

An Ethically-Guided Domain-Independent Model of Computational Emotions

Suman Ojha

Faculty of Engineering and Information Technology
University of Technology Sydney

This dissertation is submitted for the degree of
Doctor of Philosophy

Supervisor:

Dist. Prof. Mary-Anne Williams

Co-supervisor:

Dr. Jonathan Vitale

February 2020

This research is dedicated to my father, Dr. Narayan Ojha, who wanted to see me as a medical doctor like him. Although, I took a path of engineering instead of medicine, I have marched on my way to become a 'doctor' – though not in medicine but in philosophy of Computer Science.

Certificate of Original Authorship

I, Suman Ojha, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature of Student:

Production Note:
Signature removed prior to publication.

Suman Ojha
24 February 2020

Acknowledgements

My research project would not have been successful without the help and support of a number of people. I hereby take an opportunity to personally acknowledge everyone who walked with me in their own ways along the journey of this PhD research. I would like to heartily thank the following people.

- My PhD supervisor, Distinguished Professor Mary-Anne Williams, for being an exemplary advisor. She allowed me to pave my own path for the completion of this journey and supported along the way in every possible manner. She also taught me that paper rejections are the norms as an early career researcher and one should take this as an opportunity to refine one's work and become a better researcher.
- My co-supervisor, Dr. Jonathan Vitale, for being more of a colleague than just a supervisor and helping me in uncountable ways.
- My wife, Asmita Thapa, who took my research as seriously as I did, and went through everything to keep me away from the stress of my studies.
- Dr. Syed Ali Raza and Dr. Richard Billingsley for providing me useful concepts of machine learning approaches for the completion of my research project.
- My family and friends who always treated me as a person rather than just a nerdy research student and never failed to make me realise that I was alive.
- My colleagues at the Innovation and Enterprise Research Lab (The Magic Lab) who helped to improve my research by attending my presentations and providing feedback on the drafts of my papers.
- All the anonymous participants of my research surveys who contributed their valuable data for the evaluation of my computational model of emotion.
- All the anonymous reviewers of my conference and journal papers who provided valuable comments and suggestions for the improvement of my research.

I consider you all as a part of my research journey.

Author's Core Research Contributions

- Ojha, S., Gudi, S. L. K. C., Vitale, J., Williams, M.-A., and Johnston, B. (2017a). I remember what you did: A behavioural guide-robot. In *International Conference on Robot Intelligence Technology and Applications (RiTA)*.
- Ojha, S., Vitale, J., Ali Raza, S., Billingsley, R., and Williams, M.-A. (2018a). Implementing the dynamic role of mood and personality in emotion processing of cognitive agents. In *Annual Conference on Advances in Cognitive Systems (ACS)*.
- Ojha, S., Vitale, J., Ali Raza, S., Billingsley, R., and Williams, M.-A. (2019). Integrating mood and personality with agent emotions. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.
- Ojha, S., Vitale, J., and Williams, M.-A. (2017b). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *Annual Meeting of the Cognitive Science Society (CogSci)*.
- Ojha, S. and Williams, M.-A. (2016). Ethically-guided emotional responses for social robots: Should i be angry? In *International Conference on Social Robotics (ICSR)*, pages 233–242. Springer.
- Ojha, S. and Williams, M.-A. (2017). Emotional appraisal: A computational perspective. In *Annual Conference on Advances in Cognitive Systems (ACS)*.
- Ojha, S., Williams, M.-A., and Johnston, B. (2018b). The essence of ethical reasoning in robot-emotion processing. *International Journal of Social Robotics (IJSR)*, 10:211–223.

Author's Additional Research Contributions

Gudi, S. L. K. C., Ojha, S., Alam, S., Johnston, B., and Williams, M.-A. (2017a). A proactive robot tutor based on emotional intelligence. In *International Conference on Robot Intelligence Technology and Applications (RiTA)*.

Gudi, S. L. K. C., Ojha, S., Clark, J., Johnston, B., and Williams, M.-A. (2017b). Fog robotics: An introduction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (iROS)*.

Gudi, S. L. K. C., Ojha, S., Johnston, B., Clark, J., and Williams, M.-A. (2018). Fog robotics for efficient, fluent and robust human-robot interaction. In *IEEE 17th International Symposium on Network Computing and Applications (NCA)*.

