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 bheavioural 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.

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

List of figures

List of tables

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 **E**thical **E**motion **G**eneration **S**ystem) 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*.

So many of our dreams at first seem impossible, then they seem improbable, and then, when we summon the will, they soon become inevitable.

— Christopher Reeve —

1

Introduction

'The future' is near. Various techniques of artificial intelligence have led to an unprecedented advancement of the field of computer science and robotics. Intelligent agents[1] these days are able to outperform humans in certain kinds of tasks such as lip reading (Assael et al., 2016) and image recognition (He et al., 2015). Autonomous robots equipped with artificial intelligence capabilities are being employed in various areas such as factory works (Bischoff et al., 2010), medical procedures (Brief et al., 2018; Rockrohr, 2018), public entertainment (Foster et al., 2016), elderly care (Portugal et al., 2015; Yang et al., 2018), child care (Blanson Henkemans et al., 2017) and so on. As the robots get closer to human society and touch our everyday lives, it is crucial that these robots are more than just intelligent machines able to perform tasks such as face processing (Kamarolzaman et al., 2017), speech recognition (Zinchenko et al., 2017), object recognition (Browatzki et al., 2014), *etc.* In order to fuse seamlessly into human society, these robots should also be able to exhibit behavioural dynamics similar to humans in addition to the above specified capabilities.

Among several other behavioural variations (Horn and Noll, 1997), humans can feel and express emotions in response to a range of situations (Phelps, 2004). Emotions

[1]By saying 'intelligent agents', I refer to all kinds of implementations such as software agents as well as physical agents like robots. Therefore, in the remainder of the dissertation, the use of the word *agent* should be considered as a superset and the word *robot* should be considered as subset of such agents.

are often considered to serve a basis for a better social exchange and non-verbal communication between two individuals (Knapp and Daly, 2002; Knapp et al., 2013; Lawler, 2001). As such, emotions have strong influence in the scenarios of human-robot interaction as well, since, according to *media equation theory of communication*, people tend to anthropomorphise machines *i.e.* deal/interact with machines as if they were humans and expect the same in return (Reeves and Nass, 1996). Thus, robots can establish a perceived sense of belongingness with human counterparts if they are able to feel and express their emotions as humans do (Beer et al., 2011).

Unfortunately, the absolute nature of emotion in humans, the knowledge of which forms the basis of the development of computational models of emotion, is still a little understood phenomenon (Barrett, 2016; Ekman and Davidson, 1994; Plutchik, 2001). Indeed, psychology literature is abundant of several variations of emotion theories (Cannon, 1927; Damasio et al., 2000; James, 1884; LeDoux, 1995; Ortony et al., 1990; Plutchik, 2001; Russell, 1980; Schachter and Singer, 1962). Proposed emotion theories are characterised by a varying range of assumptions despite having some common grounds. Due to the lack of consensus among the researchers, there is not a single universal theory that describes how emotions are construed. Among those theories, *cognitive appraisal theory of emotion* is one notable emotion theory and frequently used by computer science researchers (Aylett et al., 2005; El-Nasr et al., 2000; Gratch and Marsella, 2004). According to cognitive appraisal theory (Ortony et al., 1990; Scherer, 2001; Smith et al., 1990), emotions result from the appraisal (evaluation) of the stimulus event. As such, different people may experience different emotions in similar situations depending on how they evaluate the event. The criteria for evaluating such events are often called *appraisal variables* in emotion theories. For example, the appraisal variable *desirability* measures if an event is desirable to an individual or not (Ortony et al., 1990). While drawing inspiration from several existing emotion theories, this research mainly builds on the postulates of cognitive appraisal theory of emotion (Ortony et al., 1990) to provide emotion generation capability in autonomous agents. This choice is mainly due to a more wider availability of other computational models (Dias et al., 2014; Gratch and Marsella, 2004) building on the cognitive appraisal theory, thus leading to a more fair and effective comparison of the results discussed in the remainder of this dissertation.

One of the most notable issues pertaining to the role of emotions in autonomous intelligent agents such as robots is that emotions of these robots can have not only positive but also negative impact on the society. Although an ability to express its emotions increases the chances of a robot for being embraced as a part of our lives, the same aspect can lead to social rejection of such robots. Indeed, an inappropriate emotional behaviour of a robot could lead to psychological damages, odd situations or misunderstandings leading to monetary loss or other issues in the society (Sharkey and Sharkey, 2010, 2011). For example, consider a robot that expresses a *very angry* emotional reaction in response to a mindless act of a young child. Will our society accept such robot in houses, at work or in other human spaces? The probability of such an incident is not deniable. If a robot is endowed with an ability to autonomously generate and express emotion as a way to enhance human-robot interaction and communication, it will not only express *joy* in pleasant situations, but will also express *anger* in adverse situations. A robot employed in human society should be emotionally and socially proficient, thus being able to discern if it is appropriate to exhibit a negative emotional behaviour in the given situation (Vitale et al., 2014; Williams, 2012). However, it should not be misunderstood that a robot should always express positive emotions. Expression of negative emotion such as disappointment may be needed in some situations. Consider a scenario, where a young child (say 12 years old) misbehaves with a robot and after a few minutes asks to play. If the robot readily agrees to play with the child happily, the child will develop a misconception that robots can be mistreated without consequences. When this develops as a habit, this may lead the child to have a persistent tendency of misbehaving with other people as well (Ojha et al., 2017a). This point is supported by the fact that children who are not acknowledged of their bad behaviour grow as an ill mannered adult (Harris et al., 1964; Martin and Pear, 2015). However, if the robot exhibits distress or some kind of fear in response to the proposal of playing from the child by making the child realise that s/he recently mistreated the robot, then the child can learn to behave well with robots and hence other people. A key question that follows is – "How can we operationalise such a socially acceptable emotional control in an artificial agent"? Although researchers have proposed several methods of *emotion regulation* (Gross, 2002; Gross and Thompson, 2007; Thompson, 1994), these approaches are mainly focused on how to prevent the harm of negative emotional experience on the 'self' without much concern about other parties involved. Similarly, some appraisal theorists suggest the mechanism of *coping*[2] as a means of emotional control (Lazarus, 1991). However, these approaches do not put much emphasis on the consequences of emotional responses to the 'others' – which is a crucial requirement in a social robot. I address this gap in the literature by seeking a *link between emotion theories and ethical theories*.

Traditional literature has extensively studied the effect of emotions on ethical decision making in a wide range of fields (Callahan, 1988; Gaudine and Thorne, 2001).

[2]Coping is the measure of the degree by which the individual experiencing the situation can deal directly with the situation or its consequences.

However, the role of ethical reasoning in the process of emotion generation has not been adequately addressed previously. In this dissertation, I suggest that the concepts of ethics (Alexander and Moore, 2007; Hooker, 1996; Pettit, 1993; Robbins and Wallace, 2007) and emotion (Ortony et al., 1990; Scherer, 2001) are interrelated and, hence, ethical reasoning mechanism can play a crucial role in the regulation of our emotions (Ojha et al., 2018). I argue that, emotions regulated in such a manner allow us to exhibit sound emotional behaviour in various social situations. This implies that a social robot whose emotions are regulated by ethical reasoning mechanism can express socially acceptable emotional responses, and it can be more easily adopted by users and society. In order to validate the above statements, in this dissertation, I develop a computational model of emotion for artificial agents called EEGS (short form for **E**thical **E**motion **G**eneration **S**ystem) and evaluate it by conducting experiments comparing the emotional responses artificially generated by my model and emotion data collected from humans (detailed discussion of these aspects will be presented in Chapter 4 and 5).

When it comes to the actual implementation of the emotion mechanism in artificial agents, several underlying ***computational problems*** arise. Although there may not be much diversion of the arguments on why emotions are needed for artificial agents, there is definitely a wide lack of consistency on the implementation of emotion mechanism in existing computational models of emotion based on appraisal theory of emotion (Lazarus, 1991; Ortony et al., 1990; Roseman et al., 1990; Scherer, 2001). This inconsistency stems from the fact that there exists a large variation even among the appraisal theorists on what should be considered as the appraisal variables. For example, Frijda (1986) proposes 7 appraisal variables while Smith et al. (1990) propose 6 different appraisal variables for the generation of emotions (see Chapter 2 for in-depth discussion of this aspect). The situation is made worse by the fact that most appraisal theories present only the functional level description of the process of appraisal and fail to provide a detailed mathematical formulation for the computation of relevant appraisals. This leads the computational emotion modelling researchers to deal with the situation with their individual assumptions for the implementation of various aspects. Most computational models of emotion implement researcher-specific hard-coded rules to perform cognitive appraisal for the evaluation of the emotion stimulus rather than using general rules that can be applied in diverse situations. This limitation is termed as *domain-dependence of the emotion models* by emotion modelling researchers (Gratch and Marsella, 2004). In other words, emotion models proposed in the literature are limited to the appraisal of situation for which they were originally designed for, or for scenarios sharing very similar assumptions (Aylett et al., 2005; Egges et al., 2004; El-Nasr et al., 2000; Velásquez and Maes, 1997). When the model has to be implemented in a new situation or domain, the appraisal rules have to be changed manually, thus leading to additional complexity and system integration issues (but, see also Gratch and Marsella (2004); Jain and Asawa (2016); Kaptein et al. (2016); Saunier and Jones (2014) for examples of previously proposed domain-independent emotion models). This very property prevents the researchers from further testing and evolving the given emotion model. More importantly, a domain-dependent emotional model does not allow a systematic comparison with other models of emotion and also hinders the integration of emotion models with diverse intelligent agents (Ojha and Williams, 2017). In this dissertation, I will present viable solutions to the above identified issues, thereby advancing computational emotion modelling research.

Another computational issue arises in the implementation when appraisals are to be quantitatively mapped into corresponding emotion intensities. Cognitive appraisal theory states that the process of appraisal is followed by the generation of emotion intensities *i.e.* the intensities of the emotions generated in response to a given situation are determined by the evaluation of several appraisal variables (Ortony et al., 1990; Scherer, 2001; Smith et al., 1990). But, how do these appraisals map to the emotion intensities? At present, there is no plausible answer to this question in the emotion literature. In existing appraisal theories, it is not clear how a quantitative association can be established between appraisal variables and emotions – mainly from implementation perspective. However, emotion research literature does provide some important insights giving a viable direction for answering this question. In fact, it has been suggested that the emotion processing mechanism is strongly modulated by aspects such as *personality* (Corr, 2008; Revelle, 1995) and *mood* (Morris, 1992; Neumann et al., 2001) of an individual among several other factors (Aleman et al., 2008; Hong et al., 2000; Scollon et al., 2004). Yet, there is not a well accepted or empirically validated approach on how these aspects can be computationally operationalised in the process of emotion generation. In other words, although personality and mood are likely to modulate the emotional responses of an individual, we do not have a computational theory demonstrating *how* those aspects can be integrated within a process of emotion generation. This is a relevant limitation for socially intelligent agents, which would likely need to integrate such information when required to express a more believable emotional state (Rusting, 1998). Some of the existing computational models of emotion integrate the notions of personality (Hudlicka, 2005; McRorie et al., 2011; Saunier and Jones, 2014) or mood (El-Nasr et al., 2000; Gratch and Marsella, 2004; Marinier III et al., 2009), but most of them do not operationalise both the characteristics (but, see also Aylett et al. (2005); Gebhard (2005); Moshkina et al. (2011)). Moreover, among the models which integrate the aspects of either mood or personality, the suggested approaches have been very simple and limited, consisting of defining some thresholds of emotion intensities based on these characteristics without crucially providing empirical evidence from human data. Instead, in this work, I insist that the quantitative association between appraisal variables and emotions should be *learned in a data-driven manner* by the model based on collected human emotion data.

This dissertation aims to provide a novel computational model of emotion capable of significantly reducing the above mentioned limitations identified from the present literature in computational emotion research.

1.1 Research Objectives

My research *aims* to address the gap identified in emotion literature and also various issues in existing emotion models by developing a new computational model of emotion (EEGS). As such, the *objectives* of this research can be summarised as follows.

1.1.1 Primary Objectives

1. To demonstrate that ethical reasoning mechanism can augment the process of emotion regulation in an artificial agent.

2. To propose and validate a computational ethical reasoning mechanism for autonomous agents by exploring the theories of ethics, in order to regulate the emotions, thereby improving the agent's social acceptance.

1.1.2 Secondary Objectives

1. To operationalise a domain-independent approach of cognitive appraisal in emotion models by proposing a new mathematical formulation for appraisal computation based on theoretical foundations.

2. To integrate the notions of mood and personality in emotion generation process by realising the mutual interaction among these aspects.

3. To learn the association between appraisal variables and emotions in relation to their intensities using machine learning technique from artificial intelligence.

1.2 Contributions

Earlier in this chapter, I identified general problems in the existing literature and also stated my objectives to be carried out in this research in order to address those issues. In this section, I will explain the contributions I will make to the universe of theoretical as well as computational emotion literature.

1.2.1 Theoretical/Conceptual Contributions

The research presented in this dissertation will be an interdisciplinary exchange between the fields of cognitive psychology and affective computing. In the process of identifying the potential issues in deploying autonomous artificial intelligent agents with emotion generation capability in the society, I will propose a new theory and develop a computational model of emotion to validate the theory. As such, there will be numerous exchange of information between these fields in this dissertation. In this section, I present the contributions of the current research on theoretical and conceptual levels.

A Comprehensive Review of Theoretical and Computational Emotion Literature

In this dissertation, I will conduct a comprehensive analysis of various theories of emotion ranging from the ones that entail the concept of physiological basis of emotion (Cannon, 1927; James, 1884) to the ones that relate the notion of emotion to anatomical regions in human brain (Damasio et al., 2000; LeDoux, 1995). I will also present a discussion of various dimensional (Plutchik, 2001; Russell and Barrett, 1999) as well as appraisal theories (Ortony et al., 1990; Scherer, 2001). These discussions will be helpful for an overall understanding of the literature in emotion sphere and related human characteristics. Moreover, I will present a critical analysis of the computational models of emotion proposed in the past two decades, compare them and identify the limitations of those models. The presented analyses will provide new insights to other researchers for the advancement of the field by addressing the issues that have still not been fully addressed by existing research.

A new Perspective on Emotion Regulation Mechanism

The topic of *emotion regulation* is relatively recent (Gross, 1998, 2002; Thompson, 1994) as compared to the emotion literature itself (Arnold, 1960; James, 1894). The existing literature in emotion regulation has mainly focused on how emotions can be managed so that the negative impact of an emotional experience on the 'self' is minimised (Gross and Thompson, 2007; Thompson, 1994). In this dissertation, I will present a new perspective on the relationship between the process of emotion regulation and ethical reasoning. I will discuss that ethical reasoning can help regulate an individual's emotions not only for the 'self-benefit' but also for the good of the society. I will present various experimental scenarios in Chapter 5, Section 5.5 demonstrating how the process of ethical reasoning can lead to the regulation of emotions elicited by a stimulus event in an agent's surrounding.

A new Relationship between Emotion and Ethics

Traditional literature has largely studied the influence of an individual's emotions on decision making process (Callahan, 1988; Gaudine and Thorne, 2001; Isen and Means, 1983). Such a decision can be judged as being ethical or not depending on the seriousness of the decision made. However, there is no clear evidence in the literature on how the process of ethical reasoning can affect emotion generation mechanism in individuals. Since the mechanism of emotion entails an evaluation (Ortony et al., 1990; Scherer, 2001) of the situation (*i.e.* some form of reasoning), it is inevitable that the notion of ethics operationalises in the process. The importance of such an influence is often underestimated in the existing literature (Ojha et al., 2017b, 2018). In this dissertation, I endeavour to reveal a reverse perspective (not as a supplement but as a compliment to the existing knowledge) on the notion of emotion and ethics and examine how the mechanism of ethical reasoning can offer a substantial influence on the emotion generation process of an artificial agent.

1.2.2 Technical/Methodological Contributions

In Section 1.2.1, I briefly stated the conceptual and theoretical contributions of my research. In the remaining of the dissertation, I will present a computational model of emotion in order to validate the thesis statement (see Section 1.3). As a consequence, on the technical and methodological level as well, current research provides some notable contributions to the field of emotion modelling and artificial intelligence as a whole. The core technical contributions of this research are presented below.

A new Domain-Independent Model of Cognitive Appraisal

In the beginning of this chapter, I identified that one of the most common problems in existing computational models of emotions is the lack of ability to perform appraisal of situations in more than one domain with the same computation mechanism (Aylett et al., 2005; Gebhard, 2005). If such a model is to be implemented in a new domain, the whole appraisal rules have to be modified because those are only suited to specific situations. In this dissertation (Chapter 4), I present my computational model of emotion, EEGS, that can perform appraisals in multiple domains with the same generic appraisal mechanism.

A new Mathematical Formulation of Cognitive Appraisal

With the goal of achieving a domain-independent cognitive appraisal of an emotion-eliciting event, I contribute with a new mathematical model for computing appraisal variables which will be discussed in detail in Chapter 4, Section 4.6.2. This is an important contribution to the emotion modelling/artificial intelligence literature because most existing computational models of emotion either do not explain the nature of the process involved in appraisal (Moshkina et al., 2011; Schneider and Adamy, 2014; Sun et al., 2016) or are not generalisable to various domains (Aylett et al., 2005; Marinier III et al., 2009) (but, see also Gratch and Marsella (2004); Jain and Asawa (2016)).

Effective Integration of the aspects of Personality and Mood with Emotion Generation Process

Although there have been previous attempts in integrating the aspects of personality (Hudlicka, 2005; Saunier and Jones, 2014), mood (Becker, 2008; El-Nasr et al., 2000; Gratch and Marsella, 2004; Marinier III et al., 2009) or both (Egges et al., 2004; Gebhard, 2005; Velásquez and Maes, 1997) in computational models of emotion, most of these models do not explicitly examine how the operationalisation of these characteristics cause difference in emotion dynamics of an agent. In this dissertation (Chapter 5, Section 5.4), I will discuss how such characteristics can play a crucial role in modulating the emotion dynamics of an intelligent agent.

Data-driven Learning of Appraisal-Emotion Association

As discussed earlier, existing computational models of emotion operationalise the association of appraisal variables to emotions using static rules based on researcher assumptions. However, I argue that such an approach is not robust enough and can hardly reflect the true relationship between the appraisals and emotions. Therefore, in this work, I employ a mechanism to learn the association of the appraisal variables with emotions based on the factors of personality and mood using machine learning technique from artificial intelligence in my computational model, EEGS. The detailed process of learning these associations will be presented in Chapter 4, Section 4.7.2.

A new Mathematical Model of Emotion Intensity Computation

One of the notable contributions of this research is the development of a new mathematical model for the computation of emotion intensities based on appraisal variables. Existing emotion modelling literature suffers from the scarcity of published mathematical/technical details explaining how emotion intensities are computed. Therefore, it is very difficult for researchers in this field to understand the computational aspects of emotion intensity generation in the provided models. In order to address this issue, I present a detailed mathematical formulation for the computation of emotion intensities in EEGS (Chapter 4, Section 4.7.3).

A new Evaluation Methodology for Computational Models of Emotion

To the best of my knowledge, there is not any standard evaluation criteria for existing computational models of emotion. Some models are evaluated examining how believable are the emotions being expressed by the artificial agent (Becker, 2008). Other models are evaluated by simulating pet (El-Nasr et al., 2000), toddler (Velásquez and Maes, 1997) or virtual conversational characters (Aylett et al., 2005; Gebhard, 2005). Among these, most models are tested as a blackbox, *i.e.* by checking the output for a given input and evaluating it based on commonsense. Such a methodology severely hinders the comparison between two models of emotion and subsequently prevents the advancement of emotion modelling research. In this dissertation, I propose a 3-stage evaluation approach for computational cognitive appraisal models of emotion where the model is evaluated by comparing the system responses at (i) appraisal computation, (ii) appraisal mapping to emotions, and (iii) emotion regulation stages directly with the data collected from humans.

Although the above identified contributions are some of the notable ones, there are several other constituent contributions – mainly from technical perspective, which are yet to be discussed in the dissertation. These contributions will provide new insights to the researchers of the field and help in the advancement of the research on computational modelling of emotion for autonomous agents. As such, current dissertation enfolds numerous implications to the researchers both from artificial intelligence as well as cognitive psychology community and also encourages a better direction on the design and development of autonomous artificial agents that can interweave seamlessly into human society.

1.3 Thesis and Methodology

In the rest of this dissertation, I will endeavour to validate the following thesis:

Thesis *The regulatory mechanism for emotional processing of an artificial agent can be enriched by an ethical reasoning mechanism enabling the selection of a more socially acceptable emotional state to express while interacting with people in a given social context.*

In other words, current research will demonstrate that computational emotions generated by the process of cognitive appraisal can be regulated by a mechanism of ethical reasoning thereby reaching to a stable emotional state. As a result, I also argue that such a regulation mechanism governed by ethical reasoning if implemented in artificial agents allows them to exhibit emotional responses and behaviour that are more acceptable in human society. As discussed earlier in the chapter, a behaviour can be considered socially acceptable if it is in-line with basic social norms.

In order to validate the above stated thesis, I will use a computer simulation of my computational model of emotion within various scenarios of social interaction between two individuals. The first step in developing a computational model of emotion is acquiring an understanding of the theoretical premises revolving around the concept of emotions as well as computational research generated as an outcome of the implications of the theories of emotion. Therefore, I will start with an in-depth exploration of the emotion related literature in psychology, cognitive science, neuroscience as well as affective computing community (see Chapter 2 for more details). I will, then, use the knowledge acquired from the theoretical literature to examine the limitations in the existing computational models of emotion and to explore the possibility of further research. As such, I will identify the crucial limitations that current computational emotion modelling research has not effectively addressed and myself endeavour to develop a new model of emotion solving the problems identified (see Chapter 3 for more details).

After the exploration of the literature and identification of potential research gaps, the next step will be to break down the general thesis statement into testable hypotheses (see Chapter 3 for the hypotheses considered in this dissertation). In order to validate the stated hypotheses, the appraisal and emotion generation mechanism of the developed model will be compared to similar data collected from human participants by asking survey questions including various interaction scenarios. Testable scenarios will be simulated in the developed model and the generated data will be recorded. Likewise, human participants will be asked to provide emotion-related data in different sets of experiments for the same scenarios. The data from humans will be compared with the data generated by system simulation which will help in validating the hypotheses thereby confirming with the overall thesis statement of the dissertation (see Chapter 5 for detailed discussion).

1.4 Scope and Limitations

I would like to clarify that the theory proposed in this dissertation as well as the associated computational implementation have several limitations. Being a time and resource constrained postgraduate research project, this dissertation does not cover the aspects of following *scope*.

i) A new direction towards *unified theory of emotion* in not offered in the current dissertation.
ii) *All the universal aspects* pertaining to the process of emotion generation are not addressed in the current discussion.
iii) Postulates of any emotion theory have not been formally *proved or disproved* in the current dissertation.

In addition to a limited scope of research within emotion domain, this research is not free from other limitations from implementation perspective as well. Some of the *main limitations* of this research project are listed below.

i) The validation of the proposed theory involves a limited set of emotions unlike infinitely many number of emotions as theorised by some researchers.
ii) Only a limited number of human characteristics that are identified as crucially important in the literature have been integrated in the process of emotion generation (*i.e.* personality and mood).
iii) Since the experiments have been simulated instead of implementing in a real physical robot, several practical issues that might emerge in real human-robot interaction situations have not been investigated in this work.

iv) Several computational assumptions have been made because of lack of concrete evidence in the existing literature. Such assumptions have been explicitly stated and justified.

v) While the actions of the human counterpart are considered in current experiments, the system operationalises only the emotional reaction of the artificial agent and does not account for the selection of its physical responses.

1.5 Dissertation Overview

The remaining of the dissertation is organised as follows.

In Chapter 2, I will present a comprehensive review of the existing emotion theories as well as the concepts of personality and mood and how these aspects are associated with one another. I will also introduce the notion of ethical reasoning and discuss its relationship with the emotion generation process.

In Chapter 3, I will perform a critical analysis of the existing computational models of emotion and identify their major limitations. I will also present the hypotheses I aim to test and validate in this research.

In Chapter 4, I will discuss about the technical/computational details of the proposed model. The chapter will offer an in-depth discussion of five major process of emotion generation (i) *emotion elicitation*, (ii) *cognitive appraisal*, (iii) *affect generation* and (iv) *affect regulation*. I will also present the overall system architecture with all the components in integration.

In Chapter 5, I will present an evaluation of the proposed computational emotion model by simulating different scenarios in the model and comparing the performance of the model with human data.

Chapter 6 will conclude the overall discussion of the dissertation with a brief recap of all the concepts, findings and contributions. I will also present a future direction of current research with possible improvements leading to the evolution of the current model.

Chapter Bibliography

Aleman, A., Swart, M., and van Rijn, S. (2008). Brain imaging, genetics and emotion. *Biological psychology*, 79(1):58–69.

Alexander, L. and Moore, M. (2007). Deontological ethics.

Arnold, M. B. (1960). *Emotion and personality.* Columbia University Press.

Assael, Y. M., Shillingford, B., Whiteson, S., and de Freitas, N. (2016). LipNet: end-to-end sentence-level lipreading.

Aylett, R., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). Fearnot!– An experiment in emergent narrative. In *Intelligent virtual agents*, pages 305–316. Springer.

Barrett, L. F. (2016). Navigating the science of emotion. *Emotion Measurement*, page 31.

Becker, C. (2008). *WASABI: Affect simulation for agents with believable interactivity*. PhD thesis, Bielefeld University.

Beer, J. M., Prakash, A., Mitzner, T. L., and Rogers, W. A. (2011). Understanding robot acceptance. Technical report, Georgia Institute of Technology.

Bischoff, R., Kurth, J., Schreiber, G., Koeppe, R., Albu-Schäffer, A., Beyer, A., Eiberger, O., Haddadin, S., Stemmer, A., Grunwald, G., et al. (2010). The kuka-dlr lightweight robot arm-a new reference platform for robotics research and manufacturing. In *Proceedings for the joint conference of ISR 2010 (41st Internationel Symposium on Robotics) und ROBOTIK 2010 (6th German Conference on Robotics)*, pages 1–8. VDE.

Blanson Henkemans, O. A., Van der Pal, S., Werner, I., Neerincx, M. A., and Looije, R. (2017). Learning with Charlie: A robot buddy for children with diabetes. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 406–406. ACM.

Brief, J., Haßfeld, S., Boesecke, R., Vogele, M., Krempien, R., Treiber, M., and Mühling, J. (2018). Robot assisted dental implantology. *Issues*, 2017:2016.

Browatzki, B., Tikhanoff, V., Metta, G., Bülthoff, H. H., and Wallraven, C. (2014). Active in-hand object recognition on a humanoid robot. *IEEE Transactions on Robotics*, 30(5):1260–1269.

Callahan, S. (1988). The role of emotion in ethical decisionmaking. *Hastings Center Report*, 18(3):9–14.

Cannon, W. B. (1927). The james-lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4):106–124.

Corr, P. J. (2008). *The reinforcement sensitivity theory of personality*. Cambridge University Press.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., and Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature neuroscience*, 3(10):1049–1056.

Dias, J., Mascarenhas, S., and Paiva, A. (2014). FAtiMA modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer.

Egges, A., Kshirsagar, S., and Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer animation and virtual worlds*, 15(1):1–13.

Ekman, P. E. and Davidson, R. J. (1994). *The nature of emotion: Fundamental questions.* Oxford University Press.

El-Nasr, M. S., Yen, J., and Ioerger, T. R. (2000). Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257.

Foster, M. E., Alami, R., Gestranius, O., Lemon, O., Niemelä, M., Odobez, J.-M., and Pandey, A. K. (2016). The MuMMER project: Engaging human-robot interaction in real-world public spaces. In *International Conference on Social Robotics*, pages 753–763. Springer.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Gaudine, A. and Thorne, L. (2001). Emotion and ethical decision-making in organizations. *Journal of Business Ethics*, 31(2):175–187.

Gebhard, P. (2005). Alma: A layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM.

Gratch, J. and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of general psychology*, 2(3):271.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, 39(3):281–291.

Gross, J. J. and Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. *Handbook of emotion regulation*.

Harris, F. R., Wolf, M. M., and Baer, D. M. (1964). Effects of adult social reinforcement on child behavior. *Young Children*, pages 8–17.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034.

Hong, Y.-y., Morris, M. W., Chiu, C.-y., and Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7):709.

Hooker, J. (1996). Three kinds of ethics.

Horn, J. L. and Noll, J. (1997). Human cognitive capabilities: Gf-gc theory.

Hudlicka, E. (2005). A computational model of emotion and personality: Applications to psychotherapy research and practice. In *Proceedings of the 10th annual cybertherapy conference: a decade of virtual reality*.

Isen, A. M. and Means, B. (1983). The influence of positive affect on decision-making strategy. *Social cognition*, 2(1):18–31.

Jain, S. and Asawa, K. (2016). Programming an expressive autonomous agent. *Expert Systems with Applications*, 43:131–141.

James, W. (1884). What is an emotion? *Mind*, 9(34):188–205.

James, W. (1894). The physical basis of emotion.

Kamarolzaman, N. A. F. M., Yeong, C. F., Su, E. L. M., Chik, S. F., Duan, F., Tan, J. T. C., Tan, P. H., and Chin, P. J. H. (2017). Smart advertising robot with image recognition for data analytics. *ELEKTRIKA-Journal of Electrical Engineering*, 16(3):11–16.

Kaptein, F., Broekens, J., Hindriks, K. V., and Neerincx, M. (2016). Caaf: A cognitive affective agent programming framework. In *International Conference on Intelligent Virtual Agents*, pages 317–330. Springer.

Knapp, M. L. and Daly, J. A. (2002). *Handbook of interpersonal communication*. Sage.

Knapp, M. L., Hall, J. A., and Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Lawler, E. J. (2001). An affect theory of social exchange. *American Journal of Sociology*, 107(2):321–352.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press on Demand.

LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annual review of psychology*, 46(1):209–235.

Marinier III, R. P., Laird, J. E., and Lewis, R. L. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, 10(1):48–69.

Martin, G. and Pear, J. J. (2015). *Behavior modification: What it is and how to do it*. Psychology Press.

McRorie, M., Sneddon, I., McKeown, G., Bevacqua, E., de Sevin, E., and Pelachaud, C. (2011). Evaluation of four designed virtual agent personalities. *IEEE Transactions on Affective Computing*, 3(3):311–322.

Morris, W. N. (1992). A functional analysis of the role of mood in affective systems.

Moshkina, L., Park, S., Arkin, R. C., Lee, J. K., and Jung, H. (2011). TAME: Time-varying affective response for humanoid robots. *International Journal of Social Robotics*, 3(3):207–221.

Neumann, R., Seibt, B., and Strack, F. (2001). The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Emotion and Cognition*, 15:725–747.

Ojha, S., Gudi, S. L. K. C., Vitale, J., Williams, M.-A., and Johnston, B. (2017a). I remember what you did: A bheavioural guide-robot. In *International Conference on Robot Intelligence Technology and Applications (RiTA)*.

Ojha, S., Vitale, J., and Williams, M. (2017b). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *Annual Meeting of the Cognitive Science Society*.

Ojha, S. and Williams, M. (2017). Emotional appraisal: A computational perspective. In *Annual Conference on Advances in Cognitive Systems*.

Ojha, S., Williams, M.-A., and Johnston, B. (2018). The essence of ethical reasoning in robot-emotion processing. *International Journal of Social Robotics (IJSR)*, 10:211–223.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Pettit, P. (1993). Consequentialism.

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current opinion in neurobiology*, 14(2):198–202.

Plutchik, R. (2001). The nature of emotions. *American scientist*, 89(4):344–350.

Portugal, D., Santos, L., Alvito, P., Dias, J., Samaras, G., and Christodoulou, E. (2015). SocialRobot: An interactive mobile robot for elderly home care. In *System Integration (SII), 2015 IEEE/SICE International Symposium on*, pages 811–816. IEEE.

Reeves, B. and Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.

Revelle, W. (1995). Personality processes. *Annual review of psychology*, 46(1):295–328.

Robbins, R. W. and Wallace, W. A. (2007). Decision support for ethical problem solving: A multi-agent approach. *Decision Support Systems*, 43(4):1571–1587.

Rockrohr, B. (2018). Robotic surgical systems, instrument drive units, and drive assemblies. US Patent App. 15/549,005.

Roseman, I. J., Spindel, M. S., and Jose, P. E. (1990). Appraisals of emotion-eliciting events. *Journal of Personality and Social Psychology*, 59(5):899–915.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.

Russell, J. A. and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805.

Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin*, 124(2):165.

Saunier, J. and Jones, H. (2014). Mixed agent/social dynamics for emotion computation. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 645–652. International Foundation for Autonomous Agents and Multiagent Systems.

Schachter, S. and Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5):379.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120):57.

Schneider, M. and Adamy, J. (2014). Towards modelling affect and emotions in autonomous agents with recurrent fuzzy systems. In *International Conference on Systems, Man and Cybernetics*, pages 31–38. IEEE.

Scollon, C. N., Diener, E., Oishi, S., and Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of cross-cultural psychology*, 35(3):304–326.

Sharkey, N. and Sharkey, A. (2010). The crying shame of robot nannies: an ethical appraisal. *Interaction Studies*, 11(2):161–190.

Sharkey, N. and Sharkey, A. (2011). The rights and wrongs of robot care. *Robot ethics: The ethical and social implications of robotics*, page 267.

Smith, C. A., Lazarus, R. S., et al. (1990). Emotion and adaptation. *Handbook of personality: Theory and research*, pages 609–637.

Sun, R., Wilson, N., and Lynch, M. (2016). Emotion: A unified mechanistic interpretation from a cognitive architecture. *Cognitive Computation*, 8(1):1–14.

Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. *Monographs of the society for research in child development*, 59(2-3):25–52.

Velásquez, J. D. and Maes, P. (1997). Cathexis: a computational model of emotions. In *Proceedings of the first international conference on Autonomous agents*, pages 518–519. ACM.

Vitale, J., Williams, M.-A., and Johnston, B. (2014). Socially impaired robots: Human social disorders and robots' socio-emotional intelligence. In *International Conference on Social Robotics*, pages 350–359. Springer.

Williams, M.-A. (2012). Robot social intelligence. In *International Conference on Social Robotics*, pages 45–55. Springer.

Yang, Y., Park, Y. J., Ro, H., Chae, S., and Han, T.-D. (2018). cARe-bot: Portable projection-based ar robot for elderly. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 384–384. ACM.

Zinchenko, K., Wu, C.-Y., and Song, K.-T. (2017). A study on speech recognition control for a surgical robot. *IEEE Transactions on Industrial Informatics*, 13(2):607–615.

Scientific knowledge is in perpetual evolution; it finds itself changed from one day to the next.

— Jean Piaget —

2

Background and Literature

In Chapter 1, I presented a general overview of my research and also the contribution my research is offering to the field. In this chapter, I will present a comprehensive review of the concepts of emotion and related literature from both theoretical and computational perspectives.

2.1 What is an Emotion?

Throughout the history of psychology, the term *emotion* has been looked at and defined with different perceptions (Kleinginna and Kleinginna, 1981; Mulligan and Scherer, 2012). Emotion theorists tend to propose a definition of emotion based on the processes they consider to take part during an emotional episode (Kleinginna and Kleinginna, 1981). Moors (2009) defines *emotional episode* as a holistic mechanism starting from the stimulus to other componential processes leading to the generation of emotion or the 'immediate' consequences of the experienced emotion. Based on the review of the historical theories of emotion causation, Moors (2009) summarises five common components of emotional episode namely (i) *Cognitive* component – pertaining to the evaluation of the stimulus, (ii) *Feeling* component – pertaining to the overall experience, monitoring or regulation of emotions, (iii) *Motivational* component – pertaining to the aspects of action tendencies and action readiness, (iv) *Somatic* component – pertaining to central and peripheral physiological responses, and (v) *Motor* component – pertaining to the expressive behaviour *i.e.* execution of actions, voice or facial expressions. Since most emotion theories deviate on the argument of the exact number, nature and occurrence of the emotion components (Moors, 2009), each theory mainly revolves around one or more components that are considered to be crucial to the understanding of the process of emotion. This means while some emotion theorists may assume one component to be operational before another, there are others who assume the opposite. Moreover, some emotion theorists may not focus on the discussion of a component at all. For example, James (1884) considers the *somatic* component to the executed before the *feeling* component suggesting that the physiological changes in reaction to a stimulus event are not the effect but the cause of emotion. However, some other theorists (such as Scherer (2001)) propose the opposite view. In the following sections, I will review some of the most notable emotion theories and identify which of the components suggested by Moors (2009) are conceptualised in those theories.

2.2 Theories of Emotion

Some researchers claim that emotion is nothing but the response of human *body* in reaction to a surrounding stimuli (Cannon, 1927; James, 1884; Lange, 1885). These theories of emotion tend to relate the process of emotion with physiological processes taking place inside human body and also with the various anatomical structures of the brain and other related parts (Cannon, 1927; James, 1884; Lange, 1885). Such emotion theories can be termed as *physiological* theories of emotion. Other researchers believe that emotion is caused by the involvement of various brain regions when a stimulus event is perceived (Damasio et al., 2000; LeDoux, 1995). Such theories can be termed as *anatomic* theories of emotion. Another class of emotion theory called *appraisal theory* offers a different take about emotions. According to this theory, emotion generation is a cognitive process and hence which emotion is elicited by an individual is determined by how the individual evaluates the given event or situation (Ortony et al., 1990; Roseman, 1984; Scherer, 2001). Unlike most appraisal, physiological or anatomic theories of emotion, which commonly consider an emotion as an individually identifiable entity, *dimensional theory* of emotion assumes that emotions are not individual (discrete) entities but rather points in a continuous dimensional space (Lövheim, 2012; Plutchik, 2001; Russell, 1980; Schlosberg, 1941). In the following sections, I will present the details of each of the above emotion theories as well as their variations. After the presentation of various emotion theories, in Section 2.2.5, I will discuss why or why not a particular theory has been considered for implementation in this dissertation.

2.2.1 Physiological Theories of Emotion

Most physiological theories of emotion are centred on *somatic* component of an emotional episode. These emotion theories mainly revolve around the physiological significance of emotional experience (Cannon, 1927, 1929; James, 1884, 1894; Schachter and Singer, 1962). James (1884) argued that emotional experience is a secondary phenomena governed by the primary changes in somatic and visceral systems in response to a surrounding stimuli. Similar view is reflected by Lange and James (1922) (original idea published in (Lange, 1885)), however with a slightly different view on physiological involvement. The major difference between the two theorists is the consideration for the parts of the body involved emotional processing event *i.e.* James (1884) considered emotion as an outcome of *somato-visceral changes*, while Lange (1885) considered emotion as a *cardiovascular event* (Lang, 1994). Yet, the core of both the theories was that emotional experience result from physiological reactions in our body – hence the combined name *James–Lange theory of emotion*. Cannon (1927) argued that physiological origin can not solely be considered the sufficient criteria for the experience of emotion. He stresses that visceral changes are merely the body's preparation to act in response to the stimuli and these changes do not show a significant difference among various emotional states (Lang, 1994). He further strengthens his statement by arguing that surgical isolation of viscera does not prevent the emotional experience (Lang, 1994) suggesting that there should be an overarching structure to organise these physiological reactions. Cannon (1929) claims that such a mechanism is controlled by diencephalon in the human brain (Lang, 1994). Partially in line with Cannon's views, Schachter (1959) disagrees with James's proposition of sufficiency of visceral changes on emotion experience arguing that different emotional experience should be governed by different patterns of visceral changes – which is not addressed by James's theory. Moreover, Jame's theory fails to describe *how* physiological changes occur as a result of exposure to a stimulus (Moors, 2009). Addressing this issue Schachter and Singer (1962) subsequently proposed that emotional state should be defined in terms of two things *viz* interaction between *physiological arousal* and *cognitive evaluation of the arousing situation* (Reisenzein, 1983). Schachter (1964) argues that arousal and cognition are necessary for emotional experience and must coexist for the process of emotion elicitation. It was a remarkable shift in the understanding of the emotion and Schachter (1964) should be credited for the introduction of *cognitive* component in the

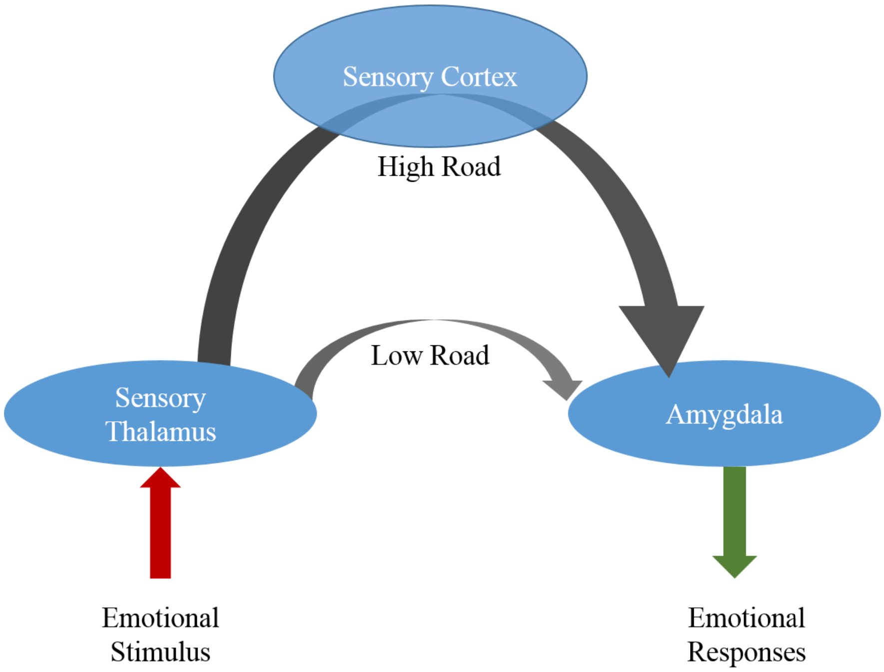


Figure 2.1 Interaction of brain regions from the Anatomic view of emotional (fear) responses. Redrawn after LeDoux (1996, p. 164).

process of emotion – be it in a preliminary form. According to his theory, cognition helps in characterising the type of emotional experience and physiological arousal helps in determining the intensity of the emotion felt. Although, the existence of physiological differences in various emotional experiences (as also advocated by James (1894) and Cannon (1927)) were later proved by empirical evidences (Prkachin et al., 1999), other researchers claim that physiological change is not the cause but the effect of emotional experience (Carter et al., 2002). While the relationship between physiology and emotion was widely discussed in the mid 20^{th} century, the notion of cognition put forward by Schachter was carried forward in the development of emotion theories – which led to the recent advancements in the understanding of emotion mechanism in humans from cognitive perspective (Lazarus, 1991; Ortony et al., 1990; Roseman, 1996; Scherer, 2001).

2.2.2 Anatomic Theories of Emotion

According to the proponents of the *anatomic theory* of emotion an experience of emotion is the result of triggering of particular regions of human brain when a sensory information reaches to those regions and an exchange of neural information takes place (Damasio et al., 2000; LeDoux, 1995, 1996; Panksepp, 1988, 2004). Although a clear line has not been drawn between the physiological and anatomic theories, both these theories can be assumed to put major focus on the *somatic* aspects of an emotional episode with physiological theories focusing more on visceral aspects and anatomic theories on neural components. Extensive research has been done along the line of this assumption in the study of conditioned fear responses (LeDoux, 1995) and their relationship with the brain areas by using neuro-imaging techniques (Penny et al., 2011) have been empirically tested. Figure 2.1 shows an overview of the relationship between various areas of brain in the process of fear elicitation. The direction of the arrows indicate the flow of neural information from one region to another. In the context of the emotion of *fear*, amygdala plays a central role in deciding whether a stimulus is danger or not (LeDoux, 1995). Amygdala receives the stimulus information from sensory processing areas in the brain i.e. *sensory thalamus* and *sensory cortex* (LeDoux, 1995). Sensory thalamus provides course representations of the information and hence reach the amygdala quickly (thus the term *low road* for the information path from thalamus to amygdala). Sensory cortex provides fine-grained information about the stimulus and hence reaches amygdala relatively late compared to *low road* becuase of more processing involved (LeDoux, 1995). This slower path going through sensory cortex is also called *high road*. According to anatomic perspective in emotion, amygdala is responsible in triggering a corresponding emotional response after receiving the required stimulus information (LeDoux, 2003, 1995, 2000).

The concept of involvement of brain regions in emotion processing mechanism has been conceptually and/or empirically supported by other researchers as well (Damasio et al., 2000; Panksepp, 2004). While historically, this notion had been mostly studied in the context of fear response (LeDoux, 1995), in their recent publication, Montag and Panksepp (2016) have presented a summary of relationship of various other emotions to specific brain areas involved in their elicitation (see Table 2.1).

Panksepp's Primary Emotional System	Brain Areas Involved	Molecular Foundations (neurotransmitter and neuropeptide systems)
FEAR	Central and lateral amygdala to medial hypothalamus and dorsal periaqueductal gray (PAG)	Glutamate (+), CRF (+), CCK (+), alpha-MSH (+)
RAGE	Medial amygdala to bed nucleus of stria terminalis (BNST). Medial and perifornical hypothalamic to PAG	Substance P (+), Ach (+), glutamate (+)
PANIC/SADNESS	Anterior cingulate, BNST and preoptic area, dorsomedial thalamus, PAG	Opioids (-), oxytocin (-), prolactin (-), CRF (+), glutamate (+)

SEEKING	nucleus accumbens – ventral tegmental area (VTA), mesolimbic and mesocortical outputs, lateral hypothalamus, PAG	DA (+), glutamate (+), opioids (+), neurotensin (+), orexin (+)
CARE	anterior cingulate, BNST, preoptic area, VTA, PAG	oxytocin (+), prolactin (+), dopamine (+), opioids (±)
LUST	cortico-medial amygdala, BNST, preoptic hypothalamus, ventromedial hypothalamus (VMH), PAG	gonadal steroids (+), vasopressin (+ male), oxytocin (+ female), LH-RH (+)
PLAY	Dorso-medial diencephalon, parafascicular area, PAG	opioids (±), glutamate (+), Ach (+), endocannabinoids (+)

Table 2.1 Summary of the relationship between Panksepp's primary emotional systems to specific brain regions. + indicates excitatory effects and - indicates inhibiting effects; CRF = orticotropin releasing hormone; CK, cholecystokinin; alpha-MSH, alpha melanocyte stimulating hormone; Ach, acetylcholine; LH-RH, luteinising hormone releasing hormone. Adapted and summarised from Montag and Panksepp (2016).

2.2.3 Dimensional Theories of Emotion

Anatomic view of emotional experience (see Section 2.2.2) considers that emotions are associated with particular brain regions (Damasio et al., 2000; LeDoux, 1995; Panksepp, 1988). This means the emotion being elicited is determined by how the stimulus information is relayed and processed in those regions. Likewise, the emotion theories focusing on physiological reactions consider bolidy changes (somatic and visceral) as the cause of emotion elicitation in humans (Cannon, 1927; James, 1884; Lange, 1885). Unlike other kinds of emotion theories, which look at a causal link underpinning the process of emotion generation, dimensional theories aim to describe the structural organisation of emotions in a certain dimensional space without answering the question of *how* various emotional states are elicited. Therefore it may not be wrong to call dimensional theories as *structural* or *representational* theories of emotion (Marsella et al., 2010) that focus solely on the *feeling* component of the overall emotional episode (see the beginning of Section 2.1 for introduction of *feeling* component). In the following sections, I will present some of the widely accepted dimensional approaches to the understanding of emotions.

