature of the application, it may be possible to allow indiscriminate unsupervised survey e.g. in social media or require a strictly supervised selection of subjects in terms of their features like age, gender, education, ethnicity or any other feature key to the social context sensitivity of the desired resulting agents. In general, scientific social simulation will require stricter selection and supervision than general agents and entertainment oriented simulations. In many cases, a reasonable compromise is a survey with subject detail gathering, allowing generation of multiple desired parameterizations depending on the filtered subject features.

6.2.2 Data format

In this section, a description of the used data format will be provided. Sample input percept representation is formally described together with the format of social context components and features.

6.2.2.1 Input samples

Regardless of the method used in the data preparation phase, there is a need to formally represent the prepared sample set S in the context analysis. Each sample may contain multiple objects and (at most) max_A agents arbitrarily chosen as most important in the given sample. $A = \{a_1, \dots, a_{maxA}\}$ denotes the agents selected in a sample.

Based on the data preparation, sets of all known object and agent percepts are represented as OP and AP . Two functions characterize the prepared sample set S using OP and AP :

$$V_{OP} : OP \times S \rightarrow [0, 1], \quad (\text{Eq. 6.1})$$

$$V_{AP} : AP \times A \times S \rightarrow [0, 1], \quad (\text{Eq. 6.2})$$

where $V_{OP}(op, s)$ represents the degree to which $op \in OP$ is present in $s \in S$ and $V_{AP}(ap, a, s)$ is the degree to which ap describes the agent $a \in AP$, in sample $s \in S$.

Example:

$$V_{OP}('ball', image_1) = 1.0,$$

$$V_{AP}('child', image_1, a_5) = 0.3$$

Finally, each sample $s \in S$ input representation $SI(s)$ takes the following form in the social context analysis :

$$SI(s) = \bigcup_{op \in OP} \{(op, V_{OP}(op, s))\} \cup \bigcup_{ap \in AP, a \in A} \{(ap, a, V_{AP}(ap, a, s))\} \quad (\text{Eq. 6.3})$$

Example:

$$SI(image_1) = \{('ball', 1.0), ('football field', 0.7), ('football uniform', a_1, 1.0), ('child', a_1, 1.0), ('play', a_1, 1.0), \dots, ('football uniform', a_5, 0.7), ('adult', a_5, 1.0), ('standing', a_5, 1.0)\}$$

6.2.2.2 Social context components

Social context components provided by the subjects in the process of crowd-sourcing belongs to three non-intersecting sets: situation (SC), role (RC) and group components (GC). Each question and answer concerning a component is represented in one of the following functions:

CHAPTER 6. SCCM/SCP CROWD-SOURCING

Table 6.1: Symbol table, crowd-sourced data

Symbol	Description
A, S	set of agents and samples
OP, AP	set of object and agent percepts
$SI(s)$	sample input for sample $s \in S$
SC, RC and GC	situation, role and group components
SF, RF and GF	situation, role and group features
$QA(n)$	n -th question-answer set (QA)
$QS(n)$	id of the sample in n -th QA
$QA_{SC}(sc)$	QA subset with $sc \in SC$ value $\neq '?'$
$QA_{RC}(rc, a)$	QA subset with $rc \in RC$ value $\neq '?'$
$QA_{GC}(gc)$	QA subset with $gc \in GC$ value $\neq '?'$
$QA_{SF}(sf)$	QA subset with $sf \in SF$ value $\neq '?'$
$QA_{RF}(rf, a)$	QA subset with $rf \in RF$ value $\neq '?'$
$QA_{GF}(gf)$	QA subset with $gf \in GF$ value $\neq '?'$
$V_{OP}(op, s)$	$op \in OP$ level in sample $s \in S$
$V_{AP}(ap, a, s)$	$ap \in AP$ level for $a \in A$ in $s \in S$
$V_{SC}(sc, n)$	n -th QA, value of $sc \in SC$
$V_{RC}(rc, a, n)$	n -th QA, value of $rc \in RC$ in $a \in A$
$V_{GC}(gc, n)$	n -th QA, value of $gc \in GC$
$SCD(n)$, $SFD(n)$	n -th QA, scene components/features
$RCD(n)$, $RFD(n)$	n -th QA, role components/features
$GCD(n)$, $GFD(n)$	n -th QA, group components/features

CHAPTER 6. SCCM/SCP CROWD-SOURCING

Table 6.2: Symbol table, analysis-generated data

Symbol	Description
$B_{RC}(ac, a, s)$	$\{ n : QA(n) \in QA_{RC}(ac, a) \text{ } QS(n)=s \}$
$B_{GC}(gc, s)$	$\{ n : QA(n) \in QA_{GC}(gc), QS(n)=s \}$
$B_{SC}(sc, s)$	$\{ n : QA(n) \in QA_{SC}(sc), QS(n)=s \}$
$B_{RF}(af, a, s)$	$\{ n : QA(n) \in QA_{RF}(af, a) \text{ } QS(n)=s \}$
$B_{GF}(gf, s)$	$\{ n : QA(n) \in QA_{GF}(gf), QS(n)=s \}$
$B_{SF}(sf, s)$	$\{ n : QA(n) \in QA_{SF}(sf), QS(n)=s \}$
$V'_{RC}(rc, a, s)$	average value for rc in agent a, sample s
$V'_{GC}(gc, s)$	average value for gc in sample s
$V'_{SC}(sc, s)$	average value for sc in sample s
$V'_{RF}(rf, a, s)$	average value for rf in agent a, sample s
$V'_{GF}(gf, s)$	average value for gf in sample s
$V'_{SF}(sf, s)$	average value for sf in sample s
$C_{X,Y}(x, y)$	correlation between $x \in X$ and $y \in Y$
$D_{RC}(r1, r2)$	correlation differences between r1 and r2
$D_{SC}(s1, s2)$	correlation differences between s1 and s2
$D_{GC}(g1, g2)$	correlation differences between g1 and g2
$\tilde{V}_X(x, a, s)$	estimation of $x \in X$ for $a \in A, s \in S$
$\tilde{V}_X(x, s)$	estimation of $x \in X$ for $s \in S$
$E_{X(x)}$	estimation error for $x \in X$
pt	component proximity tolerance

CHAPTER 6. SCCM/SCP CROWD-SOURCING

$$V_{SC} : SC \times N \rightarrow [0, 1] \cup \{?\}, \quad (\text{Eq. 6.4})$$

$$V_{RC} : RC \times A \times N \rightarrow [0, 1] \cup \{?\}, \quad (\text{Eq. 6.5})$$

$$V_{GC} : GC \times N \rightarrow [0, 1] \cup \{?\}, \quad (\text{Eq. 6.6})$$

where $V_{SC}(sc, q)$, $V_{RC}(rc, a, q)$ and $V_{GC}(gc, q)$ represent the degree to which corresponding components sc , rc and gc were found in the n -th question-answer (in agent a in case of V_{RC}). User answers are scaled into the $[-1, 1]$ range while '?' represents the situation in which component's presence was not investigated in the given question-answer set.

Example:

$$V_{SC}(\text{'private situation'}, 1) = -0.2,$$

$$V_{RC}(\text{'dominant'}, a_5, 1) = '?',$$

$$V_{GC}(\text{'friendly'}, 1) = 0.7.$$

Based on the above functions, all context components in the given sample are stored in the situation, role and group component datasets:

$$SCD(n) = \bigcup_{s \in SC} \{(s, V_{SC}(s, n))\}, \quad (\text{Eq. 6.7})$$

$$RCD(n) = \bigcup_{r \in RC, a \in A} \{(r, a, V_{RC}(r, a, n))\}, \quad (\text{Eq. 6.8})$$

$$GCD(n) = \bigcup_{g \in GC} \{(g, V_{GC}(g, n))\}. \quad (\text{Eq. 6.9})$$

Example:

$$SCD(1) = \{(\text{fun}, 0.9), (\text{football match}, 0.4), (\text{competition}, '?'), (\text{professional game}, -0.8)\}$$

CHAPTER 6. SCCM/SCP CROWD-SOURCING

$RCD(1) = \{ (a_1, \text{playing kid}, 1.0), (a_1, \text{footballer}, 0.3), (a_1, \text{referee}, -1.0), \dots, (a_1, \text{coach}, '?'), \dots, (a_5, \text{playing kid}, -0.9), (a_5, \text{footballer}, -0.4), (a_5, \text{referee}, 0.9), \dots, (a_1, \text{coach}, 0.7) \}$

$GCD(1) = \{ (\text{children}, 1.0), (\text{professional players}, -0.6), (\text{family}, -0.8), (\text{friends}, '?') \}$

6.2.2.3 Social context component features

Social context features are stored as SF , RF and GF i.e. situation, role and group features respectively. Again, their presence in the n -th question-answer set is stored as:

$$V_{SF} : SF \times N \rightarrow [0, 1] \cup \{?\}, \quad (\text{Eq. 6.10})$$

$$V_{RF} : RF \times A \times N \rightarrow [0, 1] \cup \{?\} \quad (\text{Eq. 6.11})$$

$$V_{GF} : GF \times N \rightarrow [0, 1] \cup \{?\}, \quad (\text{Eq. 6.12})$$

where $V_{SF}(\text{sf}, q)$, $V_{RF}(\text{rf}, a, q)$ and $V_{GF}(\text{gf}, q)$ represent the degree to which the feature (sf, rf or gf) is found in the n -th question-answer set (and in agent a in case of V_{RF}).