Herse, S., Vitale, J., Ebrahimian, D., Tonkin, M., Ojha, S., Sidra, S., Johnston, B., Phillips, S., Gudi, S. L. K. C., Clark, J., et al. (2018a). Bon appetit! robot persuasion for food recommendation. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 125–126. ACM.

Herse, S., Vitale, J., Tonkin, M., Ebrahimian, D., Ojha, S., Johnston, B., Judge, W., and Williams, M.-A. (2018b). Do you trust me, blindly? factors influencing trust towards a robot recommender system. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*.

Tonkin, M., Vitale, J., Ojha, S., Clark, J., Pfeiffer, S., Judge, W., Wang, X., and Williams, M.-A. (2017a). Embodiment, privacy and social robots: May i remember you? In *International Conference on Social Robotics (ICSR)*, pages 506–515. Springer.

Tonkin, M., Vitale, J., Ojha, S., Williams, M.-A., Fuller, P., Judge, W., and Wang, X. a. (2017b). Would you like to sample? Robot engagement in a shopping centre. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE.

Author's Additional Research Contributions

Vitale, J., Tonkin, M., Herse, S., Ojha, S., Clark, J., Williams, M.-A., Wang, X., and Judge, W. (2018). Be more transparent and users will like you: A robot privacy and user experience design experiment. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 379–387. ACM.

Vitale, J., Tonkin, M., Ojha, S., Williams, M.-A., Wang, X., and Judge, W. (2017). Privacy by design in machine learning data collection: A user experience experimentation. In *AAAI Spring Symposium Series*.

Table of Contents

1	Introduction	1
1.1	Research Objectives	6
1.1.1	Primary Objectives	6
1.1.2	Secondary Objectives	6
1.2	Contributions	7
1.2.1	Theoretical/Conceptual Contributions	7
1.2.2	Technical/Methodological Contributions	8
1.3	Thesis and Methodology	11
1.4	Scope and Limitations	12
1.5	Dissertation Overview	13
2	Background and Literature	21
2.1	What is an Emotion?	21
2.2	Theories of Emotion	22
2.2.1	Physiological Theories of Emotion	23
2.2.2	Anatomic Theories of Emotion	24
2.2.3	Dimensional Theories of Emotion	26
2.2.4	Appraisal Theories of Emotion	32
2.2.5	Discussion	46
2.3	Understanding Appraisal Dynamics	47
2.4	Role of Mood and Personality in Emotional Appraisal	49
2.4.1	Personality Factor	51
2.4.2	Mood Factor	54
2.4.3	Interaction among Emotion, Mood and Personality	54
2.5	Emotion and Ethics	56

Table of Contents

2.5.1	Theories of Ethics	57
2.5.2	Connecting Ethics to Emotions	59
2.6	Chapter Summary	65
3	Computational Emotion Models and Research Context	82
3.1	Computational Models of Emotion	82
3.1.1	Cathexis	84
3.1.2	FLAME	86
3.1.3	Model of Egges and Colleagues	88
3.1.4	FearNot!	90
3.1.5	ALMA	92
3.1.6	MAMID	94
3.1.7	EMA	95
3.1.8	Soar-Emote	96
3.1.9	WASABI	99
3.1.10	TAME	101
3.1.11	FAtiMA	103
3.1.12	MA/SDEC	106
3.1.13	EMIA	107
3.1.14	CAAF	108
3.1.15	Other Models of Emotion	109
3.1.16	Summary and Comparison	112
3.2	Hypotheses	117
3.3	Chapter Summary	121
4	Ethical Emotion Generation System (EEGS) Details	130
4.1	EEGS: Ethical Emotion Generation System	130
4.2	Revisiting Appraisal Dynamics	131
4.3	Overall System Architecture	132
4.3.1	Emotion Elicitation Module	133
4.3.2	Cognitive Appraisal Module	134
4.3.3	Memory Module	134
4.3.4	Characteristics Module	134
4.3.5	Affect Generation Module	135
4.3.6	Affect Regulation Module	136