Circumplex Model of Affect

According to this model, emotions align in a circular pattern in a two dimensional space. Early proposals of circumplex model were brought forward by Schlosberg (1941,

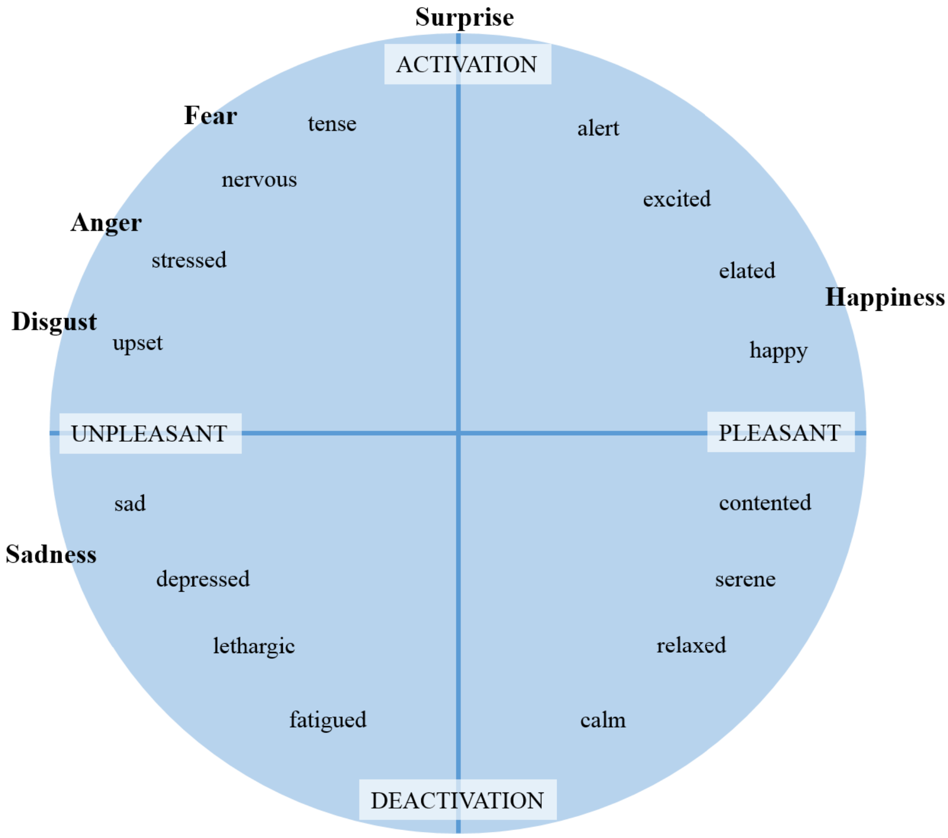


Figure 2.2 A circumplex representation of emotional states. Redrawn after Russell and Barrett (1999, p. 808).

1952). His initial description on the organisation of emotional states consisted of an oval surface with longer *pleasantness-unpleasantness* axis and shorter *attention-rejection* axis (Schlosberg, 1952). He argued that emotional states can be represented by a point in this two dimensional space. Other proponents of circumplex model of affect have suggested perfect circular alignment instead of oval in their publications (Russell, 1980). The proposal of circular alignment of affective states is also supported by the findings of other studies by contemporary researchers (Feldman Barrett and Russell, 1998; Russell and Mehrabian, 1977; Rusting and Larsen, 1995; Zevon and Tellegen, 1982). Figure 2.2 shows a form of circumplex model of emotions as explained by Russell and Barrett (1999). The horizontal axis represents the level of pleasantness. The emotional states lying on the left of the axis are characterised by negative experience while the ones on the right are characterised by positive experience. This axis can also be termed as *pleasure axis* as it reflects whether the emotional experience reflects pleasure or displeasure. Likewise, the vertical axis represents the level of activation. The higher the emotional state along the axis the higher is the activation level during the experience of that emotion. For example, in Figure 2.2, the emotional state of *excited* has high level

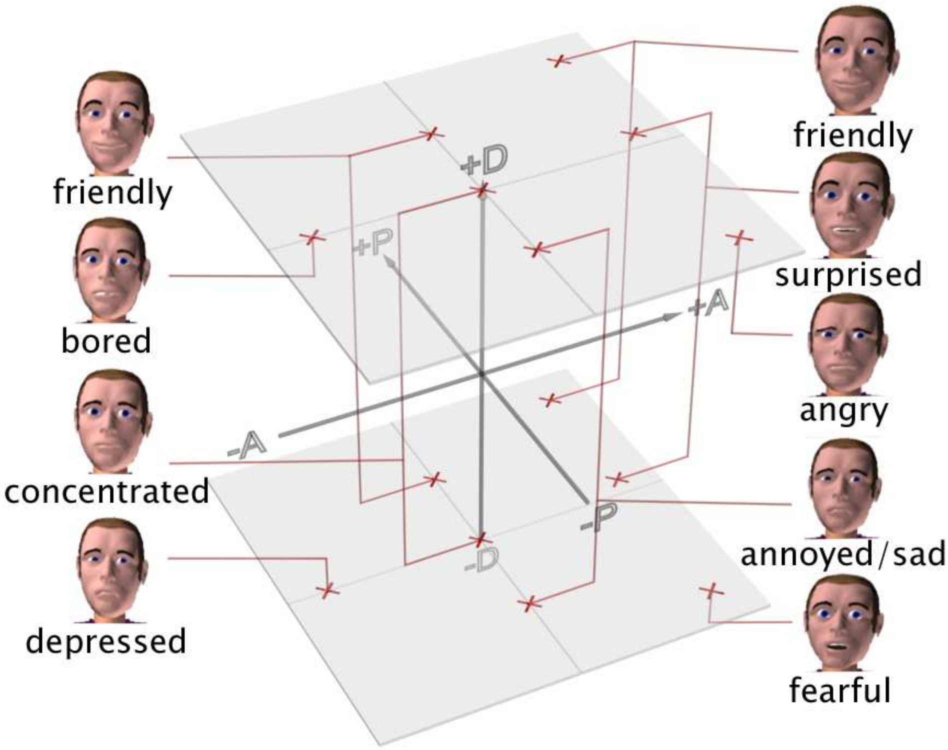


Figure 2.3 A conceptual representation of pleasure, arousal and dominance dimensions. Adapted from Becker et al. (2006) with author's permission.

of activation as compared to the emotional state of *depressed*. The axis of activation can also be called as *arousal axis* since it reflects the level of arousal experienced during a particular emotional experience.

These two dimensions are considered basic for the understanding of human emotional experience and were also supported in empirical studies carried out by other researchers (Abelson and Sermat, 1962; Engen et al., 1958; Gladstones, 1962).

Pleasure-Arousal-Dominance (PAD) Model of Emotions

Schlosberg (1954) presented a concept of three dimensions for describing the notion of emotional experience among humans by revising the initial proposals of two dimensions (Schlosberg, 1941). In addition to the initial dimensions of *pleasantness* and *activation*, he proposed that a new dimension, namely *control* is required for a more comprehensive description of the nature of emotions in dimensional approach. This proposition was supported by empirical results of Osgood (1966) as well as Williams and Sundene (1965). The dimension of *control*, which refers to the ability of the individual to control or modify the stimuli of its consequences, has also been referred as *dominance* (Russell and Mehrabian, 1977). Hence, the three dimensions of emotion are named as *pleasure*,

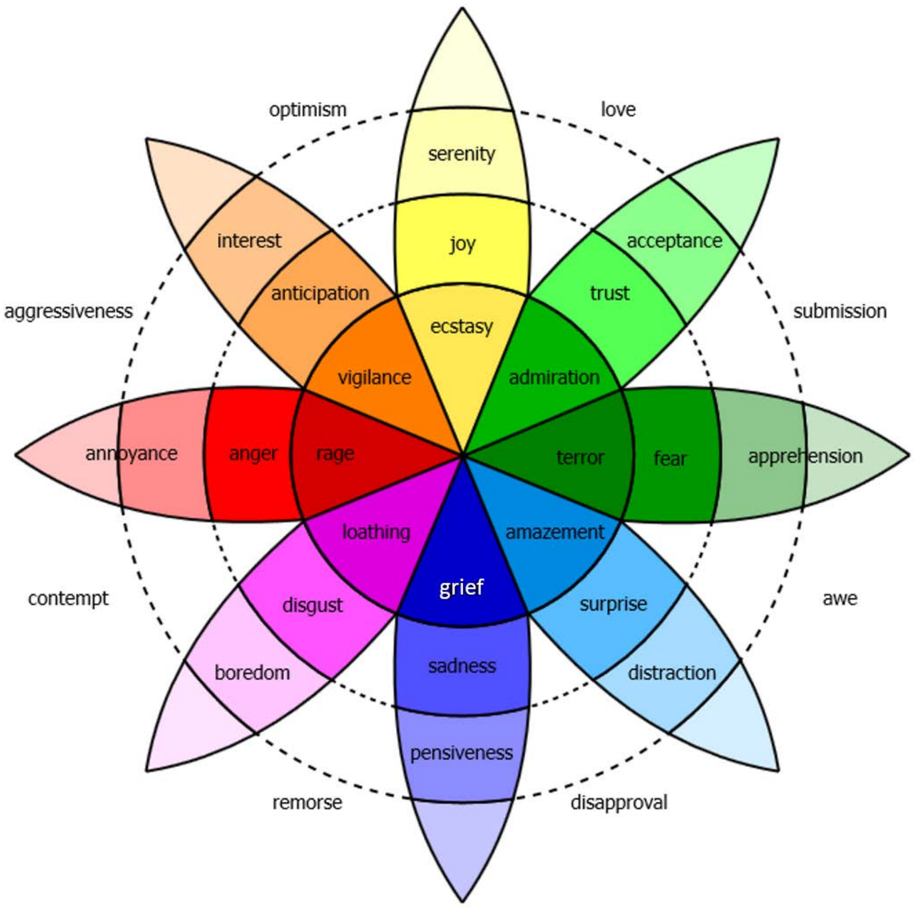


Figure 2.4 Plutchik's wheel of emotions showing different sections of varying colours representing each type of emotion. Redrawn after Plutchik (2001, p. 349)

arousal and *dominance* i.e. PAD in short form. Figure 2.3 shows how various emotions can be differentiated based on the three PAD dimensions.

Plutchik's Model of Emotions

Plutchik (1980, 2001) proposed a different kind of dimensional approach to the understanding of emotions. Influenced by the idea that parallels exist between emotions and natural colours (McDougall, 1921), this *psychoevolutionary* theory of emotions (Plutchik, 2001, p. 350) asserts that emotions can be understood as a distribution of various colours arranged in a circular pattern forming a specific pairs of opposites. Therefore, according to this theory, two or more emotions can mix together to form a more complex kind of emotions – hence also called *palette theory of emotions*. Plutchik (1980) proposes 8 primary emotions as a pair of opposites namely *joy-sorrow*,*anger-fear*,

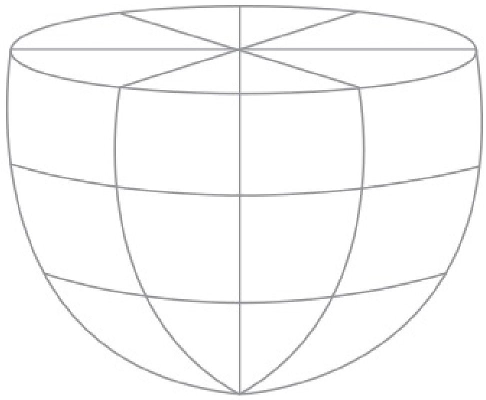


Figure 2.5 Plutchik's cone below the wheel of emotions signifying the possible intensity of each coloured section in the wheel. Redrawn after Plutchik (2001, p. 349)

acceptance-disgust and *surprise-expectancy*. Other complex emotions are considered as the mixture of two primary emotions like the combination of blue and red makes purple colour. As such, the emotion of *love* is considered as the mixture of can be considered the mixture of *joy* and *trust*. Accordingly, the emotion *love* is located between *joy* and trust in Figure 2.4. Likewise *disgust* and *anger* produces *contempt*. Such mixtures are referred to as *primary dyads*. Continuing this way and mixing of the emotion colours offers the realisation of hundreds of emotion words in the literature (Plutchik, 2001).

It should be noted that the model is not merely a two dimensional structure although it looks like a *wheel of emotions* if observed from the top – as shown by the image in Figure 2.4. Instead, Plutchik (2001) describes the model as a three dimensional conical structure – as represented by the image in Figure 2.5. Different shades used in the colour in various sections of the wheel can be regarded as the third dimension, which has been structurally represented as a conical shape. The third (vertical cone) dimension represents the intensity of a particular emotion. The maximum depth of the cone that can be touched by the vertical line drawn from each primary emotion or primary dyad (without crossing the outer boundary) signifies the highest intensity that can be felt for that emotion. Therefore, the closer the emotion lies towards the centre of the wheel (observed from above), the higher the possible intensity felt during that emotion. For example, the intensity for the emotion *grief* can be much higher than that of *sadness* i.e. intensities of emotions increase as we move towards the centre of the wheel.

Lövheim Cube of Emotions

Lövheim (2012) proposed a new explanatory model of emotions based on three monoamine neurotransmitters namely (i) *serotonin*, *dopamine* and *noradrenaline*. Ac-

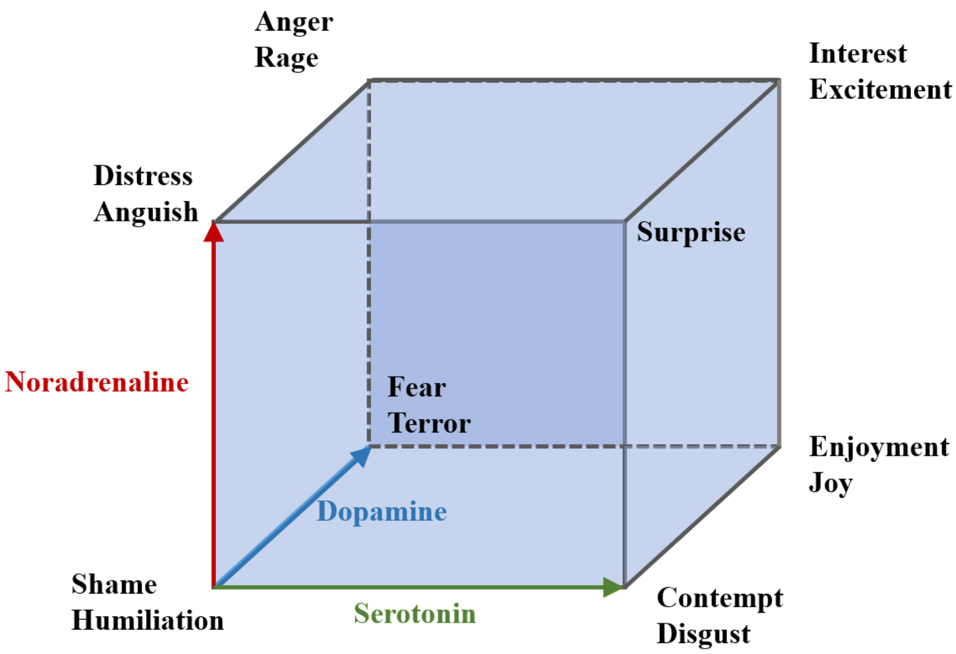


Figure 2.6 Lövheim's cube of emotion. The axes represent the three monoamine neurotransmitters – *serotonin*, *dopamine* and *noradrenaline*. A point within the cube represents the level of each of the three neurotransmitters. Redrawn after Lövheim (2012, p. 342).

cording to this model, *anger* is the result of low serotonin, high dopamine and high noradrenaline. Lövheim (2012) locates eight basic emotions proposed by Tomkins (1982) at the corners of the cube. The characterisation of the various emotions in this model of emotions is summarised in Table 2.2.

Basic Emotion	Serotonin	Dopamine	Noradrenaline
Interest/excitement	High	High	High
Enjoyment/joy	High	High	Low
Surprise	High	Low	High
Distress/anguish	Low	Low	High
Fear/terror	Low	High	Low
Shame/humiliation	Low	Low	Low
Contempt/disgust	High	Low	Low
Anger/rage	Low	High	High

Table 2.2 Levels of monoamine neurotransmitters in various emotions according to the theory of Lövheim (2012). Adapted and modified from Lövheim (2012, p. 343)

Discussion of Dimensional Theories

Although dimensional models of emotion are quite intuitive to understand how various emotions may be structurally organised, these theories are often criticised for being unable to capture the crucial differences among complex emotional states (Remington et al., 2000). For example, the emotional states of fear and anger both lie on the negative side of the axis and both exhibit high level of arousal and hence lie on the same region of the circumplex and also located close by – yet these emotions are entirely different from cognitive perspective (Larsen and Diener, 1992). Moreover, such dimensional representations fall short in explaining the origin or cause of emotional states. Such an explanation is crucial in making computational realisations of an emotion model since such implementations need to be aware of the situational context and analyse the stimuli in relation to the process of emotion generation. This kind of structural representation is of secondary importance while the identification and assessment of the surrounding stimuli is rather critical in such applications. This suggests that an understanding of the theory of emotion that focuses on the cognitive antecedents of the experience of emotion rather than a structural representation is essential to be able to realise the process of emotion generation in artificial agents. In the following sections, I will present a discussion of some of the most influential theories of emotion that consider cognition of pivotal role in the process of emotion elicitation.

2.2.4 Appraisal Theories of Emotion

In the previous sections, I discussed the propositions of physiological, anatomic and dimensional views of emotion processing in humans. I explored how according to some emotion theorists, emotions are purely associated with certain regions in the brain (Damasio et al., 2000; LeDoux, 1996). Likewise, other theorists believed that bodily reactions like heat or heartbeat are not only the consequents of the emotional experience but rather antecedents (Cannon, 1927; James, 1884; Lange, 1885). I also presented the views of other researchers who proposed various structural understanding of the relationship among various emotions (Lövheim, 2012; Plutchik, 1980; Russell and Barrett, 1999; Russell and Mehrabian, 1977; Schlosberg, 1941, 1954). In this section, I will present a different perspective on the process of emotion often called as *Appraisal theory*. *Cognitive* component of an emotional episode (as identified by Moors (2009)) lies at the heart of appraisal theory. According to appraisal theory, emotion in an individual is the result of cognitive evaluation (appraisal) of the given event or situation by the individual (Ortony et al., 1990; Roseman, 1984; Scherer, 2001; Smith et al., 1990). This theory suggests that the evaluation of a situation is performed considering a set of criteria called *appraisal variables*. Appraisal variables can be considered as the basis on which a particular event is evaluated beneficial or harmful by the individual experiencing the situation, which in turn leads to a particular emotional experience as a result of the evaluation. Thus, the resulting emotional state of an individual is determined by how s/he performs the appraisal of the given event based on various appraisal variables. In other words, appraisal theory considers emotion as a more cognitive process than a physio-anatomic one. Since the process of appraisal is person specific and hence subjective in nature, an emotion eliciting stimulus is not always guaranteed to trigger same emotional response for two different individuals (Smith and Lazarus, 1993). For example, imagine a family with children happening to be in a nudist beach without knowing that they may experience embarrassment and concern, whereas a young couple happening in the same situation may experience curiosity and laugh about it. It is because how a person perceives and analyses the environment depends on the individual assessment of the situation, which leads to various emotional experiences. However, this is not to deny that physio-anatomic processes do not at all take part in the process of emotion generation. Lambie and Marcel (2002) suggest that the process of appraisal indeed can occur in two levels: (i) early first-order phenomenological evaluation, which reflects the bodily reactions to the event, and (ii) conscious second-order cognitive appraisal, which denotes higher level awareness of the situation. This view is further supported by other researchers like Frijda (2005). The earliest proponent of appraisal theory of emotion is regarded as Magda B. Arnold (Arnold, 1960). She made a revolutionary proposal of the cognitive analysis of the process of emotional experience in individuals and hence put forward the concept of appraisal in the mechanism of emotion generation (Reisenzein, 2006). Influenced by this concept of appraisal, several other psychologists proposed variations of appraisal theory thereafter (Frijda, 1986; Ortony et al., 1990; Scherer, 2001; Smith et al., 1990). In the remaining of this section, I will present the views of various cognitive appraisal theories which will be followed by my interpretation of the similarities and differences among various appraisal theories.

Frijda's Appraisal Theory of Emotion

In his book, *The emotions*, Frijda expressed his idea that "emotional experience ... is [an] experience of the situation" (Frijda, 1986, p. 193) suggesting that emotions result based on how one evaluates the situation. Frijda is a strong proponent of the concept of appraisal in the process of emotion generation (Frijda, 1986, 1987, 1993). In his early proposal of appraisal theory, he suggested seven basic appraisal variables in relation to the process of emotion elicitation (Frijda, 1986). Table 2.3 summarises the appraisal variables proposed by Frijda (1986).

Appraisal Variable	Description
Change	It is an evaluation of whether the change in the surrounding of the appraising individual is significant/notable to the individual or not.
Familiarity	It is an evaluation of whether the event or the agent involved in the event is familiar to the individual experiencing the situation.
Valence	It is an evaluation of whether the event is likely to cause negative or positive consequences on the individual.
Focality	It is an evaluation of the degree by which the consequences of the event impacts the goals of the individual.
Certainty	It is an evaluation of the likelihood of that the event will have successful consequences.
Intent/Self-other	It is an evaluation of who should be considered responsible for the event or the consequences of the event.
Value Relevance	It is an evaluation of how much does the action of an agent in the event confirm with the norms/values of the appraising individual.

Table 2.3 Appraisal variables proposed in the theory of Frijda (1986).

The notion of *change* as an appraisal variable seems unique. It is described as an individual's ability to identify the change in the environment in response to the event. This variable, therefore, can be considered analogous to what Smith and Ellsworth (1985) describe as *attentional activity*. The appraisal variable *familiarity* measures the degree of familiarity of the appraising individual with the event or the agent involved in the event. The appraisal variable *valence* is the measure of whether the event causes positive or negative feeling on the individual. This assessment is similar to what Scherer calls as *intrinsic pleasantness* (Scherer, 1984, 2001). The appraisal variable *focality* measures how significant is the event to affect the achievement of the goals of an individual. The appraisal variable *certainty*, which is analogous to *outcome probability* in the theory of Scherer (2001) and *probability* in the theory of Roseman (1979), measures the probability of the consequence(s) of the event. The appraisal variable *intent/self-other* is the measure of the cause of the event or its consequences. This measure is analogous to *causal attribution* (Scherer, 2001) and *accountability* (Smith et al., 1990). The appraisal variable *value relevance* measures how much does the action of the agent in the event resemble the norms or values of the appraising individual. This variable is analogous to the appraisal objective of *normative significance evaluation* in the appraisal theory of Scherer (2001).

Scherer's Sequential Check Theory of Emotion

Appraisal theory of Scherer (2001) considers that the process of appraisal does not necessarily occur concurrently as others suggest (Smith and Kirby, 2001). In other words, Scherer (2001) argues that the process of appraisal follows a particular pattern in which certain evaluation criteria (appraisal variables) are measured before other criteria. Such an evaluation process is termed as *Stimulus Evaluation Checks (SEC)* in the theory of Scherer (2001). These evaluation checks offer the achievement of four appraisal objectives namely:

1. *Relevance* – How relevant is this event to me?
2. *Implications* – What will be the consequence of this event on my immediate and long-term goals?
3. *Coping* – Can I cope with the situation caused by this event?
4. *Normative Significance* – How does the event relate to my understanding of norms?

Appraisal objectives and corresponding stimulus evaluation checks in the order of processing are presented in Table 2.4. According to Scherer (2001), *Relevance* of an event to elicit an individual's emotions are determined by three checks in the following order – *novelty check*, *intrinsic pleasantness check* and *goal relevance check*. Novelty check, in case of "most primitive level of sensory-motor processing"(Scherer, 2001, p. 95) suggests whether the event in worth noticing or not. From a much higher level assessment, it refers to the "predictability of the occurrence of the event"(Scherer, 2001, p. 95). Second check for relevance detection is considered as intrinsic pleasantness. It is a measure of likelihood of a positive/negative experience following the event. Final check for relevance detection is goal relevance check which is an early assessment of the chances the event might affect the goals or needs of the appraising individual.

Appraisal Objective	Appraisal Checks	Description
Relevance Detection	Novelty Check	It is an evaluation of whether the event is significant enough to deserve an attention thereby leading to further cognitive assessment.
	Intrinsic Pleasantness Check	It is an evaluation of whether the given event is likely to induce pleasant (positive) or unpleasant (negative) experience.

	Goal Relevance Check	It is the evaluation of whether the event affects the individual's goals and/or needs.
Implication Assessment	Causal Attribution Check	It is an evaluation of the cause of the event.
	Outcome Probability Check	It is an evaluation of the likelihood of certain consequence as a result of the event.
	Discrepancy from Expectation Check	It is an evaluation of how much unexpected was the event.
	Goal/Need Conduciveness Check	It is an evaluation of how much the event helps in the attainment of one or more goals/needs.
	Urgency Check	It is an evaluation of how immediately does the individual react so that the event does not endanger the goals/needs.
Coping Potential Determination	Control Check	It is an evaluation of the extent to which the event or its outcomes can be controlled by the individual.
	Power Check	It is an evaluation of whether the individual has power (i.e. physical strength, money, etc.) to control the situation.
	Adjustment Check	It is an evaluation of whether the individual can adjust or adapt to the situation.
Normative Significance Evaluation	Internal Standards Check	It is an evaluation of whether someone's action in the event confirmed with the internal standards (personal beliefs) of the individual.
	External Standards Check	It is an evaluation of whether someone's action in the event confirmed with the standards of the society (societal beliefs and norms).

Table 2.4 Summary of Stimulus Evaluation Checks in appraisal theory of Scherer (2001).

Once the initial appraisal suggests that the event is relevant to the achievement of the goals, further assessment of the *Implication* of the event to the goal(s) happens. This assessment is composed of following checks – *causal attribution check*, *outcome probability check*, *discrepancy from expectation check*, *goal/need conduciveness check* and *urgency check*. Causal attribution check is an assessment of the cause of the event. In other words, causal attribution refers to the identification of the person or object responsible for the event. Scherer (2001) argues that the it is not the events that cause emotions but the perceived outcomes of the event. That said, probability that the event may cause some consequence on the individual's goals results in emotional experience, which is measured by the variable called outcome probability check. Another check that determines the implication of the event in this theory is discrepancy of the event from expectation of the appraising individual. It is a measure of deviation of the event from the expectations i.e. how unexpected was the event? Likewise, goal/need conduciveness check measures how much the event can support or hinder the attainment of one or more goals or or current needs of the individual. Final check for the objective of implication assessment is urgency check. It measures how critical is it to come up with adaptive actions in reaction to the event depending on the degree by which the event impacts the goals. Scherer (2001) argues that these checks in sequence can help in determining whether an event has any implication on the individual or not. The measure of *Implication* corresponds to the notion of *desirability* as described in OCC theory of emotion. Desirability in OCC theory is the measure of how desirable is the given event for the attainment of the goals of the individual (Ortony et al., 1990).

Coping potential is a capacity of an individual to handle the situation caused by the event in its environment (Scherer, 2001; Smith and Lazarus, 1993). Coping can take two forms, either: (i) being able to change the circumstances of the environment, or (ii) adapting self to the environment (Lazarus, 1991; Scherer, 2001). Appraisal theory of Scherer (2001) considers three appraisal checks to determine the coping potential of the individual in the situation of consequence of the event – *control check*, *power check*, *adjustment check*. Control check is the measure of degree by which the individual can influence or control the consequences of the event. While power check sounds similar to control check, it is operationalised in a different way (Scherer, 2001). Power basically refers to the resources available to cope with the given event. Such sources of power can range from physical strength to money and position in the society. The evaluation check of power after coping sounds incoherent to the core idea of sequential check assumption of Scherer (2001). It is because, I believe that a person would first perform an internal assessment of the power available before measuring the ability to cope with the given situation. Final check for the coping potential determination is suggested as adjustment check. As per the theory, it is a measure of how well the individual can adjust the environment or other agents in favour of the self. In another case, it is a measure of how well the individual can adapt to the current situation.

Ultimate appraisal objective in the theory of Scherer (2001) is the evaluation of *Normative Significance* of the event. Determination of normative significance of an event can be achieved by two checks – *internal standards check* and *internal standards check*. Internal standards refer to an individual's beliefs pertaining to the self. These might include a person's projection of self-image as well as assumed moral code of conduct (Scherer, 2001). Another check suggested in the theory is external standards check. This check is similar to internal standards check but considers the world (society) as opposed to oneself in the latter. The evaluation of normative significance in this theory corresponds to the description of standards and appraisal in relation to the internal standards as presented in the OCC theory of emotion (Ortony et al., 1990).

Scherer (2001) stresses the implications of this theory by arguing that although other appraisal theories may account for the checks proposed in this theory in one way or the other, enforcement of a fixed sequential checks informs a unique strength of the theory. However, although there are some empirical evidence on the partial truth of this assertion (Lanctôt and Hess, 2007), I argue that the very sequential characteristic may be problematic from a computational perspective. One problem is that the implementation of sequential appraisal mechanism in a computational model of emotion causes the emotion processing mechanism to be slower than when processed concurrently. Another problem is that if one appraisal check fails to complete, the appraisal process can not progress at all thereby leading to a deadlock situation. Moreover Scherer (2001) creates confusion through an ambiguous statement "sequence assumption [of the theory] does not deny the existence of parallel processing"(Scherer, 2001, p. 100). Author elaborates it by explaining that his theory does not impose restriction on when an appraisal check starts – it is rather on when it completes (Scherer, 2001). According to the argument in the paper, this theory requires the result of prior check to have completed before a following check can begin – which basically takes us back to the restriction on non-parallel appraisal processing. Because of above mentioned reasons I chose not to rely completely on the appraisal theory of Scherer (2001) for the computational implementation of our model. However, I have adopted some of the useful implications of this theory in my implemented model (Ojha et al., 2017; Ojha and Williams, 2017, 2016), which also draws inspiration from OCC theory of emotion (Ortony et al., 1990). I will discuss more on the computational accounts of emotion implementing appraisal theories in Section 3.1.

Smith and Lazarus's Emotion Theory of Appraisal and Coping

The appraisal theory of Smith et al. (1990) considers coping as a critical component of emotion processing. As such, Smith et al. (1990) put significant focus on *cognitive* as well as *motivational* components of an emotional episode (Moors, 2009). According to this theory how a person appraises an event and hence reaches to an emotional state is determined by the individual's assessment of the ability to cope with the consequences of the event (Smith and Lazarus, 1993; Smith et al., 1990). In line with this argument, the theory defines emotion as an adaptational feature[1] that allows an individual to either

[1]The authors clarify that emotional response is not the only contributor to adaptation – similar function can be served by components of reflexes, physiological drives, etc. (Smith et al., 1990).

adjust the environment to align with owns goals and/or needs or adapt to the change in the environment – as also agreed by Scherer (2001) in his theory of emotional appraisal. All of this has directed the theory of Smith and Lazarus (1993) to explain how emotion helps in adaptation to the environment. While Ortony et al. (1990) define emotion as *valenced reaction to an event*, Smith and Lazarus (1993) consider emotion as a *reaction to the change in person-environment relationship*. This proposition reveals that the latter theory considers an assessment of the 'probable' consequence(s) of the event in the surrounding as a cause of the emotional experience. The reason for putting the notion of 'person-environment relationship' in the centre of the theory could be to capture the influence of environmental factors (like temperature, rain, etc.) on emotion. However, this assumption may lead the theory failing to account for several other situations. For example, consider a situation where a person hears a news of death of his beloved friend who lives far away from him. If we consider this event to have occurred outside the immediate environment of the person who heard the news, the person should not feel any sadness – however that will not be the case. Therefore, the notion of person-environment relationship in their theory should be re-thought in order to eliminate the probable ambiguity in understanding of the theory.

Two levels of appraisal have been proposed by this theory: (i) *primary appraisal* and (ii) *secondary appraisal* (Lazarus, 1966; Lazarus and Smith, 1988; Smith et al., 1990). Primary appraisal evaluates whether the event is relevant to the individual's well-being and secondary appraisal evaluates the individual's ability to cope with the situation.

Primary appraisal is composed of two components: (a) *motivational relevance* and (b) *motivational congruence*. Motivational relevance is the measure of extent by which an event touches the goals of the appraising individual (Smith and Ellsworth, 1987; Smith et al., 1990). Motivational congruence reflects the degree by which the event causes the success or failure in achieving the goal. This notion is similar to the notion of *goal conduciveness* in the theory of Scherer (1984) and *motive consistency* in the theory of Roseman (1984). Secondary appraisal is composed of four components: (a) *accountability*, (b) *problem-focused coping*, (c) *emotion-focused coping* and (d) *future expectancy*. Accountability is an assessment of who shall be held as the cause of the event or its consequences. Remaining three components of the secondary appraisal mainly encapsulate the notion of *coping*. Problem-focused coping potential mainly refers to the ability of an individual to react to the situation based on its power and/or resources. Emotion-focused coping potential is the measure of how well the individual can adapt to the situation by changing its beliefs about the event (Gärdenfors, 2003).

Future expectancy is the measure of how much the event is expected by the experiencing individual in the future.

Appraisal Level	Appraisal Components	Description
Primary Appraisal	Motivational Relevance	It is an evaluation of the degree by which the given event affects the goals of the individual.
	Motivational Congruence	It is an evaluation of whether the event helps in the achievement of the goals or hinders the goal achievement.
Secondary Appraisal	Accountability	It is an evaluation of who is to be credited or blamed for the occurrence of the event.
	Problem-focused Coping	It is an evaluation of the availability of adequate resources in order to deal with the situation.
	Emotion-focused Coping	It is an evaluation of the ability of an individual to adapt psychologically to the similar situations.
	Future Expectancy	It is an evaluation of future possibility of the event as perceived by the individual.

Table 2.5 Summary of Appraisal Components in Cognitive-Motivational-Emotive theory of emotion Smith et al. (1990).

Smith et al. (1990) suggest that the various appraisal components discussed above combine to produce a cognitive entity called *core relational theme*. Each core relational theme helps in determining the type of emotion being elicited by the event (Smith et al., 1990). For example, the emotion of *anger* is elicited by an event when a core relational theme of *other-blame* is activated which is turn in composed of the constituent appraisals – motivationally relevant, motivationally incongruent and other-accountability. Table 2.6 shows a summary of the functional mapping between appraisal components, core relational theme and specific emotions as described in the theory of Smith et al. (1990). The authors have only considered a subset of widely recognised emotions in the description this theory. They admit it by stating “we have not analyzed a number of positive emotions, including happiness, pride, relief, and gratitude”(Smith et al., 1990, p. 621).

Emotion	Core Relational Theme	Appraisal Components
Anger	Other-blame	Motivationally relevant Motivationally incongruent Other-accountability
Guilt	Self-blame	Motivationally relevant Motivationally incongruent Self-accountability

Anxiety	Ambiguous danger/threat	Motivationally relevant Motivationally incongruent Low/uncertain (emotion-focused) coping potential
Sadness	Irrevocable loss	Motivationally relevant Motivationally incongruent Low (problem-focused) coping potential
Hope	Possibility of amelioration/success	Motivationally relevant Motivationally incongruent High future expectancy

Table 2.6 Functional analysis of some emotions based on core relational theme. Adapted from Smith et al. (1990).

The obvious lack of comprehensiveness heavily limits the applicability of this theory in computational realisations of emotion for artificial agents. In addition to this problem, although the association of a particular emotion and appraisal components is bridged by the core relational theme, it is not clear how the intensity of a particular emotion is determined. This explanation is extremely important for implementation of any emotion theory as a computational model because the emotional experience or expression should be differentiated on the basis of the intensity of the processed emotion. Therefore, I borrow only some of the non-ambiguous concepts from the appraisal theory of Smith et al. (1990) in the implementation of the proposed model.

OCC Theory of Emotion

The appraisal theory by Ortony et al. (1990) is one of the most prominent cognitive appraisal theories of emotion. This theory is commonly called *OCC Theory* by abbreviating the surnames of the authors i.e. Ortony, Clore and Collins who wrote the book on this emotion theory. This theory defines emotion as “valanced reactions to events, agents, or objects, with their particular nature being determined by the way in which the eliciting situation is construed” (Ortony et al., 1990, p. 13). OCC theory relates the concept of emotions to the notion of (i) *events*, (ii) *actions of agents*, and (iii) *objects*. As such emotions are considered as the outcome of how an individual evaluates the consequence of the event, nature of the action performed by the agent and attractiveness of the object in interaction (which can be living or non-living being) (Ortony et al., 1990). It should be noted that the concepts of event, action and object are not mutually exclusive and two or more of these can influence the emotion processing mechanism together. For example, consider a situation where you miss your train because one of your college classmates whom you dislike stops you on your way to the station just to ask some unreasonable questions. This event of missing the train will have negative consequence of making you (i) late for work, you find your classmate (ii) blameworthy of the action and you (iii) dislike him more than before.

Appraisal Variable	Governing Criteria	Description
Desirability	Goals	It is an evaluation of the degree by which the event helps or hinders the attainment of one or more goals.
Praiseworthiness	Standards	It is an evaluation of whether the action of the agent confirms with the standards of the appraising individual or not and by how much.
Appealingness	Attitudes	It is an evaluation of how attractive is the interacting object to the appraising individual.
Deservingness	Attitudes	It is an evaluation of whether the agent (including self) deserved what just happened or not.
Proximity (Familiarity)	Attitudes	It is an evaluation of how close is the agent or object to the individual (physical or psychological).
Expectation Deviation (Unexpectedness)	Standards/Attitudes	It is an evaluation of how unexpected the event was or how unexpected the action of the agent was to the individual.
Likelihood	Goals	It is an evaluation of how likely it is that something will happen.
Effort	Goals	It is an evaluation of the extent to which effort was applied towards the achievement of the goals.
Realisation	Goals	It is an evaluation of the extent to which the goals have been realised.
Liking	Attitudes	It is an evaluation of how much is the interacting agent or object liked by the appraising individual.
Sense of Reality (Strength of Cognitive Unit)		It is an evaluation of whether the event, action or object real or imaginary.

Table 2.7 Appraisal variables and evaluation criteria in OCC theory of emotion (Ortony et al., 1990).

Table 2.7 summarises the appraisal variables described in OCC theory of emotion. According to the theory, the appraisal of *desirability* is determined by evaluating how much does the event help in attaining the *goals* of the person; the appraisal of *praiseworthiness* of an action of agent is determined based on the *standards* of the appraising individual; and the appraisal of *appealingness* of an object is determined by the *attitudes* of the appraising individual about the interacting object. In OCC theory, these three appraisal variables are considered as central (local) variables which affect the intensity of only a particular group of emotions (Ortony et al., 1990). I will present a computational account of the relationship between appraisal variables and various emotions in Chapter 4.

Other appraisal variables like *deservingness* measures how much did the third agent or the individual him/herself deserved the recent event or action from another agent. This evaluation may be governed by the attitude of the appraising individual towards the agent which in turn might be affected by previous experience of the individual towards the agent. The appraisal variable *proximity* is the measure of how familiar the interacting agent or object is to the appraising individual. The appraisal variable *expectation deviation* measures how unexpected was the event or the action of the interacting agent in the given context and situation. Other appraisal variables namely *likelihood*, *effort*, and *realisation* reflect the probability, attempt and completion respectively in the achievement of the goals of the individual. The final appraisal variable *sense of reality* measures whether the event, action or object is real or imaginary. OCC theory (Ortony et al., 1990) explains that each of these appraisal variables are evaluated based on the goals, standards or attitudes of the individual and are associated to one or more emotions to be elicited. A detailed discussion of how appraisal variables can be associated with various emotion types will be presented in Chapter 4.

While OCC has provided a description of 22 emotion types under 6 emotion groups (see Table 2.8), I have considered only a subset (described later) of the emotions for the purpose of my computational implementation and experimentation. OCC theory is relatively comprehensive in terms of the depth of explanation in regards to the relation between appraisal variables and various emotions. However, there are several obvious ambiguity in the theory. One example is the use of the word *liking* interchangeably to denote a kind of appraisal as well as a type of emotion. This may introduce some confusion when someone attempts to implement the theory as a computational model of emotion. Moreover, although the theory relates the notions of appraisal with goals, standards and attitudes and suggests that most appraisals are performed under the premises of these concepts, the theory does not explain clearly how these concepts can be related in a practical implementation[2]. Additionally, although the theory says that appraisal variables are associated with various emotions with different degrees, a

[2]Although OCC theory describes the notion of goals in detail in the book (Ortony et al., 1990), it does not explain the computational implications of the goal structure from the perspective of appraisal computation. This problem is not a unique feature of OCC theory. Other appraisal theories also inherently carry this lack of computational adequacy.

clear explanation or direction to realise such a weighted relationship is not available in the book. I will try to address these issues in the later sections of the dissertation and present a discussion of my solution to these inherent problems in the theory.

Emotion Group	Emotion Type	Description
WELL-BEING	1) joy 2) distress	Emotions elicited after evaluation of benefit/harm of the event/action/object to self.
FORTUNES-OF-OTHERS	1) happy-for 2) resentment 3) gloating 4) pity	Emotions elicited after evaluation of benefit/harm of the event/action/object to other agent.
ATTRIBUTION	1) pride 2) shame 3) admiration 4) reproach	Emotions elicited after evaluation of attributes of actions of agent or interacting object to self and vice-versa.
ATTRACTION	1) love 2) hate	Emotions elicited after evaluation of the interacting object.
PROSPECT-BASED	1) hope 2) fear 3) satisfaction 4) fears-confirmed 5) relief 6) disappointment	Emotions elicited after evaluation of the prospect of certain events or actions to occur.
COMPOUND-EMOTIONS	1) gratification 2) remorse 3) gratitude 4) anger	Emotions elicited after a combined evaluation of WELL-BEING and ATTRIBUTION emotions.

Table 2.8 Emotion groups and emotion types in OCC theory of emotion (Ortony et al., 1990).

Roseman's Theory of Emotion

Roseman (1984) is also one of the most referred emotion theorist in psychology. The appraisal theory of Roseman (1979) originally proposed five *dimensions of appraisal*[3] for emotion elicitation. Table 2.9 presents a summary of the appraisal dimensions originally proposed[4] by the theory (Roseman, 1979). The appraisal dimension *motivational state* measures how motivationally appetitive or aversive the event is to the appraising individual. In other words, it is the measure of how desirable or undesirable is the event

[3] I will use the terms 'appraisal variable' and 'appraisal dimension' interchangeably.

[4] It should be noted that the theory was revised later by Roseman in his publication (Roseman, 1984), the discussion of which will follow in the remaining part of this section.

thereby indicating the reward or punishment offered by the event to the achievement of the motives of the individual (Roseman, 1984). This appraisal dimension refers to the similar assessment as offered by the variable *desirability* in appraisal theory of Ortony et al. (1990) the variable *motivational congruence* in the appraisal theory of Smith et al. (1990) and the variable *goal conduciveness* in the appraisal theory of Scherer (2001). Appraisal variable *situational state* measures whether the event is present or absent in the surrounding environment of the individual appraising it. In other words, this variable measures if the event is real or imaginary. The implication of this variable is comparable with what Ortony et al. (1990) call *sense of reality* or strength of cognitive unit. The appraisal variable *probability* is the measure of the likelihood of the event to happen (Roseman, 1984). This appraisal variable is analogous with the variable *likelihood* in the theory of Ortony et al. (1990), the variable *future expectancy* in the theory of Smith et al. (1990) and the variable *outcome probability check* in the theory of Scherer (2001). While the variable outcome probability check in latter theory mainly focuses on the likelihood of the consequences of the events, the variable probability in this theory focuses on the likelihood of the event irrespective of the consequences. The appraisal variable *legitimacy* measures whether the agent experiencing the consequences of the event deserved what happened or not. This notion is analogous to what Ortony et al. (1990) call as *deservingness*. The final appraisal variable *agency* described in the theory measures who is responsible for the event (Roseman, 1984). This variable is analogous with the variable *accountability* in the theory of Smith et al. (1990) and the variable *causal attribution check* in the theory of Scherer (2001).

Appraisal Dimension	Assessed Criteria	Description
Motivational State	1) Appetitive 2) Aversive	It is an evaluation of the degree by which the event rewards or punished of the current motives.
Situational State	1) Present 2) Absent	It is an evaluation of the event assessed as appetitive or aversive is actually present in the vicinity of the appraising individual.
Probability	1) Certain 2) Uncertain	It is an evaluation of an event or consequence of an event is likely to occur.
Legitimacy	1) Deserved 2) Undeserved	It is an evaluation of whether the affected agent in the given scenario really deserved what happened or not.
Agency	1) Circumstances 2) Other 3) Self	It is an evaluation of who should be held responsible for the event or its consequences.

Table 2.9 Appraisal dimensions in appraisal theory of Roseman (1979).

By admitting that the original version of "[t]he theory ... [is] incomplete in several respects" (Roseman, 1984, 1996, p. 244), he offers some changes for the reinterpretation of the description of some of the appraisal dimensions and also offers addition of new dimensions (see Roseman, 1984, 1996, on further discussion of the changes). Nonetheless, despite the fact that the theory presents a well defined structural representation relating the notions of the appraisal and emotions proposed in the theory, the available descriptive accounts do not provide adequate computational perspective capable of allowing a computer scientist to model the theory – except for some implementations which will be reviewed later (El-Nasr et al., 2000; Velásquez and Maes, 1997).

2.2.5 Discussion

Although it seems intuitive to explain the notion of emotion in terms of parts of human body (*i.e.* viscera or cardiovascular systems) or regions of the brain, it can be argued that the available accounts on this notion of emotion mechanism are not rich enough to be realised in computational implementations of emotion model. Even if one may be able to implement a close representation of the body or brain regions as explained in physiological (James, 1894; Lange, 1885) or anatomic theories (Damasio et al., 2000; LeDoux, 1995), it is not practically feasible to accurately replicate the actual mechanism of information flow that occurs within the body and brain because it still is a largely unexplored area of research and a clear and unambiguous accounts of the detailed structural and functional aspects of neural mechanism in brain are not yet available. Apart from these issues, whether the neuro-biological responses are the causes or effects of emotional reaction is still a debatable topic.

As discussed in Section 2.2.3, even if the dimensional models of emotion make it easy to understand the structural relation among various emotions, these theories fall short in explaining how complex emotions are construed (Remington et al., 2000). Moreover, dimensional representations do not take into account the causal relation between the emotion eliciting events and the emotional states, which makes it difficult to realise a computational emotion model solely based on dimensional theories. However, there have been computational implementations on the basis of dimensional theories of emtoion (see for example the work of Becker (2008); Schroder et al. (2011).

Appraisal theories of emotion are often considered capable of addressing the complex causal relation between emotion eliciting stimulus and the experience of emotions (Lazarus, 1991; Ortony et al., 1990; Scherer, 2001; Smith et al., 1990). Therefore, the computational implementation in this research is based on appraisal theory. Based on the discussion of appraisal theories in previous paragraphs, it can be inferred that most of the appraisal variables/dimensions are common among the theories, although they have been referred using different terms (Schorr, 2001). According to these theories, evaluation of the appraisal variables occurs within an individual in response to an event. The result of these evaluations determine which emotional state will be experienced by the individual. A such, the theories proposed so far are good enough to understand the mechanism of cognitive appraisal and how the realisation of appraisal can be useful in the computational modelling of the experience of emotion in artificial agents like robots and virtual characters. However, despite all this, although appraisal theories describe the relation of their appraisal variables with emotions, no clear quantitative association of individual appraisals with associated emotions have been described well in any of the existing theories – except for few examples of attempts of establishing a computational perspective (Ortony et al., 1990; Scherer, 2001). OCC theory of emotion (Ortony et al., 1990) is often considered to be a computationally extensive theory because of the well defined relationship between appraisal variables and emotions. In this dissertation, I will present a plausible computational approach of establishing this relationship between the appraisal variables and emotions inspired by OCC theory. I will present a computational model of emotion accompanied by a detailed mathematical formulation for the mechanism of computation of appraisal variables as well as a machine learning approach of establishing the relationship of appraisal variables with various emotions – solving the problem of lack of quantitative association of appraisals into emotions in the existing appraisal theories. Moreover, it is important to note that the mechanism of mapping the appraisal variables into emotion intensities in this dissertation is not limited to OCC theory – although it has been used in current experiments. The proposed mechanism (to be discussed in detail later in the dissertation) is generic enough to adjust any appraisal theory with any number of appraisal variables and any number of emotions.

2.3 Understanding Appraisal Dynamics

In the previous sections, I presented different theories of emotion available in the literature and I discussed why in this dissertation I chose to build my research on appraisal theories of emotion. Among those appraisal theories, I chose to focus mainly on OCC theory (Ortony et al., 1990), because of its computational advantages over the other available appraisal theories, and additional benefits discussed in the previous section of this chapter. In this section, I will present more in-depth details of the

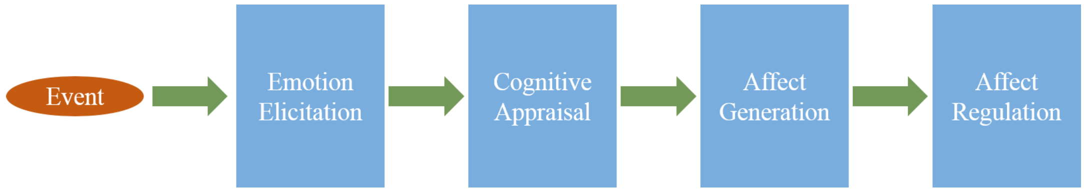


Figure 2.7 Process flow from an stimulus event to (1) emotion elicitation, (2) cognitive appraisal, (3) affect generation, and (4) affect regulation. Adapted from Ojha et al. (2018a).

association of emotion eliciting situation with the appraisal process and the subsequent emotion generation and regulation processes.