Example:

$$V_{SF}(\text{'family situation'}, 1) = -0.2,$$

$$V_{RF}(\text{'referee'}, a_5, 1) = 0.8,$$

$$V_{GF}(\text{'kids'}, 1) = 0.7$$

Again, all context component features in the given sample are stored in feature datasets:

$$SFD(n) = \bigcup_{s \in SC} \{(s, v_{SC}(s, n))\}, \quad (\text{Eq. 6.13})$$

$$RFD(n) = \bigcup_{r \in RC, a \in A} \{(r, a, v_{RC}(a, r, n))\}, \quad (\text{Eq. 6.14})$$

$$GFD(n) = \bigcup_{g \in GC} \{(g, v_{GC}(g, n))\}. \quad (\text{Eq. 6.15})$$

where $SFD(n)$, $RFD(n)$ and $GFD(n)$ are respectively scene, role and group feature datasets provided in n -th question-answer set.

6.2.2.4 Question-answer set representation

Using the prepared samples, the crowd-sourcing system acquires social context data by presenting them to the user and storing information about user's resulting context cognitions in the question-answer sets. N -th question-answer set is represented by

$$QA(n) = (SCD(n), RCD(n), GCD(n), SFD(n), RFD(n), GFD(n)). \quad (\text{Eq. 6.16})$$

The entire question-answer database is

$$QA = \{QA(1), \dots, QA(maxQA)\}, \quad (\text{Eq. 6.17})$$

where $maxQA$ is the current number of obtained question-answer sets. Subset of QA with known (i.e. $\neq '?'$) values of given components and features is defined as

$$QA_{SC}(sc) = \{QA(n) : V_{SC}(sc, n) \neq '?'\} \quad (\text{Eq. 6.18})$$

$$QA_{RC}(rc, a) = \{QA(n) : V_{RC}(rc, a, n) \neq '?'\} \quad (\text{Eq. 6.19})$$

$$QA_{GC}(gc) = \{QA(n) : V_{GC}(gc, n) \neq '?'\} \quad (\text{Eq. 6.20})$$

$$QA_{SF}(sf) = \{QA(n) : V_{SF}(sf, n) \neq '?'\} \quad (\text{Eq. 6.21})$$

$$QA_{RF}(rf, a) = \{QA(n) : V_{RF}(rf, a, n) \neq '?'\} \quad (\text{Eq. 6.22})$$

$$QA_{GF}(gf) = \{QA(n) : V_{GF}(gf, n) \neq '?'\} \quad (\text{Eq. 6.23})$$

Finally $QS(n)$ is the id of the sample n .

6.2.3 Crowd-sourcing procedure

In this section, details of the crowd-sourcing procedure are presented including context calibration with the proper component and feature collection.

6.2.3.1 Context calibration

Before the question-answer sets gathering, subjects are exposed to several samples, to achieve domain calibration in terms of scope and granularity of the context interpretation. Without such exposure, subjects may provide too general or too narrow classifications for the first few questions asked (e.g. labeling any group as 'people' etc.). Exposure provides an automatic method of contextualization of the later questions while being non-invasive i.e. not influencing subject's interpretations with directing questions or suggestions.

6.2.3.2 Component and feature collection

After the calibration, proper data gathering starts. First, subject is shown a sample and asked about interpretation of the presented sample:

- situation e.g. 'What is this situation?' and 'How well does it exemplify this label?',
- group e.g. 'Who are those people?' and 'How well do they exemplify this label?',

CHAPTER 6. SCCM/SCP CROWD-SOURCING

Figure 6.1: Example images representing domain with marked objects and agents

- role e.g. 'What is this person's role?' and 'How well does it exemplify this label?'.

The qualitative part of the question is used to introduce new context components (in one or two words), and the quantitative one contains information about the quality of the fit (-1.0 to 1.0) used in the analysis. The order of questions from the most general (situation) to detailed (roles) proved optimal, as it forced increased specificity in the description, promoting answers like the group - players, role - goalkeeper, instead of role - player, the group - players obtained otherwise.

Second part, is an investigation of the components obtained previously in other contexts:

- 'How well 'S' describes this situation?',
- 'How well 'G' describes this group? ',
- 'How well 'R' describes this person?',

where 'S', 'G' and 'R' are the situation, group and role components previously obtained by the system. The subject is asked to provide a quantitative evaluation, with 1.0 as a perfect fit, 0.0 - a neutral/irrelevant one and -1.0 as a total contradiction between the given component and the presented sample. While the goal of the first phase is to

acquire new, optimal social components, the second phase provides the data about associations between the existing ones. Finally, the third part consists of collection of the feature data. The subject is asked similar questions as in part two, but 'S', 'G' and 'R' this time are component feature candidates.

6.2.4 Data processing

This section describes data processing that takes place parallel to its gathering. It consists of correlation analysis, duplicate detection, and merger and finally, feature evaluation and SCP generation.

6.2.4.1 Sample base and averaged values

The obtained sample-component *base* data, is defined as:

$$B_{RC}(ac, a, s) = \{n : QS(n) = s, QA(n) \in QA_{RC}(ac, a)\}, \quad (\text{Eq. 6.24})$$

$$B_{GC}(gc, s) = \{n : QS(n) = s, QA(n) \in QA_{GC}(gc)\}, \quad (\text{Eq. 6.25})$$

$$B_{SC}(sc, s) = \{n : QS(n) = s, QA(n) \in QA_{SC}(sc)\}. \quad (\text{Eq. 6.26})$$

Similarly, sample-feature *base* is :

$$B_{RF}(af, a, s) = \{n : QS(n) = s, QA(n) \in QA_{RF}(af, a)\}, \quad (\text{Eq. 6.27})$$

$$B_{GF}(gf, s) = \{n : QS(n) = s, QA(n) \in QA_{GF}(gf)\}, \text{ and} \quad (\text{Eq. 6.28})$$

$$B_{SF}(sf, s) = \{n : QS(n) = s, QA(n) \in QA_{SF}(sf)\} \quad (\text{Eq. 6.29})$$

Figure 6.2: **SCP/SCCM crowdsourcing procedure.** After *sample preprocessing* (section 6.2.1.1), *questions* for the users are being generated and answers stored in the component and feature databases. Ongoing *duplicate analysis* merges the component duplicates, unless designer specifies otherwise (section 6.2.4.3) , while *feature evaluation* selects the most effective features (section 6.2.4.4). When the crowdsourcing is complete, the designer may request SCCM generation (section 6.2.4.5)

CHAPTER 6. SCCM/SCP CROWD-SOURCING

Initial goal of the system is to build a large answer base for all elements. This allows to define proper *averaged values* for components and features:

$$V'_{RC}(rc, a, s) = \sum_{n \in B_{AC}(rc, a, s)} \frac{V(rc, a, n)}{|B_{RC}(rc, a, s)|} \quad (\text{Eq. 6.30})$$

$$V'_{GC}(gc, s) = \sum_{n \in B_{GC}(gc, s)} \frac{V(gc, s)}{|B_{GC}(gc, s)|} \quad (\text{Eq. 6.31})$$

$$V'_{SC}(sc, s) = \sum_{n \in B_{SC}(sc, s)} \frac{V(sc, s)}{|B_{SC}(sc, s)|} \quad (\text{Eq. 6.32})$$

$$V'_{RF}(rf, a, s) = \sum_{n \in B_{RF}(rf, a, s)} \frac{V(rf, a, s)}{|B_{RF}(rf, a, s)|} \quad (\text{Eq. 6.33})$$

$$V'_{GF}(gf, s) = \sum_{n \in B_{GF}(gf, s)} \frac{V(gf, s)}{|B_{GF}(gf, s)|} \quad (\text{Eq. 6.34})$$

$$V'_{SF}(sf, s) = \sum_{n \in B_{SF}(sf, s)} \frac{V(sf, s)}{|B_{SF}(sf, s)|} \quad (\text{Eq. 6.35})$$

6.2.4.2 Correlation analysis

Next step is determination of the correlations between input percepts and the acquired social context components. For $X = \text{OP, AP, RF, GF, SF, RC, GC and SC}$, $\bar{V}_X$ is:

$$\bar{V}_X(x, a, n) = V'_X(x, a, n) - \bar{x}, \quad (\text{Eq. 6.36})$$

$$\bar{V}_X(x, n) = V'_X(x, n) - \bar{x}, \quad (\text{Eq. 6.37})$$

where $\bar{x}$ is a mean value of the x element in the question-answer database while U_X is:

CHAPTER 6. SCCM/SCP CROWD-SOURCING

$$U_X(x) = \sum_{QA(n) \in QA_X(x)} \bar{V}_X(x, n)^2, \quad (\text{Eq. 6.38})$$

it is possible to calculate Pearson correlation $C_{X,Y}$ between component, percept or feature sets X, Y . E.g. agent percept - role correlation $C_{AP,RC}$ is given as:

$$C_{AP,RC}(ap, rc) = \sum_{\substack{a \in A \\ q \in QA_{RC}(rc, a)}} \frac{\bar{V}_{AP}(ap, a, q) * \bar{V}_{RC}(rc, a, q)}{\sqrt{U_{AP}(ap) * U_{RC}(rc)}}, \quad (\text{Eq. 6.39})$$

while situation-group component correlation is:

$$C_{SC,GC}(sc, gc) = \sum_{q \in QA_{SC}(sc)} \frac{\bar{V}_{SP}(sp, q) * \bar{V}_{SC}(sc, q)}{\sqrt{U_{SP}(sp) * U_{SC}(sc)}}, \quad (\text{Eq. 6.40})$$

This way correlations between all component/feature combinations may be obtained.