Table of Contents

4.4	Events, Actions and Objects	137
4.4.1	Structure of Events, Actions and Objects	138
4.5	Emotion Elicitation	141
4.6	Cognitive Appraisal	142
4.6.1	Goals, Standards and Attitudes	143
4.6.2	Appraisal Variables	147
4.7	Affect Generation	156
4.7.1	Appraisal–Emotion Network	157
4.7.2	Data-driven Learning of Appraisal–Emotion Association	160
4.7.3	Implementation of Affect Generation Process in EEGS	164
4.8	Affect Regulation	182
4.8.1	Emotion Convergence in Computational Models	182
4.8.2	Ethical Reasoning for Emotion Regulation in EEGS	184
4.8.3	Reasoning Mechanism in EEGS	186
4.9	A Guideline for the Implementation of EEGS Modules	189
4.9.1	Implementing the Emotion Elicitation Module	189
4.9.2	Implementing the Cognitive Appraisal Module	190
4.9.3	Implementing the Affect Generation Module	193
4.9.4	Implementing the Affect Regulation Module	195
4.10	Chapter Summary	196
5	Model Evaluation and Thesis Validation	206
5.1	Introduction to a 3-Stage Evaluation Approach	207
5.2	Scenarios and Data Collection	208
5.2.1	Scenario Design	209
5.2.2	Data Collection	210
5.3	Stage 1: Cognitive Appraisal Evaluation	215
5.3.1	Methodology	216
5.3.2	Results	217
5.3.3	Additional Discussion	220
5.4	Stage 2: Affect Generation Evaluation	220
5.4.1	Methodology	221
5.4.2	Results	222
5.4.3	Additional Discussion	231

Table of Contents

5.5	Stage 3: Affect Regulation Evaluation	237
5.5.1	Methodology and Results	238
5.6	Justification of Thesis Validation	247
5.7	Chapter Summary	249
6	Conclusion and Future Directions	256
6.1	Contributions and Implications	257
6.2	Limitations and Future Work	261
6.3	Personal Reflection	263

List of figures

2.1	Interaction of brain regions from the Anatomic view of emotional (fear) responses	24
2.2	A circumplex representation of emotional states	27
2.3	A conceptual representation of pleasure, arousal and dominance dimensions	28
2.4	Plutchik's wheel of emotions showing different sections of varying colours representing each type of emotion	29
2.5	Plutchik's cone below the wheel of emotions signifying the possible intensity of each coloured section in the wheel	30
2.6	Lövheim's cube of emotion	31
2.7	Process flow from an stimulus event to (1) emotion elicitation, (2) cognitive appraisal, (3) affect generation, and (4) affect regulation	48
2.8	<i>Positive affect</i> and <i>Negative affect</i> dimensions of mood and their relationship with the dimensions of <i>pleasantness</i> and <i>arousal</i>	53
2.9	Interaction between emotion, mood and personality in a <i>mediation</i> approach	55
2.10	A consensual process model of emotion generation and regulation	63
3.1	Several components of Cathexis model	85
3.2	FLAME agent architecture	86
3.3	Overview of the integrated personality, mood and emotion model of Egges et al. (2004)	89
3.4	FearNot! affectively driven agent architecture	91

List of figures

3.5	Process flow in MAMID cognitive-affective architecture	94
3.6	Cognitive–motivational–emotive system architecture of EMA model	95
3.7	A basic PEACTION cycle	97
3.8	Soar-Emote’s unification of PEACTION and appraisals	98
3.9	The conceptual distinction of cognition and embodiment in WASABI architecture	100
3.10	Conceptual overview of TAME architecture	102
3.11	FAtiMA Core architecture	104
3.12	Appraisal mechanism in FAtiMA emotion architecture.	104
3.13	EMIA architecture divided into three layers	107
3.14	Graphical representation of how the validation of Hypothesis 1 and 2 will involve the evaluation of first part (<i>i.e.</i> emotion elicitation, cognitive appraisal, affect generation), and second part (<i>i.e.</i> affect regulation) of the overall computational process in EEGS.	117
4.1	Process flow from an stimulus event to (1) emotion elicitation, (2) cognitive appraisal, (3) affect generation, and (4) affect regulation revisited.	131
4.2	Overall architecture of EEGS.	133
4.3	Influence of <i>goals, standards</i> and <i>attitudes</i> in cognitive ap- praisal process as suggested by Ortony et al. (1990).	143
4.4	An example of a goal tree in EEGS based on OCC theory	144
4.5	Parallel computation of appraisals in EEGS.	155
4.6	Role of cognitive appraisal in affect generation process	157
4.7	An weighted appraisal-emotion network showing many-to- many relationship between appraisal variables and emotions	158
4.8	A general appraisal-emotion network with k appraisal variables and l emotion types.	161
4.9	Decomposition of the link between appraisal variable v_1 and emotion type e_1	161
4.10	Mechanism for mapping the angle of an emotion type into a signed valence degree	167
4.11	Cyclic interaction between emotion and mood.	175