Figure 2.7 shows the four basic stages in the process of emotion generation – (1) *Emotion Elicitation*, (2) *Cognitive Appraisal*, (3) *Affect Generation*, (4) *Affect Regulation*. It should be noted that there can be several components involved in the completion of an emotional episode as identified by Moors (2009), namely *somatic*, *cognitive*, *feeling*, *motivational* and *motor* as discussed in the beginning of the chapter. However, most appraisal theories mainly focus on *cognitive* aspect of emotional episode for the activation of *feeling* component with less emphasis given to the discussion of *somatic*, *motivational* and *motor* components. The computational model proposed in this dissertation is mainly focused on the operationalisation of *somatic*, *cognitive* and *feeling* components because these components are adequate enough to explain the process of emotion generation.*Emotion elicitation* can be defined as an early process of attending to the stimulus event and recognising that the event can have either positive or negative impact on the individual. This kind of mechanism is usually considered as a first-order phenomenological response without the involvement of conscious cognitive component (Lambie and Marcel, 2002). This mechanism can be related to the concept of *relevance detection* in the appraisal theory of Scherer (2001). When an event is determined significant enough to trigger emotional reaction, a second-order (higher level) *Cognitive Appraisal* is performed (Lambie and Marcel, 2002; Scherer, 2001). This is where the concepts of various appraisal theories come into play. The variables proposed by the appraisal theories (called *appraisal variables*) are evaluated to analyse how the event may affect the individual or any other agent or object in the environment of the appraising individual. The appraisal variables are associated with various emotions. For example, the appraisal variable *desirability* is related to the emotion *joy* because if an event is desirable then it may induce joy and the degree of joy is determined by the degree of desirability of the event. This kind of mapping leads to the process of *Affect Generation*. According to appraisal theories, an event can lead to the generation of more than one emotions at the same time – but with different intensities (Ortony et al., 1990; Scherer, 2001). Such a situation is handled by the mechanism called *Affect Regulation* (Gross, 1998b, 2002; Gross and Thompson, 2007).

2.4 Role of Mood and Personality in Emotional Appraisal

There is a long standing debate in psychology literature to the question of whether certain emotions are *basic* in nature suggesting that those emotions are universal across people and not influenced by factors like culture, experience and other individual characteristics (Ekman, 1992; Ekman and Cordaro, 2011; Izard, 1992, 2007; Johnson-Laird and Oatley, 1992; Ortony and Turner, 1990; Panksepp, 2007). Even among the advocates of basic emotions, there is no consensus thereby the relevant literature in this area includes a wide range of publications each with a different list of basic emotions. For example, Arnold (1960) presents anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love and sadness as basic emotions on the basis of relation to action tendencies; Izard (1977) proposes anger, contempt, disgust, distress, fear, guilt, interest, joy, shame and surprise as basic emotions based on the analysis that each of these experiences are linked to certain neuro-physiological changes occurring within an individual after being exposed to an stimulus event; Ekman et al. (1982) proposed a set of six basic emotions namely anger, disgust, fear, joy, sadness and surprise based on the study of universality of facial expressions of people across culture, age and regions. The list goes on and it is possible to find as many list of basic emotions as there are emotion theorists. Yet, the set of six basic emotions of Ekman et al. (1982) is the most widely accepted and less controversial compared to other theorists' proposals.

Although some researchers believe that emotions can exhibit universal nature and the processing mechanism of such emotions remains consistent across culture and societies (Ekman, 1992), there have been other studies which suggest that emotion is a multidimensional characteristic of an individual and can be influenced by a wide range of factors. One such factor is identified as *culture*. Some researchers argue that culture can have a huge impact on how an individual construes a stimulus event in relation to emotion elicitation (Hong et al., 2000; Kitayama and Markus, 1994; Lewis and Saarni, 1985; Scollon et al., 2004). Recently, over the last decade, there have been proposals that the process of emotion elicitation can also be influenced by genetic aspects considering *genetics* as a strong factor that can define the emotional behaviour of an individual (Aleman et al., 2008; Bevilacqua and Goldman, 2011; Canli et al., 2009; Hariri and Forbes, 2007). *Personality* is also considered to have a substantial impact on the emotion generation process (Corr, 2008; Revelle, 1995; Watson and Clark, 1997) suggesting that people with certain temperamental qualities are predisposed towards certain emotional experiences and are likely to experience either positive positive or negative emotions with higher degree (Revelle and Scherer, 2009). For example, a chronically pessimistic person (often characterised by the personality trait called *neuroticism*) is more likely to see negative aspects of anything than positives thereby tends to exhibit negative emotional experiences with the stimulus events. This position is also supported by the study of Scherer et al. (2004), where 1,242 participants were asked to remember and describe their emotional experiences, leading to a conclusion that behavioural disposition increases the tendency to experience a particular kind of emotion (Revelle and Scherer, 2009). In addition to the above mentioned factors, researchers believe that the emotional elicitation process is also affected by *mood* state of an individual (Ekman, 1994; Morris, 1992; Neumann et al., 2001). As a conclusion of results of three experiments, Neumann et al. (2001) suggest that "pre-existing mood increases the intensity of affectively congruent emotions while dampening the intensity of incongruent emotions independent of attributional knowledge" (Neumann et al., 2001, p. 725).

Despite several propositions and a range of opinions regarding the factors influencing the process of emotion generation, in this work, I shall limit my investigation to the influence of personality and mood in emotion mechanism. It is because personality and mood are the most discussed factors influencing emotions by several cognitive appraisal theorists (Ortony et al., 1990; Scherer, 2001) and also gather a relatively higher consensus among the computer science researchers (Aylett et al., 2005; Dias et al., 2014; Egges et al., 2004; Moshkina et al., 2011; Rasool et al., 2015; Velásquez and Maes, 1997). Additionally, studies have shown that personality of an individual is largely determined by inheritance (Bouchard Jr, 2004). Durbin et al. (2005) found that temperamental attributes of an individual are determined at birth or very young age. Therefore, the factors of personality, which are considered to be fixed after adulthood (Costa and McCrae, 1988), also capture the influence of genetic organisation (Revelle and Scherer, 2009). Other researchers have argued that personality traits of a person are shaped by long term exposure to certain socio-environmental condition or culture (Benet-Martínez and Oishi, 2008; Dwivedi, 1996; Hogan and Bond, 2009; LeVine, 1963) thereby suggesting that personality traits are the long-term reflection of a person's culture and background. All these facts suggest that considering the factors of personality and mood can help capture a wide range of influencing factors in the process of emotion generation.

2.4.1 Personality Factor

Early two decades of research and discussion on the aspects of human personality (Baumgarten, 1937; Klages and Johnston, 1926; McDougall, 1932) focused on the relation of personality to the language and words. These works laid important foundation for the systematic studies conducted by Cattell (1943, 1946, 1947, 1948). Cattell's system of personality organisation was based on factor-analytic studies of the rating provided by the peers of college students (Digman, 1990). However, the proposal of 16 primary factors and 8 second-order factors made the personality model of Cattell (1948) overly complicated (Digman, 1990) and was heavily criticised by contemporary researchers (Banks, 1948). Fiske (1949) attempted to replicate the study of Cattell (1948) only to find that there was no evidence for the existence of more than five-factors solution in the description of personality (Digman, 1990). A later study conducted by Tupes and Christal (1961) also reported that all the scales under study could be accounted for by the use of five factors namely *surgency*, *agreeableness*, *dependability*, *emotional stability*, and *culture*. Borgatta (1964) also concluded that the personality can be defined by five stable factors – *assertiveness*, *likeability*, *emotionality*, *intelligence*, and *responsibility* (Digman, 1990). Studies by other personality researchers also lead to the conclusion that human personality can be described by five distinct factors (Eysenck, 1970; Guilford, 1975; Norman, 1967; Smith, 1967). Although, most of these researchers agree that five factors are sufficient to understand the notion of personality in humans, the factors they propose show a slight variation on the definition of some of the factors presented. Table 2.10 presents the summary of the evolutionary history of five dimensions of personality.

Author	Factor I	Factor II	Factor III	Factor IV	Factor V
Fiske (1949)	social adaptability	conformity	will to achieve	emotional control	inquiring intellect
Tupes and Christal (1961)	surgency	agreeableness	dependability	emotionality	culture
Norman (1963)	surgency	agreeableness	conscientiousness	emotional	culture
Digman (1988)	extraversion	friendly compliance	will to achieve	neuroticism	intellect
Costa and McCrae (1985, 1992)	extraversion	agreeableness	conscientiousness	neuroticism	openness

Table 2.10 Evolutionary history of five factors of personality. Adapted and modified from Digman (1990).

Of the various *Five-Factor Models* of personality, the model presented by Costa and McCrae (1985, 1992) has remained the most influential model of personality. Costa and McCrae (1985) presented following five factors to be necessary as well as sufficient to analyse the personality traits of a person. There factors are also termed as *Big Five* (Goldberg, 1990; John and Srivastava, 1999).

(I) *Extraversion* – It refers to the characteristic in which a person is more outgoing (McCrae and Costa, 1987) and talkative (McCrae and John, 1992) and has more tendency of experiencing positive emotionality (Watson and Clark, 1997). The notion of extraversion was also proposed by Digman (1988). The use of the factor *surgency* by Tupes and Christal (1961) and Norman (1963) also refers to the similar characteristic of a person.

(II) *Agreeableness* – It covers the range of characteristics that makes a person friendly, compassionate, approachable and forgiving, sympathetic, kind and trusting (McCrae and John, 1992). This factor was also proposed by Tupes and Christal (1961) and Norman (1963). The factor of *friendly compliance* as in the model of Digman (1988) and the factor of *conformnity* as in the model of Fiske (1949) both hold the similar notion as implied by *agreeableness* factor in the model of Costa and McCrae (1985).

(III) *Conscientiousness* – It is the characteristic that makes a person organised (McCrae and Costa, 1987), reliable and systematic (McCrae and John, 1992). The factor *conscientiousness* was also offered by Norman (1963). The notion of *dependability* in the model of Tupes and Christal (1961) and the notion of *will to achieve* in the model of Fiske (1949) and Digman (1988) all aim to denote similar characteristic. However, I argue that the notion of *extraversion* proposed by Costa and McCrae (1985) is able to capture broader range of attributes relevant to the goal of measurement.

(IV) *Neuroticism* – It refers to the higher tendency of experiencing negative emotions like distress (McCrae and John, 1992). Highly neurotic people are also found to experience chronic negative effects (Watson and Clark, 1984). The notion of *neuroticism* is often linked to the concept of *emotionality* (Tupes and Christal, 1961) and *emotional control* (Fiske, 1949).

(V) *Openness* – It includes characteristics such as eagerness to learn new things or try something risky instead of following the regular trend or routine (McCrae and John, 1992). The personality factor of *openness* is quite unique compared to the fifth factor proposed by other researchers. For example, Fiske (1949) and

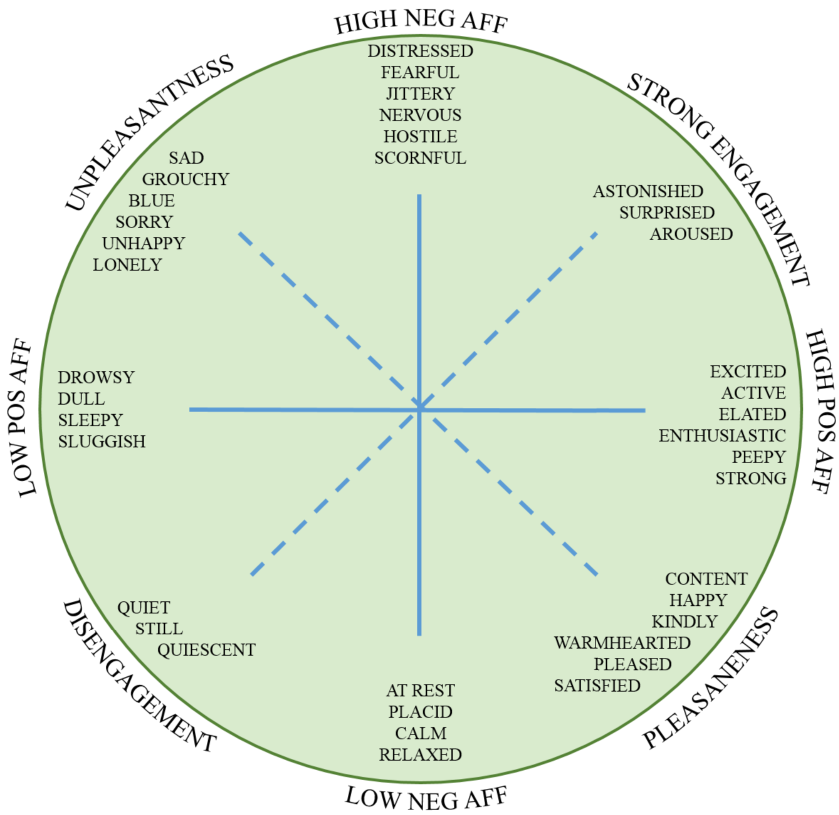


Figure 2.8 *Positive affect* (horizontal) and *Negative affect* (vertical) dimensions of mood (Watson and Tellegen, 1985) and their relationship with the dimensions of *pleasantness* and *arousal* (engagement) of Russell (1979). Redrawn after Watson and Tellegen (1985, p. 220) and Meyer and Shack (1989, p. 693).

> Digman (1988) relate the fifth factor to the intelligence of an individual while Tupes and Christal (1961) and Norman (1963) consider *culture* to be the fifth factor to determine personality of an individual. I argue that, since the notion of *openness* as proposed by Costa and McCrae (1985) also considers other aspects of intelligence like creativity (McCrae and Costa, 1987) and the aspects of culture like inclination to traditional routines McCrae and Costa (1987), the personality factor *openness* is more comprehensive compared to others.

In the remaining of the dissertation, I will refer to the personality factors proposed by Costa and McCrae (1985, 1992) when I use the term *Five-Factor Model (FFM)* of personality. Since the relationship between various emotional states with the dimensions of Five Factors is not well established (Revelle and Scherer, 2009) in the literature, I will offer a computational perspective on the identification of these relationships in this dissertation.

2.4.2 Mood Factor

Psychology literature seems sparse in the study of mood as compared to personality and emotion. Mood is often considered relatively stable and long-lasting experience compared to emotion (Beedie et al., 2005). Most researchers believe that mood is basically an accumulation of repeated emotional episodes over a course of time (Beedie et al., 2005; Ekman, 1994). This suggests that there might be some association between the experience of mood and that of emotion. Therefore, it has also been suggested that mood also follows a circumplex structure similar to that of emotions in dimensional theories of emotion (Russell, 1980; Schlosberg, 1941).

Watson and Tellegen (1985) did an empirical analysis of various mood related data and came to a conclusion that the circumplex structure of *positive affect* and *negative affect* represents a stable and robust structure in the analysis of mood. Their own research findings together with their review and reanalysis of prior works on mood research from other psychologists resulted in the development of the mood model presented in Figure 2.8. The two proposed dimensions of positive affect and negative affect were found to exist consistent in studies across culture (Watson and Clark, 1984) and rated time frames and response formats (Watson, 1988). Watson and Tellegen (1985) explain that the dimensions of *pleasantness* and *arousal* (engagement) proposed by Russell (1979) can be accounted for by rotating the axes in 45° (represented by dotted lines in Figure 2.8). Feldman (1995) has presented an interesting analysis based on empirical study and found that while judging their mood, people weigh the valence (positive vs. negative) dimension much more than the arousal dimension suggesting that the dimension of positivity and negativity is stronger to capture the notion of mood in people (Feldman, 1995). This position coincides with the views of Ekman (1994) who considers mood as less intense compared to emotions. In line with these arguments, I will consider mood as a characteristic having the dimensions of positivity and negativity in the remaining of the dissertation.

2.4.3 Interaction among Emotion, Mood and Personality

It is important to understand that the attributes of emotion, mood and personality vary significantly although some researchers might use the term *affect* to refer to more than one of these characteristics (Mayer, 1986; Petty et al., 1991). Forgas (1992) defines mood as “diffuse and relatively enduring affective state” (Forgas, 1992, p. 230) of an individual. Contrastingly, Forgas (1992) defines emotions as intense and short-lived affective states. The proposition that *emotions are short-lived and mood is long-lasting* is also supported by several other researchers (Ekman, 1994; Mayer et al., 1992;

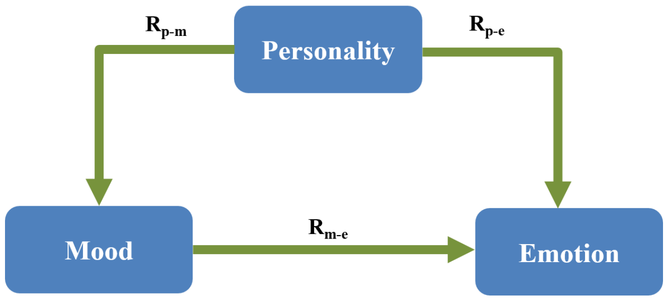


Figure 2.9 Interaction between emotion, mood and personality in a *mediation* approach (Rusting, 1998). The link R_{p-e} denotes the influence of personality factors on emotions; the link R_{m-e} denotes the influence of mood factor on emotions and the link R_{p-m} denotes the influence of personality factors on mood.

Rosenberg, 1998; Sedikides, 1992). Although researchers have proposed that mood state has significant impact in the process of emotion generation (Ekman, 1994; Morris, 1992; Neumann et al., 2001), mood is also considered as an accumulated effect of several emotional experiences with high intensity (Beedie et al., 2005; Ekman, 1994; Parkinson et al., 1996). These propositions suggest to look at the relationship between mood and emotion as a two-way interactive process i.e. *mood affects emotions and emotions in turn affect the resulting mood.*

While mood and emotion are perceived as dynamic characteristics that can change in the course of interaction with surrounding stimuli, personality is considered as a rather stable feature of an individual (Costa and McCrae, 1988; Dweck, 2008), yet it is believed to influence the mechanism of emotion processing (Corr, 2008; Revelle, 1995; Watson and Clark, 1997). Rusting (1998) argued that traditional approach of considering an independent influence of mood and/or personality on emotion may not reflect a true essence of these factors in emotion generation process. A new kind of interaction called *mediation approach* was proposed for understanding how emotion, mood and personality interact (Rusting, 1998). The idea behind this concept is that in addition to affecting emotions directly, personality traits can also influence the individual's disposition towards positive or negative mood which in turn affects emotions through another route. This proposition was supported by a subsequent experimental study study where Rusting (1999) found that personality and mood play an interactive role in modulating emotions i.e. the influence of personality and mood is not independent of each other but rather work in conjunction in the process of emotion generation. This phenomena is demonstrated in the Figure 2.9, and its implementation will be further discussed in Chapter 4 by means of the computational model of emotion provided in this dissertation.

Differentiating Emotion, Mood and Personality

From the discussion above, it can be inferred that the term ‘affect’ can be used to denote human characteristics such as emotion, mood, personality *etc.* (Mayer, 1986; Petty et al., 1991). In order to avoid confusion, I would like to clarify the meaning for each of these terms in the context of this dissertation.

Affective State	Definition
Emotion	A valenced reaction to stimulus event (Ortony et al., 1990) that stays for relatively shorter duration (Ekman, 1994).
Mood	A relatively long-lasting affective state (Ekman, 1994) resulting from predisposition to certain behavioural tendencies (Rusting, 1998) and/or repeated exposure to certain type of stimulus (Beedie et al., 2005).
Personality	A stable characteristic that does not change significantly after adulthood (Costa and McCrae, 1988; Dweck, 2008) and influences an individual’s tendency to experience certain mood or emotional states (Corr, 2008; Rusting, 1998; Watson and Clark, 1997).

Table 2.11 Definition of some affective states in the context of current dissertation.

2.5 Emotion and Ethics[5]

In the previous sections, I discussed about various theories of emotion ranging from the ones that focus on physiological changes for the understanding of the mechanism of emotion (Cannon, 1927; James, 1884; Lange, 1885) to the ones that consider higher cognition as a determinant of emotional state of an individual in reaction to a stimuli (Ortony et al., 1990; Roseman, 1996; Scherer, 2001; Smith et al., 1990). I also presented how the factors of personality and mood are associated with emotion and provide a crucial contribution in modulating the mechanism of emotion processing (Corr, 2008; Ekman, 1994; Morris, 1992; Neumann et al., 2001; Watson and Clark, 1997). In this section, I will discuss how the *emotion regulation* mechanism is related to the process of *ethical reasoning*. The use of the term *regulation* instead of *generation* is explicit because, in this work, I argue that the ethical reasoning acts as an augmenting layer of emotion regulation post the cognitive appraisal process – in line with the arguments of Gross (1998b).

[5]Parts of the discussion in this section have been previously published in Ojha et al. (2018b)

Ethical reasoning can be defined as a mechanism where an individual reasons about the appropriateness of making a decision when a set of conflicting options are available (Hooker, 1996; Robbins and Wallace, 2007). Such a situation where an individual has multiple conflicting choices and a single ethical choice needs to be made is referred to as *ethical dilemma* (Allen, 2012). It is often believed that humans are capable of regulating their emotional states by applying higher cognitive mechanisms like ethical reasoning in order to alter the emotional experience and hence action tendencies (Gross, 2002; Gross and Thompson, 2007). An artificial application that aims to incorporate such an ability of emotional control needs to properly analyse the underpinning association between the aspects of ethical theories in relation to the emotion generation process. As such, the discussion in the following sections will relate the postulates of ethical theories to the implementation of those concepts in computational models.

2.5.1 Theories of Ethics

The term *ethics* does not have a single universal definition. The discussion of ethics in human psychology and philosophy is long standing dating back to the time of Aristotle (Hooker, 1996) and may be even earlier than that. Although there is no accepted definition of ethics, several approaches have been described in literature as the means of making a choice in the scenarios of an ethical dilemma *i.e.* situations where a person has conflicting choices and the options have different impacts on people. Commonly in the literature, ethics is categorised into two types – (i) *descriptive ethics* and *normative ethics* (Robbins and Wallace, 2007).

Descriptive ethics focuses on what is right and what is wrong – in short, it says "what is an ethical action?" (Forsyth, 1980; Kohlberg, 1969). Since my objective is to construct a computational model with ethical emotional control, a descriptive form of ethics may not be adequate for the scope of my research because it is difficult for a computer/robot to process, understand and implement such descriptive information.

Normative ethics or prescriptive ethics is an approach that tells people "how to act ethically" (Hooker, 1996). An artificial agent not only needs to know what is ethical but also to know how to act ethically in social scenarios. For this type of computation, a normative approach to ethics would be a suitable solution because it can be realised in the forms by which an artificial system can be directed. The three most influential normative theories of ethics are (i) *Virtue Ethics*, (ii) *Deontological Ethics*, and (iii) *Consequentialist Ethics* (Hooker, 1996). Table 2.1 shows a summary of some selected normative ethical theories as adapted from Robbins and Wallace (2007).

Based On	Ethical Reasoning Rule	Maxim(s)
Virtue	Aristotelian Virtue (Joachim and Rees, 1953)	One should apply and develop their characteristics towards their distinctive purpose; this will naturally lead to the best actions and best results. To Aristotle, a person's distinctive purpose was to reason.
Duty	Categorical Imperative (Kant, 1993)	Act towards others as you would have them act towards you – or – Treat people as ends, not means.
Duty	Religious Rule-based (form of deontology)	Ten commandments; The Eight-Fold Path; etc.
Duty	Obligation-based (Ross, 1988)	1) Act according to our agreements, 2) Reciprocate help, 3) Be just (reward for merit, etc.), 4) Help those in need, 5) Perfect yourself, and 6) Do no harm.
Duty	Secular Rule-based (Gert, 1988)	1) Don't kill, 2) Don't cause pain, 3) Don't disable, 4) Don't deprive people of freedom, 5) Don't deprive people of pleasure, 6) Don't deceive, 7) Keep your promises, 8) Don't cheat, 9) Obey the law, and 10) Do your duty.
Consequence	Utilitarianism (Bentham, 1907)	Act to reach the highest net ratio of good to bad results for the most number of people.
Consequence	Egoism (Rand, 1964)	Act to reach the highest outcome possible for one's self, irrespective to others.

Table 2.12 Normative theories of ethics. List of selected ethical theories adapted from Robbins and Wallace (2007).

Virtue Ethics

Virtue ethics or Aristotelian ethics (Hooker, 1996) is the form of ethics that focuses on raising the level or standard of one's own character i.e. making oneself virtuous in terms of character. As per the principles of virtue ethics, one should act such that his/her character is going to be considered of high standard. Hence, virtue ethics is also considered as *character-based ethics* since an individual tailors his/her decisions to make his/her character more virtuous.

Deontological Ethics

Unlike virtue ethics, which is *character-based* (Hooker, 1996; Robbins and Wallace, 2007), deontological ethics (Alexander and Moore, 2007) is *duty-based* (Hooker, 1996) meaning it considers the number of duties that an individual is supposed to take into account before choosing an action or making a decision. *Deon* in Greek means *duty* in English. As the name suggests, deontology considers several duties that an individual has to satisfy before making some action/decision. Deontology is different from virtue ethics in that the *goal in virtue ethics is to make one-self good* but the *goal of deontology is to fulfill one's duties* thereby doing good to others. In short, it can be understood that virtue ethics approach is self-directed while deontology is others-directed.

Consequentialist Ethics

Consequentialist ethics (Pettit, 1993) – also known as teleological ethics (from Greek *telos*, meaning end or purpose) (Anderson and Anderson, 2007) or consequentialism (Allen et al., 2000) or utilitarianism (Hooker, 1996; Sen et al., 1982), is commonly called *consequences-based* or outcome-based ethics since it considers the choice of an action that provides the highest overall good consequence to all the parties involved in the decision made by an individual (Bentham, 1907).

A major limitation of virtue ethics, in my view, is that there is no concrete understanding of how a person's character is considered to be of high standard. Since virtue based ethics puts 'self' in the focus, it is less appropriate to be implemented in artificial agents because in practical implementation behaviour selection by artificial agents has ethical impact on 'others' not on the self (Anderson and Anderson, 2007). However, both deontological and consequentialist ethics approaches can be viably used to design a computational model of ethical emotion. The objective of this research is to provide an ethical ability of emotion generation and expression to a robot. Naturally, such a robot has a number of duties towards humans, which makes deontological approach a good option for implementing the desired computational mechanism. But, since the preference of social robots is based on what impact they or their actions have on humans, the ethically emotional robot would not only need to be concerned about its duties, but also need to consider the impact of its behavior on others. This suggests that an integration of deontological and consequentialist approach of ethics would be an ideal solution for the design of artificial ethical emotional agent. In this dissertation, I will propose an approach that combines the concepts of both the deontological and consequentialist ethics in my computational model of emotion for artificial agents.

2.5.2 Connecting Ethics to Emotions

Currently available literature includes a large number of studies examining and confirming that emotional state affects the decision making of an individual thereby determining whether the choice of the person was ethical or not (Callahan, 1988; Gaudine and Thorne, 2001; Isen and Means, 1983). Specifically, these studies examine how an individual's decision in the state of ethical dilemma is influenced by the emotional state of the individual (Ojha et al., 2018b). “Findings of these researches suggest that the emotional state of an individual has a huge impact on the decision s/he makes. For example, a person who never gives a spare coin to a beggar at his train station may decide to hand him a $5 note on the day of his promotion because he is in the emotional state of joy” (Ojha et al., 2018b, p. 213). In agreement with this example, researchers have found that a person in positive emotional state is likely to make more ethical decisions than the one in negative emotional state (Gaudine and Thorne, 2001).

However, in contrast to the research focusing on the effect of emotion on ethical decision making, our exploration revealed that the literature studying the effect of ethical standards, which refers to an individual’s beliefs about what is right and what is wrong, on the process of emotion generation and expression is relatively thin giving rise to several open ended questions on the matter (Ojha et al., 2018b). In the context of computational modelling of emotions, existing models do not consider the role an ethical reasoning mechanism may play in emotion processing (see, for example, the works of El-Nasr et al., 2000; Gebhard, 2005; Marsella and Gratch, 2009). Robots implementing such models able to generate and express emotions autonomously may be interacting with people in its surrounding without a control over the appropriateness of the emotions and actions selected by the underlying computational model of emotion – thereby likely to introduce unacceptable risks in human society. Some researchers (Gratch and Marsella, 2014) have proposed that it could also be possible to regulate emotions of autonomous agents based on the concept of ‘social-functional’ emotions (Gratch et al., 2016; McQuiggan et al., 2008). However, this ideology will only be able to regulate emotions in case of ‘shared goals’ and will not handle situations in multi-agent interactions where one of the ‘innocent’ agents may be a stranger to the appraising agent (i.e. no shared goals). For example, consider a situation where an agent’s companion tries to hurt an innocent stranger.The approach proposed in this dissertation allows an expansion and experimentation of the model in wide range of social situations. As suggested by some researchers, humans have a complex mechanism of regulating their emotional experience and they align their behavioural pattern to conform with the societal values (Gross and Thompson, 2007). In line with this view, in the remainder of this dissertation, I will argue that such an emotional control is strongly governed by ethical standards of an individual. I insist that a mechanism that helps a robot to perform ethical reasoning before reaching to a final emotional state is a crucial aspect to ensure an acceptable and safe robot’s behavior, and social acceptance of such robots in human spaces. My argument is that since, under the appraisal theory perspective, emotion generation is a cognitive process (Ortony et al., 1990; Scherer,

2001) it should be influenced by ethical beliefs and values of a person in one way or the other. “For example, we tend not to express anger to a stupid act of a naïve child but might be angry about the same action from an adult because our standards suggest us to do so. Similarly, a father might not be happy on receipt of a large sum of money from his son which he knows has been robbed from someone in dire need of money – say for the treatment of his ill wife in hospital” (Ojha et al., 2018b, p. 213). It is important to investigate the underlying mechanism that helps us in achieving this kind of control and regulation of emotional states. In the first scenario, one reason for controlling anger might be that we think that an angry and violent response can teach bad behaviour to a young child. In other words, as per our ethical standard, it is our duty to make sure that we do not let negative things affect children. In the second scenario, it is not considered ethical to become happy since the sorrow experienced by the person losing the money might be much more painful in magnitude than the happiness one gets by receiving the money – say his wife could die because of lack of treatment which would impact the lives their kids as they would be left orphans. In other words, negative consequences of the incident on the person (or family) losing the money would be much higher than the positive consequences on the father receiving money. These two examples show how we normally perform an ethical evaluation based on: (i) the notion of our duties and responsibilities (Alexander and Moore, 2007) and (ii) the consequences our decision has on the people involved (Mill, 1901). These two ideas relate to the well accepted ethical theories of *deontology* and *consequentialism* respectively (as discussed previously in this section).

Operationalising Ethical Reasoning in Emotion Processing of Agents

According to appraisal theory, a single event can cause the generation of more than one emotions at the same time (Ortony et al., 1990; Scherer, 2001). This means the completion of *affect generation* process (explained in Section 2.3) activates multiple emotions at a time with varying intensities. One of the challenges in computational modelling of emotions is employing an approach that allows the multiple conflicting emotions to converge to a regulated emotional state. Moreover, it is not always sufficient for an autonomous agent to just express a ‘situation congruent’ emotion which is merely *believable* (Becker, 2008; Reilly, 2006) to the interacting user (Ojha et al., 2017). I argue that believability of an agent’s emotions, although crucially necessary, is not a sufficient quality to elicit socially acceptable emotions (Ojha et al., 2018b) because “[b]elievable characters are characters that seem to be alive ... but does not mean honest, convincing, or realistic” (Reilly, 1996, pp. 8). Therefore, another challenge in emotion models is to ensure that the emotions and, hence, behavioural tendencies of an autonomous agent are not only believable but also socially acceptable.

Term	Definition
Believability	A feature of an artificial agent by which its emotions and behavioural responses seem alive (Reilly, 1996) and congruent to the situation (Becker, 2008; Ojha et al., 2018b).
Social Acceptability	A feature of an artificial agent by which its emotions and behavioural responses are in compliance with common human expectations or norms in given social situations.

Table 2.13 Definition of believability and social acceptability in the context of current dissertation.

Before I proceed further, it is necessary to explain the difference between *believability* and *social acceptability*. In the context of robots, a social robot with emotion generation capability can be considered as believable if it is behaving in situation congruent manner (Becker, 2008), *i.e.* it is exhibiting positive emotions in response to the positive actions and negative emotions in response to the negative actions of the person interacting with it, in the context of emotions, unless designed to respond otherwise (see work of Schroder et al. (2011) and McRorie et al. (2011) for examples of agents acting in incongruent ways and still considered believable). In the context of this dissertation, if a robot expresses sadness if acted rudely and expresses happiness if behaved in a nice way, then its emotion processing mechanism can be considered quite plausible and believable. I argue that it is not sufficient for a robot with emotion generation capacity to be only believable in order to be employed in human society where it has to interact with people of different age, background and nature. Emotion model in such robots should have high level of cognitive ability and should be able to distinguish what is right and what is wrong - at least in the context of emotion generation and expression. The rationale behind this position is that despite being believable, emotions of a robot sometimes may not be considered acceptable. For example, consider an interaction between a robot and a young child. Even if a young child may behave inappropriately with the robot, it should try not to express extreme anger – rather an expression of disappointment would be more socially acceptable. This is because expression of aggressive emotion or behaviour towards sensitive age group such as young children or elderly people may promote anxiety while interacting with autonomous robots, which is considered as one of the hindering factors for social acceptability of robotic technologies (Heerink et al., 2010). Additionally, Mitsunaga et al. (2008) argue that ‘circumstance-appropriateness’ in behavioural responses is a

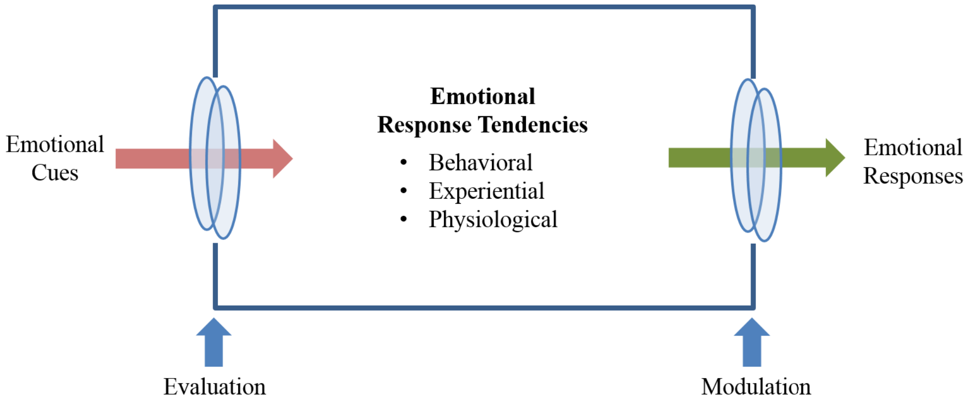


Figure 2.10 A consensual process model of emotion generation and regulation. Redrawn after (Gross, 1998b, p. 272). Original picture appears in Gross (1998a) – Copyright 1998 by the American Psychological Association.

crucial measure of social acceptability of robotic companions. In other words, how appropriately a robot behaves in a given social situation determines its acceptability. Based on these suggestions, this dissertation will consider the definitions presented in Table 2.13 when referring to the terms believability and social acceptability in the rest of the dissertation.

Psychology literature has largely explored the topic of *emotion regulation* which studies "how individuals influence which emotions they have, when they have them, and how they experience and express them" (Gross, 1998b, p. 271), the implications of which can partly enhance the social acceptability in emotions and behaviour of artificial agents. Gross (1998a,b) proposes two broad classes of emotion regulation namely (1) *Antecedent-focused* emotion regulation and (2) *Response-focused* emotion regulation. *Antecedent-focused* emotion regulation refers to the process of augmenting the process of evaluation of emotional cues (which is shown in Figure 2.10) before the generation of emotions (Gross, 1998a). *Response-focused* emotion regulation refers to the process of modulating the emotional responses after the generation of emotions from an appraisal process (Gross, 1998a).

A form of antecedent-focused emotion regulation can be the process called *cognitive change* as suggested by Gross (1998b). Cognitive change is the mechanism of altering the cognitive phenomena of evaluating the emotional stimulus (Frijda, 1986; Gross, 1998b). In other words, it is the process by which an individual alters the appraisal of an emotion eliciting event. Reappraisal (re-evaluation of the situation) is often identified by researchers as an effective form of cognitive change (Lazarus, 1991). The concept of reappraisal in the context of emotion regulation may cause problems when viewed from computational perspective. One thing that lacks clarity of the concept is the number of times the reappraisal should be performed before a regulated emotional state is expected. Alternatively, from a computational perspective, how should an emotion model decide if the cycle of reappraisal should be terminated? Besides, although the 'reappraisal' process of cognitive change may offer a considerable level of emotion regulation, it does not guarantee that the final emotional state will be well regulated since the (re)appraisal process leads to the generation of more than one emotions that need to be converged to a stable state (Ortony et al., 1990; Scherer, 2001). This argument is not to claim that the process of reappraisal is not a useful phenomenon. My point is that although reappraisal may provide a refinement of the initial appraisal process (Lazarus, 1991), it does not guarantee the regulation in the later stages of emotion processing because of other complexities involved. Moreover, since the experience of emotion is widely considered as instantaneous and short lived (Ekman, 1994; Forgas, 1992; Rosenberg, 1998), multiple rounds of reappraisal unnecessarily slow down the process of emotion generation leaving the phenomenon not being able to satisfy the notion of instantaneous and short life. For these reasons, I opted for focusing on the implementation of response-focused emotion regulation process in EEGS in order to achieve not only believability but also social acceptance of its emotional responses. It should also be noted that the idea of emotion regulation proposed in this dissertation does not solely depend on the arguments of Gross and Thompson (2007). While emotion regulation approaches offered by Gross and Thompson (2007) are more 'self-focused' *i.e.* concerned about the emotional well-being of the self, proposed emotion regulation method in this dissertation takes a social perspective (more details on this will be presented in Chapter 4, Section 4.8.2).

2.6 Chapter Summary

This chapter presented an understanding of various *historical perspectives on the phenomena of emotion* and its underlying mechanism. I reviewed some of the most influential theories of the second half of the 20^{th} century and early 21^{st} century as well as some of the theories from late 19^{th} century.

In Section 2.2, I presented a discussion of various emotion theories proposed in the literature. Theories of emotion were classified into: (i) *Physiological theories of emotion*, (ii) *Anatomic theories of emotion*, (iii) *Dimensional theories of emotion*, and (iv) *Appraisal theories of emotion*. The assumption of *physiological theories* of emotion is that the physiological changes in human body in response to a stimulus is not the effect but the cause of emotion (James, 1884; Lange, 1885). In other words, physiological theories assume that emotions occur because of *somato-visceral* (James, 1884) and/or *cardiovascular* (Lange, 1885) changes in the body of an individual. *Anatomic theories* of emotion assert that emotion is all about how a stimulus event is linked to different regions in human brain (Damasio et al., 2000; LeDoux, 1995). According to this theory, what emotion is triggered by an event depends on which brain regions are stimulated by the perception of the event (Montag and Panksepp, 2016). Another class of emotion theories called *dimensional theories* propose a non-discrete approach to understanding emotions. In other words, dimensional theorists believe that emotions should be represented in a continuous *two-dimensional* (Russell, 1980; Schlosberg, 1941) or *three-dimensional* (Lövheim, 2012; Plutchik, 2001; Russell and Mehrabian, 1977) space – allowing the generation of infinite number of emotional states. Unlike other theories of emotion, the fourth class of emotion theories called *appraisal theories* relate the concept of emotion to the notion of *cognition* (Frijda, 1986; Ortony et al., 1990; Scherer, 2001) – hence also called *cognitive appraisal theories* of emotion. According to these theories, emotion generation mechanism in humans is accompanied by how a person cognitively appraises (evaluates) the given stimulus event. Appraisal theories assert that since each individual may evaluate the same stimulus in a different way depending on personal goals, standards and attitudes (Ortony et al., 1990), different persons can have different emotional experience in response to the exact same stimulus – thereby explaining the subjective nature of the experience of emotion in humans, which is governed by several other person-specific factors. Other theories of emotion fail to account for this kind of subjectivity in the experience of emotion. For example, physiological and anatomic theories are unable to account for the modulation in the emotion generation mechanism that can be caused by the characteristics such as *mood* (Ekman, 1994; Morris, 1992; Neumann et al., 2001) and *personality* (Corr, 2008;

Revelle, 1995; Watson and Clark, 1997). Likewise, although dimensional theories of emotion are intuitive to understand the *structure* or alignment of emotions from a spatial perspective, these theories have two serious limitations. *First*, they are not efficient enough to achieve a distinction among complex emotional states (Remington et al., 2000) like fear and anger (Larsen and Diener, 1992). *Second*, these dimensional theories are more focused on the representation of emotional states and do not provide an explanation for the origin or cause of emotional states. In other words, dimensional theories do not provide a clear association of how the stimulus events are associated to their proposed dimensions of the emotion. This posits appraisal theories of emotion as relatively comprehensive from computational perspective – as also pointed by other computer scientists (Dias et al., 2014; Gratch and Marsella, 2004; Hudlicka, 2005; Marsella and Gratch, 2009).

It should be noted that appraisal is not a self-contained process for the generation of emotion. An appraisal mechanism has to interact with several other processes to reach to an ultimate emotional state. In line with this view, I introduced a generic *four-stage* process of *appraisal dynamics* namely (1) *Emotion Elicitation*, (2) *Cognitive Appraisal*, (3) *Emotion Generation*, (4) *Emotion Regulation*, in Section 2.3 (see Figure 2.7). *Emotion elicitation* is the first-order non-cognitive response in which an individual's body attends to the stimulus event in his/her surrounding (Lambie and Marcel, 2002). This process determines positive or negative significance of the event for the individual which is followed by second-order *cognitive appraisal* (Lambie and Marcel, 2002; Scherer, 2001). Once the process of cognitive appraisal is complete, an individual achieves a subjective assessment of the situation in the form of a set of criteria called *appraisal variables*. These appraisal variables are the ones that determine the emotional state of the appraising individual. Such a mapping of appraisal variables into emotions is performed by the process called *emotion generation*. Appraisal theories propose that a single event can lead to the generation of more than one emotions (Ortony et al., 1990; Scherer, 2001). The mechanism for handling the conflicting emotional states is called *emotion regulation* (Gross, 2002; Gross and Thompson, 2007).

The process of *emotion generation* in the four-stage approach presented in Figure 2.7 can be modulated by several factors (see Section 2.4 for more details). Among others, it has been empirically proved that the process of emotion generation is significantly influenced by the factors like mood and/or personality (Morris, 1992; Neumann et al., 2001; Revelle and Scherer, 2009; Scherer et al., 2004). Therefore, I presented an interaction among emotion, mood and personality based on the proposals of Rusting (1998). The work of Rusting (1998) presents an approach called *mediation* where the emotions of an individual are affected by mood and personality. Moreover, mood is also believed to be influenced by emotions – where continuous experience of congruent emotions may lead to the change of mood (Beedie et al., 2005; Parkinson et al., 1996)[6] making emotional experience as instantaneous and short-lived phenomena compared to mood (Ekman, 1994; Mayer et al., 1992; Rosenberg, 1998). Personality, which is relatively stable characteristic compared to mood and emotion (Costa and McCrae, 1988; Dweck, 2008), can influence emotions through two routes (Rusting, 1998) – one *through direct route* and another *through mood* (as shown in Figure 2.9).

The emotions resulting from the *emotion generation* process are regulated by *emotion regulation* mechanism before they influence the behavioural and/or expressive responses (Gross, 2002). In Section 2.5, I proposed that such an influence in humans is mainly achieved through a process called *ethical reasoning* (Hooker, 1996; Robbins and Wallace, 2007). Ethical reasoning is a mechanism in which an individual reasons about the appropriateness of making a decision when a set of conflicting options are available. I propose that an individual follows a similar mechanism when it comes to the experience of an emotional state when multiple conflicting emotional states are triggered. Further empirical analysis in support of this proposal will be presented in the following chapters. In light of the above proposal, I explored three kinds of ethical theories in Section 2.5.1 namely (i) *virtue ethics*, (ii) *deontological ethics*, and (iii) *consequentialist ethics* (Hooker, 1996). Since, deontological ethics (Alexander and Moore, 2007) considers the duties to be fulfilled before making a decision and consequentialist ethics (Allen et al., 2000; Pettit, 1993) considers the overall consequence to the parties involved, I concluded that an integration of both of these aspects can help in achieving a plausible model of emotion regulation in social artificial agents. Technical details of this integration will be presented in Chapter 4.

[6] Although Rusting (1998) does not consider the effect of emotions on mood in initial proposal (as also reflected in Figure 2.9), it should be noted that the emotion generation mechanism in EEGS considers this kind of interaction, which will be discussed in detail in the following chapters.

Chapter Bibliography

Abelson, R. P. and Sermat, V. (1962). Multidimensional scaling of facial expressions. *Journal of experimental psychology*, 63(6):546.

Aleman, A., Swart, M., and van Rijn, S. (2008). Brain imaging, genetics and emotion. *Biological psychology*, 79(1):58–69.

Alexander, L. and Moore, M. (2007). Deontological ethics.

Allen, C., Varner, G., and Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3):251–261.

Allen, K. (2012). What is an ethical dilemma. *New Soc Worker*, 19:4–5.

Anderson, M. and Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, 28(4):15.

Arnold, M. B. (1960). *Emotion and personality.* Columbia University Press.

Aylett, R., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). Fearnot!– An experiment in emergent narrative. In *Intelligent virtual agents*, pages 305–316. Springer.

Banks, C. (1948). Primary personality factors in women: a re-analysis. *British Journal of Mathematical and Statistical Psychology*, 1(3):204–218.

Baumgarten, F. (1937). Die charaktereigenschaften. *The Journal of Nervous and Mental Disease*, 86(1):113.

Becker, C. (2008). *WASABI: Affect simulation for agents with believable interactivity*. PhD thesis, Bielefeld University.

Becker, C., Lessmann, N., Kopp, S., and Wachsmuth, I. (2006). Connecting feelings and thoughts-modeling the interaction of emotion and cognition in embodied agents. In *Proceedings of the 7th International Conference on Cognitive Modeling*.

Beedie, C., Terry, P., and Lane, A. (2005). Distinctions between emotion and mood. *Cognition & Emotion*, 19(6):847–878.

Benet-Martínez, V. and Oishi, S. (2008). Culture and personality. *Handbook of personality: Theory and research*, 3:542–567.

Bentham, J. (1907). *Introduction to the Principles of Morals and Legislation (reprint Ed.).*

Bevilacqua, L. and Goldman, D. (2011). Genetics of emotion. *Trends in cognitive sciences*, 15(9):401–408.

Borgatta, E. F. (1964). The structure of personality characteristics. *Systems Research and Behavioral Science*, 9(1):8–17.

Bouchard Jr, T. J. (2004). Genetic influence on human psychological traits: A survey. *Current Directions in Psychological Science*, 13(4):148–151.

Callahan, S. (1988). The role of emotion in ethical decisionmaking. *Hastings Center Report*, 18(3):9–14.

Canli, T., Ferri, J., and Duman, E. (2009). Genetics of emotion regulation. *Neuroscience*, 164(1):43–54.

Cannon, W. B. (1927). The james-lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4):106–124.

Cannon, W. B. (1929). Bodily changes in pain, hunger, fear and rage.

Carter, L. E., McNeil, D. W., Vowles, K. E., Sorrell, J. T., Turk, C. L., Ries, B. J., and Hopko, D. R. (2002). Effects of emotion on pain reports, tolerance and physiology. *Pain Research and Management*, 7(1):21–30.

Cattell, R. B. (1943). The description of personality: basic traits resolved into clusters. *The journal of abnormal and social psychology*, 38(4):476.

Cattell, R. B. (1946). *The Description and Measurement of Personality*. Yonkers, NY: World Book.

Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika*, 12(3):197–220.

Cattell, R. B. (1948). The primary personality factors in women compared with those in men. *British Journal of Mathematical and Statistical Psychology*, 1(2):114–130.

Corr, P. J. (2008). *The reinforcement sensitivity theory of personality*. Cambridge University Press.

Costa, P. T. and McCrae, R. R. (1985). The neo personality inventory.

Costa, P. T. and McCrae, R. R. (1988). Personality in adulthood: a six-year longitudinal study of self-reports and spouse ratings on the neo personality inventory. *Journal of personality and social psychology*, 54(5):853.

Costa, P. T. and McCrae, R. R. (1992). NEO PI-R professional manual. *Odessa, FL: Psychological assessment resources*, 396:653–65.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., and Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature neuroscience*, 3(10):1049–1056.

Dias, J., Mascarenhas, S., and Paiva, A. (2014). FAtiMA modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer.

Digman, J. M. (1988). Classical theories of trait organization and the big five factors of personality. In *96th Annual Convention of the American Psychological Association, Atlanta, GE*.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440.