6.2.4.3 Duplicate detection and merger

Unsupervised, undirected and open nature of questions often results in multiple related or synonymous references to similar social context components provided by the different subjects e.g. 'match', 'football match' and 'football competition'. Ideally, synonyms should be automatically detected and 'merged' in the dataset. However, the question when two specific terms are synonymous and when they are similar but distinct components is not just technically hard but fundamentally subjective and domain dependent. As in each application, the answer may be different for the same terms, it cannot be based e.g. on general linguistic data but on the domain character i.e. a) regularities in the obtained data and b) the choices of system designer understanding the domain. Therefore duplicate detection uses a) proximity tolerance parameter pt to limit similarity of terms merged as synonyms and b) manual markers forcing specific terms as 'distinct' or 'duplicates' allowing the designer to impose socio-cognitive 'granularity' when needed.

CHAPTER 6. SCCM/SCP CROWD-SOURCING

Distance between two components of one type, is defined as a difference in correlation with other percepts and components, e.g.:

$$C_{SC,GC}(sc, gc) = \sum_{q \in Q_{A_{SC}(ac)}} \frac{\bar{V}_{SP}(sp, q) * \bar{v}_{SC}(sc, q)}{\sqrt{U_{SP}(sp) * U_{SC}(sc)}}, \quad (\text{Eq. 6.41})$$

System identifies duplicate pairs, as fulfilling:

$$D_{GC}(g1, g2) < pt * |GP| * |SC| * |RC|, \quad (\text{Eq. 6.42})$$

$$D_{SC}(s1, s2) < pt * |SP| * |GC| * |RC|, \text{ or} \quad (\text{Eq. 6.43})$$

$$D_{RC}(r1, r2) < pt * |RP| * |SC| * |GC|, \quad (\text{Eq. 6.44})$$

where pt is the mentioned proximity tolerance parameter. Such pairs are merged in the database i.e. all references to them are replaced by a mutual new one, which decreases the number of components. Merger influences correlation of the remaining components and may cause subsequent mergers. At any time designer has access to an ordered list of correlated component pairs and the already merged duplicates and may mark them as 'distinct', 'duplicate' or 'unmarked' (Figure 6.3).

6.2.4.4 Feature evaluation

In parallel to the component acquisition, feature evaluation takes place. Percepts are used to estimate features in linear regression. For example, in the case of situation features the equations take the following form:

$$\tilde{V}_{SF}(f, s) = \sum_{p \in OP} (\alpha(p, f) + \beta(p, f)V_{OP}(p, s)) \quad (\text{Eq. 6.45})$$

CHAPTER 6. SCCM/SCP CROWD-SOURCING

ROLE COMPONENT	BASE SIZE	MERGER PAIRS				Proximity tolerance parameter: 0.2	
referee	89 QA	COMP 1	COMP 2	STATUS	Distance		
player	95 QA	referee	ref	duplicate	0.08		
goalkeeper	72 QA	referee	arbiter	duplicate	0.13		
father	110 QA	player	football player	duplicate	0.18		
child	78 QA	referee	assistant referee	duplicate (user set)	0.28		
spectator	99 QA	player	goalkeeper	distinct (user set)	0.13		
mother	80 QA	father	dad	distinct	0.23		
See Scene Correlations		Set as automanaged		Set as distinct		Set as duplicate	
See Group Correlations		See samples		See question-answer sets		Generate Model	

Figure 6.3: Snapshot of the role component control page. Merged candidate pairs are sorted by distance (right), three pairs (red) detected as duplicates based, two set by user designer (yellow) one classified as distinct (green). Components are listed with their base size, user can investigate their feature/scene/group correlations (left), generate model, samples and answers.

$$\beta(p, f) = \frac{\sum_{QA(n) \in QA_{SF}(f)} \bar{V}_{OP}(p, n) * \bar{V}_{SF}(f, n)}{\sum_{QA(n) \in QA_{SF}(f)} \bar{V}_{OP}(p, n)^2}, \quad (\text{Eq. 6.46})$$

$$\alpha(p, f) = \bar{p} - \bar{f} * \beta_{SP, SF}(p, f) \quad (\text{Eq. 6.47})$$

to fit SCP/SCCM Gaussian function is used in feature-component estimation:

$$\tilde{V}_{SC}(c, n) = \exp \left(- \sum_{f \in SF} \frac{(V_{SF}(f, n) - \bar{f})^2}{2D_{SC, SF}(c, f)^2} \right), \quad (\text{Eq. 6.48})$$

where $D_{SC, SF}(c, f)$ is the appropriate standard deviation. Estimation errors indicate the feature fit quality. Component value error like

$$E_{SC}(c) = \sum_{QA(n) \in QA_{SC}(c)} \frac{\tilde{V}_{SC}(c, n) - \bar{V}_{SC}(c, n)}{|QA_{SC}(c)|} \quad (\text{Eq. 6.49})$$

captures degree to which current features capture component activation. Feature error, e.g.

$$E_{SF}(f) = \sum_{QA(n) \in QA_{SF}(f)} \frac{\tilde{V}_{SF}(f, n) - \bar{V}_{SF}(f, n)}{|QA_{SF}(f)|} \quad (\text{Eq. 6.50})$$

CHAPTER 6. SCCM/SCP CROWD-SOURCING

measures the degree to which system is able to extract it from input percepts. Both parameters automatically help designer modify the system accordingly.

6.2.4.5 Model generation

At this stage, all the data to parameterize the SCCM/SCP is obtained. Correlations between the percepts, features and components may be used to parametrize generated models, based on the designer's instructions. Designer may use the complete dataset or select a subgroup of subjects, to tailor the creation process, based on specific features. For example, only gender or certain age group may be used as a basis for model creation.

Regardless of the subject group scope, components obtained in the crowd-sourcing may be used directly to generate SCCM/SCP group, role and situation prototypes (section 3.2.3). Specifically, $Y_s = SC$, $Y_g = GC$, $Y_r = RC$. The same time, correlations between the components form the prototype connection weights i.e.:

$$Vsg_{i,j} = C_{SC,GC}(s_i, g_j), \quad (\text{Eq. 6.51})$$

$$Vgs_{i,j} = C_{GC,SC}(g_i, s_j), \quad (\text{Eq. 6.52})$$

$$Vsr_{i,j} = C_{SC,RC}(s_i, r_j), \quad (\text{Eq. 6.53})$$

$$Vrs_{i,j} = C_{RC,SC}(r_i, s_j), \quad (\text{Eq. 6.54})$$

$$Vgr_{i,j} = C_{GC,RS}(g_i, r_j), \quad (\text{Eq. 6.55})$$

$$Vrg_{i,j} = C_{RC,GS}(r_i, g_j), \quad (\text{Eq. 6.56})$$

CHAPTER 6. EXAMPLE APPLICATION

Similarly, prototype features are established using the component feature set. Features chosen for the selected subject subset, in feature evaluation process (possibly modified by the designer), form SCCM/SCP prototype features i.e. $X_s = SF$, $X_g = GF$, $X_r = RF$. Feature influence on the prototypes is equal to the approximations provided in Eq. 6.45, i.e.

$$Hs_{i,j} = \tilde{V}_{SF}(f_i, s_j) \quad (\text{Eq. 6.57})$$

$$Hg_{i,j} = \tilde{V}_{SF}(f_i, g_j) \quad (\text{Eq. 6.58})$$

$$Hr_{i,j} = \tilde{V}_{SF}(f_i, r_j) \quad (\text{Eq. 6.59})$$

while D_s , D_g , and D_r are equal to the standard deviation in the collected feature data.

6.3 Example application

Short description of an example application of the automatic system is presented in this section, including used domain and samples, gathered data and analysis results and the outcome of validation.

6.3.1 Domain and samples

To verify the system behavior in a new domain, perception of the social context in sport and recreation has been selected. Samples consisting of images from family life through amateur sport to professional sports games were selected. Samples were preprocessed as described in section 6.2.1.1 i.e. objects like ball were tagged, agents were attributed percepts like perceived age, uniform, expression etc. Facial expressions were automatically

CHAPTER 6. EXAMPLE APPLICATION

analyzed using facial emotion extraction software where possible, using free and high-quality facial emotion extraction tool (Microsoft's Oxford Project). Initial superset of the scene, role, and group features was established, for example, scene feature dimensions for scenes were: intimate (vs. public), leisure (vs professional), relaxed (vs. intense).

6.3.2 Validation and result analysis

For example purposes, 50 preprocessed social context samples were chosen and ten subjects were questioned by the system, as described in section 6.2.3. System obtained 18 role components, 11 situation components and 8 group components that were merged into 9 role, 7 situation and 5 group components.