List of figures

4.12	Proposed dynamic interaction between emotion, mood and personality	176
4.13	Comparison of different emotion decay functions.	180
4.14	Process of affect regulation in EEGS where conflicting emotional states are converged to a final stable and regulated emotional state based on ethical reasoning guided by ethical standards	184
4.15	Normalisation function for appraisal variables in the range [0,1].	191
4.16	Normalisation function for appraisal variables in the range [-1,1].	191
4.17	Normalisation function for emotion intensities.	194
5.1	Proposed <i>3-Stage Evaluation</i> approach for computational models of emotion.	207
5.2	Accuracy in computation of various appraisals by EEGS as compared to appraisals rated by human participants in the given scenario.	218
5.3	Accuracy of the overall appraisal of EEGS compared to the error in appraisal computation.	219
5.4	Desirability (appraisal) dynamics of EEGS for two different scenarios – (i) Two Strangers in a Park and (ii) Husband and Wife.	220
5.5	Overall accuracy in prediction of eight emotions over the 10 training-testing sessions for each of the emotions.	223
5.6	Evolution in learning of the association between appraisal variables and emotion for eight different emotions for a training session where the test data set was used for prediction of emotion intensity after each epoch of the session.	225
5.7	Difference in intensity of <i>joy</i> emotion in Scenario 1 (Two Strangers in a Park) when the personality factor of <i>extraversion</i> (E) is altered	228
5.8	Difference in intensity of <i>joy</i> emotion in Scenario 3 (Husband and Wife) when the personality factor of <i>extraversion</i> (E) is altered	229
5.9	Emotion dynamics of EEGS when initial mood is very positive in Scenario 1 (Two Strangers in a Park)	229

List of figures

5.10	Emotion dynamics of EEGS when initial mood is very negative in Scenario 1 (Two Strangers in a Park)	230
5.11	Learning trend for the association of the appraisal variable <i>desirability</i> to the emotion <i>joy</i> averaged over 10 training sessions and the variation in the learned weights across the training sessions	232
5.12	Accuracy in prediction of intensity of <i>joy</i> emotion during testing phase.	233
5.13	Evolution of the learned model with the increasing epochs. . .	234
5.14	Learning trend for the association of the appraisal variable <i>desirability</i> to the emotion <i>distress</i> averaged over 10 training sessions and the variation in the learned weights across the training sessions	235
5.15	Mirrored pattern in learning of the weights for personality factors for the association of the appraisal variable <i>desirability</i> to the emotions <i>joy</i> and <i>distress</i> . Surprisingly, the mood factor did not exhibit a mirrored pattern for <i>joy</i> and <i>distress</i>	236
5.16	Mirrored pattern obtained for the weight of mood factor (f_M) when only the mood is considered in the learning process. . . .	237
5.17	Comparison of the rank distance from the average human rating of the emotion intensity generated by (i) highest intensity, (ii) blended intensity and (iii) ethical reasoning approaches . .	239
5.18	Cumulative rank distance from the average human rating for the emotion intensity generated by (i) highest intensity, (ii) blended intensity and (iii) ethical reasoning approaches	240
5.19	Emotion dynamics in EEGS using (i) highest intensity approach, (ii) blended intensity approach, and (iii) ethical reasoning approach in Scenario 4	244
5.20	Emotion dynamics in EEGS using (i) highest intensity approach, (ii) blended intensity approach, and (iii) ethical reasoning approach in Scenario 5	246