Durbin, C. E., Klein, D. N., Hayden, E. P., Buckley, M. E., and Moerk, K. C. (2005). Temperamental emotionality in preschoolers and parental mood disorders. *Journal of abnormal Psychology*, 114(1):28.

Dweck, C. S. (2008). Can personality be changed? the role of beliefs in personality and change. *Current directions in psychological science*, 17(6):391–394.

Dwivedi, K. N. (1996). Culture and personality. A handbook for professionals.

Egges, A., Kshirsagar, S., and Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer animation and virtual worlds*, 15(1):1–13.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Ekman, P. (1994). Moods, emotions, and traits. *P. Ekman & RJ Davidson, The Nature of Emotion: Fundamental Questions*, pages 56–58.

Ekman, P. and Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, 3(4):364–370.

Ekman, P., Friesen, W. V., and Ellsworth, P. (1982). What emotion categories or dimensions can observers judge from facial behavior? In Ekman, P., editor, *Emotion in the human face*. Cambridge University Press, New York.

El-Nasr, M. S., Yen, J., and Ioerger, T. R. (2000). Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257.

Engen, T., Levy, N., and Schlosberg, H. (1958). The dimensional analysis of a new series of facial expressions. *Journal of Experimental Psychology*, 55(5):454.

Eysenck, H. (1970). *The structure of human personality*.

Feldman, L. A. (1995). Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*, 21(8):806–817.

Feldman Barrett, L. and Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, 74(4):967.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, 44(3):329.

Forgas, J. P. (1992). Affect in social judgments and decisions: A multiprocess model. *Advances in experimental social psychology*, 25:227–275.

Forsyth, D. R. (1980). A taxonomy of ethical ideologies. *Journal of Personality and Social psychology*, 39(1):175.

Frijda, N. (2005). Emotion experience. *Cognition & Emotion*, 19(4):473–497.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and emotion*, 1(2):115–143.

Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition & Emotion*, 7(3-4):357–387.

Gärdenfors, P. (2003). *Belief revision*, volume 29. Cambridge University Press.

Gaudine, A. and Thorne, L. (2001). Emotion and ethical decision-making in organizations. *Journal of Business Ethics*, 31(2):175–187.

Gebhard, P. (2005). Alma: A layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM.

Gert, B. (1988). A new justification of the moral rules.

Gladstones, W. (1962). A multidimensional study of facial expression of emotion. *Australian Journal of Psychology*, 14(2):95–100.

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216.

Gratch, J. and Marsella, S. (2004). Technical details of a domain-independent framework for modeling emotion. Technical report, University of Southern California Marina Del Rey Ca Inst for Creative Technologies.

Gratch, J. and Marsella, S. (2014). *Social emotions in nature and artifact*. Oxford University Press.

Gratch, J., Nazari, Z., and Johnson, E. (2016). The misrepresentation game: How to win at negotiation while seeming like a nice guy. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 728–737. International Foundation for Autonomous Agents and Multiagent Systems.

Gross, J. J. (1998a). Antecedent-and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of personality and social psychology*, 74(1):224.

Gross, J. J. (1998b). The emerging field of emotion regulation: An integrative review. *Review of general psychology*, 2(3):271.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, 39(3):281–291.

Gross, J. J. and Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. *Handbook of emotion regulation*.

Guilford, J. P. (1975). Factors and factors of personality. *Psychological Bulletin*, 82(5):802.

Hariri, A. R. and Forbes, E. E. (2007). Genetics of emotion regulation. *Handbook of emotion regulation*, pages 110–132.

Heerink, M., Kröse, B., Evers, V., and Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: the almere model. *International journal of social robotics*, 2(4):361–375.

Hogan, R. and Bond, M. H. (2009). Culture and personality.

Hong, Y.-y., Morris, M. W., Chiu, C.-y., and Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7):709.

Hooker, J. (1996). Three kinds of ethics.

Hudlicka, E. (2005). A computational model of emotion and personality: Applications to psychotherapy research and practice. In *Proceedings of the 10th annual cybertherapy conference: a decade of virtual reality*.

Isen, A. M. and Means, B. (1983). The influence of positive affect on decision-making strategy. *Social cognition*, 2(1):18–31.

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on psychological science*, 2(3):260–280.

James, W. (1884). What is an emotion? *Mind*, 9(34):188–205.

James, W. (1894). The physical basis of emotion.

Joachim, H. H. and Rees, D. A. (1953). Aristotle: The nicomachean ethics.

John, O. P. and Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138.

Johnson-Laird, P. N. and Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition & Emotion*, 6(3-4):201–223.

Kant, I. (1993). *Grounding for the metaphysics of morals: With on a supposed right to lie because of philanthropic concerns*. Hackett Publishing.

Kitayama, S. E. and Markus, H. R. E. (1994). *Emotion and culture: Empirical studies of mutual influence.* American Psychological Association.

Klages, L. and Johnston, W. H. (1926). The science of character.

Kleinginna, P. R. and Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4):345–379.

Kohlberg, L. (1969). Stage and sequence: the cognitive-developmental approach to socialization, goslin, da (ed.), handbook of socialization and endash; theory and research. *Chicago: McNally*.

Lambie, J. A. and Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychological review*, 109(2):219.

Lanctôt, N. and Hess, U. (2007). The timing of appraisals. *Emotion*, 7(1):207.

Lang, P. J. (1994). The varieties of emotional experience: a meditation on james-lange theory. *Psychological review*, 101(2):211.

Lange, C. G. (1885). *Om sindsbevaegelser; et psyko-fysiologisk studie*. Lund.

Lange, C. G. and James, W. (1922). *The emotions*, volume 1. Williams & Wilkins.

Larsen, R. J. and Diener, E. (1992). Promises and problems with the circumplex model of emotion.

Lazarus, R. S. (1966). *Psychological stress and the coping process.* McGraw-Hill.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press on Demand.

Lazarus, R. S. and Smith, C. A. (1988). Knowledge and appraisal in the cognition—emotion relationship. *Cognition & Emotion*, 2(4):281–300.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and molecular neurobiology*, 23(4-5):727–738.

LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annual review of psychology*, 46(1):209–235.

LeDoux, J. E. (1996). *The emotional brain: The Mysterious Underpinnings of Emotional Life*. New York, Simon & Schuster.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual review of neuroscience*, 23(1):155–184.

LeVine, R. A. (1963). Culture and personality. *Biennial review of anthropology*, 3:107–145.

Lewis, M. and Saarni, C. (1985). Culture and emotions. In *The socialization of emotions*, pages 1–17. Springer.

Lövheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses*, 78(2):341–348.

Marsella, S., Gratch, J., Petta, P., et al. (2010). Computational models of emotion. In *A Blueprint for Affective Computing-A sourcebook and manual*, volume 11, pages 21–46. Oxford University Press New York, NY.

Marsella, S. C. and Gratch, J. (2009). Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90.

Mayer, J. D. (1986). How mood influences cognition. *Advances in cognitive science*, 1:290–314.

Mayer, J. D., Gaschke, Y. N., Braverman, D. L., and Evans, T. W. (1992). Mood-congruent judgment is a general effect. *Journal of personality and social psychology*, 63(1):119.

McCrae, R. R. and Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1):81.

McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

McDougall, W. (1921). Belief as a derived emotion. *Psychological Review*, 28(5):315.

McDougall, W. (1932). Of the words character and personality. *Journal of Personality*, 1(1):3–16.

McQuiggan, S. W., Robison, J. L., Phillips, R., and Lester, J. C. (2008). Modeling parallel and reactive empathy in virtual agents: An inductive approach. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, pages 167–174. International Foundation for Autonomous Agents and Multiagent Systems.

McRorie, M., Sneddon, I., McKeown, G., Bevacqua, E., de Sevin, E., and Pelachaud, C. (2011). Evaluation of four designed virtual agent personalities. *IEEE Transactions on Affective Computing*, 3(3):311–322.

Meyer, G. J. and Shack, J. R. (1989). Structural convergence of mood and personality: Evidence for old and new directions. *Journal of personality and social psychology*, 57(4):691.

Mill, J. S. (1901). *Utilitarianism*. Longmans, Green and Company.

Mitsunaga, N., Miyashita, Z., Shinozawa, K., Miyashita, T., Ishiguro, H., and Hagita, N. (2008). What makes people accept a robot in a social environment-discussion from six-week study in an office. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3336–3343. IEEE.

Montag, C. and Panksepp, J. (2016). Primal emotional-affective expressive foundations of human facial expression. *Motivation and Emotion*, 40(5):760–766.

Moors, A. (2009). Theories of emotion causation: A review. In *Cognition and Emotion*, pages 625–662. Psychology Press.

Morris, W. N. (1992). A functional analysis of the role of mood in affective systems.

Moshkina, L., Park, S., Arkin, R. C., Lee, J. K., and Jung, H. (2011). TAME: Time-varying affective response for humanoid robots. *International Journal of Social Robotics*, 3(3):207–221.

Mulligan, K. and Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review*, 4(4):345–357.

Neumann, R., Seibt, B., and Strack, F. (2001). The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Emotion and Cognition*, 15:725–747.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, 66(6):574.

Norman, W. T. (1967). 2800 personality trait descriptors: Normative operating characteristics for a university population.

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., and Williams, M. (2017). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *Annual Meeting of the Cognitive Science Society*.

Ojha, S. and Williams, M. (2017). Emotional appraisal: A computational perspective. In *Annual Conference on Advances in Cognitive Systems*.

Ojha, S. and Williams, M.-A. (2016). Ethically-guided emotional responses for social robots: Should i be angry? In *International Conference on Social Robotics*, pages 233–242. Springer.

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.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Ortony, A. and Turner, T. J. (1990). What's basic about basic emotions? *Psychological review*, 97(3):315.

Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian journal of Psychology*, 7(1):1–30.

Panksepp, J. (1988). Brain emotional circuits and psychopathologies. In *Emotions and psychopathology*, pages 37–76. Springer.

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.

Panksepp, J. (2007). Criteria for basic emotions: Is disgust a primary "emotion"? *Cognition and Emotion*, 21(8):1819–1828.

Parkinson, B., Totterdell, P., Briner, R. B., and Reynolds, S. (1996). *Changing moods: The psychology of mood and mood regulation*. Longman.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*. Academic press.

Pettit, P. (1993). Consequentialism.

Petty, R. E., Gleicher, F., and Baker, S. (1991). Multiple roles for affect in persuasion. In Forgas, J. P., editor, *Emotion and social judgments*, pages 181–200. Blackwell Publishing, Elmsford, NY: Pergamon Press.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Theories of emotion*, 1.

Plutchik, R. (2001). The nature of emotions. *American scientist*, 89(4):344–350.

Prkachin, K. M., Williams-Avery, R. M., Zwaal, C., and Mills, D. E. (1999). Cardiovascular changes during induced emotion: An application of lang's theory of emotional imagery. *Journal of Psychosomatic Research*, 47(3):255–267.

Rand, A. (1964). The virtue of selfishness. *New York: Signet*.

Rasool, Z., Masuyama, N., Islam, M. N., and Loo, C. K. (2015). Empathic interaction using the computational emotion model. In *IEEE Symposium Series on Computational Intelligence*, pages 109–116. IEEE.

Reilly, S. N. (2006). Modeling what happens between emotional antecedents and emotional consequents. *ACE*, 2006(19).

Reilly, W. S. (1996). *Believable Social and Emotional Agents.* PhD thesis.

Reisenzein, R. (1983). The schachter theory of emotion: two decades later. *Psychological bulletin*, 94(2):239.

Reisenzein, R. (2006). Arnold's theory of emotion in historical perspective. *Cognition and Emotion*, 20(7):920–951.

Remington, N. A., Fabrigar, L. R., and Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of personality and social psychology*, 79(2):286.

Revelle, W. (1995). Personality processes. *Annual review of psychology*, 46(1):295–328.

Revelle, W. and Scherer, K. R. (2009). Personality and emotion. *Oxford companion to emotion and the affective sciences*, pages 304–306.

Robbins, R. W. and Wallace, W. A. (2007). Decision support for ethical problem solving: A multi-agent approach. *Decision Support Systems*, 43(4):1571–1587.

Roseman, I. J. (1979). Cognitive aspects of emotion and emotional behavior. In *87th Annual Convention of the American Psychological Association, New York*, volume 58.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of personality & social psychology*.

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, 10(3):241–278.

Rosenberg, E. L. (1998). *Levels of analysis and the organization of affect.*, volume 2. Educational Publishing Foundation.

Ross, W. D. (1988). The right and the good (1930). *Digital Text International*.

Russell, J. A. (1979). Affective space is bipolar. *Journal of personality and social psychology*, 37(3):345.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.

Russell, J. A. and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805.

Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294.

Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin*, 124(2):165.

Rusting, C. L. (1999). Interactive effects of personality and mood on emotion-congruent memory and judgment. *Journal of personality and social psychology*, 77(5):1073.

Rusting, C. L. and Larsen, R. J. (1995). Moods as sources of stimulation: Relationships between personality and desired mood states. *Personality and Individual Differences*, 18(3):321–329.

Schachter, S. (1959). *The psychology of association*. Stanford University Press.

Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. *Advances in experimental social psychology*, 1:49–80.

Schachter, S. and Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5):379.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. *Approaches to emotion*, 2293:317.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120):57.

Scherer, K. R., Wranik, T., Sangsue, J., Tran, V., and Scherer, U. (2004). Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction patterns. *Social Science Information*, 43(4):499–570.

Schlosberg, H. (1941). A scale for the judgment of facial expressions. *Journal of experimental psychology*, 29(6):497.

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of experimental psychology*, 44(4):229–237.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological review*, 61(2):81.

Schorr, A. (2001). The evolution of an idea. *Appraisal processes in emotion: Theory, methods, research*, pages 20–34.

Schroder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., Ter Maat, M., McKeown, G., Pammi, S., Pantic, M., et al. (2011). Building autonomous sensitive artificial listeners. *IEEE transactions on affective computing*, 3(2):165–183.

Scollon, C. N., Diener, E., Oishi, S., and Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of cross-cultural psychology*, 35(3):304–326.

Sedikides, C. (1992). Changes in the valence of the self as a function of mood. *Review of personality and social psychology*, 14(27):l–3.

Sen, A., Williams, B., and Williams, B. A. O. (1982). *Utilitarianism and beyond*. Cambridge University Press.

Smith, C. A. and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, 48(4):813.

Smith, C. A. and Ellsworth, P. C. (1987). Patterns of appraisal and emotion related to taking an exam. *Journal of personality and social psychology*, 52(3):475.

Smith, C. A. and Kirby, L. D. (2001). Toward delivering on the promise of appraisal theory. In *Series in affective science. Appraisal processes in emotion: Theory, methods, research*, pages 121–138. Oxford University Press.

Smith, C. A. and Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition & Emotion*, 7(3-4):233–269.

Smith, C. A., Lazarus, R. S., et al. (1990). Emotion and adaptation. *Handbook of personality: Theory and research*, pages 609–637.

Smith, G. M. (1967). Usefulness of peer ratings of personality in educational research. *Educational and Psychological measurement*, 27(4):967–984.

Tomkins, S. S. (1982). Affect theory. In Ekman P, Friesen W, E. P., editor, *Emotions in the human face*, page 355–395. Cambridge: Cambridge University Press.

Tupes, E. C. and Christal, R. E. (1961). Recurrent personality factors based on trait ratings. Technical report, PERSONNEL RESEARCH LAB LACKLAND AFB TX.

Velásquez, J. D. and Maes, P. (1997). Cathexis: a computational model of emotions. In *Proceedings of the first international conference on Autonomous agents*, pages 518–519. ACM.

Watson, D. (1988). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of personality and social psychology*, 55(1):128.

Watson, D. and Clark, L. A. (1984). Negative affectivity: the disposition to experience aversive emotional states. *Psychological bulletin*, 96(3):465.

Watson, D. and Clark, L. A. (1997). Extraversion and its positive emotional core.

Watson, D. and Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98(2):219.

Williams, F. and Sundene, B. (1965). Dimensions of recognition: Visual vs. vocal expression of emotion. *Educational Technology Research and Development*, 13(1):44–52.

Zevon, M. A. and Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, 43(1):111.

The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.

— Sir William Bragg —

3

Computational Emotion Models and Research Context

In Chapter 2, I presented an analysis of various theoretical concepts ranging from the *theories of emotion* to the *theories of ethics*. The discussion in the previous chapter was mostly inclined to the understanding of various psychological and philosophical research pertaining to the attainment of the goals of current work. Unlike in the previous chapter, where only a limited associations of the theoretical premises with computational accounts were presented, this chapter shall deal with a review of existing computational models of emotion. Moreover, based on the review of the existing models, appropriate hypotheses will be established which will guide the upcoming discussion throughout the dissertation.

3.1 Computational Models of Emotion

My review of the existing computational models of emotion will start with a brief overview of the model followed by a brief discussion of its strengths as well as a critical analysis of the limitations of the model particularly examining the five crucial properties explained below.

(i) **Is the model *domain-independent*?**
By saying domain-independent, I refer to the property of a computational model of emotion where the same model can be used in the appraisal of an event in various domains without changing the computational details (Ojha and Williams, 2017). In other words, a computational model of emotion can be considered domain-independent, if it does not implement user-defined domain-specific rules for the computation of appraisals, and hence the appraisal rules can be applied in more than one domain. This property is particularly important to make a model applicable in disparate situations (Castellanos et al., 2018; Gratch and Marsella, 2004a; Ojha and Williams, 2017).

(ii) **Does the model integrate the aspects of *mood*?**
This criteria examines whether a computational model of emotion realise the concept of mood. Mood may be used to influence the emotion intensities and emotion intensities can, in turn, be used to influence the generation of mood as suggested by previous researchers (see Section 2.4 for more details on the role of mood in emotion processing).

(iii) **Does the model integrate the aspects of *personality*?**
Similar to mood, this criteria examines if a computational model of emotion integrates the notion of personality. This realisation can be made by modelling the influence of personality directly on emotion intensities and/or also through an indirect route of changing the mood and then emotion intensities (see Section 2.4 for further discussion on the influence of personality on emotion).

(iv) **Does the model offer *data-driven mapping* of the appraisals into emotion intensities?**
By saying data-driven mapping, I refer to the process in which the relationship of the appraisal variables with various emotions (as suggested conceptually in the OCC appraisal theory Ortony et al., 1990) is determined utilising human data instead of using user-defined static parameters or rules. My argument is that implementing ad-hoc rules to map the appraisal variables into emotion intensities may not provide plausible association between these quantities. Therefore, it is important to have a mechanism that allows the model to establish this relationship based on the empirical data supplied.

(v) **Does the model implement *ethical reasoning* mechanism to regulate the emotional state of the agent?**
Although researchers have suggested ways to regulate emotions (see, for example Lazarus, 1991, for the discussion of the concept of 'coping'), the proposals do not

explicitly address the notion of ethical reasoning. In this dissertation, I introduce an ethical reasoning mechanism for emotion regulation (to be discussed in detail in Chapter 4, Section 4.8.2). This criteria examines whether a computational model of emotion employs an explicit mechanism of ethical reasoning to regulate its emotions in order to reach to the most socially acceptable emotional state – and hence action tendencies.

It is important to note that for the purposes of real life human-agent and human-robot interaction situations the above criteria may not be sufficient to qualify an emotional model as suitable for effective deployment. There are other questions that could be asked such as (1) Does the model learn the mappings from actual (human-agent) interaction data? (2) Is the model (artificial emotion generation) fully autonomous? (3) Does the model (emotion generation) run in real-time (in an online manner)? (4) Has the model (artificial emotion generation) been tested/evaluated in an actual human-agent interaction context? I have to admit that answering these questions is not in the current scope of the dissertation and attempting so could over-complicate the current issue. As such, this dissertation will deal with laboratory-like experiments for the purposes of evaluation of the proposed model (see Chapter 5 for details).

In the following sections, I will review the historical computational models of emotion in chronological order of their occurrence. I will mainly focus my discussion on the most influential models proposed in the last two decades. A reader who rather intends to get the gist of the reviewed models and does not want to get into the computational details of each of the model reviewed may jump to Section 3.1.16, where I have presented the summary of various computational emotion models as well as comparison of several models based on above five criteria in Table 3.1 and Table 3.2 respectively.

3.1.1 Cathexis

Cathexis (Velásquez, 1997; Velásquez and Maes, 1997) is a computational model of emotion mainly influenced by the concepts of Izard (1993) and Tomkins (1962). The model stands as a strong foundation in the emotion modelling literature and forms a basis of inspiration for several models proposed afterwards. The Cathexis system as a whole is an interaction between *perceptual system*, *emotion generation system*, *drive system*, *behavior system* and *motor system* (Velásquez, 1999).

Perceptual system perceives the surrounding stimulus and provides the information to *emotion generation system* directly as well as to other systems like *motor system* and *behaviour system*, which are responsible for physical or behavioural responses. In Cathexis, drives play a significant role in biasing the actions of agent through the

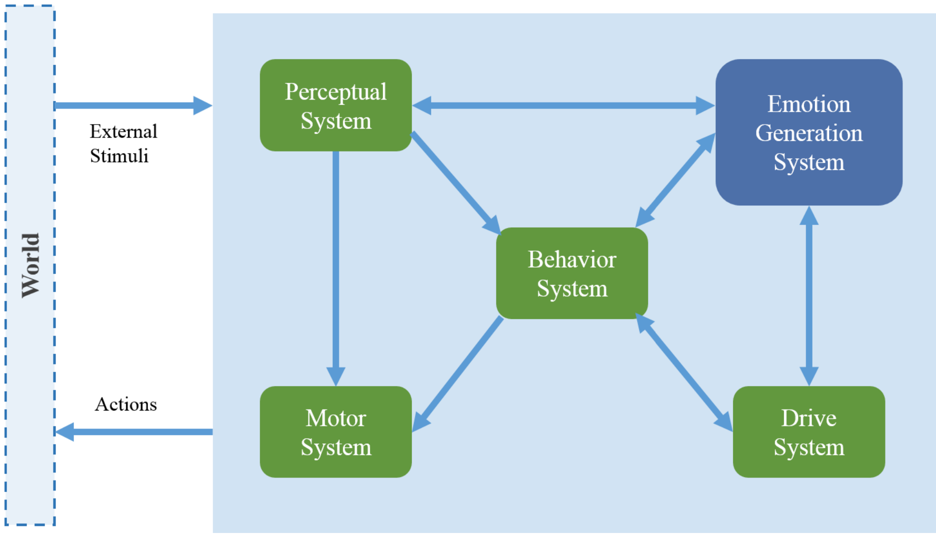


Figure 3.1 Several components of Cathexis model. Redrawn after Velásquez (1999, p. 236).

implemented *drive system*. For example, a version of the Cathexis model implemented in Yuppy robot (Velásquez, 1999) implements four different drives like *recharging-regulation* which monitors the battery level, *imitation* & *dance* which aim to entertain the interacting people and *fatigue* which checks the robot's activity level (Velásquez, 1999). Core emotion processing mechanism occurs in *emotion generation system*. Although, Velásquez (1997) considers *neural*, *sensorimotor* and *motivational* in addition to *cognitive* antecedents of emotions, how these concepts are integrated together is not clearly explained.

One issue with Cathexis model evident from its implementation, in a virtual toddler (Velásquez, 1997) and Yuppy robot (Velásquez, 1999), is that the concepts of motivation and drives have to be handcrafted to work well independently in each of these as well as any other scenario. Despite the author mentions that the cognitive aspect of emotions in Cathexis is based on the theory of Roseman et al. (1990), he does not provide details on how this notion is computationally realised. It seems that the author uses pre-defined appraisal rules to realise this kind of relationship. Moods and temperaments (personality) in Cathexis are implemented using the concepts proposed by Minsky (1988). Mood is modelled as a low arousal characteristic compared to emotions (Velásquez, 1997). However, temperament is modelled by setting user-defined threshold for various emotions – which, however, was not modelled upon empirical evidence or theories available from emotion research. Furthermore, the details of how the appraisals were mapped into emotion intensities were not sufficiently provided, thus introducing additional obstacles when attempting to assess the model's performance

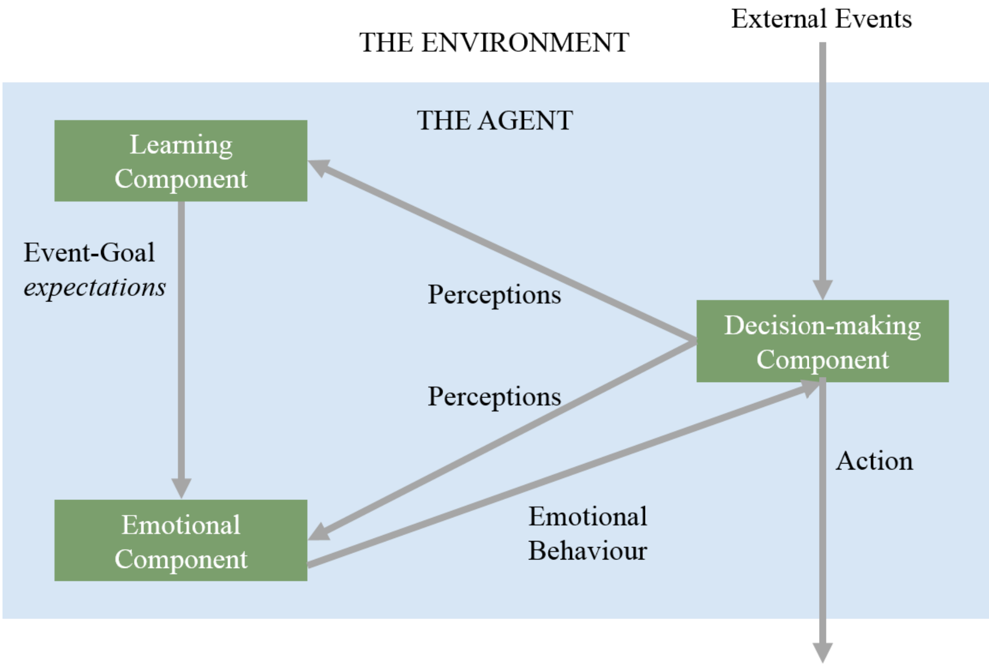


Figure 3.2 FLAME agent architecture. Redrawn after El-Nasr et al. (2000, p. 227).

within different scenarios. Finally, behaviour system consists of a pool of possible behaviours like *startle*, *dance*, *imitate*, *greet-person*, *etc.* (Velásquez, 1999) which hardly cover a broad range of social scenarios, and those actions are not regulated by any mechanism assessing their social impact, in contrast to the model I will present in this dissertation.

3.1.2 FLAME

FLAME (**F**uzzy **L**ogic **A**daptive **M**odel of **E**motions) is a computational emotion model of El-Nasr et al. (2000) strongly based on Fuzzy Logic (Zadeh, 1996). Unlike the traditional set theory concept of binary inclusion (*i.e.* an element either lies in a set or does not lie in it), fuzzy set theory allows a partial membership of an element (Zadeh, 1996). FLAME uses similar concepts in mapping an emotion eliciting event to emotion intensities as well as to action selection process.

Figure 3.2 shows an overall agent architecture of FLAME emotion model (El-Nasr et al., 2000). *External events* from the environment are turned into *perceptions* by the *Decision-making Component*. According to the authors, the evaluation of perceptions takes place in two steps. In the first step, experience determines which *goals* are influenced by the recent event and in the second step, mapping rules are used to determine the *desirability* of the event. The computation of desirability relies on two factors: one is (i) the degree by which the event affects the goal, and another is (ii) the importance of the affected goal (El-Nasr et al., 2000). FLAME is inspired by the appraisal theory of Ortony et al. (1990) and Roseman (1996). In addition to *desirability*, FLAME also uses another appraisal variable – *expectation*, to generate emotional reactions (El-Nasr et al., 2000). These variables are first calculated as labels using Fuzzy logic and then converted to numeric values with a defuzzification process (El-Nasr et al., 2000). After this process, emotions are quantified with intensity using the formulae proposed by Price et al. (1985). In the scenario of conflicting (or mixed) emotions, motivational states (Bolles and Fanselow, 1980) such as hunger, thirst, pain, fatigue, *etc.* are used to filter a situation congruent emotional state. Mood is modelled in FLAME as merely a consequent of emotions where mood depends on the intensity of the recent emotional experiences (El-Nasr et al., 2000). Unlike other models, mood is modelled as a binary value which can either be *positive* or *negative* instead of continuous value. This is probably a reason FLAME does not model the influence of mood on emotion intensities.

One limitation of FLAME model is the lack of task segregation among various components. From the details and formulae provided, it is not clear enough the functions of each component and their inter-dependencies. To mention just an example, it is not clear enough if the goals are influenced by an event by means of the Decision-making Component, the Learning Component or both. This and other ambiguities introduce additional barriers limiting replication studies of the FLAME model, thus leading to major research gaps. With this limitation in mind, in this dissertation, I will provide exhaustive details of the computational mechanisms used in the provided model, motivated by appropriate literature and assumptions. This will facilitate replicating the model and its adoption for future research and comparative studies. In addition to the above mentioned ambiguities, the appraisal and mapping rules presented in the paper are strongly domain dependent. In fact, by being heavily characterised by domain-dependent knowledge, FLAME model requires to define new appraisal rules when situated in different scenarios. Obviously, defining such new rules is not an easy task, and details on how to proceed with this task are not provided by the author. While the model presents the use of Fuzzy logic as a strength, it can actually introduce computational limitations. Indeed, the process of fuzzification and defuzzification result in the loss of a lot of useful information thereby altering the emotional and behavioural tendencies of the agents running the model. I target to address this issue by presenting my loss-less approach of obtaining continuous values of the appraisal variables as well as emotion intensities later in this dissertation (see Chapter 4, Section 4.7.3). Besides these issues, authors state their awareness that “mood may also aid in filtering the mixture of emotions developed" (El-Nasr et al., 2000, p. 234) but do not realise this in their model. Likewise, FLAME fails to integrate the notion of personality in the process of emotion generation which they agree is "one of the most important features that can enhance believability of animated characters" (El-Nasr et al., 2000, p. 253). The authors admit that it is hard to effectively integrate the notion of mood and personality in their version of FLAME and requires further research to make such a realisation. Additionally, although the model uses a behaviour selection strategy (also based on Fuzzy logic), their focus is on increasing the believability of FLAME model instead of focusing on the regulatory mechanisms to ensure social implications of the emotional and behavioural responses.

3.1.3 Emotion, Mood and Personality Model of Egges et al. (2004)

Egges et al. (2004) proposed a generic model for the integration of the concepts of personality, mood and emotions to obtain a believable responses of virtual conversational characters. This is an improved version of their previous works (Egges et al., 2003; Kshirsagar, 2002). Figure 3.3 shows an overview of the generic personality, mood and emotion framework of Egges et al. (2004). It considers a scenario where a user interacts with an intelligent agent (conversational character) through some interface. The user provides *perceptive data* which can be in the form of speech, text, images, *etc.* (Egges et al., 2004), which is stored in the form of a *semantic interpretation* by the agent. The perceptive data is also appraised using an appraisal theory (in this case the authors show an example of Ortony et al., 1990). The evaluation of the perceptive data is based on the *goals*, *standards* and *attitudes* of the intelligent agent. This appraisal is affected by personality factors of the agent (Egges et al., 2004). *Personality* factors also affect the *Emotional State* in this model. First the personality influences the emotions of the agent based on what the authors call "*personality-emotion influence matrix*" (Egges et al., 2004, p. 5), which contains the author defined parameters that link each personality factors to each emotion type modelled in the framework. In the similar manner, authors model the influence of personality factors on mood based on "*personality-mood influence matrix*" (Egges et al., 2004, p. 7). While the influence of emotions on mood is determined by the "*emotion-mood influence matrix*" (Egges et al., 2004, p.7), how the mood affects emotions is not clearly explained – although the authors present a high level functional representation (see Egges et al., 2004, eqn. (14), p. 6). Once the emotions and mood get updated, personality, mood state and

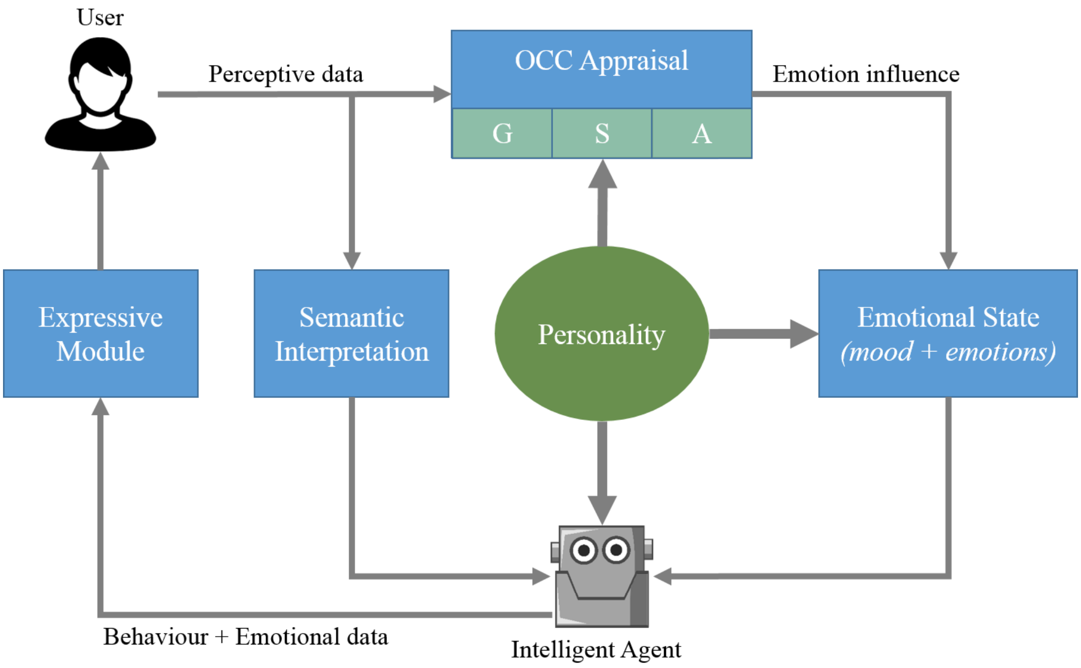


Figure 3.3 Overview of the integrated personality, mood and emotion model of Egges et al. (2004). Redrawn after Egges et al. (2004, p. 2).

emotions altogether define the behaviour of the intelligent agent which in turn influence the expressive patterns operated through an *expressive module*.

Although at a glance, it seems that the model proposed by Egges et al. (2004) is comprehensive enough to capture the mutual interaction among personality factors, mood and emotion, there are some critical issues in their framework. The first issue emerges from the *ad hoc* realisation of *personality-emotion influence matrix*, *personality-mood influence matrix*, and *emotion-mood influence matrix* without any empirical or theoretical premises. Authors merely mention about the use of these matrices in operationalising the interaction among emotion, mood and personality but do not provide details on how these matrices can be obtained. The authors conclude their study by admitting that they are unsure about how the appraisal data is conceptualised in the emotion processing mechanism (Egges et al., 2004), and merely pass an m dimensional vector[1] called *emotion influence* from the appraisal component. The paper does not explain how the vector is computed, thus leaving the reader without enough details to replicate such mechanism. This makes it difficult to assert if the model is indeed domain-independent or not – since it cannot be easily replicated and tested in new and different scenarios. Besides, personality and mood do not take part in mapping the appraisals into emotional states. In addition to these limitations, although the framework

[1]The dimension m represents the number of emotions modelled based on the theory used.

incorporates the behavioural and expressive modules which are controlled by personality and emotional state, authors do not explicitly employ emotion regulation mechanisms that ensure the social acceptance of the behavioural responses of the agent using their framework.

3.1.4 FearNot!

FearNot! short name for "Fun with Empathetic Agents Reaching Novel Outcomes in Teaching" (Aylett et al., 2005, p. 306) is a computational model of emotions implemented in a virtual storytelling (Cavazza et al., 2002) framework. The application was originally intended to teach anti-bullying lessons to young children in schools (Aylett et al., 2005). Children would be asked to participate in an emergent narrative environment (Aylett, 2000) and provide advice to the victims of bullying in the simulation scenario, where they would act as 'invisible friend' of the victim and empathise with him/her. Figure 3.4 shows the overall agent architecture implemented in the FearNot! application. *Sensors* perceive the agent's environment and receives the stimulus event. This stimulus event is then evaluated (appraised) based on the features of objects, agents and the event itself – as described in the theory of OCC (Ortony et al., 1990). The resulting emotional state from the appraisal process is used for action selection at two levels: (i) *action-tendencies* and (ii) *coping behaviour* (Aylett et al., 2005) – inspired by the theory of Lazarus (1991). *Schematic Level* controls the action tendencies and triggers the impulsive actions of the agent. *Coping Level* controls the problem-focused and emotion-focused behavioural responses. Action tendencies affect the body, speech and facial expressions through a unit called *Effectors*. Problem-focused coping refers to deliberated actions taken as a measure to control the consequences of the stimulus event and emotion-focused coping involves the changing the self belief or interpretation of the environment (world).

The description of the architecture makes it apparent that FearNot! agent architecture is inspired by the appraisal theory of Ortony et al. (1990) and Lazarus (1991) – which are well accepted theories of emotional appraisal. However, the major issue with the FearNot! architecture is its strong domain-specificity. The authors tested their architecture in only few specific anti-bullying situations and generalising this model to a broader range of applications and scenarios can become quite hard. Indeed, this limitation is acknowledged by the authors, suggesting that – "it was envisaged that if the advice was to 'tell a teacher' or 'tell a parent', then this would happen off-stage to avoid the difficult issues involved in representing teachers and parents as (possibly less than perfect) story characters." (Aylett et al., 2005, p. 308). Moreover, although

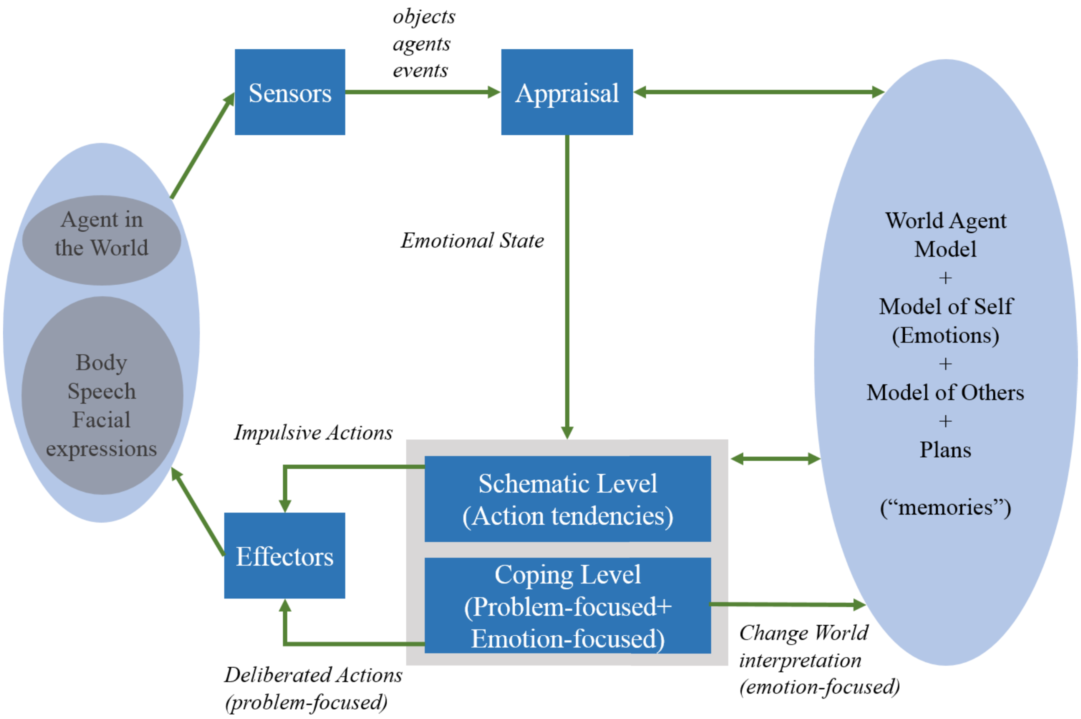


Figure 3.4 FearNot! affectively driven agent architecture. Redrawn after Aylett et al. (2005, p. 309).

the architecture shows that event, object and agent are considered while appraising the sensory information, it is not clear how these concepts of OCC theory (Ortony et al., 1990) are actually operationalised in their model. Another limitation of FearNot! architecture is that the "character's personality is ... strongly based [o]n OCC" (Dias and Paiva, 2005, p. 131). OCC theory is primarily the theory of emotional appraisal which only presents a very brief discussion of the notion of personality in relation to emotion generation leaving the reader to explore the depth of the topic from other sources. By using the limited understanding of personality provided by OCC theory, the authors may lose the opportunity to reach more significant results. Personality research itself consists of a vast analysis of a number of factors, for which I provided a discussion in Section 2.4. FearNot! also integrates the notion of mood in the emotion generation process. Mood is considered as the cumulative effect of several emotional experiences – in line with the views of Ekman (1994). Based on the proposals of Picard (1997), the mood also affects the possibility of experiencing the positive or negative emotions. For example, if the agent is in a good mood state, it is more likely to experience positive emotions. As discussed earlier in the dissertation, mood and personality can be strong modulating factors that affect the process of emotion generation and hence the overall emotional experience of an individual (see Section 2.4).

But, psychology literature does not provide sufficient details to computationally realise such relationships. FearNot! and other emotion models have addressed this limitation by offering pre-defined rules or static parameters to implement such a relationship between personality, mood and emotions (Dias and Paiva, 2005; El-Nasr et al., 2000; Moshkina et al., 2011). However use of this approach cannot ensure whether the interaction between personality, mood and emotion resembles what actually happens in humans. I propose that a computational model of emotion should be allowed to identify this kind of relationship based on personality and mood data collected from humans. I shall discuss more on this in Section 3.1.16. Although the coping strategies have been simulated in limited pre-defined domains for anti-bullying applications in experimental scenarios, the model is not able to autonomously reason for emotion regulation in absence of the '*stage manager*' (Aylett et al., 2005, p. 308) who is a user supposed to determine the characters' actions within the simulated scenario. Apart from these, the authors do not provide some of the implementation details, such as the type and number of appraisal variables used and how appraisal variable were mapped into emotional state.

3.1.5 ALMA

ALMA (**A** **L**ayered **M**odel of **A**ffect) is a computational model of emotion that aims to integrate the concepts of emotion, mood and personality in a single system (Gebhard, 2005). The model is implemented to influence the emotions and hence behavioural responses of virtual 3D conversational characters. Gebhard (2005) argues that emotions can influence the dialog selection strategies by such virtual characters thereby leading to different dialogues and facial expressions based on its current emotional state. The author uses the notions of emotions from the OCC theory of emotional appraisal (Ortony et al., 1990), borrows the dimensions of pleasure(P), arousal(A) and dominance(D) from Mehrabian (1996a,b) for the generation of mood, and reflect the concepts of the Five Factor Model of personality to realise the implementation of personality aspects (Digman, 1990; McCrae and John, 1992).

The mechanism of appraisal in ALMA is unclear as the model does not provide sufficient computational details on how a stimulus event is mapped into various appraisal variables – although the paper mentions that it is based on appraisal theory of Ortony et al. (1990). The model is hardly generalisable and very domain-specific because "it has to be defined how each kind of affect will be computed and how they interact with each other" (Gebhard, 2005), so not being able to achieve this mapping through a more generic and domain-independent appraisal computation mechanism. Although mood is modelled in ALMA, it is considered solely as a function of personality factors and emerging emotional states. Similar to other models (Gratch and Marsella, 2004a; Marinier III et al., 2009) mood is influenced by the aggregated emotional experiences (Gebhard, 2005). One unique aspect of this model commonly not seen in other computational models of emotion is the determination of the initial mood state based on the personality factors. The model uses eight types of mood based on the work of Mehrabian (1996a).

- **Exuberant – Bored** $\rightarrow$ (+P+A+D) vs. (-P-A-D)
- **Dependent – Disdainful** $\rightarrow$ (+P+A-D) vs. (-P-A+D)
- **Relaxed – Anxious** $\rightarrow$ (+P-A+D) vs. (-P+A-D)
- **Docile – Hostile** $\rightarrow$ (+P-A-D) vs. (-P+A+D)

The PAD dimensions of initial mood are determined, as shown below, by the personality factors (O = openness, C = conscientiousness, E = extraversion, $\mathbb{A}$[2] = agreeableness, and N = neuroticism) as suggested by Mehrabian (1996a).

$$P = 0.21 * E + 0.59 * \mathbb{A} + 0.19 * N$$

$$A = 0.15 * O + 0.30 * \mathbb{A} - 0.57 * N$$

$$D = 0.25 * O + 0.17 * C + 0.60 * E - 0.32 * \mathbb{A}$$

In summary, one limitation of ALMA is that the mood of a virtual agent is not used to modulate the emotion intensities which is also admitted by the author in the conclusion section of the paper. Also, the appraisal mechanism has a strong domain dependence and the appraisal rules need to be handcrafted for each new scenario or interaction domain. Additionally, despite the lack of emotion modulation by personality factors, the use of the personality traits in the explanation of model's implementation is ambiguous. For example, the author uses the terms 'severe versus tolerant' and 'dry versus flowery' but does not provide the necessary links describing how these terms are related to the trait theory and the notion of five factors mentioned in the paper. Besides, the model does not explain how the elicitation of conflicting emotional states is managed and how the emotional and behavioural regulation occurs in ALMA.

[2]The notation for agreeableness will be used as $\mathbb{A}$ only in this section to avoid ambiguity with *arousal*(A) dimension of PAD space.

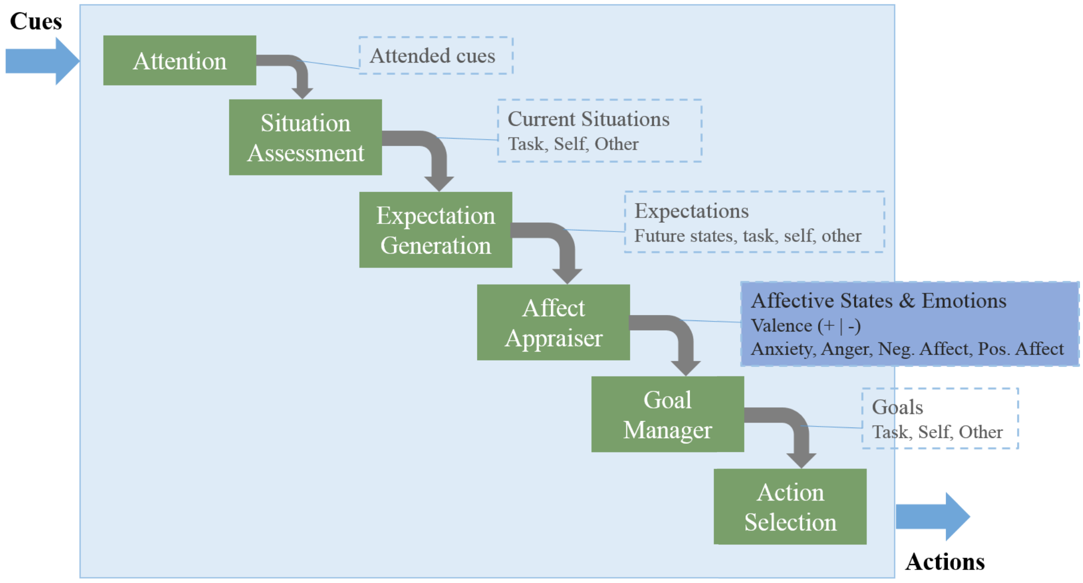


Figure 3.5 Process flow in MAMID cognitive-affective architecture. Redrawn after Hudlicka (2008).

3.1.6 MAMID

MAMID (Hudlicka, 2005) is a *cognitive-affective* architecture with a particular focus of role of emotion on cognition (Hudlicka, 2008). MAMID operates in a *see-think-do* processing sequence where a stimulus event (*cue*) is processed by a *thinking* (cognitive) unit and the situation assessment is applied into *actions* selection. *Attention* module filters the incoming cues and selects only a subset to be processed. The *attended cues* are fed into the *situation assessment* module which performs a semantic appraisal of the current event and then provides information to the *expectation generation* module, which maps the current situation into probable future states. These new expectations lead to the generation of affective states and emotions in *affect appraiser* module. The generated affective and emotional states affect the process of goal selection in *goal manager* module and the choice of actions in *action selection* module.

One limitation of MAMID is evident from the application of the model in a *clinical practice* scenario. Hudlicka (2005) uses a set of very specific *if....then* rules which make the process of appraisal as well as action selection heavily domain-dependent. Moreover, although the model integrates the notion of personality to adjust the responses of the model (Hudlicka, 2002, 2005), it does not account for the aspects of mood. Likewise, the model does not employ a regulatory/control mechanism to ensure that an agent implementing it responds with socially plausible actions or behaviours.