Initial system validation has three aspects: feature and component error evaluation; evaluation based on the samples unused in the parameter generation and finally, evaluation of the generated models in continuous stream stability tests.

In the samples used in parametrization, average feature estimation error was 0.19, with standard deviation = 0.079, while component estimation error was 0.12, with standard deviation = 0.032. This means that for the roles, scene and group components, the system performance was similar to human subject performance and on average slightly better in the case of features. Especially features proved unequally difficult to capture, e.g 'intimate' dimension was much harder to capture from selected percepts than e.g. 'leisure'. This may suggest that more percepts are required or nonlinear percept-feature relations must be included in the future.

Additional ten preprocessed samples were excluded from the parametrization and used in second validation that checked generalization. System estimation of context features and components was compared with subject provided data. Here, average feature estimation error was 0.27, with a standard deviation of 0.068, while component estimation error was 0.19, with standard deviation = 0.031. The system performed reasonably, but slightly weaker than the previous test.

CHAPTER 6. SUMMARY

The final stage was a generation of 50 SCPs based on the analysis. They were subjected to several prepared scenarios consisting of percept values changing dynamically with subject-based data acquired at several time-points. The purpose of this part was to evaluate if SCPs generated by the system retain context recognition in the continuous stream of context classification.

Comparing time-point data points with system estimation of context features gave and average feature estimation error of 0.25, with standard deviation = 0.075 and component estimation error of 0.22, with standard deviation = 0.036. This performance is comparable to results of static data classification from the second test, suggesting stability of system-generated SCPs.

6.4 Summary

This chapter has presented a semi-automatic method for parameterizing the social context cognition model SCP [KTT15a]. It discussed the need for automatic methods of social context cognition model parametrization e.g. for personality or culture modeling. It described a new method of SCCM and SCP parametrization based on crowd-sourcing. Results were validated and the system showed the ability to recreate social context cognition in the presented samples, good generalization for unknown samples and stability in a continuous stream of dynamic data. The presented system may allow broader application of much needed social context cognition models whose troublesome parametrization was their main issue, contributing to the quality of believable agents and social simulations.

Chapter 7

Conclusions

7.1 Summary of contributions

This thesis presented a contribution to computational models of human-like non-rationality. A critical, interdisciplinary analysis of the current trends in non-rationality modeling was presented, containing social context cognition, personality and culture modeling in domains including Social Signal Processing, Multi-Agent Simulations, Cognitive Affective Architectures and other fields was presented with constructive suggestions. Based on this analysis, models of cognitively rich agents with stereotypic, habitual and irrational human-like characteristics were proposed. The introduced models of 'System 1' [Kah11] social context cognition were used to introduce a novel approach to personality, culture, and non-appraisal affect generation. More importantly, they were used in social agent architectures in Multi-Agent Simulations, placed in a continual stream of sensory input and tested in a series of simulations. Simulations confirmed their capability to generate believable 'System 1' social context cognition and behavior, and their ability to recreate many patterns observable in human behavior. Finally, as the proposed models are parameter-heavy, a semi-automatic, crowd-sourcing based method of their parametrization was presented. The main contributions of this dissertation work are summarized as follows.

CHAPTER 7. SUMMARY OF CONTRIBUTIONS

- (i) A comprehensive interdisciplinary review of the state-of-the-art and the current trends in non-rationality modeling. Its scope included works related to social context cognition, personality and culture modeling, in various domains including Social Signal Processing, Multi-Agent Simulations, Cognitive-Affective Architectures and other fields.
- (ii) Social Context Cognition Model (SCCM) - a connectionist model and the network activations it establishes reflect agent's informal, 'System 1' [Kah11] social associations induced by a perceived situation. SCCM captures non-conceptual, habitual and associative connotations and their behavioral consequences, rather than rule and reasoning based, calculative ones. As an implementation of the social dimension of 'System 1', it may simulate its various effects on cognition discovered in humans [Kah11]. SCCM is usable in simulations oriented both on entertainment (e.g. games) and scientific, social simulation and may be parameterized based on the sociological data, by hand or by semi-automatic methods.
- (iii) A Dramaturgical Dissonance Model (DDM) - an alternative to SCCM, a component usable in social situation modeling, inspired by the Dramaturgical [BE05, Gof59], and Cognitive Dissonance [Fes62] theories. Simulations were used to depict DDM's capabilities to produce believable behaviors based on the relevant psychological theories. Example scenarios and simulations were provided, to show how habitual, social role and scene patterns could be used to generate emotional responses or plans in a believable way.

Believability of the provided solutions was tested and a DDM-SCCM comparison is provided with an analysis of their relative strengths and weaknesses and practical suggestions about their use. DDM shows its advantage when a) dissonances and

CHAPTER 7. LIMITATIONS OF THE PRESENTED MODELS AND FUTURE WORK

other formal relations between partial context cognitions are relevant b) personalized processing of known agents is relevant, and individualized data is available c) subjective believability, rather than scientific accuracy is the target. Conversely, SCCM is better in cases where a 'black box' solution providing social context cognition is sufficient.

- (iv) Social Context based Personality (SCP) model - a direct specialization of Social Context Cognition Model (SCCM) to the problem of personality modeling. SCP is an attempt to incorporate contextual aspect into personality models, to overcome the limitation of currently dominating, simplistic 'trait' models. It uses a connectionist representation grounded in the personality theories largely ignored in the computational modeling. Results show that SCCM/SCP can recreate various regularities observed in the real data and is promising in both entertainment and social simulation oriented applications
- (v) The semi-automatic method of parametrization of the socio-cognitive models based on crowd-sourcing. As proposed models depend on data validity, are parameter-heavy, and their parametrization is challenging, such automatization was crucial for practical purposes. The generated models are used in simulations and validated. Such semi-automatic SCP and SCCM generation may be used in similar human-like behavior modulation or personality models but also in multi-agent socio-cultural simulations.

7.2 Limitations of the presented models and future work

Like any others, the presented models make simplifications to capture complex phenomena and have numerous limitations. Those limitations include:

CHAPTER 7. LIMITATIONS OF THE PRESENTED MODELS AND FUTURE WORK

- lack of on-line learning and exchange of the cultural contents although a frame for such on-line learning is provided with an automatic but offline model generation.
- lack of large-scale simulation of cultural exchange of modeled subconscious material, based on the above-mentioned on-line learning.
- abstract level of stereotype processing not linked to the neural, biological level and purely pragmatic, behavioral similarity to human subsymbolic cognition that may be a limitation for some applications.
- contextuality limited to roles, situations, and groups. Although useful, this distinction is most probably less clear-cut and limited in human cognition and their context elements or even machine learning methods could be useful in capturing such elements.
- assumed linear relations between some of the model parameters (e.g. features and prototypes) that may have been better modeled with higher order functions.
- concrete and rather narrow domain of the example contexts and need expansion for a new context that may demand time requiring the acquisition of role-situation-group prototype sets with their connections, even with the presented semi-automatic method.

For this reason, there are many possible directions for further research on the modeling of non-rationality. As mentioned, one fundamental issue, is the introduction of the on-line learning into the proposed models, instead of the current off-line generation. While off-line generated models of human-like non-rationality are a relevant first step, culturally obtained and personally developed idiosyncrasies accumulate over time, based on the specific events, experiences, observed patterns, social feedback, and communication from other agents. Such diverse, heterogenic sources of non-rationality must be

CHAPTER 7. LIMITATIONS OF THE PRESENTED MODELS AND FUTURE WORK

reflected by several types of unsupervised, supervised, and reinforced on-line learning based on penalization, observation, communication or imitation. Such learning would not only be useful in itself, as allowing model and simulation-based formulation and testing of hypotheses concerning the individual development of non-rationality but could also open the possibility to dynamically, 'organically' grow 'System 1' informal cultures. Social groups, with 'System 1' culture and stereotypes resulting from the interactions and events in the society, would be a powerful tool in socio-cultural simulations and, even more so, in entertainment. One example would be games in which interactions of thousands of artificial and real agents shape cognitive contents of the populations, their in-group and out-group stereotypes, militaristic vs. pacifistic, democratic vs. authoritarian or religious vs. secular character. This kind of solutions are imaginable from a perspective of the proposed models, and they would provide a very new quality compared to the currently existing products and solutions.

Another challenge is the development of more sophisticated classifiers. As shown in chapter 6, one of the bigger limitations of SCP/SCCM was the linear relation between the observable input and the perceived context features. Exchanging them for a higher order feature extractors based on machine learning techniques could significantly improve the quality of the results. This problem is related to further automatization of the data acquisition process, especially in the low-level, machine learning based feature extraction.

Next broad issue, is a tighter linking of the current social context cognition to the human-like goal and motivation structures, including social awareness and theory of the mind. While SCCM provides a natural mechanism of influence of social context perception on the own motivation structures and those attributed to others, this influence should be mutual. The problem of human-like non-rationality in motivation structures is also closely related to the evolutionary or memetic perspective i.e. one seeing people not only as autonomous agents but also ones realizing agenda of other entities, be it biological (genes) or ideological ones (religions, nations, tribes, beliefs, ideologies). Networks

CHAPTER 7. LIMITATIONS OF THE PRESENTED MODELS AND FUTURE WORK

of those objects survive and replicate through their 'manipulation' of human values, motivation structures (like social goals), habits, stereotypes and other aspects of cognition. Social simulation based on this assumption would not only go against the notion of rationality but also autonomy as it is currently understood in the computational approach to agents as such. It may then not only require new models compatible with current agent architectures but a reconceptualization of the notion of agency itself.