List of tables

2.1	Summary of the relationship between Panksepp's primary emotional systems to specific brain regions	26
2.2	Levels of monoamine neurotransmitters in various emotions according to the theory of Lövheim (2012)	31
2.3	Appraisal variables proposed in the theory of Frijda (1986)	34
2.4	Summary of Stimulus Evaluation Checks in appraisal theory of Scherer (2001)	36
2.5	Summary of Appraisal Components in Cognitive-Motivational-Emotive theory of emotion Smith et al. (1990)	40
2.6	Functional analysis of some emotions based on core relational theme	41
2.7	Appraisal variables and evaluation criteria in OCC theory of emotion	42
2.8	Emotion groups and emotion types in OCC theory of emotion	44
2.9	Appraisal dimensions in appraisal theory of Roseman (1979)	45
2.10	Evolutionary history of five factors of personality	51
2.11	Definition of some affective states in the context of current dissertation.	56
2.12	Normative theories of ethics. List of selected ethical theories adapted from Robbins and Wallace (2007)	58
2.13	Definition of believability and social acceptability in the context of current dissertation.	62
3.1	List of computational models of emotion	114

List of tables

3.2	Comparison of several computational implementations of emotion in artificial agents over the last two decades	116
4.1	An example of some events	139
4.2	An example of some actions	139
4.3	An example of some objects	140
4.4	An example of some emotion and action related standards in EEGS	146
4.5	Mapping of the angles of the circumplex into valence degree for various emotions.	168
4.6	Association of various appraisal variables with different emotions as suggested in the OCC theory	170
4.7	A summary of different emotion decay mechanisms used in various computational models of emotion.	179
4.8	An example of a set of ethical standards for <i>anger</i> emotion . .	186
4.9	Input(s) and output(s) of cognitive appraisal module.	190
4.10	Appraisal variables in EEGS and their value ranges.	192
4.11	Input(s) and output(s) of affect generation module.	193
4.12	Input(s) and output(s) of affect regulation module.	195
5.1	Summary of the scenarios considered.	210
5.2	Error in appraisal computation of EEGS.	217
5.3	Paired t-Test to compare the appraisals computed by EEGS to the ratings provided by human participants.	219
5.4	Overall accuracy in prediction of various emotion intensities. .	223
5.5	Best classification accuracy obtained by Meuleman and Scherer (2013)	227
5.6	Comparison of median distances from the human assessment for (i) highest intensity, (ii) blended intensity and (iii) ethical reasoning approaches in EEGS.	238
5.7	Quantified emotion values of (i) highest intensity approach, (ii) blended intensity approach, and (iii) ethical reasoning approach in response to various actions of Rose (Dementia patient) to Lily (service robot) in Scenario 4	243

List of tables

5.8	Quantified emotion values of (i) highest intensity approach, (ii) blended intensity approach, and (iii) ethical reasoning approach in response to various actions of Andrew (little boy) to Robert (companion robot) in Scenario 5	245
6.1	A summary of contributions and implications of the presented dissertation.	261

Abstract

Advancement of artificial intelligence research has supported the development of intelligent autonomous agents. Such intelligent agents, like social robots, are already appearing in public places, homes and offices. Unlike the robots intended for use in factories for mechanical work, social robots should not only be proficient in capabilities such as vision and speech, but also be endowed with other human skills in order to facilitate a sound relationship with human counterparts.

Phenomena of emotions is a distinguishable human feature that plays a significant role in human social communication because ability to express emotions enhances the social exchange between two individuals. As such, artificial agents employed in social settings should also exhibit adequate emotional and behavioural abilities to be easily adopted by people.

A critical aspect to consider when developing models of artificial emotions for autonomous intelligent agents is the likely impact that the emotional interaction can have on the human counterparts. For example, an *emotional robot* that shows an angry expression along with a loud voice may scare a young child more than a *non-emotional robot* that only denies a request. Indeed, most modern societies consider a strong emotional reaction towards a young child to be unacceptable and even unethical.

How can a robot select a socially acceptable emotional state to express while interacting with people? I answer this question by providing an association between emotion theories and ethical theories – which has largely been ignored in the existing literature.

A regulatory mechanism for artificial agents inspired by ethical theories is a viable way to ensure that the emotional and behavioural responses of the agent are acceptable in a given social context. As such, an intelligent agent with emotion generation capability can establish social acceptance if its emotions are regulated by ethical reasoning mechanism.

In order to validate the above statement, in this work, I provide a novel computational model of emotion for artificial agents – EEGS (short name for **Ethical Emotion Generation System**) and evaluate it by comparing the emotional responses of the model with emotion data collected from human participants. Experimental results support that *ethical reasoning mechanism can indeed help an artificial agent to reach to a socially acceptable emotional state*.