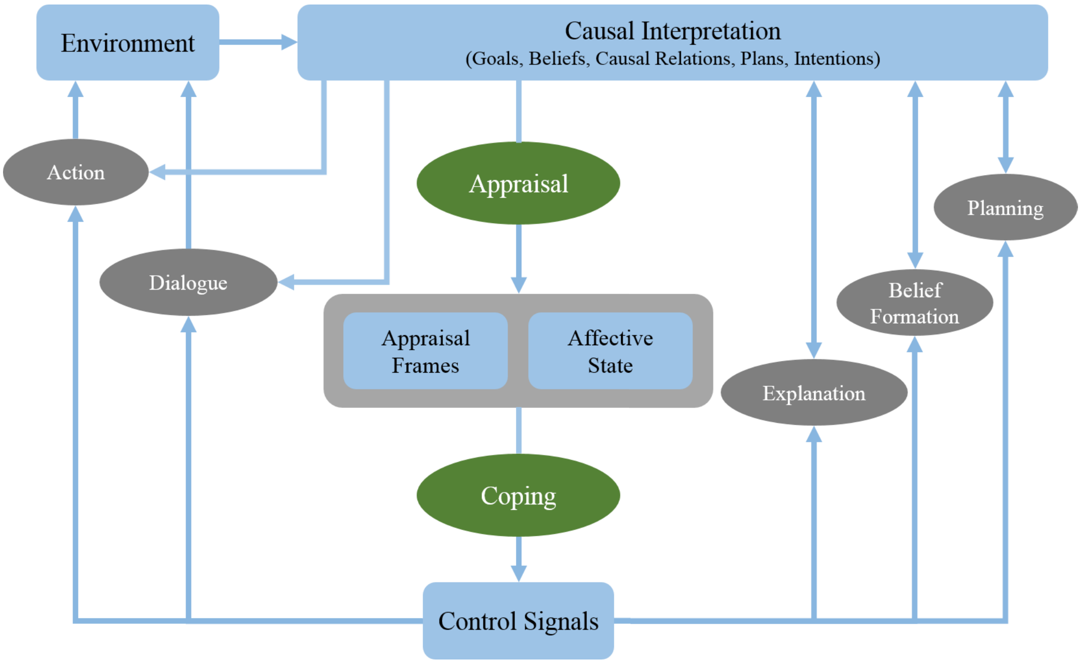


Figure 3.6 Cognitive–motivational–emotive system architecture of EMA model. Redrawn after Gratch and Marsella (2004a, p. 278).

3.1.7 EMA

EMA, short form for **EM**otion and **A**daptation, is a computational model of emotions hevaily based on the appraisal theories of Smith et al. (1990) and Lazarus (1991). EMA is a representative example of an emotion model considering coping at the heart of emotion processing. According to Gratch and Marsella (2004a), an analysis of coping potential by an individual has a strong influence on emotion regulation and action tendencies (Gratch and Marsella, 2004a; Marsella and Gratch, 2009). EMA considers a single-level three stage loop of *appraisal* $\rightarrow$ *coping* $\rightarrow$ *reappraisal*. A heavy reliance on the notion of coping and reappraisal reflects the influence of Lazarus (1991) and Ellsworth (1991).

Figure 3.6 shows the overall structure of EMA model. *Environment* represents the relationship of the appraising individual with other agents or objects in the surrounding of the individual. Gratch and Marsella (2004a) term this relationship as *person-environment relationship* based on the work of Smith et al. (1990). When an emotion eliciting event occurs within the person-environment relationship, an interpretation of the event is required. Hence, the information about the event is sent to the *causal interpretation* component of the model. The primary role of this component is to interpret the event based on the agent's internal *goals*, *beliefs*, *causal relations*, *plan* and *intentions*, which then assists in the *appraisal* of the event. An appraisal of an event leads to the instantiation of multiple *appraisal frames* which are then mapped into corresponding *affective states* using the mapping rules of Elliot (1992). These affective states are the drivers of the behavioural responses of the agent experiencing the emotion inducing situation. This is where the notion of *coping* comes into play. The coping mechanism proposed by Gratch and Marsella (2004a) helps to control or alter various subsequent processes that are likely to be influenced by the emotional state of an agent. The coping mechanism in EMA (Gratch and Marsella, 2004a) can be *problem-focused* or *emotion-focused* as explained in the theory of Smith et al. (1990). Problem-focused coping promotes action tendencies (action or dialogue) that help to change the person-environment relationship in favour of the appraising individual. Emotion-focused coping promotes the change in the causal interpretation by providing further mental explanation to the interpretation of the event, forming/revising the beliefs and (re)planning the course of actions. These coping strategies then might lead to the reappraisal of the situation.

The concept of appraisal is a common aspect of appraisal theories although they differ on the type, number and sequence of the appraisal process. While most appraisal theories do not deny to the notion of coping as an important aspect in emotion regulation (Ortony et al., 1990; Scherer, 2001), they do not also put a strong emphasis on the crucial role the coping strategies might play – for example in the case of reappraisal of a situation (Lazarus, 1991). EMA efficiently realises this notion of coping based on the theory of Lazarus (1991). Although this is an important position, EMA model and also the subsequent descriptions of the model (see Marsella and Gratch, 2009, for more details), it is difficult to understand 'how' these coping strategies can be generalised to various situations and/or other computational models. Gratch and Marsella (2004a) are strong advocates of domain-independent emotion modelling and EMA is one of the most influential emotion models being able to perform appraisals in domain-independent manner. When dealing with additional factors interacting with emotion, EMA only models the effect of mood. In EMA, mood is computed as an aggregation of emotion instances passed through a sigmoid function (Gratch and Marsella, 2004a). This mood is then used to alter the intensities of the emotion instances in the next appraisal round.

3.1.8 Soar-Emote

Soar-Emote (Marinier III and Laird, 2004) is a computational unification of theory of cognition with theory of emotion (mainly appraisal theory – which itself embraces the notion of the role of cognition) (Marinier III and Laird, 2007; Marinier III et al., 2009). This model argues that the gaps in the PEACTIDM theory of cognitive control

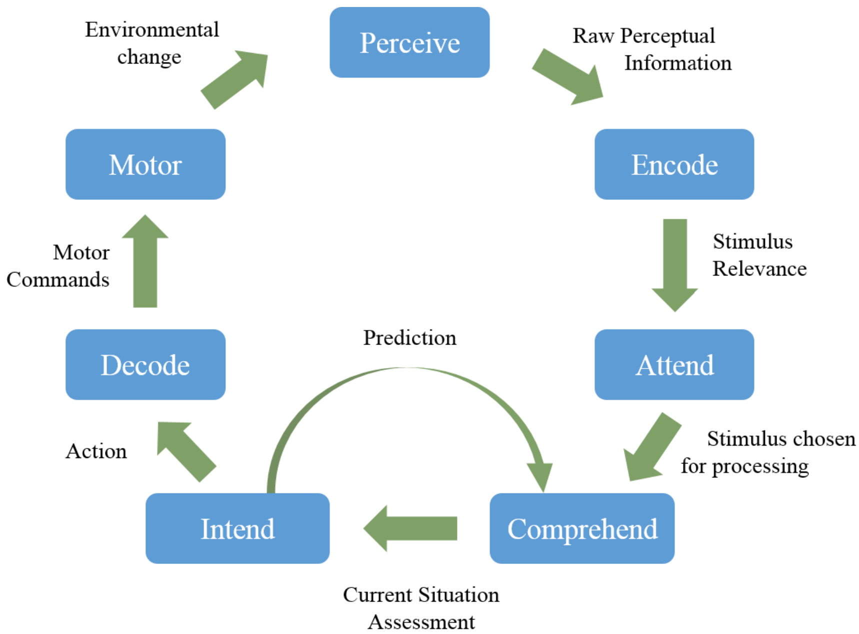


Figure 3.7 A basic PEACTIDM cycle. Arrows indicate the direction of the flow of information from one process to another. Output of *Intend* can go to *Decode* as well as into the next cycle's *Comprehend*. *Tasking*, which is not shown in the figure, competes with *Attend*. Tasking deals with the goals and makes necessary changes, which also serves as an input to the *Encode* and Comprehend cycles. Redrawn after Marinier III et al. (2009, p. 50).

proposed by Newell (1994) can be filled by the aspects of appraisal theories (Marinier III et al., 2009). Moreover, the authors stress that the functional operations missing in the description of most appraisal theories can be provided by the PEACTIDM cognitive theory (Marinier III et al., 2009). PEACTIDM is the short form used to denote the steps in the cognitive theory of Newell (1994) namely *Perceive*, *Encode*, *Attend*, *Comprehend*, *Tasking*, *Intend*, *Decode*, and *Motor* (Newell, 1994).

Perceive component receives the sensory information required for emotion and subsequent processing. This raw perceptual information needs to be converted in the format that can be analysed by a computational system. This is achieved with the help of *Encode* unit. *Attend* process determines if it is useful to attend to the stimulus event or not. This phenomenon can be considered similar to what Scherer (2001) calls *relevance detection* (see Section 2.2.4 for more details on this). *Comprehend* understands the stimulus information and converts it into the required task-specific representation. *Tasking* (not shown in Figure 3.8 for simplicity) converts the task representation to goals to be executed. Tasking helps comprehend to learn about the next task plan and also helps *Intend* to choose an action. *Decode* helps to translate the computational action representation to motor commands. Finally, *Motor* unit executes a

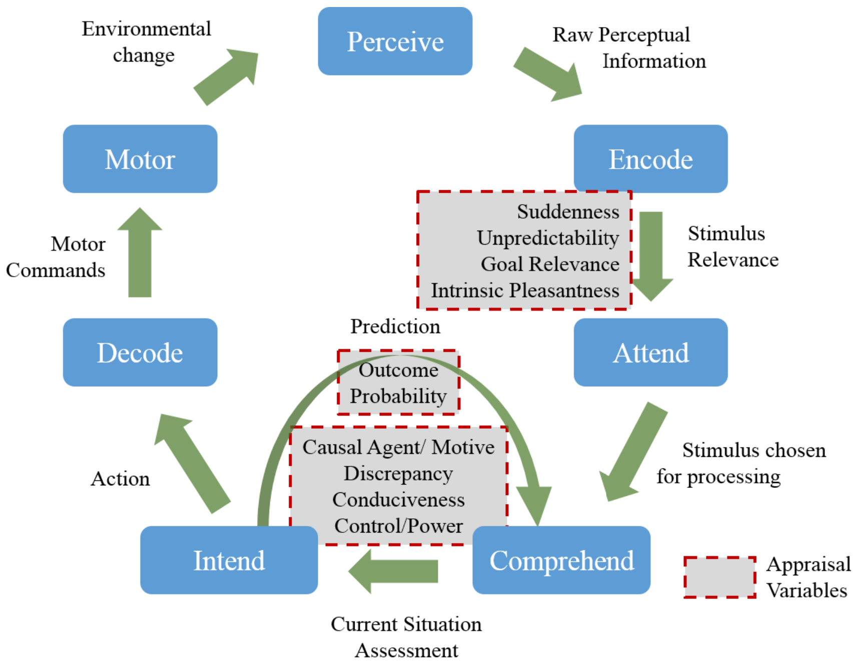


Figure 3.8 Soar-Emote's unification of PEACTIDM (Newell, 1994) and appraisals (Scherer, 2001). Redrawn after Marinier III et al. (2009, p. 54).

physical action which can range from facial, verbal or even the movement of physical parts.

Marinier III et al. (2009) propose an unification of PEACTIDM (Newell, 1994) with the appraisal theory of Scherer (2001). Figure 3.8 shows a unified picture of the PEACTIDM theory of cognition and appraisal theory of emotions. Authors describe that in Soar-Emote, the *relevance appraisal variables* like *suddenness*, *unpredictability*, *goal relevance* and *intrinsic pleasantness* are generated by the Perceive and Encode components and are used by the Attend component (Marinier III et al., 2009). The Comprehend component generates the *assessment appraisal variables* like *causal agent/motive*, *discrepancy*, *conduciveness*, and *control/power*. Intend component generates the *outcome probability* appraisal variable which is used by the Comprehend component in another cycle (Marinier III et al., 2009).

The work proposed by (Marinier III et al., 2009) is based on the theory of Scherer (2001), thus "imposing sequential constraints" (Marinier III et al., 2009, p. 55) in relation to the computation of appraisals. However, some of the authors' choices conflict with Scherer's theory. For example, when organising the appraisal variables in their unified model, the appraisal variables *causal attribution* (used as Causal Agent/Motive here) and *outcome probability* are used in later stage in Soar-Emote (Marinier III et al., 2009), whereas in Scherer's theory they are posed to check the relevance of the stimulus event during early stages of emotion appraisal Scherer (2001).

In addition to the notion of appraisals, Soar-Emote also embraces the aspects of mood and proposes an approach to integrate the mood and emotions to generate feelings of an agent (Marinier III and Laird, 2007). Although, other researchers point that the approach used in Soar-Emote (Marinier III et al., 2009) does not effectively combine the notion of mood and emotion to generate plausible feelings of an agent Becker (2008), in Soar-Emote, mood is computed as an aggregated effect of several emotional experiences following the theoretical assertions of Ekman (1994) and computational implementation of Gratch and Marsella (2004a). Authors explain that a mechanism called *Appraisal Detector* (Smith and Kirby, 2001) maps an active appraisal frame to the current emotion but do not describe if mood takes part in the process. This integrated model of emotion and cognition does not explicitly model the effects of personality factors in the process of emotion generation. Moreover, the approach used in the integration of the theory of emotion and cognition in Soar-Emote may not be completely domain-independent – despite the claims of the authors (Marinier III et al., 2009), because computational details for the operationalised appraisal variables is not provided.

3.1.9 WASABI

WASABI (Becker, 2008; Becker and Wachsmuth, 2009) is an *Affect Simulation Architecture for Believable Interactivity* in which emotions are modelled in a continuous three dimensional space of *Pleasure*, *Arousal* and *Dominance* which are also called PAD space (Russell and Mehrabian, 1977) (see Section 2.2.3 for more discussion of PAD).

Figure 3.9 shows an integrated architecture for cognition and embodiment in WASABI. The agent can *perceive* a stimulus event which is then processed by parallel *conscious* and *non-conscious* appraisals (Becker and Wachsmuth, 2009). Non-conscious appraisal is determined by what Scherer (2001) calls *intrinsic pleasantness* resulting into *emotional impulses*. This creates a 'low road' (LeDoux, 1996) by which the stimulus affects the *emotion dynamics* of the agent, thereby contributing to the determination of *mood*, *Pleasure*(P) and *Arousal*(A). At the same time, a 'high road' influence occurs through the conscious appraisal process which appraises the *goal conduciveness* (Scherer, 2001) of the stimulus event using "BDI-based cognitive reasoning abilities" (Becker and Wachsmuth, 2009) and updating the *memory* and generating *expectations* at the same time. This conscious appraisal helps in determining the level of *Dominance*(D) based on the situation and also generate *secondary emotions*. These computed PAD

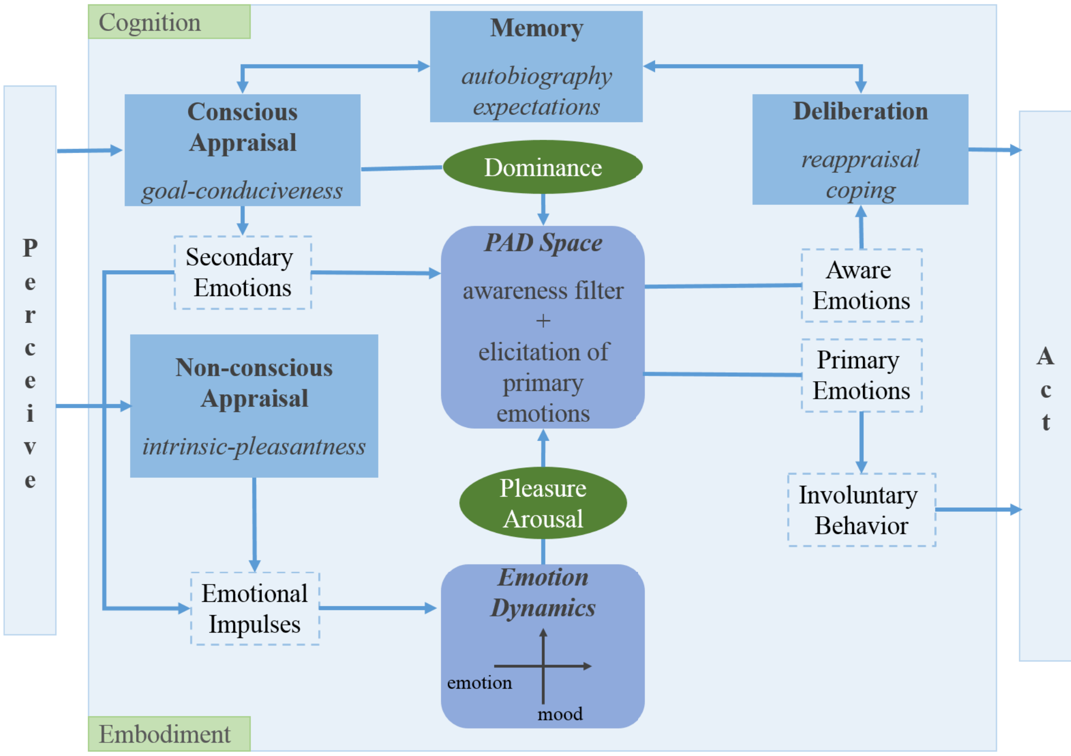


Figure 3.9 The conceptual distinction of cognition and embodiment in WASABI architecture. Redrawn after Becker and Wachsmuth (2009).

values generate *primary emotions* (through direct route) and *aware emotions* (which consists of both the primary and secondary emotions). Primary emotions triggered through direct route can cause *involuntary behavior* of the agent like facial expressions (*i.e. Act*). *Awareness filter* is used to filter the aware emotions which in turn take part in the the *deliberation* process of *reappraisal* and *coping*.

Although WASABI architecture uses the notion of appraisal to evaluate the perceived stimulus, clear computational details explaining how the stimulus information is processed to produce appraisals are missing. Moreover the use of domain-specific designs for the validation of the model such as using a virtual character MAX as a museum guide and a scenario of playing Skip-Bo game can pose serious limits to generalise the proposed appraisal process to other domains. Besides, the use of only one appraisal variable *goal conduciveness* (Scherer, 2001) may not be sound enough when dealing with more complex and general situations, where more appraisal variables are likely to be necessary. WASABI also models the influence of mood on emotions and collective effect of emotional experiences on mood, where mood is considered as a bipolar quantity ranging from -100 to +100 (-100 representing very negative mood and +100 representing very positive mood). However, the involvement of personality factors in the process of emotion generation is not effectively modelled in WASABI. Integration of the aspects of personality may be useful for other practical applications of this model since virtual conversational characters need to exhibit variations in behaviour based on the personality of the self was well as the person interacting with it. The authors also model the processes of reappraisal and coping which are merely used to alter the behaviour of their virtual character MAX such that the actions are in favour of the self. This emotion regulation mechanism, though useful for self defence or self benefit, may not be useful enough to regulate the emotions and behaviour of an agent implementing WASABI in other social contexts where more complex social interactions can happen. In fact, the proposed mechanism does not ensure if the behaviour of MAX is socially acceptable in any given context or not.

3.1.10 TAME

TAME (**T**raits, **A**ttitudes, **M**oods, and **E**motions), is an effective architecture implemented in an embodied robot (Moshkina et al., 2011). As the name suggests, authors claim that their framework models the notion of all of the four aspects namely *traits*, *attitudes*, *moods*, and *emotions*. By using the word 'trait' authors refer to the personality factors of the agent implementing the model. In addition to the notion of emotions and mood, authors introduce another human characteristic called 'attitudes'. According to their definition, attitudes refer to the "enduring, positive or negative, feelings about an object, a person or a issue" (Moshkina et al., 2011, p. 209). This notion is analogous to what Ortony et al. (1990) suggest on OCC theory of emotion. Authors have considered attitudes as more dynamic than personality factors and less dynamic than the mood followed by emotions, where emotions are the most dynamic and constantly changing characteristic.

Figure 3.10 shows the overall architecture of TAME. *Affective Module* represents the core of the TAME architecture (Moshkina et al., 2011), which is responsible for the mechanisms related to mood and emotion generation. Affective module is composed of *Dispositions* and *Affective State*. Dispositions constitute *Personality Traits* and *Affective Attitudes*, which represent a tendency to experience emotion in a certain ways and also to exhibit a typical behavioural pattern (please see Section 2.4 for more discussion on this). Affective State module consists of *Emotion Component* and *Mood Component*, which are relatively dynamic and frequently changing aspects compared to personality and attitude. Affective module generates overall emotional state in response to an event with the influence of personality factors and attitudes. This emotional state is then used to select a behaviour that is sent to *Motor Module* through *Behaviour Coordination*.

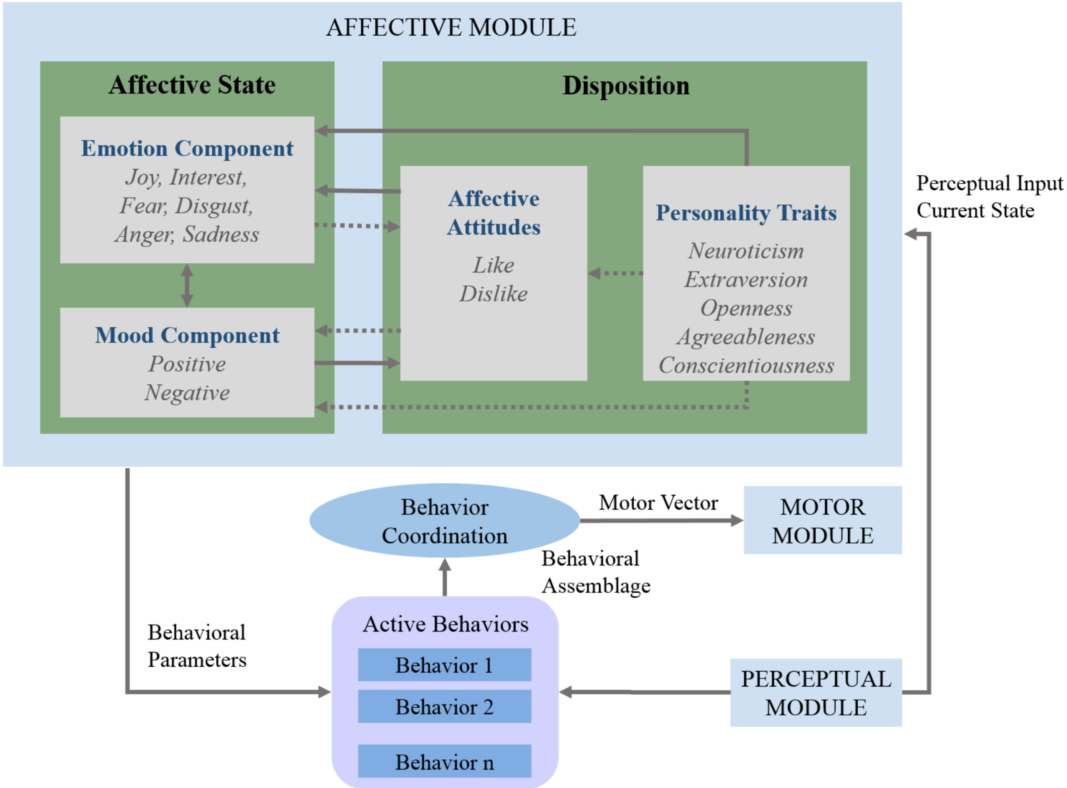


Figure 3.10 Conceptual overview of TAME architecture. Redrawn after Moshkina et al. (2011, p. 210).

Although a separate module called *Perceptual Module* is shown in the architecture, it is not explained what is the module responsible for and how it is linked to rest of the architecture. Previous publications presenting the slightly different version of the architecture also does not explain about the perceptual module (Moshkina, 2006).

Emotions in TAME are dependent on personality traits and mood (Moshkina et al., 2011). Personality traits are used to alter the threshold (activation point), peak (amplitude) value as well as the rise time (slope) of the emotions (Moshkina et al., 2011). However, since the authors use a *linear mapping* to model the influence of personality on emotions thereby affecting the above mentioned values, possibly using some matrix, which is user-defined and not based on data. In regards to modelling mood, authors mention that mood influences the activation point, where positive mood makes it easier to experience positive emotions and negative mood makes it easier to experience negative emotions (Moshkina et al., 2011). Yet, there is no clear explanation on how this relationship is computationally realised in TAME. As a result, it is uncertain whether the model allows the learning of this mapping relationship. In this dissertation, I shall present my model of emotions that is able to learn this relationship with machine learning based on the data provided.

Besides the above limitations, although the authors cite some of the theories of emotions mentioning that they take "inspiration from a large number of theories and findings from personality, emotion, mood and attitude [studies]" (Moshkina et al., 2011, p. 208), they have not made it explicit which of these emotion theories is realised in their implementation. This explanation could be useful for a reader to gain a better understanding of possible implementations of the existing emotion theories. Moreover, there is no evidence of discussion of appraisal variables even if the authors cite the work of Ortony et al. (1990) as their theoretical foundation. Likewise, the authors do not consider discussing how their model handles the situations of conflicting emotions and how overall emotion regulation mechanism augments the process of behavioural selection – despite they talk about the aspects of behaviour.

3.1.11 FAtiMA

FAtiMA (**F**earnot **A**ffec**ti**ve **M**ind **A**rchitecture) (Dias et al., 2014) is an extension of FearNot! affective architecture (Aylett et al., 2005). As per the authors, FAtiMA model allows the integration of various components that may affect the appraisal of a situation by an artificial agent namely *cultural behaviour* (Mascarenhas et al., 2010) and *drives* (Lim et al., 2012). Moreover, authors argue that their architecture can be used to apply the appraisal mechanism for multiple appraisal theories particularly citing the theories of Ortony et al. (1990) and Scherer (2001). The incorporation of various behavioural and appraisal components is handled by a core layer which the authors name as *FAtiMA Core* (Dias et al., 2014), the structure of which has been inspired from the model of Marsella et al. (2010). Figure 3.11 shows the core structure of the FAtiMA agent architecture.

An agent implementing the architecture is able to receive the perceptions of the environment i.e. event through the *perceptions* component, which information is then sent to the *appraisal derivation* component. The information about the triggering event is also stored in the memory for future behavioural management. Appraisal derivation component performs the evaluation of the event and provides the appraisal information to *affect derivation* component which then produces affective states like emotions and mood (Dias et al., 2014). These affective states are also stored in the agent's memory for future reference and are in turn used to affect the *action selection* process that finally help in the execution of an *action* in reaction to the stimulus event (Dias et al., 2014).

An interesting aspect of this emotion architecture is the introduction of *multi-appraisal components* that can contribute to each appraisal variable (based on the appraisal theory considered) (Dias et al., 2014). They also use the term *appraisal*

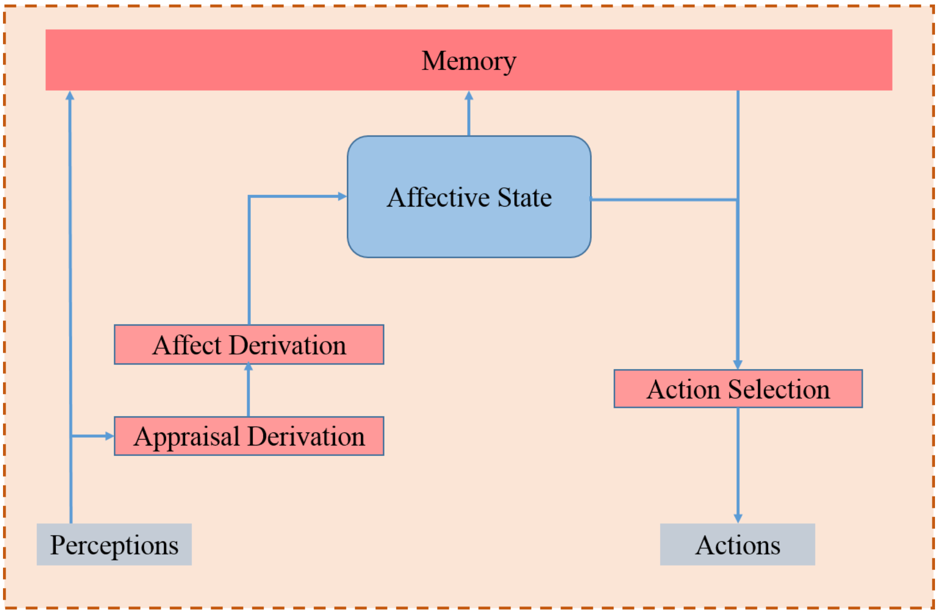


Figure 3.11 FAtiMA Core architecture. Redrawn after Dias et al. (2014, p. 45).

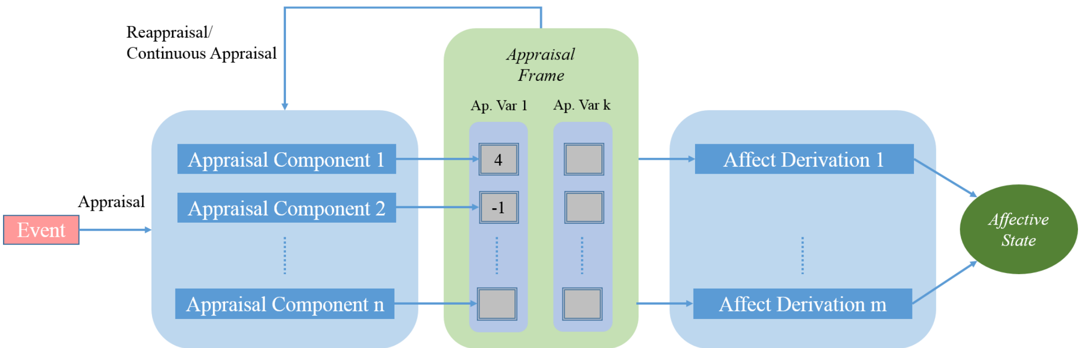


Figure 3.12 Appraisal mechanism in FAtiMA emotion architecture. Redrawn after Dias et al. (2014, p. 48).

frame instead of appraisal variables – probably following the notions of Gratch and Marsella (2004b). Figure 3.12 shows the mechanism of appraisal and mapping to affective state in FAtiMA. When an event occurs, appraisal process is triggered and the event information is provided to the set of appraisal components. There can be multiple appraisal components as per the need of the application (Dias et al., 2014). In this framework "any [appraisal] component can contribute to any appraisal variable" (Dias et al., 2014, p. 47). For example, the *appraisal component 1* in Figure 3.12 can affect the values of the appraisal variable 1 to the variable k (i.e. Ap. Var 1 to Ap. Var k in Figure 3.12). Instead of a single value, each appraisal variable is represented as an array containing as many values as the number of appraisal components so that each appraisal components can have a value for that appraisal variable. The ultimate value for an appraisal variable out of all the values in the appraisal frame is determined by applying a heuristic that best suits the appraisal theory used. Authors mention that the value determined by considering the contribution of the latest component or a priority mechanism allowed an optimal realisation of the appraisal theory of Scherer (2001). This model also realises the idea of *reappraisal* put forward by Lazarus (1966), which is shown by an arrow from appraisal frames to the appraisal components in Figure 3.12. The completion of an appraisal mechanism is marked by the completion of all the appraisal components and any necessary reappraisal cycles which is then followed by the process of affect derivation.

In FAtiMA, each set of appraisal variables in appraisal frame leads to generation of its oen set of affective states (emotions and mood) as the authors state “affect derivation components are independent from each other” (Dias et al., 2014, p. 48) indicating that one affect derivation component is determined by one set of appraisal variables and another affect derivation component is determined by a different set of appraisal variables. In addition, the authors explain that the model generates a final value of an appraisal variable suggesting that an appraisal variable contributes to the intensities of all the emotions independent of other appraisal variables and these emotional states are added to the *affective state* of that emotion, if not already present in the affective state. If an emotion caused by exactly same appraisal variable and same event is being added to the affective state (possibly due to appraisal of another component), then the emotion previously present in the affective state is replaced (Dias et al., 2014). This is a significant problem because the emotion intensity of a particular emotion is completely ignored when a new intensity of the same emotion is supplied by another appraisal process. Although the authors embrace the idea of Scherer (2001) that appraisal process is incremental, this principle is not completely integrated in their model when it comes to estimating the emotion intensities. Besides these limitations, the authors do not explain how the model reaches to the final emotional state when the affective state has multiple, and possibly contrasting, emotions with significant intensities (which the authors term as *mixed emtoions* (Dias et al., 2014, p. 48)).

Apart from the above mentioned limitations, the authors explain that in their model, “predefined emotional reaction rules [are used] to determine the value of ... OCC appraisal variables” (Dias et al., 2014, p. 50). This detail implies that the model can be hardly used in disparate situations, thus significantly limiting its use in applications like carer robots. Moreover, the authors do not provide comprehensive references to emotion theories justifying their design choices. Although this is not a major limitation to assess the performance of the computational model, it is a non-negligible barrier preventing experiments to test, proof or disproof emotion theories and advance the understanding of human cognition. For example, the absence of such references raise some crucial research questions, such as, *Can multiple appraisal theories be applied in a model of emotion processing simultaneously?* If yes, *how are the overlapping appraisal variables of the theories handled and integrated?* Authors also mention about the use of personality factors to modulate the mapping of appraisal variables to emotions without further explanation than what was done for the FearNot! model (Aylett et al., 2005). Similar to the case of FearNot!, it is not clear how the coping and other emotion regulation strategies are applied in FAtiMA – although these aspects are mentioned in their work (Aylett et al., 2005; Dias et al., 2014).

3.1.12 MA/SDEC

MA/SDEC model (Saunier and Jones, 2014) intends to simulate the dynamics of social contagion (Hatfield et al., 1994) in a model of emotions within a *multi-agent system*. In MA/SDEC, emotions are computed as a combination of three major functions: (i) *event dynamics*, which occurs within the agent's mind, (ii) *internal dynamics*, which occurs in the agent's body, and (ii) *external dynamics* which occurs in the environment (Saunier and Jones, 2014). Internal dynamics involves the process of perceiving a stimulus event and performing situation assessment based on *beliefs*, *intentions* and *personality* of the agent. This information processed in the mind is fed into the *internal dynamics* mechanism within the body of the agent which is then responsible for the generation of emotions. The internal dynamics of emotion is heavily modulated by the personality of the agent which is further optimised by a *control* module in the body of the agent. The emotions processed so far interact with the environment (other agents) through the *external dynamics* mechanism. All of these three processes in totality determine the final emotional state of the agent (Saunier and Jones, 2014). A clear issue with the model is the use of various formulas for the computation of emotional states. Each of the event dynamics, internal dynamics, and external dynamics at time $t+1$ depends on the emotion at time t. However, authors do not make it clear how the emotion at time t is computed in the first place or if the initial state is set to a default value. Moreover, the external dynamics at time $t+1$, which is responsible to capture the notion of emotional contagion, is a function of personality and emotional state at time t (Saunier and Jones, 2014). How the notion of physical/psychological closeness is operationalised in the mechanism of contagion is unclear in the paper. Likewise, although the concepts of personality is adapted from the Five Factor model (Digman, 1990), MA/SDEC does not integrate the notion of mood in the emotion generation process. Additionally, since the main goal of the model is to simulate the process of emotion contagion, there are no

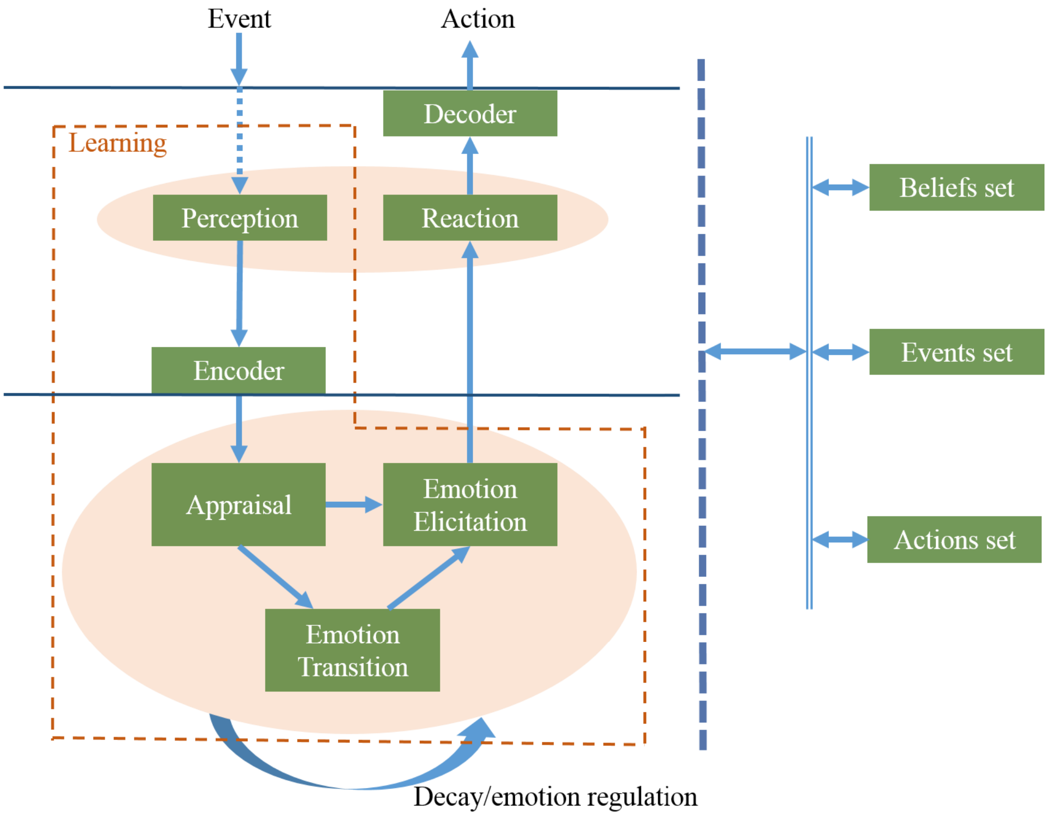


Figure 3.13 EMIA architecture divided into three layers. Redrawn after Jain and Asawa (2015, p. 454).

references to an emotion regulatory mechanism which can prevent the agent to express an unacceptable emotional response.

3.1.13 EMIA

EMIA (**E**motion **M**odel for **I**ntelligent **A**gent) is one of the recent *domain-independent* computational models of emotions proposed by Jain and Asawa (2015, 2016). It integrates the concepts of *emotional appraisal* with *emotion regulation* and *emotion transition*. This model is based on the appraisal theories of Ortony et al. (1990), Scherer (2001), and Roseman et al. (1990) and emotion regulation theory of Gross and Thompson (2007). The model implements five out of six basic emotions proposed by Ekman (1992) using five appraisal variables (Jain and Asawa, 2015).

Figure 3.13 shows a *three-layered* organisation of the EMIA model of emotions (Jain and Asawa, 2015). The first (topmost) layer deals with the environment and is responsible to receive the stimulus event as act upon the surrounding. In the second (middle) layer, the perceived stimulus is encoded and sent to the *emotional appraisal* module in the third layer, which is also responsible for *emotion transition* and *emotion regulation*. Based on the work of Atkinson and Shiffrin (1968), the authors embrace three types of memory namely (i) *perceptual memory*, (ii) *working(short-term) memory*, and (iii) *long-term memory*. In the long-term memory, *belief set*, *events set* and *actions set* are stored (see Figure 3.13). As the names suggest, beliefs set stores the beliefs of the agent, events store the experienced events and actions set stores a collection of possible actions to act on the environment. Appraisal of an event is performed based on beliefs and goals of the agent. Then, fuzzy methods are applied to convert the appraisal into emotion intensities (Jain and Asawa, 2015). EMIA does not let the model to use the raw emotions for expressive or behavioural actions. Rather, an emotion regulation mechanism based on the theory of Gross and Thompson (2007) is applied to achieve *antecedent-focused emotion regulation* – reappraisal and *response-focused emotion regulation* – suppression Jain and Asawa (2015). However, the authors do not examine the effect of the generated emotions on decision making and action selection. As admitted by the authors themselves in their subsequent publication (Jain and Asawa, 2016), their model does not integrate the notions of personality and/or mood by stating that their model "can further be broadened by investigating the effects of personality [and mood] on emotion elicitation, emotion regulation and plan selection processes" (Jain and Asawa, 2016, p. 141). In regards to the use of fuzzy logic in emotion elicitation process, authors support the idea that fuzzy logic is a highly desirable feature to obtain continuous values of the emotion intensities. However, I shall show in this dissertation that continuous values of emotion intensities can be obtained without using fuzzy logic.

3.1.14 CAAF

CAAF (**C**ognitive **A**ffective **A**gent Programming **F**ramework) (Kaptein et al., 2016) is a computational unification of *belief-desire theory of emotion* (Reisenzein, 2001) and *cognitive agent-programming framework* (Hindriks, 2009). The goal of their contribution is to enable the development of a cognitive agent with affective capabilities (Kaptein et al., 2016). The mechanism of computation in the model keeps *beliefs–desires–intentions* (BDI) approach to the centre of the framework. The authors seem to be motivated to use the *belief-desire theory of emotion* (Reisenzein, 2001) instead of appraisal theories (Ortony et al., 1990; Roseman et al., 1990; Scherer, 2001). According to the authors, this choice was made because the *belief-desire theory of emotion* is conceptually close to the concepts of BDI and appraisal theories are difficult to be realised computationally since they introduce a large number of appraisal variables without an indication on how those variables could be feasibly realised mathematically/computationally (Kaptein et al., 2016). In fact, their computational realisation ends up with a serious limitation of having "only two appraisals" (Kaptein et al., 2016, p. 4). Moreover, their realisation allows only the use of maximum seven emotions while appraisal theories suggest the presence of larger number of possible emotions – as also supported by other dimensional theorists Plutchik (1980). Although the model has a positive aspect of being somewhat domain-independent, it does not account for the notions of personality and mood. Likewise, the authors do not explicitly discuss about the emotional regulation mechanism in their computational model.

3.1.15 Other Models of Emotion

There have been a lot of other proposals for modelling emotion in artificial agents in addition to the ones discussed above. KARO (Meyer, 2006; van der Hoek et al., 1999) is "an alloy of modal logic, dynamic logic, and accessory operators of motivation" (Kowalczuk and Czubenko, 2016, p. 7). Meyer (2006) has tried to integrate the aspects of emotional experience in their original rational agent – KARO (van der Hoek et al., 1999). However, the depth of the notion of emotion itself seems shadowed in the understanding of the model as it has other core goals rather than expressing emotionality. By realising the BDI-like reasoning they "only looked at *some* particular aspects of emotions" (Meyer, 2006, p. 17). The author also admits the limitations of the model saying that it is not adequate enough to be a complete logical theory of emotions. The notions of personality and mood are not integrated into the model – probably because it is difficult to achieve such an integration with the current semantics of the model. Although the extensions of the model logically formalise the OCC theory of emotions (Ortony et al., 1990) plausibly, the proposed solutions (Steunebrink et al., 2007, 2008) still fall short in various aspects. For example, even if the later versions attempt to model the quantitative aspects like threshold and intensity of emotions, they fail to describe how the appraisals actually occur in the model – which are only presented as high level functions (Gluz and Jaques, 2017). Other researchers also have independently taken the logical approach to modelling emotions (Adam et al., 2009). However, as admitted by the authors they also inherit a problem common to the work of Meyer (2006) *i.e.* the lack of integration of aspects like mood and personality. Besides, since the authors are not clear how they can achieve the computation of various appraisals, the mechanism is highly likely to be heavily domain-dependent. However, AfPL (**Af**fective **P**robabilistic **L**ogic) (Gluz and Jaques, 2017) recently offered several solutions to the problems in previous works of logical formalisation of OCC theory of emotions. AfPL is a "formalization of the appraisal process of emotions in a probabilistic BDI logic"

(Gluz and Jaques, 2017, p. 661). While previous proposals were unable to provide a clear computational mechanism to obtain the values of the appraisal variables, AfPL offers logic based calculation of certain appraisal variables as described in OCC theory (Ortony et al., 1990). Yet, this model still inherits some of the limitations of previous ones *i.e.* it neither models the aspects of mood and personality in modulating the emotional responses nor explicitly defines an emotion regulation mechanism. Recently, Alfonso et al. (2017) proposed an integration of emotion, mood and personality in GenIA[3] (**Gen**eral-purpose **I**ntelligent **A**ffective **A**gent **A**rchitecture) by extending the BDI agent language *AgentSpeak* (Bordini and Hübner, 2010) to include affective component. However, the model still falls short is addressing the role played by mood and personality in mapping appraisals to emotion intensities as well the mechanism to regulate conflicting emotional states.

Schneider and Adamy (2014) proposed a fuzzy logic based computational model of affect and emotion generation inspired by the work of El-Nasr et al. (2000). Their model is simulated in a "memoryless, myopic, autonomous, embodied, reactive [virtual] agent living in an unstructured environment" (Schneider and Adamy, 2014, p. 33). The agent constitutes *inner states* to store the agent's *energy* and *integrity* levels, a *perceptual system* to receive the surrounding stimulus. A *motivation system* tends to maintain the *homeostatis* of the internal variables i.e. energy and integrity for the achievement of a pre-defined *high-level task*. One issue with the model is the use of fuzzy approach to generate emotions which, as argued earlier in case of El-Nasr et al. (2000) (see Section 3.1.2), causes the loss of a large amount of cognitive information. Moreover, as admitted by the authors at the end of paper, since the model is simulated in an extremely simplified environment it is "not capable of producing higher order affective states" (Schneider and Adamy, 2014, p. 38), and hence not suitable for applications involving complex human-robot interactions. The latter argument is further supported by the fact that the model does not incorporate the concepts of mood and personality and hence will not allow an agent to adapt dynamically to the person interacting with it and become socially acceptable.

Rasool et al. (2015) proposed a model for the expression of *empathic* emotions based on the recognition of the emotional expressions of interacting users as well as the mood and personality of the artificial agent. The model operates in three sequential stages/components: (i) *Perception* – which is responsible to recognise and represent the emotional state of the interacting user into the system structure (using a third party framework called Grimace[3]), (ii) *Empathic appraisal* – which is responsible

[3]http://www.grimace-project.net/

to generate the emotional responses which are congruent to the emotional state of the interacting user based on the agent's mood and personality, and (iii) *Empathic expression* – which deals with the expression of the felt emotions through a virtual interface (Rasool et al., 2015). From the discussion of the model and computational implementations of various aspects, the model is discussed as presenting a domain-independent nature for generating emotions. However, the authors neither explain how the appraisal actually occurs in the model nor provide a link to the concepts of appraisals realised in the model. The model implements an interaction among emotion, mood and personality where personality affects mood and emotion, mood affects emotions and seemingly emotions also affect the mood. But, the authors do not discuss how the factors of personality and mood help in converting appraisals into emotion intensities – in fact, this is simply obtained by using a fuzzy mapping of the coordinates from a 2D *pleasure-arousal plane* to numeric emotion intensities. Besides, while the model deals with the emotion regulation mechanism based on the concept of *empathy*, it does not consider the social appropriateness of expressing a particular emotional response – say, for example, whether it is appropriate to express empathy to someone that just started crying because of fear of prosecution after having committed a crime or an unethical act.

Sun et al. (2016) present a mechanistic viewpoint where emotions emerge as a multifaceted phenomena within a cognitive architecture CLARION (Sun, 2006). In the proposed framework, emotions are generated by a multimodel interaction of the four subsystems of CLARION: (i) *action-centered subsystem* (ACS), (ii) *non-action-centered subsystem*(NACS), (iii) *motivational subsystem*(MS), and (iv) *metacognitive subsystem* (MCS). The overall process of emotional experience is divided into three stages namely *reactive affect* referring to lower order non-conscious emotional experience, *deliberative appraisal* referring to more conscious appraisal of stimulus event and *coping* referring to metacognitive process of emotion regulation (Sun et al., 2016). Authors mention that the first-order '*automatic appraisals*'(Sun et al., 2016, p. 8) is handled by ACS, MS and MCS subsystems and *deliberative appraisal* is performed by NACS subsystem. The implemented model is inspired by the appraisal theory of Scherer (2001) when it comes to realising the notion of cognitive appraisal in their model. The link between the computed appraisals and coping strategies is largely determined by the *goals* of the agent. Although, Sun et al. (2016) use *normative significance* as one of the appraisal variables, it is not clear how this variable takes part in the process of coping thereby leaving an unanswered question of whether the final behavioural responses of the agent implementing this model confirm with social norms or not. Besides, authors do not integrate the notions of mood and personality in their model. Since the mechanism of appraisal computation is not clearly explained in the paper (Sun et al., 2016), it is unfortunately not clear if their model can perform or not a domain-independent cognitive appraisal.