Finally, parallel to the challenges presented above it is necessary to expand the application of similar models to practical simulations. It is crucial both as a continuation of the validation process, and further investigation of the actual boundaries of this approach, but also as a part of the effort of the creation of the social context 'libraries' reusable in new simulations.

Journal Publications

- [KTT15] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Dramaturgical and dissonance theories in explicit social context modeling for complex agents. *Computer Animation and Virtual Worlds*, 26(3-4):247–257, 2015.
- [KTT16a] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Social context cognition crowd-sourcing and semi-automatic parametrization. *Computer Animation and Virtual Worlds*, 27(3-4):330–339, 2016.
- [KTT16b] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Social context cognition model and personality generation. *Submitted*, 2016.

Conference Papers

- [KTT12] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Dramaturgical metaphor in believable social agent modeling. In *Proceedings of the 2012 Conference on Computer Animation and Social Agents (CASA)*, 2012.
- [KTT13] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Enabling generative, emergent artificial culture. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2013.
- [KTT15a] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Beyond traits: Social context based personality model. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1529–1538. International Foundation for Autonomous Agents and Multiagent Systems, 2015.
- [KTT15b] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Dramaturgical and dissonance theories in explicit social context modeling for complex agents. *Computer Animation and Virtual Worlds*, 26(3-4):247–257, 2015.

References

- [AA14] Mostafa Z Ali and Noor H Awad. A novel class of niche hybrid cultural algorithms for continuous engineering optimization. *Information Sciences*, 267:158–190, 2014.
- [ABG⁺17] Reyhan Aydoğan, Tim Baarslag, Enrico Gerding, Catholijn M Jonker, Vicente Julian, and Victor Sanchez-Anguix. *Conflict Resolution in Decision Making: Second International Workshop, COREDEMA 2016, The Hague, The Netherlands, August 29-30, 2016, Revised Selected Papers*, volume 10238. Springer, 2017.
- [ACCC10] Giulia Andrighetto, Marco Campenn, Federico Cecconi, and Rosaria Conte. *The complex loop of norm emergence: A simulation model*, pages 19–35. Springer, 2010.
- [ACCP07] Giulia Andrighetto, Marco Campennì, Rosaria Conte, and Marco Paolucci. On the immergence of norms: a normative agent architecture. In *Proceedings of AAAI symposium, social and organizational aspects of intelligence*, page 19, 2007.
- [AHL09] Carole Adam, Andreas Herzig, and Dominique Longin. A logical formalization of the occ theory of emotions. *Synthese*, 168(2):201–248, 2009.
- [AKG⁺00] Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. Integrating models of personality and emotions into life-like characters. In *Affective interactions*, pages 150–165. Springer, 2000.
- [AL14] Carole Adam and Emiliano Lorini. A bdi emotional reasoning engine for an artificial companion. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 66–78. Springer, 2014.

REFERENCES

- [And07] John Anderson. *How can the human mind occur in the physical universe?* Oxford University Press, 2007.
- [Arn60] Magda Arnold. *Emotion and personality*. 1960.
- [AVA⁺09] Ruth Aylett, Natalie Vannini, Elisabeth Andre, Ana Paiva, Sibylle Enz, and Lynne Hall. But that was in another country: agents and intercultural empathy. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 329–336, 2009.
- [Axe86] Robert Axelrod. An evolutionary approach to norms. *American political science review*, 80(04):1095–1111, 1986.
- [Bar08] Lawrence Barsalou. Grounded cognition. *Annu. Rev. Psychol.*, 59:617–645, 2008.
- [BC10] Cristina Bicchieri and Alex Chavez. Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178, 2010.
- [BDH⁺01] Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang, and Leendert van der Torre. The boid architecture: conflicts between beliefs, obligations, intentions and desires. In *Proceedings of the fifth international conference on Autonomous agents*, pages 9–16. ACM, 2001.
- [BDK08] Joost Broekens, Doug Degroot, and Walter Kusters. Formal models of appraisal: Theory, specification, and computational model. *Cognitive Systems Research*, 9(3):173–197, 2008.
- [BE05] Dennis Brissett and Charles Edgley. *Life as theater: a dramaturgical sourcebook*. Transaction Publishers, 2005.
- [Bic06] Cristina Bicchieri. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2006.
- [Ble09] Wiebke Bleidorn. Linking personality states, current social roles and major life goals. *European Journal of Personality*, 23(6):509–530, 2009.
- [BM11] Cristina Bicchieri and Ryan Muldoon. *Social norms*. 2011.

REFERENCES

- [BNR06] Melissa Bateson, Daniel Nettle, and Gilbert Roberts. Cues of being watched enhance cooperation in a real-world setting. *Biology letters*, 2(3):412–414, 2006.
- [Bou05] Pierre Bourdieu. *The social structures of the economy*. Polity, 2005.
- [BR95] Cristina Bicchieri and Carlo Rovelli. Evolution and revolution the dynamics of corruption. *Rationality and Society*, 7(2):201–224, 1995.
- [Bra87] Michael Bratman. Intention, plans, and practical reason. 1987.
- [Bro08] Melissa Brown. *Explaining culture scientifically*. University of Washington Press, 2008.
- [BTDH11] John Mc Breen, Gennaro Di Tosto, Frank Dignum, and Gert Jan Hofstede. Linking norms and culture. In *Culture and Computing (Culture Computing), 2011 Second International Conference on*, pages 9–14. IEEE, 2011.
- [BU92] Michael Boulton and Kerry Underwood. Bully/victim problems among middle school children. *British Journal of Educational Psychology*, 62(1):73–87, 1992.
- [BvdT03] Guido Boella and Leendert van der Torre. Attributing mental attitudes to normative systems. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. ACM, 2003.
- [BVDTV08] Guido Boella, Leendert Van Der Torre, and Harko Verhagen. Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17(1):1–10, 2008.
- [CB98] Nicki Crick and Maureen Bigbee. Relational and overt forms of peer victimization: a multiinformant approach. *Journal of consulting and clinical psychology*, 66(2):337, 1998.
- [CC94] James Coleman and James Samuel Coleman. *Foundations of social theory*. Harvard University Press, 1994.

REFERENCES

- [CDJT00] Cristiano Castelfranchi, Frank Dignum, Catholijn Jonker, and Jan Treur. Deliberative normative agents: Principles and architecture. In *Intelligent Agents VI. Agent Theories, Architectures, and Languages*, pages 364–378. Springer, 2000.
- [CFK⁺09] Amy Cuddy, Susan Fiske, Virginia Kwan, Peter Glick, Stephanie Demoulin, Jacques-Philippe Leyens, Michael Harris Bond, Jean-Claude Croizet, Naomi Ellemers, Ed Sleebos, et al. Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1):1–33, 2009.
- [CHK⁺05] Mark Chapell, Stefanie Hasselman, Theresa Kitchin, Safiya Lomon, Kenneth MacIver, and Patrick Sarullo. Bullying in elementary school, high school, and college. *Adolescence*, 41(164):633–648, 2005.
- [CL03] Cristiano Castelfranchi and Emiliano Lorini. Cognitive anatomy and functions of expectations. In *IJCAI03 Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions*, pages 9–11. Citeseer, 2003.
- [CM98] Kevin Corcoran and Jennifer Michels. A prototype analysis of psychological situations through the lens of alcohol expectancies and gender. *Addictive behaviors*, 23(5):685–691, 1998.
- [CMS82] Nancy Cantor, Walter Mischel, and Judith Schwartz. A prototype analysis of psychological situations. *Cognitive psychology*, 14(1):45–77, 1982.
- [CNCM12] Francesco Chirico, Mattias Nordqvist, Gianluca Colombo, and Edoardo Mollona. Simulating dynamic capabilities and value creation in family firms: Is paternalism an asset or a liability? *Family Business Review*, 25(3):318–338, 2012.
- [Col04] Randall Collins. *Interaction ritual chains*. Princeton university press, 2004.
- [CRK90] Robert Cialdini, Raymond Reno, and Carl Kallgren. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015, 1990.