Besides the computational models, architectures, frameworks, characters or hybrid cognitive-affective implementations of artificial emotions discussed above, there are numerous others that were proposed. For example, AR (Elliot, 1992), EM (Reilly and Bates, 1992), ACRES/WILL (Moffat et al., 1993), CyberCafe (Rousseau et al., 1996), Émile (Gratch, 2000), H-CogAff (Sloman, 2001), ALEC (Gadanho and Custódio, 2002), ParleE (Bui et al., 2002), TABASCO (Petta, 2003), CBI (Marsella, 2003), EM-ONE (Singh et al., 2005), Thespian (Si et al., 2006), and ActAffAct (Rank, 2009) are some of the models which were not discussed in detail in this dissertation. It is mainly because of two reasons: (i) *these models share a lot of common properties with the models discussed so far*, and (ii) *the exploration of the emotion models critically analysed so far already include limitations also presented by those models* (which is also summarised in Table 3.2).

3.1.16 Summary and Comparison of Computational Models of Emotion

In Table 3.1, I present a convenient summary of various computational models of emotion. The table was originally adapted from the work of Lin et al. (2012) and has been amended based on my recent knowledge to include some of the latest computational realisations of emotion in various artificial forms. The summary explores the *base cognitive theory* used by the model, *emotion theory* on which the emotion elicitation and generation mechanism is based on and some of the notable cognitive/behavioural *effects modelled* in the proposed emotion model.

Model	Base Cognitive Theory	Emotion Theory	Effects Modelled
EM (Reilly and Bates, 1992)	Oz architecture	OCC	Plan change
ACRES/WILL (Moffat et al., 1993; Swagerman, 1987)	BDI, Planning, Decision Theory, Agents	Frijda	Coping: goal shift, attention shift
Cathexis (Velásquez, 1997, 1999)	BDI	Izard, Tomkins, Roseman	Behaviour modulation, habituation and sensitisation, emotional conditioning
Émile (Gratch, 2000)	Strips Planning	OCC, Sloman	Plan change, plan selection criteria

FLAME (El-Nasr et al., 2000)	Fuzzy logic, Planning, Decision Theory, Q-Learning	OCC, Roseman	Choice and inhibititon of plans, emotion-based learning and conditioning
H-CogAff (Sloman, 2001)	BDI, Cognition and Affect	Sloman, OCC	Attention shift (alarms), decision biases, precognitive reaction
ALEC (Gadanho and Custódio, 2002)	CLARION	Sloman, Damasio	Decision rules learned based on past experience
TABASCO (Petta, 2003)	ACT, BDI	Scherer, Lazarus, Smith	Plan updates
EM-ONE (Singh et al., 2005)	Minsky-Sloman	Minsky, Sloman	Modification of “narratives”: plans, desires, or beliefs. Modifications of “critic” processes
KARO (Meyer, 2006)	BDI	OCC, Logic of Emotional Agents (LEA)	Plan/agenda changes; Fear causes cautious planning
MAMID (Hudlicka, 2005, 2007, 2008)	Belief Net, Decision Theory	Scherer, Smith & Kirby, Sloman, OCC	Biases mental constructs (data) based on emotional state; Working memory capacity, speed; attention
ActAffAct (Rank, 2009)	Agents, BDI, Unified Cognition	Frijda, Scherer	Coping: choice of Relational Action Tendency
EMA (Marsella and Gratch, 2009)	BDI, Agents, Decision Theory, Planning	Lazarus, Smith, OCC	Coping: attention shift, plan changes, BDI changes, action tendency changes
Soar-Emote (Marinier III et al., 2009)	Soar, PEACTIDM	Scherer, Roseman	Attention shift, goal shift, reinforcement learning biases (both encoding and recall)
WASABI (Becker and Wachsmuth, 2009)	BDI	PAD, Scherer	Plan utility valuation process biased towards optimism or pessimism, mapping of emotions as beliefs, action biases
TAME (Moshkina et al., 2011)	BDI	?	Emotion, mood and behavioural dynamics
FAtiMA (Dias et al., 2014)	BDI	OCC, Scherer, Lazarus	Coping: plan and goal changes
Schneider and Adamy (2014)	Fuzzy logic	?	Motivation based goal achievement
MA/SDEC (Saunier and Jones, 2014)	BDI, Social contagion	Hatfield, Cacioppo, Rapson	Effect of social contagion in multi-agent interaction scenario
Rasool et al. (2015)	Fuzzy logic	Russel, Mehrabian	Expression of empathic responses
EMIA (Jain and Asawa, 2015, 2016)	BDI, Fuzzy logic, Gross’s Emotion Regulation	OCC, Scherer, Roseman	Emotion transition, emotion regulation
Sun et al. (2016)	CLARION	Scherer	Coping, action/behaviour selection

CAAF (Kaptein et al., 2016)	Belief-Desire Theory of Emotion (BDTE), Cognitive Agent programming Frameworks(CAF)	Reisenzein	Goal directed behaviour
AfPL (Gluz and Jaques, 2017)	BDI, Probabilistic modal logic	OCC	Goal directed behaviour
GenIA[3] (Alfonso et al., 2017)	BDI, Logic	OCC	Goal directed behaviour

Table 3.1 List of computational models of emotion. Adapted and amended from Lin et al. (2012, pp. 63-64).

In Table 3.2, I show a comparative assessment between the computational models of emotion reviewed in Section 3.1.1 to 3.1.15. The comparison is based on five main questions: (i) Is the model *domain-independent*, (ii) Does the model integrate the aspects of *mood*?, (iii) Does the model integrate the aspects of *personality*?, (iv) Do the mood and personality play a *role in mapping* the appraisals into emotion intensities?, and (v) Does the model implement *ethical reasoning* mechanism to regulate the emotional state of the agent? as mentioned in the beginning of Section 3.1. A *tick mark* (✓) indicates that the model satisfies the given property; a *cross mark* (✗) indicates that the model does not satisfy the given property; and a *question mark* (?) suggests that I was not able to retrieve enough information to identify whether particular property is satisfied or not. As evident from the summary table as well the detailed discussion of all the models in previous sections, it can be inferred that existing computational models of emotion largely fail to account for most of the properties listed in Table 3.2. Importantly, most of them do not model emotions in a domain-independent manner with only a few existing accounts being domain-independent (see Table 3.2). Also, while some of the models have tried to integrate the notion of personality and/or mood, the majority of those approaches make use of static pre-defined parameters that are used to fine tune the role of mood and personality on emotion intensities. Of the models which integrate the notion of mood and/or personality, almost all of them do not describe if the mapping of their appraisal variables into emotion intensities is data-driven – or determined in ad-hoc manner (but, see also Egges et al., 2004). Above all, although some of the models have offered *coping mechanism* to regulate the emotional states generated based on various appraisals (Marsella and Gratch, 2009; Sun et al., 2016), the proposed models in the history of computational emotion modelling *do not integrate the aspect of ethical reasoning* as a deliberate mechanism of regulating emotional state by handling the conflicting emotions (although this mechanism could have been implemented with certain adjustments in the models). All these facts suggest the presence of important research gaps not yet addressed which, with this work, I will advance. In the following chapter, I shall present the details of my proposed computational model of emotion, EEGS, which effectively offers viable solutions to the gaps identified above.

Model	Domain-Independent	Models Mood	Models Personality	Data-driven Mapping	Ethical Reasoning
Cathexis (Velásquez, 1997, 1999)	✗	✓	✓	✗	✗
FLAME (El-Nasr et al., 2000)	✗	✓	✗	✗	✗
Model of Egges et al. (2004)	✓[4]	✓	✓	✓	✗
FearNot! (Aylett et al., 2005)	✗	✓	✓	✗	✗
ALMA (Gebhard, 2005)	✗	✓	✓	✗	✗
KARO (Meyer, 2006)	✗	✗	✗	✗	✗
MAMID (Hudlicka, 2005, 2008)	✗	✗	✓	✗	✗
EMA (Gratch and Marsella, 2004a; Marsella and Gratch, 2009)	✓	✓	✗	✗	✗
Soar-Emote (Marinier III and Laird, 2004; Marinier III et al., 2009)	✗	✓	✗	✗	✗
WASABI (Becker, 2008; Becker and Wachsmuth, 2009)	✗	✓	✗	✗	✗
TAME (Moshkina et al., 2011)	?	✓	✓	✗	✗
FAtiMA (Dias et al., 2014)	✗	✓	✓	✗	✗
Schneider and Adamy (2014)	?	✗	✗	✗	✗
MA/SDEC (Saunier and Jones, 2014)	✓	✗	✓	✗	✗
Rasool et al. (2015)	✓	✓	✓	✗	✗

[4] Authors claim the model to be domain-independent, but enough details to replicate it in different scenarios are not provided and thus the validity of this property cannot be fully verified.

EMIA (Jain and Asawa, 2015, 2016)	✓	✗	✗	✗	✗
Sun et al. (2016)	?	✗	✗	✗	✗
CAAF (Kaptein et al., 2016)	✓	✗	✗	✗	✗
AfPL (Gluz and Jaques, 2017)	✓	✗	✗	✗	✗
GenIA^3 (Alfonso et al., 2017)	✓	✓	✓	✗	✗
Proposed EEGS Model	✓	✓	✓	✓	✓

Table 3.2 Comparison of several computational implementations of emotion in artificial agents over the last two decades based on (i) whether the appraisal mechanism in the model can be achieved in *domain-independent* manner or not, (ii) whether the model integrates the notion of *mood* with emotions or not, (iii) whether the model integrates the notion of *personality*, (iv) whether the model has a *data-driven mapping* of the appraisal variables into emotion intensities based on the learned relationship between appraisal variables and emotions, and (v) whether the model implements a emotion regulation mechanism based on *ethical reasoning* or not. Last row shows the strengths of my proposed EEGS model.

In the previous sections, I discussed the advancements and the limitations in the computational emotion modelling research in the last two decades by exploring the most influential models of emotions proposed in the past. I examined the existing models based on five assessment criteria. The first criteria was whether the model is domain-independent or not *i.e.* being able to appraise in different interaction domains without changing the computation mechanism. Second criteria was whether the model integrates the aspects of mood or not. Likewise, the third criteria was whether the model integrates the aspects of personality or not. Fourth criteria was whether the model allows a data-driven mapping from appraisals to emotion intensities or not. And, the fifth criteria was whether the model employs ethical reasoning mechanism to converge to a stable emotional state and regulate the activated emotions or not. In the following section, I will present the hypotheses this research aims to test in line with the identified limitations in the existing models and set the research methodology to address the identified research gaps and validate the proposed hypotheses. The hypotheses will constitute the testable statements used to validate the thesis statement presented in Chapter 1. Each hypothesis will be validated by the research methodology proposed in this section.

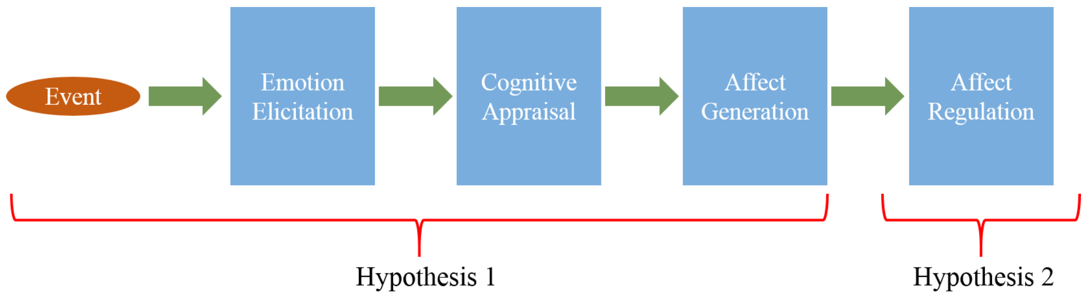


Figure 3.14 Graphical representation of how the validation of Hypothesis 1 and 2 will involve the evaluation of first part (*i.e.* emotion elicitation, cognitive appraisal, affect generation), and second part (*i.e.* affect regulation) of the overall computational process in EEGS.

3.2 Hypotheses[5]

Before stating the hypotheses of this dissertation, I would like to present a functional segmentation of the computational model of emotions for artificial agents being developed in this research by revisiting the appraisal dynamics discussed in Chapter 2, Section 2.3. In Figure 3.14, *emotion elicitation* represents the first-order lower level appraisal process (Lambie and Marcel, 2002); *cognitive appraisal* represents the second-order higher level cognitive evaluation of the emotion eliciting event; *affect generation* represents the process of mapping the appraisals into emotion intensities; and *affect regulation* represents the process of converging to a stable emotional state if there are conflicting emotions generated by a given situation. As shown in figure, validation of Hypothesis 1 will involve the evaluation of the first part of the overall computational process and validation of Hypothesis 2 will involve the evaluation of the second (final) part of the overall computational process. Since an objective of this research is to demonstrate that ethical reasoning mechanism can augment the process of emotion regulation in an artificial agent, it is also necessary to have a mechanism of generating the candidate emotions for regulation. This is why, the process of cognitive appraisal and affect generation (evaluated for validation of Hypothesis 1) precedes the affect regulation process (evaluated for the validation of Hypothesis 2).

Consider a social interaction between two subjects. Let us use *sender* to denote the subject producing a behavioural response directed to the other subject, which I shall call as *receiver*. Denote with $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ a value determining how negative or positive the behaviour $\mathbf{B}$ of the sender is perceived by the receiver in a given context $\mathbf{C}$. $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ is a plausible computational representation summarising within a single valenced value

[5]This section contains extracts from Ojha et al. (2017) and Ojha et al. (2018).

the somato-visceral reactions of the body to the given situation $(\mathbf{B},\mathbf{C})$ following the first-order phenomenological stage of emotional processing[6] (Bechara et al., 2000).

Denote with $\mathcal{C}(\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver})$ a cognitive appraisal process able to appraise the intensities $\mathcal{I} = \{\hat{i}_{e_1}, \ldots, \hat{i}_{e_l}\}$ of a set of l considered emotional states $\{e_1, \ldots, e_l\}$ given the first-order phenomenological reaction of the receiver $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$. Thus, my first hypothesis is that:

> **Hypothesis 1** *Assumed a given somato-visceral reaction $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ of the receiver in a given context $\mathbf{C}$ for a received behaviour $\mathbf{B}$, $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ allows a computational cognitive appraisal process $\mathcal{C}$ to compute appraisal of the situation. In turn, the resulting appraisals can be used to compute the emotional intensities (resembling human-like emotions) for the given context.*

The validation of Hypothesis 1 will constitute the evaluation of first part of the proposed computational model, which deals will the appraisal computation and generation of emotion intensities. However, according to appraisal theories, an appraisal can lead to the generation of more than one emotions (Ortony et al., 1990; Scherer, 2001). Therefore, a computational model of emotion should also be equipped with an ability to converge to a final stable emotional state that can promote in further decision making or action selection by the agent. Denote with $\mathcal{E}(\mathcal{I}, \theta^{ethics})$ and with $\hat{\mathcal{E}}(\mathcal{I})$ two processes able to select a final emotional state given the set of the elicited emotion intensities $\mathcal{I}$ realised by the cognitive appraisal process $\mathcal{C}$. $\mathcal{E}(\mathcal{I}, \theta^{ethics})$ includes parameters, θ^{ethics}, operationalising ethical theories, whereas $\hat{\mathcal{E}}(\mathcal{I})$ uses strategies without considering the ethical impact of the selected emotion to express during the current situation. As such, my second hypothesis is that:

> **Hypothesis 2** *A computational cognitive appraisal process converged by ethical reasoning mechanism $\mathcal{E}(\mathcal{I}, \theta^{ethics})$ more accurately resembles human-like emotion mechanism compared to generic convergence mechanisms $\hat{\mathcal{E}}(\mathcal{I})$. This, in turn, supports the generation of socially appropriate emotional responses in autonomous agents.*

By validating Hypothesis 1, I will have a viable tool to generate emotions accurately predicting human data. By validating Hypothesis 2, I will demonstrate that my ethical reasoning mechanism can better predict what people consider appropriate emotional responses in a specific social context. Therefore, by validating both my hypotheses, I will validate my thesis argument, which states.

---

[6] It should be noted that this dissertation will not address the detailed discussion or implementation of first-order phenomenological processes. Any reader interested in further understanding of the concept is directed to Lambie and Marcel (2002).

> *The regulatory mechanism for emotional processing of an artificial agent can be enriched by an ethical reasoning mechanism enabling the selection of a more socially acceptable emotional state to express while interacting with people in a given social context.*

Hypotheses Validation Methodology

In order to validate Hypothesis 1, I will develop a computational model of emotion that computes appraisals of a situation based on theoretical premises and empirical data. The evaluation will occur in two steps. First, *cognitive appraisal* component will be evaluated which will lead to the validation of the first part of Hypothesis 1. Second, *affect generation* component will be evaluated which will lead to the validation of second part of Hypothesis 1.

The value of $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ in Hypothesis 1 will be feed into the model as a set of knowledge about the interaction situation. The interaction situation will be simulated by carrying out actions from *sender* (some user interacting with the agent in the model) and the *receiver* (agent in the model). The model will compute the relevant appraisals using the domain-independent appraisal rules that can perform appraisals irrespective of the domain knowledge (see Chapter 4 for more details of the implemented domain-independent appraisal rules). In order to test the accuracy of the appraisal computation, these appraisal variables will be compared with the data collected from humans, where they will provide appraisal ratings on the same scenario. Moreover, I will demonstrate the domain-independence of the proposed model by using same appraisal computation rules in two entirely different interaction scenarios.

Next, in order to validate the second part of Hypothesis 1, I will operationalise a mechanism of mapping the computed appraisals into emotion intensities. I will achieve this by taking inspiration from appraisal theories as well as mood and personality research. Instead of traditional approaches where researchers fine-tune the personality and mood parameters to estimate emotion intensities based on commonsense knowledge, I will use a machine learning approach to establish the relationship between appraisal variables and emotion intensities based on mood and personality factors. As such, I will demonstrate that appraisal-emotion association established in such a manner can lead to a high level of mapping accuracy and also allows the modulation of the emotional responses of an agent based on the difference of these characteristics.

In order to validate Hypothesis 2, I will apply three different convergence mechanisms to regulate the emotions of the model – namely highest intensity approach, blended intensity approach and ethical reasoning approach, and compare the converged responses to the data collected from humans (see Chapter 5 for more details). Second part of Hypothesis 2 states that emotion regulation mechanism augmented by ethical reasoning mechanism allows autonomous agents to exhibit socially appropriate emotional responses. In order to validate this statement, I will simulate scenarios of interaction between two individuals that involve ethical concerns and demonstrate how emotions regulated by ethical reasoning mechanisms help in achieving social appropriateness of emotional responses.

3.3 Chapter Summary

This chapter presented a discussion of the computational models of emotion proposed in the past. Computational models, frameworks or architectures for artificial emotion generation were critically analysed to examine if they satisfy the *five crucial properties* that a computational model should have – (i) *domain-independence*, (ii) *integration of mood*, (iii) *integration of personality*, (iv) *data-driven mapping of appraisals into emotion intensities*, and (v) *ethical reasoning* for emotion regulation. It was concluded that none of the computational models of emotion proposed so far satisfy all of the aforementioned properties (see Table 3.2 for a summary). While each of the existing models embraces one or more of these aspects, with many of those trying to integrate the notion of mood and personality (Aylett et al., 2005; El-Nasr et al., 2000; Gebhard, 2005; Velásquez and Maes, 1997), most of them fail to propose a domain-independent approach of obtaining a cognitive appraisal of the stimulus event (except for a few examples Gratch and Marsella, 2004a; Jain and Asawa, 2015; Saunier and Jones, 2014). When it comes to the mapping of appraisals to emotion intensities, none of the models proposed provide a data-driven establishment of this relationship modulated by various factors, except the model proposed by Egges et al. (2004). However, it is still unclear how the various *relationship matrices* described in the model of Egges et al. (2004) are obtained. Interestingly, even if some of the models propose an emotion regulation mechanism called *coping* (Marsella and Gratch, 2009; Rank, 2009), none of the contemporary models implement an explicit ethical reasoning mechanism for emotion regulation. Additionally, I established testable hypotheses as well as the methodology that I will use to validate the stated hypotheses in the rest of the dissertation.

In the following chapter, I will present my computational model of emotion, EEGS, that computes appraisals in a domain-independent manner with the integration of the notion of personality and mood that modulate the relationship of appraisals with emotion intensity in a data-driven manner which is also augmented by an ethical reasoning mechanism for the regulation of the resulting emotional states.

Chapter Bibliography

Adam, C., Herzig, A., and Longin, D. (2009). A logical formalization of the occ theory of emotions. *Synthese*, 168(2):201–248.

Alfonso, B., Vivancos, E., and Botti, V. (2017). Toward formal modeling of affective agents in a bdi architecture. *ACM Transactions on Internet Technology (TOIT)*, 17(1):5.

Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, 2:89–195.

Aylett, R. (2000). Emergent narrative, social immersion and "storification". In *Proceedings of the 1st International Workshop on Narrative and Interactive Learning Environments*, pages 35–44.

Aylett, R., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). Fearnot!– An experiment in emergent narrative. In *Intelligent virtual agents*, pages 305–316. Springer.

Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, 10(3):295–307.

Becker, C. (2008). *WASABI: Affect simulation for agents with believable interactivity*. PhD thesis, Bielefeld University.

Becker, C. and Wachsmuth, I. (2009). Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, 20(1).

Bolles, R. C. and Fanselow, M. S. (1980). A perceptual-defensive-recuperative model of fear and pain. *The Behavioral and Brain Sciences*, 3:291–301.

Bordini, R. H. and Hübner, J. F. (2010). Semantics for the jason variant of agentspeak (plan failure and some internal actions). In *ECAI*, pages 635–640.

Bui, T., Heylen, D., Poel, M., and Nijholt, A. (2002). ParleE: an adaptive plan based event appraisal model of emotions. *KI 2002: Advances in Artificial Intelligence*, pages 1–9.

Castellanos, S., Rodríguez, L.-F., Castro, L. A., and Gutierrez-Garcia, J. O. (2018). A computational model of emotion assessment influenced by cognition in autonomous agents. *Biologically Inspired Cognitive Architectures*.

Cavazza, M., Charles, F., and Mead, S. J. (2002). Character-based interactive storytelling. *IEEE Intelligent systems*, 17(4):17–24.

Dias, J., Mascarenhas, S., and Paiva, A. (2014). FAtiMA modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer.

Dias, J. and Paiva, A. (2005). Feeling and reasoning: A computational model for emotional characters. In *EPIA*, volume 3808, pages 127–140. Springer.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440.

Egges, A., Kshirsagar, S., and Magnenat-Thalmann, N. (2003). A model for personality and emotion simulation. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 453–461. Springer.

Egges, A., Kshirsagar, S., and Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer animation and virtual worlds*, 15(1):1–13.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Ekman, P. (1994). Moods, emotions, and traits. *P. Ekman & RJ Davidson, The Nature of Emotion: Fundamental Questions*, pages 56–58.

El-Nasr, M. S., Yen, J., and Ioerger, T. R. (2000). Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257.

Elliot, C. (1992). *The Affective Reasoner: A process model of emotions in a multi-agent system*. PhD thesis, Northwestern University, Institute for the Learning Sciences.

Ellsworth, P. C. (1991). Some implications of cognitive appraisal theories of emotion. *International review of studies on emotion*, 1:143–161.

Gadanho, S. C. and Custódio, L. (2002). Asynchronous learning by emotions and cognition. In *Proceedings of the seventh international conference on simulation of adaptive behavior on From animals to animats*, pages 224–225. MIT Press.

Gebhard, P. (2005). Alma: A layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM.

Gluz, J. and Jaques, P. A. (2017). A probabilistic formalization of the appraisal for the occ event-based emotions. *Journal of Artificial Intelligence Research*, 58:627–664.

Gratch, J. (2000). Emile: Marshalling passions in training and education. In *Proceedings of the fourth international conference on Autonomous agents*, pages 325–332. ACM.

Gratch, J. and Marsella, S. (2004a). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306.

Gratch, J. and Marsella, S. (2004b). Technical details of a domain-independent framework for modeling emotion. Technical report, University of Southern California Marina Del Rey Ca Inst for Creative Technologies.

Gross, J. J. and Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. *Handbook of emotion regulation*.

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1994). *Emotional contagion*. Cambridge, UK: Cambridge University Press.

Hindriks, K. V. (2009). Programming rational agents in goal. *Multi-Agent Programming*, pages 119–157.

Hudlicka, E. (2002). This time with feeling: Integrated model of trait and state effects on cognition and behavior. *Applied Artificial Intelligence*, 16(7-8):611–641.

Hudlicka, E. (2005). A computational model of emotion and personality: Applications to psychotherapy research and practice. In *Proceedings of the 10th annual cybertherapy conference: a decade of virtual reality*.

Hudlicka, E. (2007). Reasons for emotions: Modeling emotions in integrated cognitive systems. *Integrated models of cognitive systems*, pages 1–37.

Hudlicka, E. (2008). Modeling the mechanisms of emotion effects on cognition. In *AAAI Fall Symposium: Biologically inspired cognitive architectures*, pages 82–86.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological review*, 100(1):68.

Jain, S. and Asawa, K. (2015). Emia: emotion model for intelligent agent. *Journal of Intelligent Systems*, 24(4):449–465.

Jain, S. and Asawa, K. (2016). Programming an expressive autonomous agent. *Expert Systems with Applications*, 43:131–141.

Kaptein, F., Broekens, J., Hindriks, K. V., and Neerincx, M. (2016). Caaf: A cognitive affective agent programming framework. In *International Conference on Intelligent Virtual Agents*, pages 317–330. Springer.

Kowalczuk, Z. and Czubenko, M. (2016). Computational approaches to modeling artificial emotion–an overview of the proposed solutions. *Frontiers in Robotics and AI*, 3:21.

Kshirsagar, S. (2002). A multilayer personality model. In *Proceedings of the 2nd international symposium on Smart graphics*, pages 107–115. ACM.

Lambie, J. A. and Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychological review*, 109(2):219.

Lazarus, R. S. (1966). *Psychological stress and the coping process.* McGraw-Hill.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press on Demand.

LeDoux, J. E. (1996). *The emotional brain: The Mysterious Underpinnings of Emotional Life*. New York, Simon & Schuster.

Lim, M. Y., Dias, J., Aylett, R., and Paiva, A. (2012). Creating adaptive affective autonomous NPCs. *Autonomous Agents and Multi-Agent Systems*, 24(2):287–311.

Lin, J., Spraragen, M., and Zyda, M. (2012). Computational models of emotion and cognition. In *Advances in Cognitive Systems*. Citeseer.

Marinier III, R. P. and Laird, J. E. (2004). Toward a comprehensive computational model of emotions and feelings. In *ICCM*, pages 172–177.

Marinier III, R. P. and Laird, J. E. (2007). Computational modeling of mood and feeling from emotion. In *Proceedings of the Cognitive Science Society*, volume 29.

Marinier III, R. P., Laird, J. E., and Lewis, R. L. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, 10(1):48–69.

Marsella, S., Gratch, J., Petta, P., et al. (2010). Computational models of emotion. In *A Blueprint for Affective Computing-A sourcebook and manual*, volume 11, pages 21–46. Oxford University Press New York, NY.

Marsella, S. C. (2003). Interactive pedagogical drama: Carmen's bright ideas assessed. In *International Workshop on Intelligent Virtual Agents*, pages 1–4. Springer.

Marsella, S. C. and Gratch, J. (2009). Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90.

Mascarenhas, S., Dias, J., Prada, R., and Paiva, A. (2010). A dimensional model for cultural behavior in virtual agents. *Applied Artificial Intelligence*, 24(6):552–574.

McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

Mehrabian, A. (1996a). Analysis of the big-five personality factors in terms of the pad temperament model. *Australian Journal of Psychology*, 48(2):86–92.

Mehrabian, A. (1996b). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292.

Meyer, J.-J. C. (2006). Reasoning about emotional agents. *International journal of intelligent systems*, 21(6):601–619.

Minsky, M. (1988). *Society Of Mind*. Simon and Schuster.

Moffat, D., Frijda, N., Phaf, R., et al. (1993). Analysis of a model of emotions. In *Prospects for artificial intelligence: proceedings of AISB93*, pages 219–228.

Moshkina, L. (2006). An integrative framework for affective agent behavior. In *International Conference on Intelligent Systems and Control*.

Moshkina, L., Park, S., Arkin, R. C., Lee, J. K., and Jung, H. (2011). TAME: Time-varying affective response for humanoid robots. *International Journal of Social Robotics*, 3(3):207–221.

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Ojha, S., Vitale, J., and Williams, M. (2017). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *Annual Meeting of the Cognitive Science Society*.

Ojha, S. and Williams, M. (2017). Emotional appraisal: A computational perspective. In *Annual Conference on Advances in Cognitive Systems*.

Ojha, S., Williams, M.-A., and Johnston, B. (2018). The essence of ethical reasoning in robot-emotion processing. *International Journal of Social Robotics (IJSR)*, 10:211–223.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Petta, P. (2003). The role of emotions in a tractable architecture for situated cognizers. *Emotions in humans and artifacts*, pages 251–288.

Picard, R. W. (1997). *Affective computing*, volume 252. MIT press Cambridge.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Theories of emotion*, 1.

Price, D. D., Barrell, J. E., and Barrell, J. J. (1985). A quantitative-experiential analysis of human emotions. *Motivation and Emotion*, 9(1):19–38.

Rank, S. (2009). *Behaviour coordination for models of affective behaviour*. PhD thesis, Vienna University of Technology.

Rasool, Z., Masuyama, N., Islam, M. N., and Loo, C. K. (2015). Empathic interaction using the computational emotion model. In *IEEE Symposium Series on Computational Intelligence*, pages 109–116. IEEE.

Reilly, W. S. and Bates, J. (1992). Building emotional agents. *Technical Report CM UC S-92-143*.

Reisenzein, R. (2001). Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions. In *Appraisal processes in emotion: Theory, methods, research.*, Series in affective science., pages 187–201. Oxford University Press, New York, NY, US.

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, 10(3):241–278.

Roseman, I. J., Spindel, M. S., and Jose, P. E. (1990). Appraisals of emotion-eliciting events. *Journal of Personality and Social Psychology*, 59(5):899–915.

Rousseau, D. et al. (1996). Personality in computer characters. In *Proceedings of the 1996 AAAI Workshop on Entertainment and AI/A-Life*, pages 38–43.

Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294.

Saunier, J. and Jones, H. (2014). Mixed agent/social dynamics for emotion computation. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 645–652. International Foundation for Autonomous Agents and Multiagent Systems.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120):57.

Schneider, M. and Adamy, J. (2014). Towards modelling affect and emotions in autonomous agents with recurrent fuzzy systems. In *International Conference on Systems, Man and Cybernetics*, pages 31–38. IEEE.

Si, M., Marsella, S., and Pynadath, D. (2006). Thespian: Modeling socially normative behavior in a decision-theoretic framework. In *Intelligent Virtual Agents*, pages 369–382. Springer.

Singh, P. et al. (2005). *EM-ONE: an architecture for reflective commonsense thinking*. PhD thesis, Massachusetts Institute of Technology.

Sloman, A. (2001). Varieties of affect and the cogaff architecture schema. In *Proceedings of the AISB'01 symposium on emotions, cognition, and affective computing.*, volume 58.

Smith, C. A. and Kirby, L. D. (2001). Toward delivering on the promise of appraisal theory. In *Series in affective science. Appraisal processes in emotion: Theory, methods, research*, pages 121–138. Oxford University Press.

Smith, C. A., Lazarus, R. S., et al. (1990). Emotion and adaptation. *Handbook of personality: Theory and research*, pages 609–637.

Steunebrink, B. R., Dastani, M., Meyer, J.-J. C., et al. (2007). A logic of emotions for intelligent agents. In *Proceedings of the National Conference on Artificial Intelligence*, volume 22, page 142. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Steunebrink, B. R., Meyer, J.-J. C., and Dastani, M. (2008). A formal model of emotions: Integrating qualitative and quantitative aspects. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Sun, R. (2006). The clarion cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction*, pages 79–99.

Sun, R., Wilson, N., and Lynch, M. (2016). Emotion: A unified mechanistic interpretation from a cognitive architecture. *Cognitive Computation*, 8(1):1–14.

Swagerman, J. (1987). *The Artificial Concern Realization System ACRES: A Computer model of Emotion*. PhD thesis.

Tomkins, S. (1962). *Affect imagery consciousness: Volume I: The positive affects*. Springer publishing company.

van der Hoek, W., van Linder, B., and Meyer, J.-J. C. (1999). An integrated modal approach to rational agents. In *Foundations of rational agency*, pages 133–167. Springer.

Velásquez, J. (1997). Modeling emotions and other motivations in synthetic agents. *AAAI/IAAI*, pages 10–15.

Velásquez, J. D. (1999). An emotion-based approach to robotics. In *International Conference on Intelligent Robots and Systems (IROS)*, volume 1, pages 235–240. IEEE.

Velásquez, J. D. and Maes, P. (1997). Cathexis: a computational model of emotions. In *Proceedings of the first international conference on Autonomous agents*, pages 518–519. ACM.

Zadeh, L. A. (1996). Fuzzy sets. In *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*, pages 394–432. World Scientific.

Simple can be harder than complex: You have to work hard to get your thinking clean to make it simple.

— Steve Jobs —

4
Ethical Emotion Generation System (EEGS) Details

In Chapter 2, I presented a detailed discussion of the theoretical premises on which my research stands as well as the existing computational accounts from which I gain inspiration and motivation to propose a new computational model of emotion – EEGS. In this chapter, I will present a detailed mathematical and computational discussion of the proposed *Ethical Emotion Generation System (EEGS)*. I have to admit that, in the process of explaining the details of the proposed model and its working mechanism, I may use examples that sound exaggerated and unlikely to happen in real human-robot interaction (HRI) settings. However, the aim of these examples is not to predict scenarios that will be reailsed by future HRI applications but rather to give the reader a better understanding of the dynamics of emotional processing and ethical reasoning with exaggerated examples that represent clear episodes of unethical or problematic human-human social interactions.

4.1 EEGS: Ethical Emotion Generation System

EEGS stands for **E**thical **E**motion **G**eneration **S**ystem which is a computational model of emotion intended to provide an ability to generate ethically-guided emotional responses

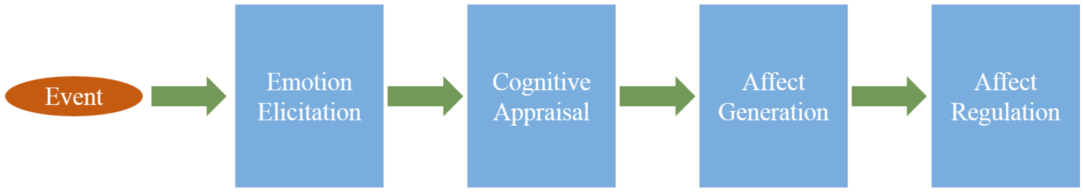


Figure 4.1 Process flow from an stimulus event to (1) emotion elicitation, (2) cognitive appraisal, (3) affect generation, and (4) affect regulation revisited.

and behavioural tendencies in social robots and virtual agents. By saying *ethically-guided emotional responses*, I refer to the process of emotion regulation that allows an intelligent agent to have adequate control over the computed emotions and regulate the emotions based on its ability to reason ethically thereby reaching to an emotional state that is socially acceptable. Before proceeding to the structural and computational details of the proposed model, I would like to set some background for the discussion by revisiting the mechanism of appraisal dynamics and overall emotion generation and regulation process introduced in Chapter 2, Section 2.3.

4.2 Revisiting Appraisal Dynamics

In Chapter 2, Section 2.3, I presented an overview of the flow of various processes that occur when a stimulus event occurs in the surrounding of an individual. Figure 2.7 from Chapter 2 has been provided below for a reader's convenience (see Figure 4.1). The whole process of emotion generation and regulation from the occurrence of an event was divided into four main phases namely (1) *Emotion Elicitation*, (2) *Cognitive Appraisal*, (3) *Affect Generation*, and (4) *Affect Regulation* (as shown in Figure 4.1). *Emotion elicitation* phase was defined as the process of being aware of the occurrence of the event and realising whether the event has positive or negative impact on the agent (natural or artificial). This process is defined as first-order phenomenological experience by Lambie and Marcel (2002). Once an individual becomes aware of the stimulus event through the non-cognitive first-order evaluation, then a second-order *cognitive appraisal* of the situation is performed (Lambie and Marcel, 2002). Once the mechanism of cognitive appraisal is completed, the appraisals performed then determine the emotion intensities which may also be affected by individual specific factors such as personality (Corr, 2008; Revelle, 1995) and mood (Morris, 1992; Neumann et al., 2001). This process is called *affect generation*. Since according to appraisal theories, an event can lead to the generation of more than one emotions, the generated emotions should be regulated to allow an artificial agent to reach a single optimal emotional state. This kind of regulatory mechanism is important because emotions of the agent may have significant effect on the decision it makes – as suggested by the literature (Callahan, 1988; Gaudine and Thorne, 2001). It has been suggested by the researchers that humans normally perform a conscious and deliberative process of handling conflicting emotional states and reach to a stable and situation-congruent emotional state (Gross and Thompson, 2007). This process is called *affect regulation*. A computational model of emotion should be able to account for these basic processes underlying the emotional episode. In line with this assertion, the following sections shall discuss how EEGS performs the above mentioned processes from a computational perspective. It is important to note that, in this work, the emotion elicitation process has not been computationally realised. It has been assumed that such a process occurs with an individual as suggested by Lambie and Marcel (2002) and has been applied to infer signed scores to indicate the negative or positive connotation of the surrounding event. This information is then provided to the cognitive appraisal process which provides necessary ingredients for the affect generation process and ultimately to the process of affect regulation. The latter three stages in Figure 4.1 have been investigated and their computational details presented in this dissertation.

4.3 Overall System Architecture

Figure 4.2 shows the overall relationship of various processes involved in the process of emotion generation in EEGS. The four core processes involved in the emotion generation mechanism are handled by (i) *Emotion Elicitation Module*, (ii) *Cognitive Appraisal Module*, (iii) *Affect Generation Module*, and (iv) *Affect Regulation Module*. These modules should depend on other modules for their effective functioning. For example, since cognitive appraisal process needs information about the goals, standards, and attitudes of the agent (Ortony et al., 1990), (v) *Memory Module* helps in the completion of this process by providing the data related to goals, standards and attitudes of the agent. Moreover, as discussed in Chapter 2, Section 2.4.3, mood and personality can play a significant role in the process of emotion generation. Such person specific factors are handled by (vi) *Characteristics Module* in EEGS. Although literature suggests that the experience of emotion may be affected by several other factors (Aleman et al., 2008; Canli et al., 2009; Corr, 2008; Hong et al., 2000; Morris, 1992; Neumann et al., 2001; Revelle, 1995; Scollon et al., 2004), I have considered the influence of only the mood and personality factors in EEGS since these factors are more widely agreed to have influence on the process of emotion. A detailed description of the working of each

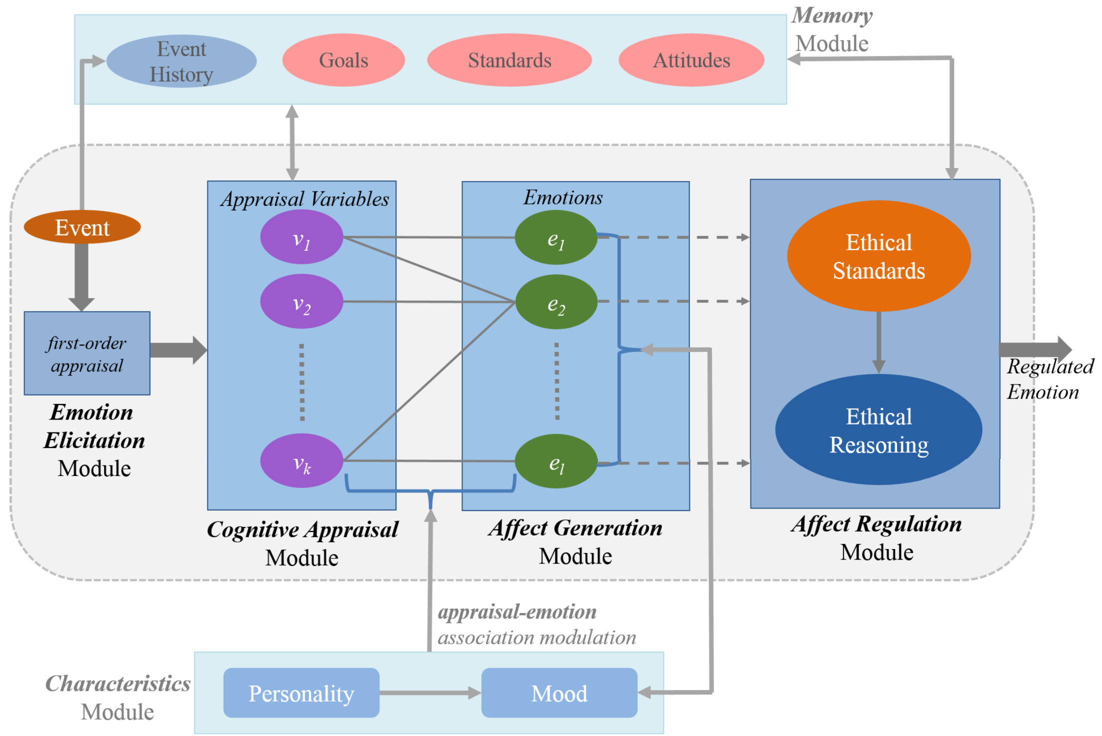


Figure 4.2 Overall architecture of EEGS.

module and its relationship with other modules in an agent running EEGS is presented below.

4.3.1 Emotion Elicitation Module

An event in the environment of an agent is first perceived by the *emotion elicitation module*. This module performs the early first-order (Lambie and Marcel, 2002) lower-level Scherer (2001) phenomenological appraisal of the situation based on the situational context. This is an instantaneous non-cognitive process where an individual evaluates the positivity or negativity of the perceived event as an experienced valenced bodily reaction (James, 1884; Lange, 1885). Such a first-order non-cognitive evaluation of the event provides initial information about the event to perform the conscious cognitive appraisal of the situation in relation to the agent's emotions generation. As mentioned earlier in Section 4.2, this research does not fully realise the process of first-order reaction. Such a reaction has been operationalised as numerical scores based on the data collected from humans for specific scenarios. The details of this process is presented in Chapter 5.

4.3.2 Cognitive Appraisal Module

The first-order phenomenological information (Lambie and Marcel, 2002) received from the *emotion elicitation module* is utilised by the *cognitive appraisal module* where the event is assessed based on the *goals*, *standards*, and *attitudes* of the agent – as suggested by Ortony et al. (1990). Once the first-order appraisal is received, *cognitive appraisal module* performs a parallel processing of several appraisals (see Section 4.6.2 to understand the parallel processing mechanism). Each of these processes compute an *appraisal variable*, which is an assessment criteria to evaluate the event. Each of the appraisal variables is associated with one or more emotions and the values of these appraisal variables determine the intensities of the associated emotions (see Section 4.7.1 for more details). This association between appraisal variables and emotions is weighted and may be affected by several individual-specific characteristics such as personality and mood (Corr, 2008; Neumann et al., 2001; Revelle, 1995). In this research, the association between appraisal variables and emotions has been influenced by only the factors of mood and personality to limit the scope of investigation.

4.3.3 Memory Module

Memory module in EEGS, plays a central role in managing the goals, standards and attitudes of the agent. Additionally, an *event history* is also maintained which keeps track of the previously experienced events from a particular external agent. The goals, standards and attitudes of the agent not only affect the cognitive appraisal process but are also affected by the resulting appraisals of the agent (as indicated by double pointing arrow between *memory module* and *cognitive appraisal module*). In the course of interaction, the goals, standards and attitudes are updated based on the how the external agent interacts with the agent (see Section 4.6.1 for more details).

4.3.4 Characteristics Module

In EEGS, the role of *characteristics module* is to manage the interaction of the emotion generation process with other person-specific factors such as personality (Corr, 2008; Revelle, 1995) and mood (Watson and Tellegen, 1985). Although there may be other factors that influence the process of emotion generation mainly in terms of determining the intensities of various emotions (Aleman et al., 2008; Bevilacqua and Goldman, 2011; Hong et al., 2000; Scollon et al., 2004), this research implements only the effect of personality and mood in EEGS. However, this is not to disagree that others factors do not participate in this process. By providing a specific *characteristics module* to accommodate factors affecting emotion elicitation process, EEGS can be extended in future works to enable interactions with other person-specific factors that have not been the focus of current research. In EEGS, mood and personality factors take part in the process of mapping the appraisal variables generated in *cognitive appraisal* module to the emotions in *affect generation module*. These factors determine to what degree an appraisal variable is associated with an emotion thereby influencing the affect of an appraisal variable on the intensity of an emotion (see Section 4.7.1 for more details).

Since the personality of a person is considered a non-dynamic characteristic (Costa and McCrae, 1988; Dweck, 2008), the personality of an agent running EEGS is not expected to change with time. Yet, mood is believed to change over the course of time with the influence of multiple emotional experiences (Beedie et al., 2005; Ekman, 1994; Parkinson et al., 1996). Therefore, emotions and mood in EEGS affect one another in the course of interaction (as indicated by double-headed arrow between emotions and mood in Figure 4.2). As such, mood affects emotions in two ways – one (i) *while mapping the appraisal variables to emotion intensities*, and another (ii) *by modulating the intensities of the emotions depending on whether the mood is congruent with the valence of the emotion*. This helps to ensure that mood congruent emotions are more likely to be activated instead of mood incongruent (see Section 4.7.3 for technical details of this process in EEGS).

4.3.5 Affect Generation Module

The central role in generating emotions in EEGS is played by *affect generation module*. I use the term 'affect' instead of 'emotion' because this module is responsible for the generation of emotion intensities as well as the mood of an agent. While the mood state is handled by *characteristics module*, its dynamics is influenced by the emotion intensities generated in the *affect generation module* (as indicated by double sided arrow between emotions and mood). Since the appraisal of an event can lead to the generation of more than one emotions with effective intensities (Ortony et al., 1990; Roseman et al., 1990; Scherer, 2001), EEGS may have more than one active emotion after the completion of affect generation process. As discussed earlier in Section 2.5.2, emotional state of an individual influences the decision s/he makes (Callahan, 1988; Gaudine and Thorne, 2001; Isen and Means, 1983). Therefore, an artificial agent with an ability to generate emotions should reach to a final stable emotional state if it has to be inolved in a decision making task that may be influenced by its emotional state. In such a situation, if there are multiple conflicting emotions active within the agent, the agent may not be able to make a right decision. Hence, it is necessary to converge the active emotions to a stable and regulated emotional state. This process is handled by *affect regulation module* discussed below.

4.3.6 Affect Regulation Module

When the *affect generation module* generates more than one conflicting emotions, these emotions should be converged to a final regulated emotional state for an agent to influence its decision based on its emotional state. This is because, the agent should also ensure that the emotional state triggered in response to a surrounding event is also socially acceptable in the given context (see Section 4.8 for more discussion on this). Although, researchers have previously proposed either *highest intensity approach* (Gratch and Marsella, 2004a) or *blended intensity approach* (Marinier III and Laird, 2007; Reilly, 2006) (see Section 4.8.1 for more discussion on this), I propose a better selection of emotional state can be achieved by *ethical reasoning mechanism*. In EEGS, ethical reasoning process is supported by its ethical standards which is constructed by the agent in the course of interaction with external agents (see Section 4.8.2 for technical details of this mechanism in EEGS). With the help of ethical reasoning capability, EEGS allows an agent to reach to a single emotional state that, when expressed, reduces the risks of socially unacceptable behavioural responses by the agent. I shall present a detailed evaluation of ethical reasoning mechanism in EEGS in Chapter 5.

It is important to note that the *goals*, *standards* and *attitudes* of an agent not only influence the process of appraisal (hence, emotion generation) and ethical reasoning, but are also influenced by these processes in return (as indicated by double-sided arrow between *memory module* and *cognitive appraisal module* and also between *memory module* and *affect regulation module*). The rationale behind this choice is that goals, standards, and attitudes of an individual are dynamic aspects and tend to change with the change in the environmental conditions and situations around the individual.

Although I do not present other modules beyond *affect regulation module* in the overall architecture of EEGS presented in Figure 4.2, it should be understood that the regulated emotional state output from the *affect regulation module* can be used as an input to any other expressive, behavioural or decision making unit an agent may have. I have left the choice open to any researcher who may further wish to extend my emotion model and integrate with other intelligent systems. This flexibility allows EEGS to be used as a plug-and-play component of other autonomous agents where selecting a socially acceptable emotional state may be needed as an input to other cognitive capabilities.

In the remaining of the chapter, I will discuss each of the modules presented in the overall architecture (Figure 4.2) in detail. Since this research is heavily inspired by the appraisal theory of Ortony et al. (1990), which considers events, actions of agents and objects at the core of appraisal mechanism, I will start the discussion with the introduction of these concepts.

4.4 Events, Actions and Objects

An event can be defined as a stimulus that has the potential to cause emotion elicitation in an individual. There can be numerous information associated with an event. For example, there can be several questions associated with an event: what happened in the event? was it likely to cause positive or negative impact on the agent? who is responsible for the incidence? when did the event occur? who was affected by the event? and many more. From the computational perspective, the notion of *event* should contain sufficient information that allows an artificial agent to perform situation assessment leading to emotional and behavioural responses. In the scenario of interaction between a human and an artificial agent or in the case of interaction between an agent and another or even in the situation of multi-agent interaction, the necessary information that should be present in an event include (i) *agent/person initiating the event*, (ii) *details about what was involved in the event*, (iii) *when did the event occur*, (iv) *who was primarily affected by the event*. The information about an event allows an individual (or an artificial agent) to perform cognitive appraisal of the situation before generating situation congruent emotional states (Ortony et al., 1990). The structure of an event operationalised in EEGS is presented in detail in Section 4.4.1.

The notion of events is closely related to the concepts of *actions* and *objects* (Ortony et al., 1990). It is because an event is usually accompanied by some action(s) and objects (persons, animals, vehicles, *etc.*) involved in its occurrence. An action performed by an agent A toward an agent B can have a positive or negative impact on agent B, where A and B may coincide as the same agent. For example, an action involving *helping* someone is considered positive by the person being helped while an action of *scolding* is considered negative by the individual being scolded. However, same action can be considered as positive or negative depending on the situation. For example, the action of *helping* may also be considered negative in some situations. Consider a situation where your best friend helps your worst enemy in a difficult task. In this case, even the action of helping which would normally be considered as positive is considered negative and it is likely to cause you anger. What determines this low-level (Scherer,

2001) first-order assessment of an action (Lambie and Marcel, 2002) is its relationship with the object involved in the event.

As mentioned before, an *object* can be a person, animal, vehicle, tree, wind, *etc*. An individual may have certain level of *familiarity* about an object. For example, in case of human object, an agent might have interacted with the person quite a number of times before thereby having a high degree of familiarity with the person. In case of non-human objects, an agent might not have recognised an object before thereby having a low degree of familiarity about the object. Moreover, an agent can build its own *perception* about an object based on its experience with the object. For example, a person who continuously misbehaves with an agent for someone liked by the agent, then the agent will gradually develop negative perception about the person.

4.4.1 Structure of Events, Actions and Objects[1]

In the previous section, I presented an understanding of the relationship among the events, actions and objects – drawing the inspirations from OCC theory (Ortony et al., 1990). In this section, I shall present the structural organisation of events, actions and objects in EEGS.

Event Structure

Events in EEGS are stored in the following form:

```
(<Source>,<Action>,<Target>,<DateTime>,<OtherInformation[]>)
```

`Source` denotes the object/person involved in the `Action` within the event. It should be noted that an `Action` is itself an object and can contain other parameters (to be dicussed soon after). `Target` denotes the object/person who is primarily affected by the action of the `Source`. `DateTime` denotes the date and time the event occurred. `OtherInformation[]` is an array that allows the storage of several other event/action/object related information. For example, familiarity and perception of the source and target after the occurrence of the event, overall impact of the event on the agent after the completion of the appraisal process, and so on. What information shall be contained in `OtherInformation[]` depends on the needs of the agent on which my emotion model is implemented. I offer this flexibility to allow broad range of applications of the proposed system.

---

[1] Parts of the discussion in this section have been previously published in my paper (Ojha and Williams, 2017)

Source	Action	Target	DateTime	OtherInformation[]
PAUL	Kick	DAVID	30/12/2017 10:30	[...]
JOHN	Help	KATE	25/09/2017 20:45	[...]
JULIE	Scold	NICK	11/07/2016 00:25	[...]

Table 4.1 An example of some events. The last column may contain other information like perception and familiarity of the agent with the source and target of action in the event, impact of the event on the agent, *etc.*

The first row in Table 4.1 shows a record of an event in agent's memory where a person named PAUL Kicked another person named DAVID on 30^{th} of December 2017 at 10:30 AM. It should be noted that although `Action` in Table 4.1 is shown as a string (*e.g.* "Kick"), an action is itself an object with more than one parameters. The internal structure of an action in EEGS is explained below.

Action Structure

Actions in EEGS are structured in the following form:

```
(<ActionName>,<ActionValence>,<ActionDegree>)
```

`ActionName` is an identifier used to denote type of action that occurred in the surrounding of the agent. For example, an action of kicking can be denoted by an identifier `"Kick"`. `ActionValence` denotes whether the action is considered positive or negative. It is a binary identifier that can be `POSITIVE` or `NEGATIVE`. `ActionDegree` signifies the degree of positivity or negativity of the action from `Source` to the `Target`. Its value can lie in the range [-1, +1], where -1 denotes extremely negative action and +1 denotes extremely positive action.

ActionName	ActionValence	ActionDegree
Greet	POSITIVE	0.31
Start Conversation	POSITIVE	0.28
Ignore	NEGATIVE	-0.17
Kick	NEGATIVE	-0.74

Table 4.2 An example of some actions. The `ActionValence` and `ActionDegree` presented in tha table are based on the data obtained from a survey (to be discussed later).

The `ActionValence` and `ActionDegree` associated with an action may have different values depending on the context. For example, the act of greeting from someone you like is likely to impact positively on the greeted person while the same action from someone you hate is likely to impact negatively. How an autonomous agent or robot would receive such a contextual information? How a robot would recognise complex physical actions like kicking in the first place, when it does not have human-like sensory channel to receive such an information? In general, how an autonomous agent would be able to appraise such actions and lead to the emotion generation? I shall address these questions and offer a solution in Section 4.5.

Object Structure

Objects in EEGS are structured in the following form:

```
(<ObjectName>, <Familiarity>, <Perception>)
```

`ObjectName` denotes the name of the object, `Familiarity` is a numerical representation showing how familiar is the agent to the object in interaction and `Perception` denotes the degree of positivity or negativity of the agent towards another agent/object. `Familiarity` can range from 0.0 to 1.0, where 1.0 denotes complete stranger and 0.0 denotes a very familiar person (considering the notion of distance). This choice was made with an analogy that close person would not be far in distance, hence the number '0' for more familiar person. Perception can range from -1.0 to +1.0, where -1.0 denotes very negative perception and +1.0 denotes very positive perception. For example, `("PAUL", 0.5, -0.4)`, denotes a person named `PAUL` who is somewhat familiar (0.5) to the agent and the agent has a negative (-0.4) perception about him. It is important to understand that the values of familiarity and perception are dynamic and change during the interaction with the person. This change applies to the case where the agent interacts with another person/agent or the interaction involves two other persons/agents leaving the model as a mere observer of the situation (Ojha and Williams, 2017).

ObjectName	Familiarity	Perception
PAUL	0.5	-0.4
JOHN	1.0	0.0
ROBERT	0.0	1.0
JESSICA	0.3	0.6
ALEX	0.1	-0.9

Table 4.3 An example of some objects. Adapted from Ojha and Williams (2017).

Table 4.3 shows an example of list of persons in the memory of EEGS. The person in the first row is named “PAUL” about whom the agent has a familiarity of 0.5 and perception of -0.4 towards the person. Initially, when a person is first introduced with the model, the person is considered stranger (i.e. familiarity = 1.0) and the model has a neutral perception (i.e. perception = 0.0) about the person. This design choice was made not to bias the model when a new person is introduced. The perception and familiarity about object changes in the course of interaction. This change is the result of continuous interaction of the object (person) with the agent thereby affecting the goals, standards and attitudes of the system, which will be explained in Section 4.6.1.

It would be unwise to argue that the above described notion of event/action/object are complete structures that can represent all the information about a event/action/object that might be relevant to the elicitation of emotions. I acknowledge that this is only one of the many ways an event/action/object can be structured. However, in-depth discussion of the structure events, actions or objects is not in the scope of this dissertation because the representations presented above are sufficient to validate my hypotheses and thesis argument.

4.5 Emotion Elicitation

In Section 4.4, I presented the structural representation of events, actions, and objects in the memory of EEGS. I also established how events, actions and objects are related to each other and augment their understanding. In this section, I will discuss how the occurrence of an event leads to the elicitation of simulated emotional experience in EEGS.

The stage of *emotion elicitation* should be considered as an early process where first-order assessment of the surrounding stimulus is performed (Lambie and Marcel, 2002). This kind of phenomenological experience can be commonly witnessed in case of humans. For example, when someone slaps you, it triggers an instant negative experience *because it is painful to you*. However, in case of autonomous agents (like robots), they are not able to have such an experience because they do not have adequate sensory organs as in humans. How can robots generate phenomenological response in reaction to an stimulus event? One approach is to provide a large amount of sample data with specified labels for various actions and train the system to define an optimal values for the `ActionValence` and `ActionDegree` for each `ActionName` (see action structure in Section 4.4.1). However, the issue with this approach is that the system will not be able to store the contextual information since these values will get updated with every new data sample. This creates problem when same action has to be simulated in different scenarios with different context, where the emotional responses and behavioural tendencies of the agent might be non-congruent with the situation. Since, the process of emotion elicitation is not the major focus of this research, I have only provided a viable alternative to the above mentioned problem, the details of which will be presented in Chapter 5.

The valenced score[2] (*i.e.* `ActionDegree`) associated with each action in the given context serve as the first-order phenomenological experience (Lambie and Marcel, 2002) of emotion before the occurrence of deliberate and conscious cognitive appraisal (Lambie and Marcel, 2002). This notion may also be related to the instant bodily sensations in response to a stimulus event as explained by James (1884). For example, a positive action (*i.e.* the action with positive value of `ActionDegree`) can lead to bodily sensations that provide a pleasant feeling and a negative action can lead to a bad feeling. This process of attending to the stimulus event is something similar to what Scherer (2001) calls as *intrinsic pleasantness check i.e.* the process of determining whether the event is inherently pleasant or unpleasant. Such an experience provides an opportunity to the agent to become aware of the stimulus event and determine whether the event may be significant to affect its current goals (Scherer, 2001). Any event that involves a significant impact on the goals of the agent would then promote a second-order deliberate and conscious process of cognitive appraisal (Lambie and Marcel, 2002; Scherer, 2001).

4.6 Cognitive Appraisal

Appraisal theory is the emotion theory in psychology that relates the process of emotion generation in humans to the cognitive aspects of mind (Lazarus, 1991; Ortony et al., 1990; Roseman, 1996; Scherer, 2001). According to the theory, generation of emotion in an individual is a cognitive process and is determined by the way the emotion inducing situation is appraised (evaluated) by the individual. Cognitive appraisal is usually considered as *second-order* (Lambie and Marcel, 2002) *higher-level* (Scherer, 2001) *deliberate* (Lazarus, 1982) and *conscious* (Frijda et al., 1989) process of evaluating a surrounding stimulus event. Since it is a complex cognitive process, appraisal mechanism is governed by a variety of factors. According to Ortony et al. (1990), the process of appraisal is determined by the *goals*, *standards* and *attitudes* of the appraising individual. As per the theory, how an event is appraised is determined by

[2]The use of the word *valenced* denotes a quantity that can be positive or negative *i.e.* has a sign.

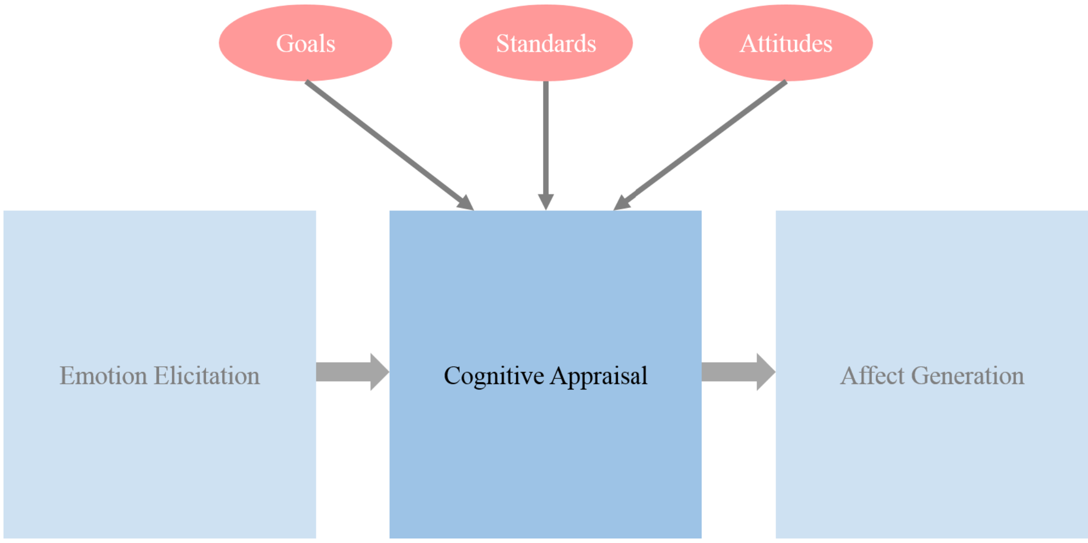


Figure 4.3 Influence of *goals*, *standards* and *attitudes* in cognitive appraisal process as suggested by Ortony et al. (1990).

the goals, standards and attitudes of the agent. This suggests that the computation of appraisal variables (variables used for the evaluation of a situation) is affected by goals, standards and attitudes of an individual (Ortony et al., 1990). Hence, to provide my model inspired by cognitive appraisal theory, I argue that it is crucial to understand the link between appraisal variables and goals, standards and attitudes as suggested by OCC theory. Although OCC theory describes the relationship of the goals, standards and attitudes to different variables, it does not provide explicit mathematical relation to compute the values of appraisal variables, which are necessary to build a computational model.

In the following sections, I will present the structure of the goals, standards and attitudes in EEGS and their relationships with various appraisal variables.

4.6.1 Goals, Standards and Attitudes[3]

According to Ortony et al. (1990), goals, standards and attitudes are the crucial driving factors in emotion generation because these are the "three major ingredients of appraisal [as] they constitute respectively the criteria for evaluating events, actions of agents, and objects" (Ortony et al., 1990, p. 13). In other words, events, actions of agents and objects can be effectively appraised in relation to emotion generation based on what the goals of the agent are, what are the standards of the agent and how is the attitude of the agent with interacting agent.

[3] Most of the content of this section has been adapted from Ojha and Williams (2017)

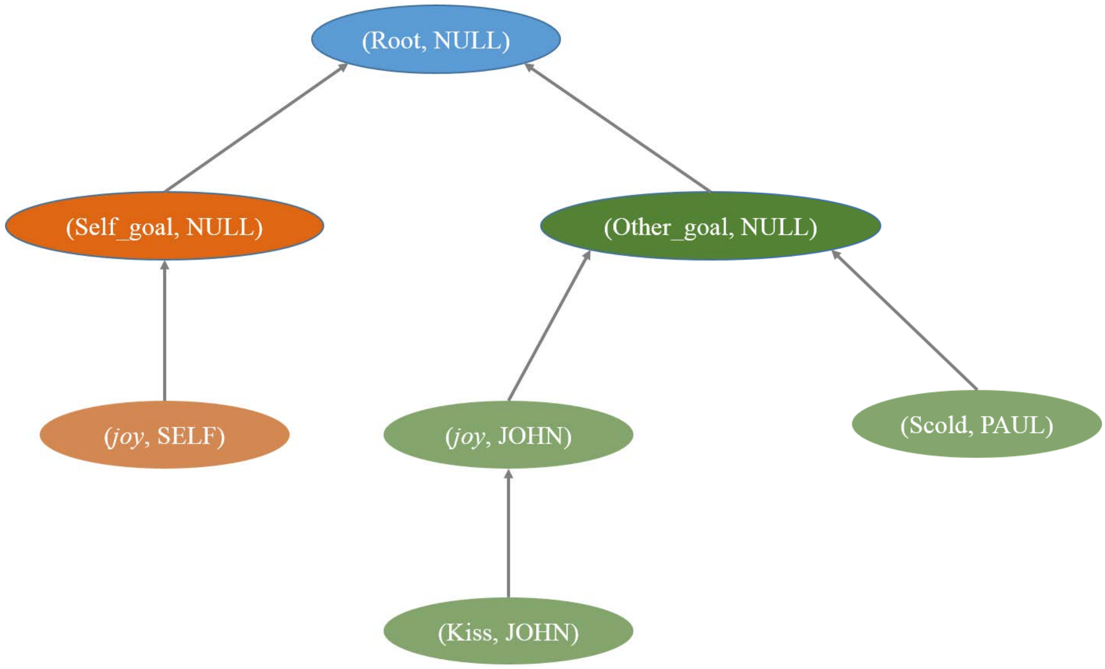


Figure 4.4 An example of a goal tree in EEGS based on OCC theory (Ortony et al., 1990). Redrawn after Ojha and Williams (2017, p. 6).

Goals

Goals represent a set of states that an individual wants to achieve. I use the term "set of states" because there can be more than one goal that an individual aims to accomplish. Moreover, accomplishment of one goal might help in achieving another goal. In EEGS, goals are represented in a hierarchy where a goal that helps in accomplishing another goal lies in the lower level of the hierarchy. I have represented the goals of the system as a tree structure in line with the proposal of the OCC theory. Each node of the tree is a goal node and a node may be linked to one or more lower level (child) nodes.

OCC theory of emotion has described three types of goals: (1) *Active-pursuit (A) goals*, (2) *Interest (I) goals* and (3) *Replenishment (R) goals* (Ortony et al., 1990). *A-goals* refer to the goals that an individual "wants to accomplish". As per the theory, such goals are discarded once they are accomplished since the person has no more desire or possibility once it is accomplished. For example, "goal of a computer scientist to win Turing Award". *I-goals* refer to the goals that an individual "wants to see happen", but does not accomplish it on its own. I-goals are also discarded once they are accomplished. *R-goals* refer to such goals that an individual "wants to accomplish" or achieve but does not stop willing to accomplish it once it is achieved. Such goals refer to regular and routing requirements of an individual. For example, a goal (desire) of eating, sleeping, etc. Such goals (desires) are not discarded once they are achieved but are needed routinely (Ortony et al., 1990). My computational model of emotion will be mainly focused on how the R-goals affect the appraisal variables that are related to goals of an individual. However, it should be noted that my mathematical formulation (to be presented following sections) can be adjusted to account for A-goals and I-goals as well.

Figure 4.4 shows an example of a goal tree in EEGS. The goals shown in the goal tree are in the form (<Action/Emotion>, <Object>), where Action/Emotion denotes the action to be done or emotional state to be attributed to a particular Object (Person)[4]. For example, goal node, (*joy*, JOHN) aims to bring "JOHN" in state of "*joy*". The root node (Root, NULL) has two children nodes (Self_goal, NULL) and (Other_goal, NULL), which denote the goals intended for self and for others respectively. Children of Self_goal node are the goals that are aimed for the benefit of oneself while the children of Other_goal node are aimed for the benefit of others. "NULL" Person for these goal nodes indicates that there is no specific target person – they just divide the goals into two categories. Lower level goals are useful for the accomplishment of the higher level goals. For example, the goal (Kiss, JOHN) helps in the accomplishment of the goal (*joy*, JOHN). It should be noted that although, here I show Action/Emotion and Person as strings, the actual computational representation stores them as objects. Similar is the case with the examples in standards and attitudes to be introduced in the following sections.

Standards

Standards maintain a collection of norms and values of an individual which may be shaped by the social context or learned concepts. In EEGS, I structure standards in the form:

(<Action/Emotion>, <Source>, <Target>, <Approval>)

which stores a belief that an Action/Emotion performed/expressed by the Source upon the Target has certain level of Approval as per the standard. Approval is further broken down into the structure (<Preference>, <ApprovalDegree>), where Preference indicates if the action from source to target is preferred or not and

[4]Since EEGS is currently intended to interact with humans only, the goals can either be an action performed to a person or an emotional state that the model wants to see in a person. But, it should be noted that this notion of goals can be extended beyond this scope without changing the computational mechanism of EEGS. Also, in line with this assumption, Object, now on-wards, shall be considered as Person only.

`ApprovalDegree` indicates the degree of that preference. For example, `(Slap, PAUL, NEIL, (NO, 0.8))` means “`PAUL` is `NOt` supposed to `Slap` `NEIL` and the degree of this preference is `0.8`”. `ApprovalDegree` denotes how strong is the belief that an individual has on the standard. Its value can range from 0 (exclusive) to 1 (inclusive). An `ApprovalDegree` of 1 indicates a very strong belief for that standard and a value close to 0 indicates a very weak belief for that standard. An example of some of the standards of EEGS is shown in Table 4.4.

Action/Emotion	Source	Target	Preference	ApprovalDegree
Slap	PAUL	NEIL	NO	0.8
Scold	SELF	ROBERT	YES	0.5
joy	SELF	JASMINE	YES	0.9
distress	SELF	NEIL	NO	0.4

Table 4.4 An example of some emotion and action related standards in EEGS. Adapted from Ojha and Williams (2017).

The structure of standard presented above may not be the only way to represent the related aspects. However, in the context of OCC theory as well as in the interaction of an agent with people/agents in social contexts, this structure of standard provides necessary ingredients for the computation of appraisal variables and hence in the generation of emotions.

Since standards contain a set of beliefs, the notion of standard should be dynamic as beliefs of a person might change in the course of life experience. For example, let us consider the example we presented in the previous paragraph. The standard `(Slap, PAUL, NEIL, (NO, 0.8))` might be changed if NEIL does some severely bad action to PAUL. In such a case, the standard might rather change to `(Slap, PAUL, NEIL, (YES, 0.2))`, which means “it is okay for `PAUL` to `Slap` `NEIL` and the degree of this preference is `0.2`”. I have maintained this idea of dynamic standards in my computational model that change in the course of interaction with human agents.

It should be noted that an individual (and hence the presented emotion model) can have many recognised persons, many recognised actions and many possible emotions. An individual’s standards should account for all of those aspects. The list of standards in Table 4.4 is not exhaustive, it only shows a few representative examples for the understanding of how the standards are structured in EEGS. Moreover, when EEGS runs for the first time, it starts with empty standards. It keeps on building and updating the standards as it interacts with various persons. This makes EEGS completely independent of the implementation domain and can build on its own as per the environmental context.

Attitudes

Attitudes defined in OCC theory (Ortony et al., 1990) can be considered as perception of an individual regarding persons or objects. But unlike the standards, attitudes in EEGS have a slightly different structure. An attitude is structured as `(<Person/Object>, <Perception>)`, where `Person/Object` refers to the person or object about whom the attitude is and `Perception` is the perception about the `Person/Object`. For example, the attitude `(JOHN,0.8)` means "the model has positive perception about `JOHN` and the degree of the positivity is `0.8`". It is important to note that even if the person "JOHN" is shown as string in the above example, it is actually represented as an object of class `Person` within computational model. As denoted earlier in the discussion about the structure of an object, `Perception` about a `Person` in EEGS can range from -1 to +1, where -1 indicates an extremely negative perception and +1 indicates extremely positive perception.

4.6.2 Appraisal Variables

Cognitive appraisal of an event is achieved through a set of criteria called appraisal variables (Ortony et al., 1990; Roseman, 1984; Scherer, 2001; Smith et al., 1990) which are governed by the goals, standards and attitudes, which were discussed in previous sections. Different theorists have proposed a different set of appraisal variables (see Chapter 2, Section 2.2.4 for a review). Moreover, there are a lot of common appraisal variables among various appraisal theories (as identified in Chapter 2, Section 2.2.4). The goal of this research is not to determine which set of appraisal variables is valid or sufficient for cognitive appraisal of a situation. So, I will not indulge into the discussion of that matter. In this dissertation, I aim to present a domain-independent approach of computing appraisal variables irrespective of the appraisal theory used. Although, appraisal theories have recently been recognised as being able to provide a theoretical foundation for achieving domain independence in computational modelling of emotions (Gratch et al., 2015), most existing computational emotion models based on appraisal theory have implemented rule-based domain specific designs (Aylett et al., 2005; Dias and Paiva, 2005; El-Nasr et al., 2000; Velásquez and Maes, 1997) to achieve cognitive appraisal for the generation of emotion. This is a significant problem because using domain specific pre-defined rules makes it difficult to reuse the model in other domains. I aim to resolve this issue through my research and present a mechanism to appraise events in domain-independent manner.

Computation of Appraisal Variables

According to appraisal theory, emotions are the result of appraisal of a particular situation or event happening in an individual's surrounding (Ortony et al., 1990). Thus, whenever an event occurs, an agent does the evaluation of the situation using several appraisal variables (based on the theory used) and the resulting values of the appraisal variables cause the generation of various emotions. The numeric value of most appraisal variables in EEGS range from the value of -1.0 to +1.0, which is only a design choice and I believe other alternatives should be equally effective (say for example, -100 to +100). The value of +1.0 for appraisal variable, say *desirability*, indicates that a particular event is extremely desirable while the value of -1.0 indicates that the event is extremely undesirable. In the following sections, I will discuss in detail how the numerical values of various appraisal variables are calculated in EEGS. Currently, EEGS is able to compute seven appraisal variables namely *goal conduciveness*, *desirability*, *praiseworthiness*, *appealingness*, *deservingness*, *familiarity* and *unexpectedness*. Since these appraisal variables have been proposed by OCC theory of emotional appraisal (Ortony et al., 1990), the computational mechanism presented in this dissertation borrows the proposals of the theory.

Desirability

Desirability is the measure of how desirable a particular situation or event is to the appraising individual. In order to evaluate the desirability of an event, it is compared to the goals of the individual (Ortony et al., 1990). If the event is likely to help in achieving goals, then the event is said to be desirable. However, if the event is likely to hinder the accomplishment of the goals, then the event is said to be undesirable. The degree of desirability or undesirability depends on the degree the event helps or hinders the achievement of the goals. An event may not be related to all the goals in the current goal tree (see Section 4.6.1 for the detailed structure of goals and goal tree considered in this research). Desirability of an event in EEGS is computed based on the overall effect the event has on the accomplishment of all the goals in the goal tree. This is determined by considering whether the event is relevant to the goal or not. Before calculating the OCC appraisal variable desirability, I compute a value called *goal conduciveness*, which is an appraisal variable adapted from Scherer's theory of appraisal (Scherer, 2001) for calculating the degree to which the event helps or hinders the achievement of a particular goal node that is related to the event. When the conduciveness of each goal is calculated, then the numerical value of desirability is computed as the average conduciveness of all the goals in the goal tree.

Suppose there are N goal nodes in the goal tree. If we denote the degree of the action[5]/emotion defined in the i^{th} goal node as $d_{g_i} \in$ [-1, 1]; the degree of the action in the recent event that is relevant to the i^{th} goal node as $d_{e_i} \in$ [-1, 1]; height of the i^{th} goal node from root node in the goal tree as h_i, then conduciveness (GC_i) of i^{th} goal in the goal tree is given by the following equation.

$$GC_i = \begin{cases} 1 - \frac{||d_{g_i}|-|d_{e_i}||}{h_i} & \text{if } sign(d_{g_i}) = sign(d_{e_i}), \\ & \text{or } d_{g_i} = d_{e_i} = 0 \\ \frac{||d_{g_i}|-|d_{e_i}||}{h_i} - 1 & \text{if } sign(d_{g_i}) \neq sign(d_{e_i}) \\ -\frac{|d_{e_i}|}{h_i} & \text{if } d_{g_i} = 0 \;\&\; d_{e_i} \neq 0 \\ -\frac{|d_{g_i}|}{h_i} & \text{if } d_{g_i} \neq 0 \;\&\; d_{e_i} = 0 \end{cases} \tag{4.1}$$

Where,
$sign(.)$ is a sign function.

The signed numeric value of d_{e_i}, which denotes the degree of the action in the event, is the input received by the cognitive appraisal component when an event occurs. This phenomenon represents what Lambie and Marcel (2002) consider as the first-order (lower-level) evaluation of an emotion inducing situation, which is performed by *emotion elicitation* component (see Section 4.5 for more discussion on emotion elicitation). The signed numeric value of d_{g_i} in a goal node is analogous to the *expected utility* of the achievement of the goal (Gratch and Marsella, 2004b). The formula in Equation (4.1) results in a numeric value between -1 and 1 which indicates the degree by which the event helps in attaining the i^{th} goal in the goal tree. A positive value of GC_i indicates that the event helps in achieving the i^{th} goal while a negative value indicates that the event hinders the accomplishment of the goal. Goal conduciveness basically computes the signed deviation of the event from the goal. The reason for dividing this quantity by the height of the node from the root, in Equation 4.1, is the assumption that if a goal node is closer to the root, its achievement will have more effect on the desirability than a goal node which is farther from the root node. When the conduciveness of each goal in the goal tree is computed, the value of desirability, here denoted with *desi*, is computed as the average goal conduciveness using the equation

[5]In EEGS, an action like *slapping* is considered to have negative degree and an action like *appreciating* is considered to have positive degree. The numeric value of degree of an action depends on how positive or negative the action is. This input value is considered as the result of first order evaluation in line with the arguments of Lambie and Marcel (2002). See Section 4.5 for more details on this.

below. The numeric value of *desi* indicates the degree by which the event is considered (un)desirable in relation to the goals of the individual.

$$desi = \frac{\sum_{i=1}^{N} GC_i}{N} \tag{4.2}$$

Where,
N is the total number of goal nodes in the goal tree.

Praiseworthiness

While the appraisal variable *desirability* is measured based on goals, the variable *praiseworthiness* is computed based on the standards (Ortony et al., 1990). An action is considered praiseworthy if it matches closely with the standards of the individual and blameworthy (negative value of the variable praiseworthiness) if it deviates from the standard(s). Praiseworthiness is evaluated by comparing the degree of an action performed by an external agent with the approval degree of that particular action from the given source to the target in the standards of the agent (see Table 4.4 for an idea on how an standard is denoted in this discussion context). If we denote the degree of the action in the event as $d_e \in$ [-1, 1]; the approval degree for the action in a given standard as $d_a \in$ (0, 1], then, praiseworthiness, denoted here with *prai*, is computed using the formula in Equation 4.3.

$$prai = \begin{cases} \text{for } d_e < 0; & \begin{cases} -(d_e * d_a) & \text{if } pref = YES \\ d_e * d_a & \text{if } pref = NO \end{cases} \\ \text{for } d_e > 0; & \begin{cases} d_e * d_a & \text{if } pref = YES \\ -(d_e * d_a) & \text{if } pref = NO \end{cases} \\ \text{for } d_e = 0; & \begin{cases} d_a & \text{if } pref = YES \\ -d_a & \text{if } pref = NO \end{cases} \end{cases} \tag{4.3}$$

Where,
$pref$ is the preference of the action in the standard. $pref$ can be “YES” if the action from the given source to target is preferred and “NO” if the action is not preferred.

The formula in Equation 4.3 considers the degree of the action performed, and preference and approval degree for the action from the source to the target in the standard. The value of degree (d_e) can range from -1 to +1, where -1 indicates very negative action and +1 indicates very positive action. The value of approval degree (d_a) can lie in the range (0, 1]. Approval degree is said to be maximum if it is equal to 1 and minimum if it has a value just greater than 0. $pref$ indicates if the action is preferred action or non-preferred action as per the standard. The formula in Equation 4.3 provides a signed numeric value *prai* which denotes how praiseworthy is the action of an external agent as per the standards of the appraising agent.

Appealingness

The appraisal variable *appealingness* measures how appealing (likeable) is the person/object to the appraising individual. In EEGS, *appealingness* is determined based on the perception of the model about the person interacting with it. A person who has done nice things in the past is considered to be appealing while a person who has done bad things is not and the degree by which such a perception is maintained depends on the frequency and intensity of the experience. AS such the numeric value of *appealingness*, denoted here with *appe* is given by the following formula.

$$appe = object\_perception \tag{4.4}$$

Where,

$object\_perception \in$ [-1, 1] is the numeric value of perception about the person/object the model has in its memory (see Section 4.4 for details about how perception of object is stored in agent's memory).

Deservingness

Deservingness is the measure of whether someone deserved to experience an incident or receive an action from someone else (Ortony et al., 1990). People normally determine someone's deservingness of something based on their past deeds. For example, a person who has done numerous good deeds in the past is considered to deserve something good while a person who is evil and does a lot of harm is considered non-deserving for experiencing good things in return. I apply similar assumptions in calculating the appraisal variable *deservingness*. Attribution plays a critical role in the computation of deservingness. For example, anyone would believe to deserve good things or at least thinking to not expecting bad things from happening. Therefore, when the target of the action in an interaction is the agent itself, the agent would consider the good actions to be deserving and bad actions to be non-deserving to itself. Likewise, when it comes to the determination of deservingness of an action/event in relation to another agent (considering a scenario of interaction between two external agents/humans), the model would consider the past actions of the interaction history in its memory. Accordingly, if an external agent has done good deeds to the agent in the model or other external agents liked by the agent in the model, then the model would consider that the external agent deserves to be treated well. If the external agent had done bad instead of good in the past, then the reverse would happen. This discussion leads to the inference that from the agent's perspective: (i) whether '*it*' deserved something to happen depends only on *how positive or negative the event/action is*, and (ii) whether an '*external agent*' deserved something to happen depends on *how positive or negative the event/action is in relation to the past deeds of the external agent*. I suggest the following formula in Equation 4.5 for deserviness, denoted by *dese*:

$$dese = \begin{cases} \text{if } target = \text{“}SELF'', & d_e \\ Otherwise, & d_e + (\ pi_{pos} + pi_{neg}) + \\ & (\ pi_{SELF_{pos}} + pi_{SELF_{neg}}) \quad \text{if } v_a = POSITIVE \\ & d_e - (\ pi_{pos} + pi_{neg}) - \\ & (\ pi_{SELF_{pos}} + pi_{SELF_{neg}}) \quad \text{if } v_a = NEGATIVE \end{cases} \tag{4.5}$$

Where,
$target$ is the target of action in the interaction,
d_e is the degree of action in the event,
pi_{pos} is the aggregate impact (positive value) of past positive actions from current *target* to the *source* of action,
pi_{neg} is the aggregate impact (negative value) of the past negative actions from *target* to the *source* of action,
$pi_{SELF_{pos}}$ is the aggregate impact (positive value) of past positive actions from current *target* to the *SELF*,
$pi_{SELF_{neg}}$ is the aggregate impact (negative value) of the past negative actions from *target* to the *SELF*, and
v_a is the valence of the action in current event.

A reader may be confused about the use of the quantities $pi_{SELF_{pos}}$ and $pi_{SELF_{neg}}$ for the calculation of deservingness even if the target is not 'SELF' (as shown in Equation 4.5). However, these quantities are necessary to signify a natural bias of personal experience a person might feel while measuring deservingness for someone. For example, consider a situation where Paul and Carl work in a company. Carl gets a promotion for his extraordinary outcomes during the year. In a situation where Paul and Carl do not know each other, Paul, by looking at Carl's achievements, will think that Carl deserves the promotion. However, if Paul and Carl know each other and they are in constant competition, with Carl having prevented Paul for getting such promotion by means of dishonest practices, Paul will think Carl does not deserve such promotion. To understand this phenomenon, let us look at the formula in Equation (4.5). In the case where *target* is not the SELF, and the valence of the action v_a is 'POSITIVE', considering the above situation of *rewarding the guy* would have a positive value for degree of action in the event d_e and the value of $(\ pi_{pos} + pi_{neg})$ will be zero because Carl and Paul have not met before. But, since Carl's deceiving behaviour has been unhelpful towards the appraising individual (Paul in this case), the value of $pi_{SELF_{pos}}$ will be zero and the value of $pi_{SELF_{neg}}$ will be some negative value. Now, the resulting value of *dese* will be $d_e + pi_{SELF_{neg}}$. If the absolute value of d_e is higher than or equal to that of $pi_{SELF_{neg}}$, the resulting deservingness will be a non-negative value; however, if that is not the case then the resulting deservingness will be a negative value signifying that Carl did not deserve to get the reward.

Familiarity

The appraisal variable *familiarity* measures how well known the person (or object) is to the model. In case of humans, if a person interacts with somebody never encountered before, that person would have no familiarity at all. Indeed, people use the word *stranger* to denote an individual about whom the appraising individual has no familiarity at all. Our familiarity with a person increases as we interact with that person or observe that person interacting with others.

Unlike, the appraisal variable *appealingness* which shows bidirectional dynamics, *i.e.* increases with the positive experience with a person and decreases with the negative experience with a person, *familiarity* change is unidirectional, *i.e.* you become more and more familiar no matter the experience was positive or negative. In EEGS, familiarity, denoted here with *fami* is determined by the following formula.

$$fami = relationship\_distance \tag{4.6}$$

Where,

$relationship\_distance \in [0, 1]$ is the numeric value that denotes how closely the model has known the person/object.

Unexpectedness

Appraisal variable *unexpectedness* measures how expected/unexpected a particular event was. In the context of interaction between two persons, a person would expect a behaviour that matches with how the other person behaved him/her in the past. As such, a sudden slapping from a best friend with a lot of fury is an example of an extremely unexpected event when one does not have any idea of wrongdoing. In EEGS, like *familiarity*, the value of the variable *unexpectedness* also lies in the range [0, 1]. But, unlike *familiarity*, where 0 means extremely familiar, in case of *unexpectedness*, 0 means fully expected and 1 means extremely unexpected. In EEGS, *unexpectedness*, denoted here with *unex*, is computed using the formula below.

$$unex = |d_{e_{avg}} - d_e| \tag{4.7}$$

Where,
unex is the unexpectedness of the event,
$d_{e_{avg}} \in$ [-1, 1] is the numeric value that denotes average degree of all the actions from *source* to the *target* in the past, and
$d_e \in$ [-1, 1] is the numeric value that denotes degree of action in the current event.

As such, $d_{e_{avg}}$ is calculated using the following formula.

$$d_{e_{avg}} = \frac{\sum_{i=1}^{N} d_{e_i}}{N} \tag{4.8}$$

Where,
d_{e_i} is the degree of the action from *source* to the *target* in the i^{th} event in the memory,
N is the number of events in memory that represent an interaction from *source* to the *target*.

Let us consider the above example to understand the working of Equation 4.7. Since the action of slapping has an inherent negative degree associated with it, the value of d_e will be negative. But, a best friends would have done a lot of help and given positive experience to a person. So, the value of $d_{e_{avg}}$ will be positive. Therefore, the absolute difference of $d_{e_{avg}}$ and d_e will be a larger quantity signifying a high degree of unexpectedness. However, if the action was something positive – say hugging, which would have positive degree. Since, both values are positive the subtraction will decrease the value of the quantity making the event more expected compared to the previous situation.

In this dissertation, I have presented a detailed computation mechanism of seven appraisal variables. It should be noted that this list of appraisal variables is not exhaustive – not even in relation to OCC theory (Ortony et al., 1990). The question of "how many appraisal variables does a model implement?" is not important. Instead the question of "whether an autonomous agent implementing an emotion model shows coherent

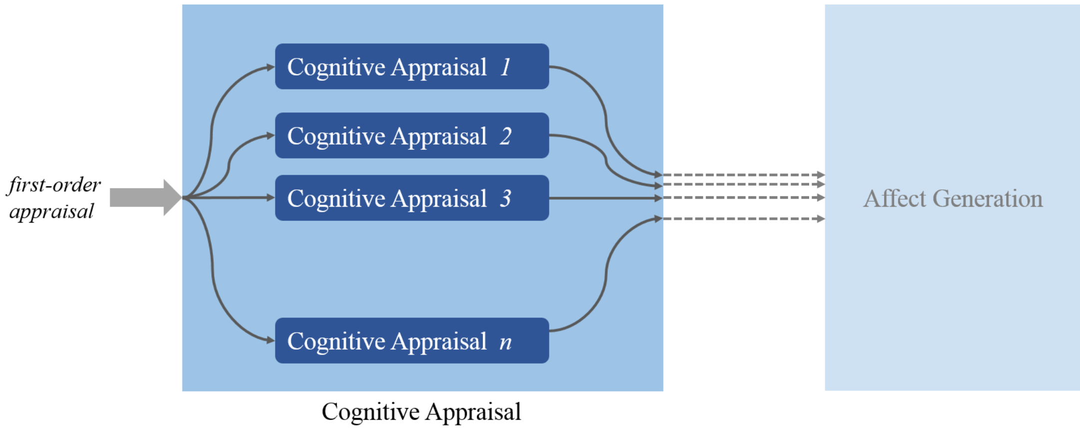


Figure 4.5 Parallel computation of appraisals in EEGS.

and consistent appraisals" is more important. The number and types of appraisal variables implemented in a model depends on the intended application of the model. For example, EMA (Gratch and Marsella, 2004a; Marsella and Gratch, 2009) models only three appraisal variables *i.e. desirability*, *likelihood* and *coping potential* (Gratch and Marsella, 2004b) but it is still considered one of the most plausible and influential models of emotion because of its ability to perform cognitive appraisal in a domain-independent manner (Becker, 2008; Dias and Paiva, 2005). Likewise, EMIA (Jain and Asawa, 2015) models only five appraisal variables namely *desirability*, *expectedness*, *outcome probability*, *suddenness*, *cause harm* which were sufficient to achieve the goals of the scenarios simulated in their model. Although, currently, I have implemented only seven appraisal variables, EEGS is flexible enough to account for additional appraisal variables should the mechanism for computation of those variables be available (I will discuss more on this in Section 4.7.1).

Parallel Processing of Appraisals

Scherer (2001) suggests that appraisal computation follows a sequential pattern and one particular appraisal can not commence until a pre-requisite appraisal check is not completed. For example, the theory of Scherer (2001) assumes that the appraisal of *control check* can only be performed after *goal/need conduciveness check*. However, this proposition is often criticised by psychologists (Ortony et al., 1990; Smith and Kirby, 2001) as well as computer scientists in affective computing field (Marsella and Gratch, 2009). In EEGS, I follow the assumptions of Smith and Kirby (2001) that appraisals run in parallel and which appraisal is computed first depends on the complexity of the situation and related history in the appraising individual's memory.

The *first-order* non-cognitive appraisal performed by *emotion elicitation* component (not shown in Figure 4.5) is received by the *cognitive appraisal component*. The cognitive appraisal component performs various appraisals in parallel (as shown in Figure 4.5). As stated earlier, which appraisal variable gets computed first depends on the complexity of current *goals*, *standards*, and *attitudes* of EEGS. The individual appraisals then directly influence the corresponding emotions and their intensities in *affect generation* component. Where more than one appraisal variables influence the same emotion, an incremental effect is applied on the emotion intensity. Since, different appraisals may complete at different times, the emotional states keep fluctuating until all the appraisals are determined for the given stimulus event. Such intermediate emotions may cause conflicting affective states to be experienced by the agent which need to converge to a stable emotional state. How such a regulation of emotional states is achieved in EEGS shall be discussed in Section 4.8.

4.7 Affect Generation

In Section 4.6, I presented how the mechanism of cognitive appraisal occurs in EEGS and how various appraisal variables are computed. According to appraisal theories, appraisal variables are responsible for the determination of emotions and their intensities in response to the stimulus event (Ortony et al., 1990; Roseman et al., 1990; Scherer, 2001; Smith et al., 1990). However, emotion is not an isolated phenomenon. Literature suggests that the mechanism of emotion processing is influenced by individual-specific factors such as *personality* (Corr, 2008; Revelle, 1995; Watson and Clark, 1997) and *mood* (Morris, 1992; Neumann et al., 2001). It is desirable to model the influence of such factors in emotion generation mechanisms of artificial agents, since in practical applications, an intelligent agent should exhibit some difference in emotional response and hence action tendency if it is to be employed in wide range of human-centred situations. For example, an intelligent agent intended to be employed as a personal development assistant is desirable to have an “organised and systematic” characteristic (conscientiousness). As such the agent might have to express disappointment or similar emotions if the person under training ignores some routine activity. However, if the agent is to be deployed as an emotional support companion, then it is preferred to forgive such a minor ignorance – hence it is desirable to have an easy going nature (agreeableness). Similarly, mood can also play an important role in modulating the emotional responses of the agent. For example, if an agent is in a very good mood state (based on recent experience), it might easily forgive an insult from the human interacting

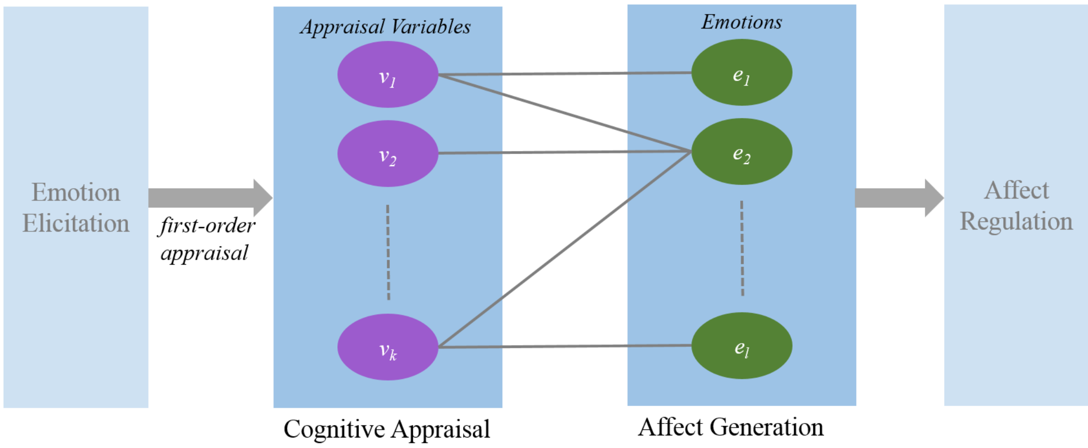


Figure 4.6 Role of cognitive appraisal in affect generation process. The appraisal variables computed as a result of the cognitive appraisal process help in determining the intensities of the associated emotions.

with it, while it may not do so if it is in very bad mood. The above examples show only a few of most situations where personality and mood can substantially influence the process of emotion generation. Hence, it is crucial to integrate the relationship of mood and personality in emotion processing mechanism of an intelligent agent. In order to realise such an effective integration of personality and mood in a model of emotion generation, I propose an *appraisal–emotion network* modulated by the factors of personality and mood, which will be detailed in Section 4.7.1.

4.7.1 Appraisal–Emotion Network

Once the appraisal variables are computed, the next step for an agent is to exhibit the situation congruent emotional response(s). This objective can be achieved by considering the previously computed appraisal variables for that situation. Studies in emotion research widely support the idea that an individual may generate more than one emotion with varying intensities in response to a single event (Ortony et al., 1990; Scherer, 2001). This means that an event can cause both *joy* and *distress* at the same time but with different intensities. For example, consider a situation in which you hear a news that your friend met with a fatal car accident in which the friend was injured (but alive) and the car was completely destroyed. This news might give you a mixed feeling. Distress that your friend met with an accident, joy that your friend is alive and safe while at the same time some distress knowing that your friend's new car is completely damaged. Now, suppose an artificial agent faces a situation in which conflicting emotions may be triggered. Although the agent can evaluate the situation

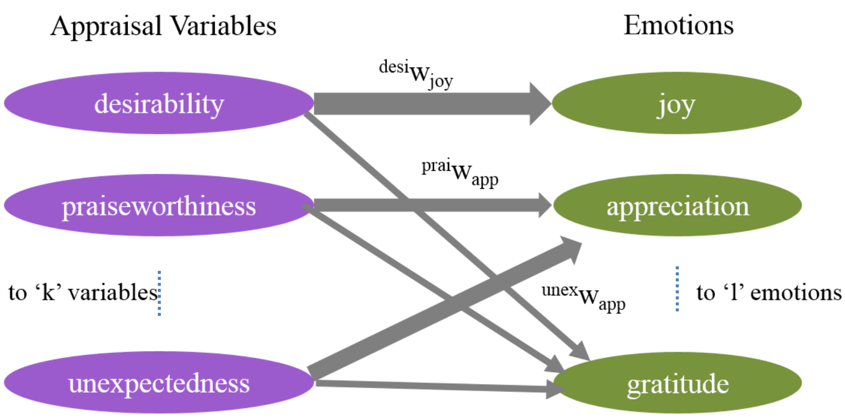


Figure 4.7 An weighted appraisal-emotion network showing many-to-many relationship between appraisal variables and emotions

using the appraisal functions that return a set of appraisal variables, the agent does not yet know how those variables may be associated with various possible emotions.