REFERENCES

- [CRR07] Gary Charness, Luca Rigotti, and Aldo Rustichini. Individual behavior and group membership. *The American Economic Review*, 97(4):1340–1352, 2007.
- [CS99] Daniel Cervone and Yuichi Shoda. *The coherence of personality: Social-cognitive bases of consistency, variability, and organization*. Guilford Press, 1999.
- [CSJ⁺10] Sherol Chen, Adam M Smith, Arnav Jhala, Noah Wardrip-Fruin, and Michael Mateas. Rolemodel: towards a formal model of dramatic roles for story generation. In *Proceedings of the Intelligent Narrative Technologies III Workshop*, page 17. ACM, 2010.
- [CTN07] Hui-Qing Chong, Ah-Hwee Tan, and Gee-Wah Ng. Integrated cognitive architectures: a survey. *Artificial Intelligence Review*, 28(2):103–130, 2007.
- [CWM05] Damon Centola, Robb Willer, and Michael Macy. The emperor’s dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, 110(4):1009–1040, 2005.
- [DGE⁺00] Jeffrey Dean, George Gumerman, Joshua Epstein, Robert Axtell, Alan Swedlund, Miles Parker, and Stephen McCarroll. Understanding anasazi culture change through agent-based modeling. *Dynamics in human and primate societies*, pages 179–205, 2000.
- [Dic04] Oxford English Dictionary. Oxford english dictionary online. *Mount Royal College Lib., Calgary*, 14, 2004.
- [DL12] Mehdi Dastani and Emiliano Lorini. A logic of emotions: from appraisal to coping. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 1133–1140. International Foundation for Autonomous Agents and Multiagent Systems, 2012.
- [DMHvdB12] Jeroen De Man, Annerieke Heuvelink, and Karel van den Bosch. A cognitive model for social role compliant behavior of virtual agents. In *International Conference on Intelligent Virtual Agents*, pages 303–310. Springer, 2012.

REFERENCES

- [DMP14] Joao Dias, Samuel Mascarenhas, and Ana Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer, 2014.
- [DOP08] Wlodzislaw Duch, Richard Jayadi Oentaryo, and Michel Pasquier. Cognitive architectures: Where do we go from here? *Frontiers in artificial intelligence and applications*, 171:122, 2008.
- [dSNdSdL13] Baldoino dos Santos Neto, Viviane Torres da Silva, and Carlos de Lucena. *Developing Goal-Oriented Normative Agents: The NBDI Architecture*, pages 176–191. Springer, 2013.
- [Dur95] Emile Durkheim. The rules of sociological method. *Durkheim: The Rules of Sociological Method and Selected Texts on Sociology and Its Method*, pages 31–163, 1895.
- [EA96] Joshua Epstein and Robert Axtell. *Growing artificial societies: social science from the bottom up*. MIT press, 1996.
- [EG01] Nicholas Epley and Thomas Gilovich. Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12(5):391–396, 2001.
- [Els89] Jon Elster. Social norms and economic theory. *The Journal of Economic Perspectives*, 3(4):99–117, 1989.
- [EP09] Micah Eckhardt and Rosalind Picard. A more effective way to label affective expressions. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–2. IEEE, 2009.
- [Eps79] Seymour Epstein. The stability of behavior: I. on predicting most of the people much of the time. *Journal of personality and social psychology*, 37(7):1097, 1979.
- [Eys91] Hans Jurgen Eysenck. Dimensions of personality: 16, 5 or 3? criteria for a taxonomic paradigm. *Personality and individual differences*, 12(8):773–790, 1991.

REFERENCES

- [FCG07] Susan T Fiske, Amy JC Cuddy, and Peter Glick. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2):77–83, 2007.
- [FCPM13] Caroline Faur, Céline Clavel, Sylvie Pesty, and Jean-Claude Martin. Perseed: a self-based model of personality for virtual agents inspired by socio-cognitive theories. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 467–472. IEEE, 2013.
- [Fes62] Leon Festinger. *A theory of cognitive dissonance*, volume 2. Stanford university press, 1962.
- [FJ13] Tim Faber and Kai Jonas. Perception in a social context: Attention for response-functional means. *Social Cognition*, 31(2):301–314, 2013.
- [FKTS89] Nico Frijda, Peter Kuipers, and Elisabeth Ter Schure. Relations among emotion, appraisal, and emotional action readiness. *Journal of personality and social psychology*, 57(2):212, 1989.
- [FN08] William Fleeson and Erik Nettle. The end of the person–situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, 2(4):1667–1684, 2008.
- [For97] Jay Wright Forrester. Industrial dynamics. *Journal of the Operational Research Society*, 48(10):1037–1041, 1997.
- [Geb05] Patrick Gebhard. Alma: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM, 2005.
- [GLR11] Navid Ghaffarzadegan, John Lyneis, and George P Richardson. How small system dynamics models can help the public policy process. *System Dynamics Review*, 27(1):22–44, 2011.
- [GM04] Jonathan Gratch and Stacy Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004.

REFERENCES

- [GOD06] Frits Goossens, Olthof, and Peter Dekker. New participant role scales: Comparison between various criteria for assigning roles and indications for their validity. *Aggressive Behavior*, 32(4):343–357, 2006.
- [Gof59] Erving Goffman. *The presentation of self in everyday life*. Doubleday, Garden City, N.Y., 1959.
- [GP09] Daniel Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12):1775–1787, 2009.
- [Gro82] Stephen Grossberg. How does a brain build a cognitive code? In *Studies of Mind and Brain*, pages 1–52. Springer, 1982.
- [Gun10] Hatice Gunes. Automatic, dimensional and continuous emotion recognition. 2010.
- [GZKB14] Thomas Gumpel, Vered Zioni-Koren, and Zvi Bekerman. An ethnographic study of participant roles in school bullying. *Aggressive behavior*, 40(3):214–228, 2014.
- [HD12] Ross A Hammond and Laurette Dubé. A systems science perspective and transdisciplinary models for food and nutrition security. *Proceedings of the National Academy of Sciences*, 109(31):12356–12363, 2012.
- [HHC09] Brian Heath, Raymond Hill, and Frank Ciarallo. A survey of agent-based modeling practices (january 1998 to july 2008). *Journal of Artificial Societies and Social Simulation*, 12(4):9, 2009.
- [HK04] Geoffrey Hodgson and Thorbjørn Knudsen. The complex evolution of a simple traffic convention: the functions and implications of habit. *Journal of Economic Behavior & Organization*, 54(1):19–47, 2004.
- [HNM05] Martie Haselton, Daniel Nettle, and Damian Murray. The evolution of cognitive bias. *The handbook of evolutionary psychology*, 2005.
- [Hof01] Geert Hofstede. *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage, 2001.

REFERENCES

- [Hua09] Lei Huang. *The investigation of personality states and situational characteristics in customer service jobs*. ProQuest, 2009.
- [Hud04] Eva Hudlicka. Two sides of appraisal: Implementing appraisal and its consequences within a cognitive architecture. In *Proceedings of the AAAI Spring Symposium 2004, Architectures for Modeling Emotion*, 2004.
- [JADT12] Jie Jiang, Huib Aldewereld, Virginia Dignum, and Yao-Hua Tan. A context-aware normative structure in mas. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1371–1372. International Foundation for Autonomous Agents and Multiagent Systems, 2012.
- [JBFW⁺04] W Lewis Johnson, Carole Beal, Anna Fowles-Winkler, Ursula Lauper, Stacy Marsella, Shrikanth Narayanan, Dimitra Papachristou, and Hannes Vilhjálmsson. Tactical language training system: An interim report. In *Intelligent Tutoring Systems*, pages 336–345. Springer, 2004.
- [JS44] Mill John Stuart. On the definition of political economy, and on the method of investigation proper to it. *Essays on some unsettled questions of Political Economy*, page 326, 1844.
- [Kah11] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
- [Kam13] Gal Kaminka. Curing robot autism: A challenge. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 801–804. International Foundation for Autonomous Agents and Multiagent Systems, 2013.
- [KASA08] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, and Mo Nours Arab. Human-computer interaction: Overview on state of the art. *International Journal on Smart Sensing and Intelligent Systems*, 1(1):137–159, 2008.
- [KBB13] Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *Affective Computing, IEEE Transactions on*, 4(1):15–33, 2013.

REFERENCES

- [KC09] Yongjun Kim and Sung-Bae Cho. A hybrid cultural algorithm with local search for traveling salesman problem. In *Computational Intelligence in Robotics and Automation (CIRA), 2009 IEEE International Symposium on*, pages 188–192. IEEE, 2009.
- [KCCS05] Edward Chao-Chun Kao, Paul Hsueh-Min Chang, Yu-Hung Chien, and Von-Wun Soo. *Using Ontology to Establish Social Context and Support Social Reasoning*, volume 3661 of *Lecture Notes in Computer Science*, chapter 29, pages 344–357. Springer Berlin Heidelberg, 2005.
- [KDL⁺08] Amit Kagian, Gideon Dror, Tommer Leyvand, Isaac Meilijson, Daniel Cohen-Or, and Eytan Ruppín. A machine learning predictor of facial attractiveness revealing human-like psychophysical biases. *Vision research*, 48(2):235–243, 2008.
- [KF04] Daniel Kahneman and Shane Frederick. Attribute substitution in intuitive judgment. *Models of a man: Essays in memory of Herbert A. Simon*, pages 411–432, 2004.
- [Kit93] James Kittock. Emergent conventions and the structure of multi-agent systems. *L. Nadel and D. Stein, eds*, 1993.
- [KM97] David Kieras and David Meyer. An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction*, 12(4):391–438, 1997.
- [KN03] Martin J Kollingbaum and Timothy J Norman. Noa-a normative agent architecture. In *IJCAI*, pages 1465–1466, 2003.
- [KR05] Ziad Kobti and Robert G Reynolds. Modeling protein exchange across the social network in the village multi-agent simulation. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 4, pages 3197–3203. IEEE, 2005.
- [KRK04] Ziad Kobti, Robert G Reynolds, and Tim Kohler. The effect of kinship cooperation learning strategy and culture on the resilience of social systems in the village multi-agent simulation. In *Evolutionary Computation, 2004. CEC2004. Congress on*, volume 2, pages 1743–1750. IEEE, 2004.

REFERENCES

- [Ksh02] Sumedha Kshirsagar. A multilayer personality model. In *Proceedings of the 2nd international symposium on Smart graphics*, pages 107–115. ACM, 2002.
- [KT77] Daniel Kahneman and Amos Tversky. Intuitive prediction: Biases and corrective procedures. Technical report, DTIC Document, 1977.
- [KTT12] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Dramaturgical metaphor in believable social agent modeling. In *Proceedings of the 2012 Conference on Computer Animation and Social Agents (CASA)*, 2012.
- [KTT13] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Enabling generative, emergent artificial culture. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2013.
- [KTT15a] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Beyond traits: Social context based personality model. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1529–1538. International Foundation for Autonomous Agents and Multiagent Systems, 2015.
- [KTT15b] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Dramaturgical and dissonance theories in explicit social context modeling for complex agents. *Computer Animation and Virtual Worlds*, 26(3-4):247–257, 2015.
- [KTT16a] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Social context cognition crowd-sourcing and semi-automatic parametrization. *Computer Animation and Virtual Worlds*, 27(3-4):330–339, 2016.
- [KTT16b] Jaroslaw Kochanowicz, Ah-Hwee Tan, and Daniel Thalmann. Social context cognition model and personality generation. *Submitted*, 2016.
- [Lai12] John Laird. *The Soar cognitive architecture*. MIT Press, 2012.

REFERENCES

- [LC] Pat Langley and Dongkyu Choi. A unified cognitive architecture for physical agents. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 1469. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
- [LL01] Alessandro Lomi and Erik R Larsen. *Dynamics of organizations: Computational modeling and organization theories*. Mit Press, 2001.
- [LLR09] Pat Langley, John Laird, and Seth Rogers. Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10(2):141–160, 2009.
- [LPER05] Mark Levine, Amy Prosser, David Evans, and Stephen Reicher. Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, 31(4):443–453, 2005.
- [LS11] Emiliano Lorini and François Schwarzentruher. A logic for reasoning about counterfactual emotions. *Artificial Intelligence*, 175(3-4):814–847, 2011.
- [LSZ12] Jerry Lin, Marc Spraragen, and Michael Zyda. Computational models of emotion and cognition. In *Advances in Cognitive Systems*. Citeseer, 2012.
- [McA95] Dan McAdams. What do we know when we know a person? *Journal of personality*, 63(3):365–396, 1995.
- [MDBC06] Zopito Marini, Andrew Dane, Sandra Bosacki, and YLC Cura. Direct and indirect bully-victims: differential psychosocial risk factors associated with adolescents involved in bullying and victimization. *Aggressive Behavior*, 32(6):551–569, 2006.
- [Mes11] Alex Mesoudi. *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences*. University of Chicago Press, 2011.
- [Mey08] John-Jules Meyer. Reasoning about emotional agents. *Artificial Intelligence Preprint Series*, 44, 2008.

REFERENCES

- [MG06] Stacy Marsella and Jonathan Gratch. Ema: A computational model of appraisal dynamics. *Agent construction and emotions*, pages 601–606, 2006.
- [MGP10] Stacy Marsella, Jonathan Gratch, and Paolo Petta. Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, pages 21–46, 2010.
- [MIL04] Robert P Marinier III and John E Laird. Toward a comprehensive computational model of emotions and feelings. In *ICCM*, pages 172–177, 2004.
- [Mis68] Walter Mischel. *Personality and assessment*. 1968.
- [Mit96] Steven J Mithen. *The prehistory of the mind a search for the origins of art, religion and science*. 1996.
- [MJ92] Robert McCrae and Oliver John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.
- [MJ97] Helen Mynard and Stephen Joseph. Bully/victim problems and their association with eysenck’s personality dimensions in 8 to 13 year-olds. *British Journal of Educational Psychology*, 67(1):51–54, 1997.
- [MLL09] Robert P Marinier, John E Laird, and Richard L Lewis. A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, 10(1):48–69, 2009.
- [MLT09] Matthias Meyer, Iris Lorscheid, and Klaus Troitzsch. The development of social simulation as reflected in the first ten years of jasss: a citation and co-citation analysis. *Journal of Artificial Societies and Social Simulation*, 12(4):12, 2009.
- [MM14] Morris and Daniel McDuff. Crowdsourcing techniques for affective computing. *Handbook of Affective Computing*, 2014.
- [MMR13] Ignacio J Martinez-Moyano and George P Richardson. Best practices in system dynamics modeling. *System Dynamics Review*, 29(2):102–123, 2013.
- [Mof97] Dave Moffat. Personality parameters and programs. In *Creating personalities for synthetic actors*, pages 120–165. Springer, 1997.

REFERENCES

-
- [MR81] James McClelland and David Rumelhart. An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological review*, 88(5):375, 1981.
 - [MRHU11] Alexis Morris, William Ross, Hadi Hosseini, and Mihaela Ulieru. Modelling culture with complex, multi-dimensional, multi-agent systems. *Integrating Cultures: Formal Models and Agent-Based Simulations (In Submission)*. Springer, 2011.
 - [MS95] Walter Mischel and Yuichi Shoda. A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological review*, 102(2):246, 1995.
 - [MS12] Mario Mikulincer and PHILIP R SHAVER. An attachment perspective on psychopathology. *World Psychiatry*, 11(1):11–15, 2012.
 - [MSM⁺12] Margaret McRorie, Ian Sneddon, Gary McKeown, Elisabetta Bevacqua, Etienne De Sevin, and Catherine Pelachaud. Evaluation of four designed virtual agent personalities. *Affective Computing, IEEE Transactions on*, 3(3):311–322, 2012.
 - [MW10] François Mairesse and Marilyn Walker. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, 20(3):227–278, 2010.
 - [NC84] Paula Niedenthal and Nancy Cantor. Making use of social prototypes: From fuzzy concepts to firm decisions. *Fuzzy sets and systems*, 14(1):5–27, 1984.
 - [Neu08] Martin Neumann. Homo socionicus: a case study of simulation models of norms. *Journal of Artificial Societies and Social Simulation*, 11(4):6, 2008.
 - [Neu10] Martin Neumann. A classification of normative architectures. *Simulating Interacting Agents and Social Phenomena*, pages 3–18, 2010.
 - [Neu12] Martin Neumann. The cognitive legacy of norm simulation. *Artificial Intelligence and Law*, 20(4):339–357, 2012.

REFERENCES

- [Ng09] Gee Wah Ng. *Brain-mind machinery: brain-inspired computing and mind opening*. World Scientific, 2009.
- [NH11] Muaz Niazi and Amir Hussain. Agent-based computing from multi-agent systems to agent-based models: a visual survey. *Scientometrics*, 89(2):479–499, 2011.
- [NTB⁺11] Michael Neff, Nicholas Toothman, Robeson Bowmani, Jean Fox Tree, and Marilyn Walker. Dont scratch! self-adaptors reflect emotional stability. In *Intelligent Virtual Agents*, pages 398–411. Springer, 2011.
- [OCC90] Andrew Ortony, Gerald Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.
- [Olw94] Dan Olweus. Bullying at school: basic facts and effects of a school based intervention program. *Journal of child psychology and psychiatry*, 35(7):1171–1190, 1994.
- [ØR01] Kristina Østvik and Floyd Rudmin. Bullying and hazing among norwegian army soldiers: Two studies of prevalence, context, and cognition. *Military Psychology*, 13(1):17, 2001.
- [PCD⁺11] Maja Pantic, Roderick Cowie, Francesca D’Errico, Dirk Heylen, Marc Mehu, Catherine Pelachaud, Isabella Poggi, Marc Schroeder, and Alessandro Vinciarelli. Social signal processing: the research agenda. In *Visual analysis of humans*, pages 511–538. Springer, 2011.
- [Per90] Lawrence Pervin. A brief history of modern personality theory. 1990.
- [PI01] Helmut Prendinger and Mitsuru Ishizuka. Social role awareness in animated agents. In *Proceedings of the fifth international conference on Autonomous agents*, pages 270–277. ACM, 2001.
- [Pia13] Fabio Pianesi. Searching for personality [social sciences]. *Signal Processing Magazine, IEEE*, 30(1):146–158, 2013.
- [Pic95] Rosalind Picard. Affective computing. 1995.

REFERENCES

-
- [PMP⁺02] KR Pattipati, C Meirina, A Pete, GM Levchuk, S Ruan, and DL Kleinman. Decision networks and command organizations. *Systems Engineering and Management for Sustainable Development, Encyclopedia of Life Support Systems*, 2002.
- [PSA⁺51] Talcott Parsons, Edward Shils, Gordon Allport, Clyde Kluckhohn, Henry Murray, Robert Sears, Richard Sheldon, Samuel Stouffer, and Edward Tolman. Some fundamental categories of the theory of action: A general statement. *Toward a general theory of action*, pages 3–29, 1951.
- [PT05] Mike Poznanski* and Paul Thagard. Changing personalities: towards realistic virtual characters. *Journal of Experimental & Theoretical Artificial Intelligence*, 17(3):221–241, 2005.
- [Puj06] Josep Pujol. *Structure in artificial societies*. PhD thesis, 2006.
- [PZN⁺08] Fabio Pianesi, Massimo Zancanaro, Elena Not, Chiara Leonardi, Vera Falcon, and Bruno Lepri. Multimodal support to group dynamics. *Personal and Ubiquitous Computing*, 12(3):181–195, 2008.
- [RA10] George P Richardson and David F Andersen. Systems thinking, mapping, and modeling in group decision and negotiation. In *Handbook of group decision and negotiation*, pages 313–324. Springer, 2010.
- [RAB⁺07] Matthias Rehm, Elisabeth André, Nikolaus Bee, Birgit Endrass, Michael Wissner, Yukiko Nakano, Toyooki Nishida, and H Huang. The cube–g approach–coaching culture–specific nonverbal behavior by virtual agents. In *Proceedings of*, volume 38, pages 313–322, 2007.
- [RBDV13] Paul Rauwolf, Tina Balke, and Marina De Vos. Modelling normative awareness: First considerations. *Social Coordination: Principles, Artefacts and Theories, SOCIAL. PATH 2013*,, pages 65–71, 2013.
- [RDW05] J Deckro Robbins, RF Deckro, and VD Wiley. Stabilization and reconstruction operations model (srom). In *Proceedings of 4th Workshop on Critical Issues in Information Fusion*, 2005.
- [RG⁺95] Anand S Rao, Michael P Georgeff, et al. Bdi agents: From theory to practice. In *ICMAS*, volume 95, pages 312–319, 1995.

REFERENCES

-
- [RHD⁺13] Rainer Reisenzein, Eva Hudlicka, Mehdi Dastani, Jonathan Gratch, Koen Hindriks, Emiliano Lorini, and John-Jules Ch Meyer. Computational modeling of emotion: Toward improving the inter-and intradisciplinary exchange. *Affective Computing, IEEE Transactions on*, 4(3):246–266, 2013.
 - [RMB⁺10] Stephen Read, Brian Monroe, Aaron Brownstein, Yu Yang, Gurveen Chopra, and Lynn Miller. A neural network model of the structure and dynamics of human personality. *Psychological review*, 117(1):61, 2010.
 - [Rob07] Brent Roberts. Contextualizing personality psychology. *Journal of personality*, 75(6):1071–1082, 2007.
 - [Rom55] Ragnar Rommetveit. Social norms and roles; explorations in the psychology of enduring social pressures. 1955.
 - [ROR11] Laurel Riek, Maria Oconnor, and Peter Robinson. Guess what? a game for affective annotation of video using crowd sourcing. In *Affective computing and intelligent interaction*, pages 277–285. Springer, 2011.
 - [ROSR03] Nestor Rychtycky, David Ostrowski, George Schleis, and Robert G Reynolds. Using cultural algorithms in industry. In *Swarm Intelligence Symposium, 2003. SIS'03. Proceedings of the 2003 IEEE*, pages 187–192. IEEE, 2003.
 - [RP05] Stefan Rank and Paolo Petta. Appraisal for a character-based story-world. In *IVA*, pages 495–496, 2005.
 - [RZSH⁺14] Avi Rosenfeld, Inon Zuckerman, Erel Segal-Halevi, Osnat Drein, and Sarit Kraus. Negochat: a chat-based negotiation agent. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 525–532. International Foundation for Autonomous Agents and Multiagent Systems, 2014.
 - [Sav11] Bastin Tony Roy Savarimuthu. Norm learning in multi-agent societies. 2011.
 - [SC11] Bastin Tony Roy Savarimuthu and Stephen Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.

REFERENCES

- [Sco71] John Scott. Internalization of norms: A sociological theory of moral commitment. 1971.
- [SDM09] Bas R Steunebrink, Mehdi Dastani, and John-Jules Ch Meyer. A formal model of emotion-based action tendency for intelligent agents. In *Portuguese Conference on Artificial Intelligence*, pages 174–186. Springer, 2009.
- [SFC⁺10] Alberto Signoretti, Antonino Feitosa, André Campos, Anne Canuto, and Sergio Fialho. Increasing the eciency of npcs using a focus of attention based on emotions and personality. In *Games and Digital Entertainment (SBGAMES), 2010 Brazilian Symposium on*, pages 171–181. IEEE, 2010.
- [SFV09] Hugues Salamin, Sarah Favre, and Alessandro Vinciarelli. Automatic role recognition in multiparty recordings: Using social affiliation networks for feature extraction. *Multimedia, IEEE Transactions on*, 11(7):1373–1380, 2009.
- [She36] Muzafer Sherif. The psychology of social norms. 1936.
- [SKB⁺05] Mechthild Schäfer, Stefan Korn, Felix Brodbeck, Dieter Wolke, and Henrike Schulz. Bullying roles in changing contexts: The stability of victim and bully roles from primary to secondary school. *International Journal of Behavioral Development*, 29(4):323–335, 2005.
- [SL90] Craig A Smith and Richard S Lazarus. Emotion and adaptation. 1990.
- [SLB⁺96] Christina Salmivalli, Kirsti Lagerspetz, Kaj Björkqvist, Karin Österman, and Ari Kaukiainen. Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive behavior*, 22(1):1–15, 1996.
- [SLL98] Christina Salmivalli, Miia Lappalainen, and Kirsti Lagerspetz. Stability and change of behavior in connection with bullying in schools: A two-year follow-up. *Aggressive behavior*, 24(3):205–218, 1998.
- [SMP10] Mei Si, Stacy C Marsella, and David V Pynadath. Modeling appraisal in theory of mind reasoning. *Autonomous Agents and Multi-Agent Systems*, 20(1):14–31, 2010.

REFERENCES

- [SPN⁺12] Barry Silverman, David Pietrocola, Ben Nye, Nathan Weyer, Oleg Osin, Dan Johnson, and Ransom Weaver. Rich socio-cognitive agents for immersive training environments: case of nonkin village. *Autonomous Agents and Multi-Agent Systems*, 24(2):312–343, 2012.
- [SPZ06] Jennifer Sandercock, Lin Padgham, and Fabio Zambetta. Creating adaptive and individual personalities in many characters without hand-crafting behaviors. In *Intelligent Virtual Agents*, pages 357–368. Springer, 2006.
- [SS09] Derek Sollenberger and Munindar Singh. *Architecture for affective social games*, pages 79–94. Springer, 2009.
- [SS16] Robert J Sternberg and Karin Sternberg. *Cognitive psychology*. Nelson Education, 2016.
- [ST] Yoav Shoham and Moshe Tennenholtz. On the synthesis of useful social laws for artificial agent societies (preliminary report). In *Proceedings of the National Conference on Artificial Intelligence*, pages 276–276. JOHN WILEY & SONS LTD.
- [Sun06] Ron Sun. The clarion cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction*, pages 79–99, 2006.
- [Sun09] Ron Sun. *Cognitive architectures and multi-agent social simulation*, pages 7–21. Springer, 2009.
- [Sun13] Ron Sun. Dual-process theories and cognitive architectures. *Rensselaer Polytechnic Institute, Troy, NY, USA*, 2013.
- [Taf81] Henri Tajfel. *Human groups and social categories: Studies in social psychology*. CUP Archive, 1981.
- [TE05] Paul Tremblay and Laura Ewart. The buss and perry aggression questionnaire and its relations to values, the big five, provoking hypothetical situations, alcohol consumption patterns, and alcohol expectancies. *Personality and Individual Differences*, 38(2):337–346, 2005.

REFERENCES

- [Tho20] Edward Thorndike. A constant error in psychological ratings. *Journal of applied psychology*, 4(1):25–29, 1920.
- [TK59] John Thibaut and Harold Kelley. The social psychology of groups. 1959.
- [TK74] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.
- [TTK05] Will Taggart, Sherry Turkle, and Cory D Kidd. An interactive robot in a nursing home: Preliminary remarks. In *Towards social mechanisms of android science: a COGSCI workshop*, 2005.
- [UM77] Edna Ullmann-Margalit. *The emergence of norms*. Clarendon Press Oxford, 1977.
- [VPB09] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.
- [VRB⁺12] Laura Pfeifer Vardoulakis, Lazlo Ring, Barbara Barry, Candace L Sidner, and Timothy Bickmore. Designing relational agents as long term social companions for older adults. In *International Conference on Intelligent Virtual Agents*, pages 289–302. Springer, 2012.
- [WM88] Jack Wright and Walter Mischel. Conditional hedges and the intuitive psychology of traits. *Journal of personality and social psychology*, 55(3):454, 1988.
- [WW95] Adam Walker and Michael Wooldridge. Understanding the emergence of conventions in multi-agent systems. In *ICMAS*, volume 95, pages 384–389, 1995.
- [You04] Stephen Younger. Reciprocity, normative reputation, and the development of mutual obligation in gift-giving societies. *JASSS-THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION*, 7(1), 2004.
- [ZMVH08] Greg Zacharias, Jean MacMillan, and Susan Van Hemel. *Behavioral modeling and simulation: from individuals to societies*. National Academy Press, 2008.