Appraisal theories of emotion propose that appraisal variables and emotions maintain a many-to-many relationship. This means that an appraisal variable may affect more than one emotion and an emotion may be affected by more than one appraisal variable (Ortony et al., 1990). For example, Figure 4.7 shows a generic network of appraisal variables linked to emotions. On the left of the network are appraisal variables. Please note that the network can have k number of appraisal variables (depending on the theory used) although only three appraisal variables are currently shown in Figure 4.7. This choice has been made for the simplicity of explanation. Likewise, on the right of the network are emotions of the artificial agent. There can be l number of emotions based on which emotion theory is implemented in the agent. As previously mentioned, each appraisal variable may influence one or more emotions. Also, an association between an appraisal variable and an emotion is weighted showing the degree by which the appraisal variables affect on the intensity of the emotion. For example, in Figure 4.7, the emotion *joy* is determined by the appraisal variable *desirability* and the weight of association between them is denoted by $^{desi}w_{joy}$. This suggests that a desirable event tends to cause the emotion of joy in the agent. Unlike the emotion of *joy*, the emotion of *appreciation* is affected by two appraisal variables. Appraisal variables *praiseworthiness* and *unexpectedness* affect the intensity of emotion *appreciation*. This means if the action of a human actor is praiseworthy from the perspective of the artificial agent and if the action was not expected, this combination may lead the agent to reach to an emotional state of appreciation. The weights of association of appraisal variable *praiseworthiness* and *unexpectedness* with emotion *appreciation* are denoted by ${}^{prai}w_{app}$ and ${}^{unex}w_{app}$ respectively.

Complex emotions like *gratitude* offers the option of two different kinds of treatment. As explained by most appraisal theories, the emotion of *gratitude* can be considered a combination of the emotions *joy* and *appreciation*. In other words, *gratitude* is a state of emotion in which an agent appraises an event to be desirable to achieve its goals and the action of the human counterpart involved in the event was praiseworthy as per the standards of the agent and this action was quite unexpected for the agent. Therefore, the intensities of complex emotions (including *gratitude*) can be determined by two different ways: (i) *by combining the intensities of simpler constituent emotions* or (ii) *by considering the association with the appraisal variables linked to the constituent emotions*. For example, intensity of the emotion *gratitude* can be determined by combining the intensities of the emotions *joy* and *appreciation*. Alternatively, the same thing can be achieved by considering that the emotion *gratitude* is affected by appraisal variables *desirability*, *praiseworthiness* and *unexpectedness* (as denoted by the arrows from these variables to the emotion *gratitude* in Figure 4.7).

However, it should be noted that none of the existing appraisal theories present a clear explanation of how the intensities of simpler emotions can be combined to produce the intensity of more complex emotions, with the exception of only a few computational accounts (Reilly, 2006). A very simple approach could be to perform an average of the intensities of constituent emotions. But, I anticipate that this approach may not provide an accurate simulation of emotion generation mechanism in artificial agents, in line with the views of Hudlicka (2008). Whether the contribution of the constituent emotion intensities to complex emotions coincides (or closely resembles) with the intensity given by considering the individual appraisal variables does not have a universal consensus. This shall involve a separate research work the discussion of which lies outside the scope of this section. In this section, I am mainly concerned about presenting a *theory-independent* approach to generate intensities of emotions from appraisal variables in an artificial agent. By saying 'theory-independent', I mean that my approach can be applicable in an artificial agent implementing any appraisal theory of emotion. One should also note that the appraisal-emotion network presented in Figure 4.7 is an example only. The approach I am proposing in this dissertation can be used for any appraisal theory. Following the mainstream belief on the existing appraisal theories, I assumed that there exists a weight for each association of an appraisal variable to an emotion (as denoted by solid arrows in Figure 4.7). But, how do we operationalise such weights in artificial agents? Although the emotion theories suggest that the process of mapping the appraisal variables to emotion intensities is modulated by factors like *personality* and *mood* (Ortony et al., 1990), only descriptive accounts are available, which cannot be readily implemented in computational models. I shall further discuss how the factors of personality and mood can help in establishing a data-driven relationship of appraisal variables with emotion intensities.

4.7.2 Data-driven Learning of Appraisal–Emotion Association

In the previous sections, I discussed that (i) *an artificial agent implementing appraisal theory for emotion generation needs to have a mechanism to map the computed appraisal variables into various emotion intensities*; (ii) *each appraisal variable may be associated with one or more emotions* and (iii) *each emotion may be influenced by one or more appraisal variables*, and (iv) *these associations are weighted indicating the degree by which the given appraisal variable affects the given emotion*. Additionally, based on the existing literature, I also stated that the weight of association of appraisal variables to emotions is modulated by the factors of personality and mood (Corr, 2008; Morris, 1992; Neumann et al., 2001). In this section, I expand the appraisal–emotion association weight (${}^{var}w_{emo}$) into a simple linear form as shown in (4.9).

$$ {}^{var}w_{emo} = f_O * O + f_C * C + f_E * E + f_A * A + f_N * N + f_M * M \qquad (4.9) $$

Where, ${}^{var}w_{emo}$ denotes the final weight of association between an appraisal variable and an emotion. O denotes the personality factor of *openness* and f_O denotes the weighting for the factor O. C denotes the personality factor of *conscientiousness* and f_C denotes the weighting for the factor C. E denotes the personality factor of *extroversion* and f_E denotes the weighting for the factor E. A denotes the personality factor of *agreeableness* and f_A denotes the weighting for the factor A. N denotes the personality factor of *neuroticism* and f_N denotes the weighting for the factor N. M denotes the *mood* factor and f_M denotes the weighting for the factor M.

In Equation (4.9), the weighing factors (f_O, f_C, f_E, f_A, f_N and f_M) can have positive as well as negative value. This is because a personality factor may have positive, negative or no influence in the relationship between an appraisal variable and an emotion. For example, we can anticipate that the personality of *extroversion* has positive effect on the emotion *joy* and the personality of *neurotocism* may have negative effect on *joy*. Hence, for the link ${}^{des}w_{joy}$ (in Figure 4.7), the value of f_E should be positive while the value of f_N should be negative. Although, we can make use of commonsense to estimate the plausible range for the values for weighing factors, we can not specify a singular

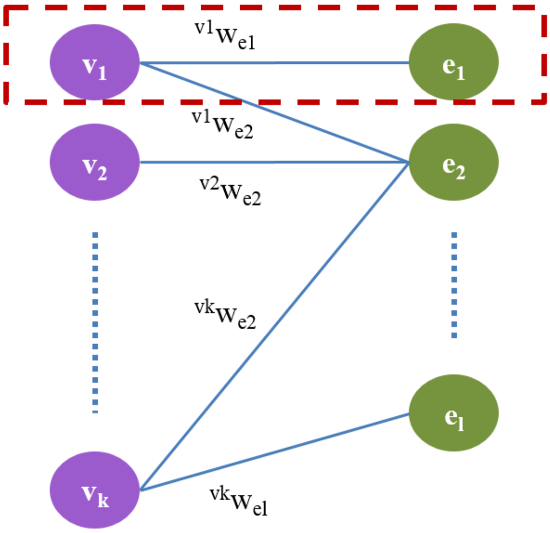


Figure 4.8 A general appraisal-emotion network with k appraisal variables and l emotion types.

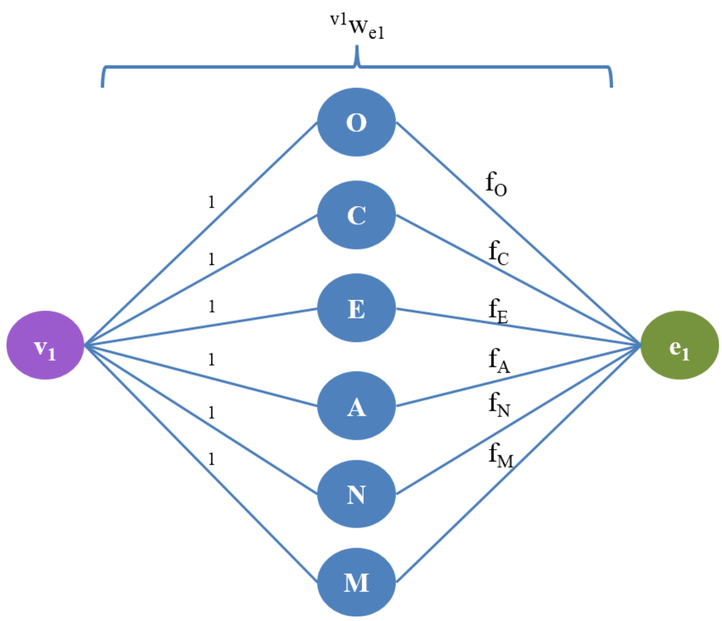


Figure 4.9 Decomposition of the link between appraisal variable v_1 and emotion type e_1.

value with certainty. This is the reason I employed a machine learning technique to determine these weighing factors for each appraisal-emotion relationship that can be applied to map any number of appraisal variables to any number of emotions depending on the appraisal theory used to simulate the emotion mechanism in artificial agent. The following section presents the learning algorithm used to establish a data-driven relationship among appraisal variables and emotions.

Learning Algorithm

Considering we have k appraisal variables and l emotion types, the weighted appraisal-emotion network in Figure 4.7 can be represented in more generalised form as in Figure 4.8. ${}^{v_1}w_{e_1}$ denotes the weight of association between the appraisal variable v_1 and the emotion type e_1 and so on. As previously mentioned, this weight is affected by the personality factors and mood of the appraising individual. In order to better understand this phenomenon, we can break down the link between v_1 and e_1 (denoted by the dashed rectangle in Figure 4.8), and represent as shown in Figure 4.9.

The contribution of an appraisal variable to an emotion intensity is given by the product of the value of the appraisal variable and the weight of association of the appraisal variable to the emotion type. For example, the contribution of the appraisal variable v_1 to the emotion type e_1 (see Figure 4.8) can be defined as the value $v_1 * {}^{v_1}w_{e_1}$. This implies the following.

$$\begin{aligned}
\hat{e}1_{v_1} &= v_1 * {}^{v_1}w_{e_1} \\
&= v_1 * (f_O * O + f_C * C + f_E * E + f_A * A + f_N * N + f_M * M) \\
&= v_1 O f_O + v_1 C f_C + v_1 E f_E + v_1 A f_A + v_1 N f_N + v_1 M f_M \\
&= p_O f_O + p_C f_C + p_E f_E + p_A f_A + p_N f_N + p_M f_M
\end{aligned} \tag{4.10}$$

Where p_O denotes the product of v_1 and the personality factor O i.e. $v_1 * O$; p_C denotes $v_1 * C$; p_E denotes $v_1 * E$; p_A denotes $v_1 * A$; p_N denotes $v_1 * N$; p_M denotes $v_1 * M$.

Alternatively, the contribution of the appraisal variable v_1 to the emotion type e_1 i.e. $\hat{e}1_{v_1}$ can be denoted as follows.

$$\begin{aligned}
\hat{e}1_{v_1} &= \sum_x p_x * f_x \\
&\text{where, } x \in \{O, C, E, A, N, M\} \\
&\qquad p_x \in \{p_O, p_C, p_E, p_A, p_N, p_M\} \\
&\qquad f_x \in \{f_O, f_C, f_E, f_A, f_N, f_M\}
\end{aligned} \tag{4.11}$$

The formulation in Equation (4.11) considers p_x as the input value and f_x as the weight to be learned for the corresponding p_x. If the expected value for the contribution of the appraisal variable v_1 to the emotion e_1 is denoted by $e_{1_{v_1}}$ and estimated value is denoted by $\hat{e}_{1_{v_1}}$, then the learning mechanism for each f is given by the equation below.

$$f_x' = f_x + \eta(e_{1_{v_1}} - \hat{e}_{1_{v_1}}) * p_x \tag{4.12}$$

In Equation (4.12), η is decaying learning rate.

Formalisation of the Learning Algorithm

In the previous section, I presented a demonstration of how a link between an appraisal variable and an emotion can be formulated to determine the weights associated with personality and mood factors for that link. In this section, I shall present a formal model of the algorithm used.

Most emotions $\mathbf{E} = \{e_1, e_2, .., e_l\}$ could be predicted directly from the appraisal variables $\mathbf{V} = \{v_1, v_2, ..., v_k\}$ using a linear model $\mathbf{W}$ having less than or equal to $k \times l$ links (weights) such that $\hat{e}_l = \sum_k w_{lk} v_k$. However, as previously mentioned, it is believed that the weights W are modulated by factors F of personality and mood $\hat{M} = \{m_O, m_C, m_E, m_A, m_N, m_M\}$ so that

$$w_{lk} = \sum_x (f_{lk})_x m_x : x \in \{O, C, E, A, N, M\} \tag{4.13}$$

So, given the data set $D = \{d_1, d_2, .., d_l\}$ of samples d, where each sample $d = \{E, V, \hat{M}\}$ is a set of an emotional state E, an appraisal V and personality/mood factors $\hat{M}$, we can define the model to predict E from V and $\hat{M}$ as follows:

$$\hat{e}_l = \sum_k \left(\sum_x (f_{lk})_x m_x \right) v_k \tag{4.14}$$

We can learn the parameters of this linear model using stochastic gradient descent (Bottou, 2010) by minimising the squared error $L = \sum_l (e_l - \hat{e}_l)^2$ summed over the dataset D.

Since:

$$\frac{dL}{df_{lkx}} = 2(e_l - \hat{e}_l)\frac{d\hat{e}_l}{df_{lkx}} = 2(e_l - \hat{e}_l) m_x v_k \tag{4.15}$$

we can minimise $\sum_{d \in D} L_d$ by iteratively performing the update using individual lines of data d:

$$f_{lkx} \leftarrow f_{lkx} + \eta(e_l - \hat{e}_l)m_x v_k \tag{4.16}$$

where η is a decaying learning rate.

The learning mechanism presented in the previous sections helps in determining (i) *how each of the personality and mood factors quantitatively affect the process of emotion generation* (personality factors as described in Five Factor model (Digman, 1990) were introduced in Chapter 2, Section 2.4). Running the above mentioned algorithm allows EEGS to (ii) *determine the weights of association for each of the factors in the link between an appraisal variable and an emotion*. Combining the sum of the product of these weights and corresponding factors gives the overall weight of association (${}^{var}w_{emo}$) between an appraisal variable (*var*) and an emotion (*emo*) as presented in Equation (4.9). This weight is then multiplied by the quantitative value of the appraisal variable in order to determine the intensity contribution of the appraisal variable to the emotion as in Equation (4.10) or (4.11). The final intensity of an emotion is determined by summing the contribution from all the appraisal variables associated, which shall be detailed in Section 4.7.3. Moreover, in Chapter 5, I will present a detailed evaluation of the learning mechanism presented above, where I will also explain about the methodology collect the dataset D. The data collection methodology used in this research can be used for any emotion model inspired by EEGS system, despite having different appraisals and emotional states.

4.7.3 Implementation of Affect Generation Process in EEGS

In Sections 4.7.1 and 4.7.2, I (i) presented how different appraisal variables can be linked with various emotions with different degrees *i.e.* weights, and also (ii) explained the machine learning approach used that allows EEGS to learn the parameters of this proposed model, namely the weights between appraisals and emotions. Most importantly, the discussed approach is general enough to consider any combination of appraisal variables and emotional states the system has to consider. As such, the only requirements is an appropriate dataset D including enough samples for such appraisals and emotions (see Chapter 5, Section 5.2.2 for the details on the data collection methodology). In the following paragraphs, I will describe how emotions are modelled in EEGS.

Emotion Types in EEGS

EEGS is currently able to generate and express ten emotions described in OCC theory (Ortony et al., 1990) which are listed below.

- *Joy* : A feeling of pleasure or happiness.
- *Distress* : A feeling of anxiety, sorrow, or pain.
- *Happy_For* : A feeling of happiness for someone's desirable situation.
- *Sorry_For* : A feeling of sadness for someone's undesirable situation.
- *Appreciation* : A feeling when one recognises the good qualities or actions of someone.
- *Reproach* : To express to (someone) one's disapproval of or disappointment in their actions.
- *Gratitude* : The state of being grateful to someone.
- *Anger* : A strong feeling of annoyance, displeasure, or hostility.
- *Liking* : A feeling when you see someone appealing or interesting.
- *Disliking* : A feeling when you see someone unappealing or uninteresting.

To effectively process emotions in a computational model like EEGS, an appropriate structure representing various aspects of an emotion is necessary. In this paragraph, I will provide the structure used in EEGS to represent emotional states. According to literature, an emotion can be categorised with a name for its type (Ortony et al., 1990) and, therefore, each emotion is addressed by a specific word in a language. This label is used as a proxy to refer to the feeling the person experiences when under the influence of such emotional state. For example, the emotional label *joy*, in the above list, is used to refer to a feeling of internal pleasure. Since my computational model has been heavily inspired by OCC theory (Ortony et al., 1990), my representation considers the assumption of the theory that *emotions are valenced reactions to situations*. Hence, I assume that the emotional state can be perceived as leading to either a negative or positive experience for the subject, *i.e.* the experience underlying the considered emotion is valenced. For example, the emotion *gratitude* is positively valenced, because it leads to a pleasant experience for the subject, whereas the emotion *anger* is negatively valenced. Importantly, to effectively describe and differentiate emotional states, it is not enough to assign them a discrete state of valence either being negative or positive. Indeed, the valenced reaction of the emotion can be situated within different degrees of a valence scale. For example, the emotion *anger* has higher degree of negativity compared to the emotion *reproach*. Detailed discussion about how emotions are differentiated with varying values for the degree of their positivity or negativity will be presented in the following section. In addition to type (name), valence and degree, emotion theories believe that there is a threshold associated with each emotion which represents the minimum intensity required for that emotion to be active, or in other words, to reach the subject's awareness (Ortony et al., 1990; Scherer, 2001). However, what should be the threshold of a particular emotion from computational perspective is still an unanswered question. Summarising these sentences, an emotion can be described as a valenced reaction to the situation. This reaction can have either negative or positive valence and emotions belonging to the same valenced class can have different degrees of valence within that class. In addition, the intensity of the emotional reaction represents how strongly the valenced reaction is perceived by the subject. As such, in this research, I have assumed that an emotion always has either a positive or negative valence. However, since the degree of valence of such emotional states can be very mild (*i.e.* close to zero), that emotional state may be experienced as neither positive or negative, though still being mildly positive or negative valenced. In addition to these aspects, commonly in emotion modelling literature, the notion of decay time is also evident (Marreiros et al., 2010; Padgham and Taylor, 1996). Decay time denotes the time needed for a particular emotion to reach to the level of 0 (zero) intensity.

Based on the existing literature, I have considered the aspects that are essential to define a data structure of emotion and represented an emotion in EEGS in the form of:

`(Name, Valence, Degree, Threshold, Intensity, DecayTime)`

where, `Name` denotes the name for the type of the emotion, `Valence` specifies whether the emotion is positive or negative, `Degree` represents the extent of the positivity and negativity of the emotion, `Threshold` represents the minimum intensity required to trigger the emotion, `Intensity` represents the strength of the emotional experience and `DecayTime` denotes the time required to drop the emotion intensity back to 0. For example, the emotion structure `(`*`distress`*`, NEGATIVE, -0.8090, 0.0, 0.5, 10)` denotes the emotion of `Name` *`distress`* which has `NEGATIVE Valence` with `Degree` of `-0.8090`, `Threshold` of `0.0`, `Intensity` of `0.5`, and `DecayTime` of `10` seconds. In EEGS, `Valence` is either "POSITIVE" or "NEGATIVE"; `Degree`[6] is a number in the range [-1, +1], where -1 denotes extremely negative emotion and +1 denotes extremely positive emotion; `Threshold` is a number in the range [0, 1); `Intensity`

[6]While the signed value of `Degree` was sufficient to specify the `Valence` as POSITIVE or NEGATIVE, I chose to consider `Valence` as an explicit parameter for the ease of computational mechanism in some conditional checks.

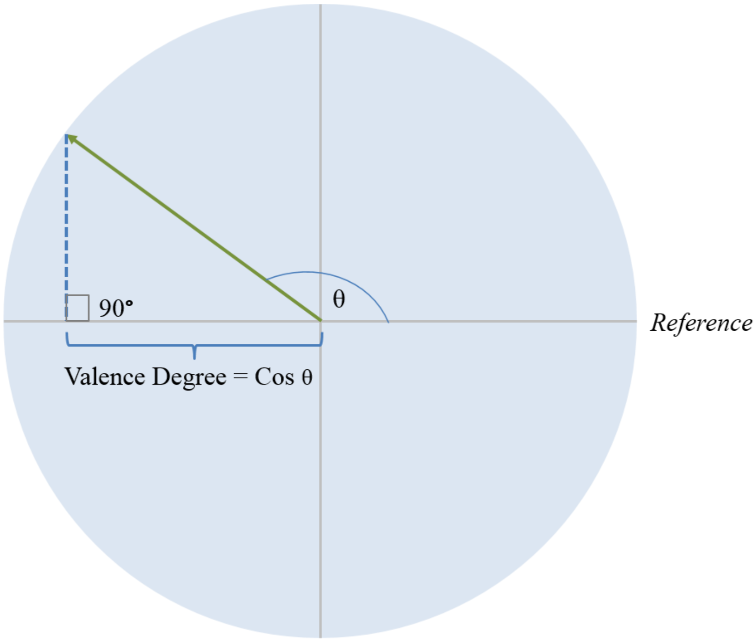


Figure 4.10 Mechanism for mapping the angle of an emotion type into a signed valence degree. θ is the average angle for an emotion type derived from Remington et al. (2000). Horizontal axis in the upper right quadrant is considered as reference for the measurement of angle.

is a number in the range [0, 1] and `DecayTime` is a number which currently has been considered between 0 and 10 seconds. I could not find strong evidence on how long the decay time should be considered for an emotion. However, some emotion models were found to use the decay time of less than 10 seconds (Becker, 2008). It should be noted that the quantities `Name`, `Valence`, `Degree`, `Threshold`, and `DecayTime` in the above defined emotion structure are constants because a particular type of emotion has a fixed name, fixed valence and degree as well as fixed threshold and decay time since these are inherent properties of an emotion. The only thing that changes in the course of interaction is the `Intensity` of a particular emotion. We can use commonsense to determine the `Valence` of an emotion. For example, it is obvious that an emotion of *joy* has POSITIVE `Valence` and an emotion of *distress* has NEGATIVE `Valence`. However, we cannot say with certainty what would be the `Degree` of the `Valence` of the given emotion *i.e.* how much positive or negative is the experience of the emotion quantitatively. In order to provide a viable solution to this issue, in Section 4.7.3, I propose a mechanism to determine the plausible values for the valence of degrees of each emotion based on the work of Remington et al. (2000).

Derivation of Emotion Valence Degree

Remington et al. (2000) present a work that performs a deep reanalysis of the circumplex model of affect (see Chapter 2, Section 2.2.3 for more discussion on circumplex model) and summarises the locations of various affective states in the circumplex based on 10 correlation matrices – 3 matrices from Feldman (1995), 1 matrix from Barrett (1996), 2 matrices from Mayer and Gaschke (1988), 3 matrices from Mayer et al. (1988), and 1 matrix from Rusting and Larsen (1995). Each of these matrices provided an angle for the location of an affective state from the 0°. Since not each matrix provided the location of all the affective states (*i.e.* emotions in the context of this research) considered in EEGS, for some emotions the angles available were less than 10 in number. In order to obtain a more precise angle (location) of an emotion, I calculated the average of the angles provided by various correlation matrices. Since, Remington et al. (2000) only provided the list of angles offered in various correlation matrices, I calculated the corresponding `Valence Degree` using a geometric mapping of the angle to its projection on the horizontal axis as shown in Figure 4.10. It should be noted that the emotion *joy* is considered as the baseline in the correlation matrices provided in Remington et al. (2000). Therefore, the location angle for *joy* is defined to be 0° which leads its `Valence Degree` to be a positive value of 1.0 (as shown in Table 4.5).

Emotion	Angle	Valence Degree
joy	0°	1.0
distress	144°	-0.8090
happy_for	58°	0.5299
sorry_for	122°	-0.5299
appreciation	26°	0.8988
reproach	153.33°	-0.8936
gratitude	8°	0.9903
anger	164.75°	-0.9648
liking	14.5°	0.9681
disliking	165.5°	-0.9681

Table 4.5 Mapping of the angles of the circumplex into valence degree for various emotions.

In Table 4.5, the valence degree of the emotion *anger* is determined by first calculating the angle of projection from baseline. For this, four correlation matrices were considered – Matrix 2 from Mayer and Gaschke (1988) and Matrix 1, 2 and 3 from Mayer et al. (1988). The angles considered were 166°, 171°, 158° and 164° respectively which gave an average angle of 164.75°. The angle provided by Rusting and Larsen

(1995) was excluded because it was an outlier as it deviated significantly from normal range *i.e.* only 83°. Applying cosine to the obtained angle gave a valence degree of -0.9648 for the emotion of *anger*.

From Appraisal to Emotion Intensities

In the previous sections, I discussed how emotions are structured in EEGS and how the `Valence` and `Degree` of various emotion types are determined. In this section, I will present the details of how the values of appraisal variables are mapped into different emotion intensities based on the weights of association identified in Sections 4.7.1 and 4.7.2. As previously discussed, each appraisal variable may affect the intensity of more than one emotions and each emotion may be affected by more than one appraisal variables. Table 4.6 shows how each appraisal variable in EEGS is linked to different emotions. This association has been defined as per the suggestion of Ortony et al. (1990). It should be noted that the weight of association between the appraisal variable and an emotion lies in the range of [-1, 1], where a positive value of the weight indicates that the variable affects the emotion intensity positively and a negative value of weight indicates that the variable affects the emotion intensity negatively. For example, the appraisal variable *desirability* which measures how desirable an event is in relation to its goals, should have positive weight of association with emotion *joy* and a negative weight of association with the emotion *distress*. As such, if the value of *desirability* is positive meaning the event is desirable, this appraisal will increase the intensity of *joy* emotion while decreasing the intensity of *distress* emotion at the same time.

Appraisal Variable	Associated Emotions
desirability	*joy* *distress* *happy_for* *sorry_for* *gratitude* *anger*
praiseworthiness	*appreciation* *reproach* *gratitude* *anger*
appealingness	*liking* *disliking*
deservingness	*happy_for* *sorry_for*
familiarity	*liking*

	disliking
unexpectedness	*appreciation*
	reproach
	gratitude
	anger

Table 4.6 Association of various appraisal variables with different emotions as suggested in the OCC theory (Ortony et al., 1990).

The weights associated with each appraisal–emotion pair contributes in determining the degree by which the appraisal variable affects the intensity of the emotion (Ortony et al., 1990). This implies that the 'effect' of an appraisal variable on an emotion is the function ($\mathcal{I}$) of the quantitative value of the appraisal variable (v_i) and weight of association of the appraisal variable (${}^{v_i}w_{e_j}$) with the emotion (e_j). By saying 'effect' I refer to the contribution an appraisal variable makes to the intensity of a particular emotion because its intensity may also be affected by the values of other appraisal variables, as represented by the equation below.

$$\begin{aligned} \hat{i}_{e_{j_i}} &= \mathcal{I}_e(v_i, {}^{v_i}w_{e_j}) : \; i \in [1,k] \;\&\; j \in [1,l] \\ &= v_i * {}^{v_i}w_{e_j} \end{aligned} \tag{4.17}$$

where, $\hat{i}_{e_{j_i}}$ denotes the contribution of the i^{th} appraisal variable to the intensity of j^{th} emotion. If there are n appraisal variables related to an emotion, then, the final intensity of each emotion ($\hat{i}_{e_j}$) is determined by the cumulative effect of all the appraisal variables linked to the emotion, as shown in Figure 4.7. This phenomenon is represented by (4.18).

$$\hat{i}_{e_j} = \sum_{i=1}^{n} \hat{i}_{e_{j_i}}, \forall\, j \in [1,l] \tag{4.18}$$

These computations are performed for all the emotions, it results in a set of emotions $\mathbf{E} = \{e_1, \ldots, e_l\}$ with respective intensities $\mathbf{I} = \{\hat{i}_{e_1}, \ldots, \hat{i}_{e_l}\}$. Hence, the appraisal–emotion network presented in Figure 4.7 helps in the computation of the intensities of various emotions of the model based on the Equations 4.17 and 4.18. However, it should be noted that not all emotions exhibit a linear combination of product of appraisal values and corresponding weight of association of appraisal variable with the emotion. The computation of the final intensity of some of the emotions like *appreciation*, *reproach*, *gratitude*, *anger*, *liking* and *disliking* may follow non-linear combination as will be discussed in the following sub-sections.

joy

As suggested by Ortony et al. (1990), the emotion *joy* is determined by only one appraisal variable – *desirability*. Therefore the intensity of *joy* emotion is determined by the value of appraisal variable *desirability* and the weight of association of the variable with *joy* emotion.

$$\hat{i}_{joy} = desi * ({}^{desi}w_{joy}) \tag{4.19}$$

distress

Like *joy*, the emotion *distress* is also affected by only one appraisal variable *desirability* (Ortony et al., 1990). Therefore, the intensity of *distress* emotion is also given by the formula similar to that of *joy* emotion.

$$\hat{i}_{dist} = desi * ({}^{desi}w_{dist}) \tag{4.20}$$

happy_for

The emotion *happy_for* denotes a feeling of happiness because something desirable happened to someone and the person deserved what happened. As such, the emotion *happy_for* is determined by the appraisal variables *desirability* and *deservingness* (Ortony et al., 1990).

$$\hat{i}_{hpy\_for} = desi * ({}^{desi}w_{hpy\_for}) + dese * ({}^{dese}w_{hpy\_for}) \tag{4.21}$$

sorry_for

The emotion *sorry_for* denotes a feeling of sadness because something undesirable happened to someone and the person did not deserve what happened. As such, the emotion *sorry_for* is also determined by the appraisal variables *desirability* and *deservingness* (Ortony et al., 1990).

$$\hat{i}_{sry\_for} = desi * ({}^{desi}w_{sry\_for}) + dese * ({}^{dese}w_{sry\_for}) \tag{4.22}$$

appreciation

Appreciation is the feeling one experiences when someone does a praiseworthy action from the viewpoint of standards of the assessing person and that action was not expected to happen (Ortony et al., 1990). As such, the emotion of *appreciation* in EEGS, is determined by two appraisal variables – *praiseworthiness* and *unexpectedness*. If an action of other agent is praiseworthy, then the appraising individual experiences some degree of appreciation. However, how much the action was unexpected largely affects the degree of praiseworthiness as well thereby defining the intensity of appreciation experienced. For this reason, I propose to combine the contributions of the appraisal variables *praiseworthiness* and *unexpectedness* in a non-linear fashion instead of linearly combining their individual contributions to the intensity of *appreciation*. If we denote the contribution of the appraisal variable *praiseworthiness* to emotion *appreciation* as $\hat{i}_{appr_{prai}} = prai * {}^{prai}w_{appr}$ and the contribution of the appraisal variable *unexpectedness* as $\hat{i}_{appr_{unex}} = unex * {}^{unex}w_{appr}$, then the overall intensity of the emotion *appreciation* ($\hat{i}_{appr}$) is given by the following formula.

$$\hat{i}_{appr} = \begin{cases} -\left|\hat{i}_{appr_{prai}}\right|^{1-\hat{i}_{appr_{unex}}} & \text{if } \hat{i}_{appr_{prai}} < 0 \\ \left(\hat{i}_{appr_{prai}}\right)^{1-\hat{i}_{appr_{unex}}} & \text{if } \hat{i}_{appr_{prai}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.23}$$

reproach

The feeling of reproach arises when a person disapproves someone's blameworthy (not praiseworthy) action (Ortony et al., 1990). The degree of reproach is also affected by unexpectedness of the event in addition to the degree of blameworthiness. As such, the emotion *reproach* in EEGS is determined by two appraisal variables – *praiseworthiness* and *unexpectedness* (as in the case of *appreciation* emotion). If we denote the contribution of appraisal variable *praiseworthiness* to emotion *reproach* as $\hat{i}_{repr_{prai}} = prai * {}^{prai}w_{repr}$ and the contribution of the appraisal variable *unexpectedness* as $\hat{i}_{repr_{unex}} = unex * {}^{unex}w_{repr}$, then the overall intensity of the emotion *reproach* ($\hat{i}_{repr}$) is given by the following formula.

$$\hat{i}_{repr} = \begin{cases} -\left|\hat{i}_{repr_{prai}}\right|^{1-\hat{i}_{repr_{unex}}} & \text{if } \hat{i}_{repr_{prai}} < 0 \\ \left(\hat{i}_{repr_{prai}}\right)^{1-\hat{i}_{repr_{unex}}} & \text{if } \hat{i}_{repr_{prai}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.24}$$

gratitude

Gratitude is a feeling one experiences in response to an unexpected praiseworthy action that is desirable for the achievement of ones goal(s) (Ortony et al., 1990). Therefore, the emotion *gratitude* in EEGS is determined by the appraisal variables *desirability*, *praiseworthiness*, and *unexpectedness*. If we denote the contribution of appraisal variable *desirability* to emotion *gratitude* as $\hat{i}_{grat_{desi}} = desi * {}^{desi}w_{grat}$, contribution of appraisal variable *praiseworthiness* as $\hat{i}_{grat_{prai}} = prai * {}^{prai}w_{grat}$ and the contribution of the appraisal variable *unexpectedness* as $\hat{i}_{grat_{unex}} = unex * {}^{unex}w_{grat}$, then the overall intensity of the emotion *gratitude* ($\hat{i}_{grat}$) is given by the following formula.

$$\hat{i}_{grat} = \hat{i}_{grat_{desi}} + \begin{cases} -\left|\hat{i}_{grat_{prai}}\right|^{1-\hat{i}_{grat_{unex}}} & \text{if } \hat{i}_{grat_{prai}} < 0 \\ \left(\hat{i}_{grat_{prai}}\right)^{1-\hat{i}_{grat_{unex}}} & \text{if } \hat{i}_{grat_{prai}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.25}$$

anger

Anger is a feeling one experiences in response to an unexpected blameworthy action that is undesirable for the achievement of ones goal(s) Ortony et al. (1990). Therefore, the emotion *anger* in EEGS is determined by the appraisal variables *desirability*, *praiseworthiness*, and *unexpectedness*. If we denote the contribution of appraisal variable *desirability* to emotion *anger* as $\hat{i}_{angr_{desi}} = desi * {}^{desi}w_{angr}$, contribution of appraisal variable *praiseworthiness* as $\hat{i}_{angr_{prai}} = prai * {}^{prai}w_{angr}$ and the contribution of the appraisal variable *unexpectedness* as $\hat{i}_{angr_{unex}} = unex * {}^{unex}w_{angr}$, then the overall intensity of the emotion *gratitude* ($\hat{i}_{angr}$) is given by the following formula.

$$\hat{i}_{angr} = \hat{i}_{angr_{desi}} + \begin{cases} -\left|\hat{i}_{angr_{prai}}\right|^{1-\hat{i}_{angr_{unex}}} & \text{if } \hat{i}_{angr_{prai}} < 0 \\ \left(\hat{i}_{angr_{prai}}\right)^{1-\hat{i}_{angr_{unex}}} & \text{if } \hat{i}_{angr_{prai}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.26}$$

liking

Liking someone depends on how much an individual thinks the person is appealing (Ortony et al., 1990). This feeling is also affected by the degree of familiarity between the two individuals (Ortony et al., 1990). In line with this, in EEGS, the intensity of the emotion *liking* is determined by the appraisal variables *appealingness* and *familiarity*. If we denote the contribution of appraisal variable *appealingness* to emotion *liking* as $\hat{i}_{lkng_{appl}} = appl * {}^{appl}w_{lkng}$ and the contribution of appraisal variable *familiarity* as $\hat{i}_{lkng_{fami}} = fami * {}^{fami}w_{lkng}$, then the overall intensity of the emotion *liking* ($\hat{i}_{lkng}$) is given by the following formula.

$$\hat{i}_{lkng} = \begin{cases} -\left|\hat{i}_{lkng_{appl}}\right|^{\hat{i}_{lkng_{fami}}} & \text{if } \hat{i}_{lkng_{appl}} < 0 \\ \left(\hat{i}_{lkng_{appl}}\right)^{\hat{i}_{lkng_{fami}}} & \text{if } \hat{i}_{lkng_{appl}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.27}$$

disliking

Similar to liking, disliking someone also depends on how much an individual thinks the person is appealing (Ortony et al., 1990). This feeling is also affected by the degree of familiarity between the two individuals (Ortony et al., 1990). As such, in EEGS, the intensity of the emotion *disliking* is determined by the appraisal variables *appealingness* and *familiarity*. If we denote the contribution of appraisal variable *appealingness* to emotion *disliking* as $\hat{i}_{dlkg_{appl}} = appl * {}^{appl}w_{dlkg}$ and the contribution of appraisal variable *familiarity* as $\hat{i}_{dlkg_{fami}} = fami * {}^{fami}w_{dlkg}$, then the overall intensity of the emotion *disliking* ($\hat{i}_{dlkg}$) is given by the following formula.

$$\hat{i}_{dlkg} = \begin{cases} -\left|\hat{i}_{dlkg_{appl}}\right|^{\hat{i}_{dlkg_{fami}}} & \text{if } \hat{i}_{dlkg_{appl}} < 0 \\ \left(\hat{i}_{dlkg_{appl}}\right)^{\hat{i}_{dlkg_{fami}}} & \text{if } \hat{i}_{dlkg_{appl}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.28}$$

It should be noted that although the formulas presented above are based on the theoretical suggestions of (Ortony et al., 1990) some assumptions were made to define the proposed formulas. Eventhough some researchers have made some suggestions for the computation of very few emotions (El-Nasr et al., 2000; Gratch and Marsella, 2004b), it is not possible to directly compare the accuracy of the currently proposed formulas with the previous proposals because the benchmark tests for previous models is not available. The formulas presented above exhibit a high level of accuracy in predicting emotion intensities collected from human participants as will be discussed in Chapter 5, which will also set a new benchmark for such evaluations for future researchers.

Emotion Intensity Threshold

As discussed earlier in Section 4.7.3, an emotion can have a threshold that specifies a minimum level of activation required for an emotion to be actually present (Ortony et al., 1990; Scherer, 2001). OCC theory also introduces a concept called *emotion–potential*. Emotion potential is the measure of the extent to which an event can trigger a particular emotion at the given time (Ortony et al., 1990). If emotion–potential of a particular emotion in reaction to an event is below the *emotion–threshold* of that emotion, then the emotion is assumed not to become active at all (Ortony et al., 1990). As such, Ortony et al. (1990) suggest to compute *emotion–potential* ($\hat{i}_e$) as a function

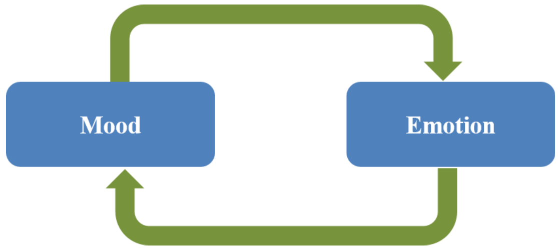


Figure 4.11 Cyclic interaction between emotion and mood.

of appraisal variables. This makes the emotion intensities computed in above sections as emotion–potential. OCC theory suggests to subtract the *emotion–threshold* ($thres_e$) of that emotion from the computed emotion–potential in order to obtain the effective *emotion–intensity* ($\hat{i}_e^{effective}$). As such the effective intensity for an emotion as suggested by Ortony et al. (1990) would be given as below.

$$\hat{i}_e^{effective} = \hat{i}_e - thres_e \tag{4.29}$$

Interestingly Ortony et al. (1990) consider that the value $thres_e$ is a function of time meaning the activation threshold of same emotion may vary with time. However, the authors do not specify how such a variation occurs. Also, the lack of well defined thresholds for various emotions in the first place makes it difficult to decide what should be the activation threshold of each emotion – if any. Unlike some models that try to realise the notion of threshold for various emotions (Becker, 2008; Dias et al., 2014; Dias and Paiva, 2005; Velásquez and Maes, 1997), I consider any emotion with emotion–potential above the value of zero to be active for the moment and consider the same value as the instantaneous intensity of that emotion. As such, in EEGS, the values of the appraisal variables directly determine the intensities of emotions depending on the weight of association by considering the activation threshold to be zero. However, I provide the flexibility of using non-zero intensity threshold in emotions in EEGS to allow future research and investigation by other researchers.

Revisiting Interaction among Emotion, Mood and Personality[7]

In Sections 4.7.1 and 4.7.2, I presented how the factors of personality and mood take part in the mapping of appraisal variables into emotion intensities. In Figure 2.9, I presented an approach called *mediation* where personality affects emotion (R_{p-e}), mood affects emotion (R_{m-e}) and mood is also affected by personality (R_{p-m}). Researchers consider mood as accumulated effect of multiple emotional episodes (Beedie et al., 2005; Ekman,

[7]Most of the content in this section is adapted from my published work (Ojha et al., 2018a)

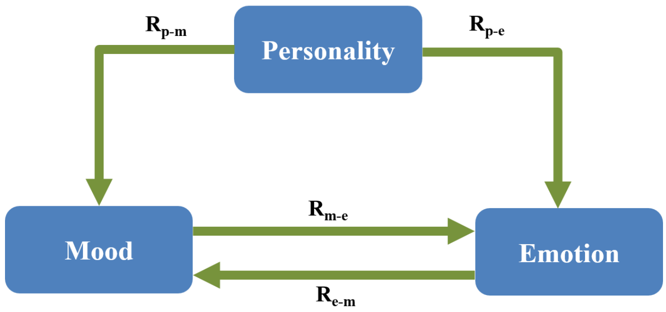


Figure 4.12 Proposed dynamic interaction between emotion, mood and personality. Adapted from Ojha et al. (2018a).

1994; Parkinson et al., 1996). Moreover, mood is also said to have significant effect on emotion intensities (Morris, 1992; Neumann et al., 2001). As previously mentioned (see Chapter 2, Section 2.4), these facts suggest a cyclic interaction between emotion and mood as opposed to one-way suggested by Rusting (1998). Therefore, I employ such cyclic interaction between emotion and mood as represented by Figure 4.11. Combining the interactions in Figures 2.9 and 4.11, I suggest the dynamic interaction among emotion, mood and personality as shown in Figure 4.12. A new relationship (R_{e-m} in addition to the ones offered by Rusting (1998) can be seen in Figure 4.12.

Since personality factors do not change significantly with time (Costa and McCrae, 1988; Dweck, 2008), the substantial effect of personality factors on mood (R_{p-m}) only occurs at the initialisation of the model *i.e.* the initial mood state of EEGS is determined by the personality factors, which is functionally represented in Equation 4.30. If we denote the personality dimensions of the model by $\mathbf{P}_d \in \{O, C, E, A, N\}$, the five personality dimensions of the model, as described in Five Factor Model of personality (Costa and McCrae, 1992; Digman, 1990), and mood state by M, the initial mood state of EEGS ($M^{initial}$) is given by the function $\mathcal{M}$ below.

$$\begin{aligned} M^{initial} &= \mathcal{M}(\mathbf{P_d}) \\ &= \mathcal{M}(O, C, E, A, N) \end{aligned} \tag{4.30}$$

The initial mood as obtained from Equation 4.30 not only takes part in mapping the appraisal variables into emotion intensities (as discussed in Sections 4.7.1 and 4.7.2), but also directly modulates the emotion intensities. Let us denote an emotion as e and its intensity as $\hat{i}_e$, mood state as M (as previously mentioned) and extent to which a mood affects emotion as $\mathbf{M}_c$, say *mood compensation*, which is a fraction of the current mood state by which the emotion intensities are influenced (see Equation 4.32).

Now, the resulting emotion intensity after the compensating effect of mood is given by Equation 4.31.

$$\tilde{\hat{i}}_e = \begin{cases} \hat{i}_e + |\mathbf{M}_c| & \text{if } Sign(e) = Sign(M) \\ \hat{i}_e - |\mathbf{M}_c| & \text{if } Sign(e) \neq Sign(M) \end{cases} \tag{4.31}$$

The value of mood compensation ($\mathbf{M}_c$) used in Equation 4.31 is given by the following formula.

$$\mathbf{M}_c = \alpha\, M \tag{4.32}$$

The ration behind the formula in Equation 4.31 is that positive mood tends to increase the intensity of positive emotions and negative mood tends to increase the intensity of negative emotions. Moreover, positive mood tends to decrease the intensity of negative emotions and negative mood tends to decrease the intensity of positive emotions (Morris, 1992; Neumann et al., 2001).

The parameter α, in Equation 4.32 should be defined as per the system need and application of the modelled system. In EEGS, I used the value of α as 10% (i.e. 0.1) because this value provided the most plausible result. It is not always mandatory to use 10% scaling factor while modelling this phenomenon. Yet, other related work also commonly demonstrate similar scaling (see, for example the work of Marinier III and Laird (2007)).

Equation 4.31 only shows how the mood state affects the emotion intensities. Two-way (cyclic) interaction of mood and emotion in EEGS (as discussed earlier – see Figure 4.11) should also account for the effect of the resulting emotional state on the mood. Majority of the literature has defined mood as the aggregate effect of numerous continuous emotional experiences (Morris, 1992; Parkinson et al., 1996). Positive emotional experience tends to shift mood towards positive scale and negative emotional experiences tend to shift mood towards negative scale. EEGS follows this notion of mood in order to model the effect of emotional experience on the subsequent mood. First of all, the impact of the emotional experience on the simulated agent (denoted as Im), which is a quantification of the overall influence of the emotional situation, is considered for modelling this relation. Then, an aggregate value is calculated using intensities of all the emotions that are congruent to the impact caused *i.e.* if the impact is positive, the intensities of positively valenced emotion are considered and if the impact is negative, intensities of negatively valenced emotions are considered for the calculation of aggregate intensity. Given n emotions whose valance is congruent to the impact, then, aggregate intensity ($\hat{i}_{agg}$) can be computed as in Equation (4.33).

$$\hat{i}_{agg} = \begin{cases} \sum_{j=1}^{n} \hat{i}_{e_j} & \text{if } Im > 0 \\ -\sum_{j=1}^{n} \hat{i}_{e_j} & \text{Otherwise} \end{cases} \tag{4.33}$$

Once the aggregate intensity is calculated, it is converted to mood factor (M_f) by passing through a modified Logistic function. M_f is a value in the range [-1, +1] given by Equation 4.34.

$$M_f = \frac{2}{1 + e^{-\hat{i}_{agg}}} - 1 \tag{4.34}$$

The value in the first part of Equation (4.34) is subtracted by 1 to shift the curve 1 step down so that the minimum output value of the function becomes -1 i.e. minimum value of the mood factor. Likewise, making the maximum value in the equation as 2 allows the resulting maximum value to be 1 since the curve is shifted one step down. This results in the value of M_f to lie in the range [-1, +1].

When the mood factor is calculated, new mood is given by the formula in (4.35).

$$M' = M + \beta \, M_f \tag{4.35}$$

The quantity β in Equation (4.35) was chosen to be 10% in EEGS. New mood state (M') computed by above equation affects the emotion intensities in subsequent emotional experience, thereby maintaining a cyclic interaction, as shown in Figure 4.11.

The discussion so far in this section, demonstrates how EEGS implements the dynamic interaction among emotion, mood and personality.

Emotion Decay

Since the experience of emotion is believed to be instantaneous and short-lived (Ekman, 1994; Forgas, 1992; Mayer et al., 1992; Rosenberg, 1998), the emotions should decay over time (Hudlicka, 2016; Picard, 1997). As such, researchers have proposed various mechanisms to model the decay of emotion intensities in their computational realisations. Most common emotion decay mechanisms proposed in the literature can be categorised as (i) *Linear Decay*, (ii) *Exponential Decay*, (iii) *Logarithmic Decay*, and (iv) *Tan-Hyperbolic Decay* as summarised in Table 4.7.

Function	Rationale	Suggested by
Linear	Emotion decays in a constant rate over the period	Becker (2008); Egges et al. (2004); El-Nasr et al. (2000); Gebhard (2005); Gebhard et al. (2003); Marinier III and Laird (2007); Thagard and Nerb (2002)
Exponential	Emotion intensity stays strong just after its experience and quickly decays after a short while	Becker (2008); Dias and Paiva (2005); Gebhard et al. (2003)
Logarithmic	Emotion starts to decay rapidly as soon as it is experienced	Hudlicka (2016)
Tan-Hyperbolic	Emotion intensity stays strong just after its experience and quickly decays after a short while and tends to stay stable thenafter	Gebhard et al. (2003)

Table 4.7 A summary of different emotion decay mechanisms used in various computational models of emotion.

Majority of the emotion modelling proposals seem to have adopted linear decay function (Egges et al., 2004; El-Nasr et al., 2000; Marinier III and Laird, 2007). For example, El-Nasr et al. (2000) uses a decay function of the form $\hat{i}_e(t+1) = \phi.\hat{i}_e(t)$ to regulate the dynamics of positive emotions, where, $\hat{i}_e(t+1)$ represents the intensity of an emotion at time $t+1$, $\hat{i}_e(t)$ represents the intensity of the emotion at time t, and the quantity ϕ represents the decay constant that determines the slope of the decay function. For the decay of negative emotions, El-Nasr et al. (2000) uses a different decay constant δ, where $\phi < \delta$. The rationale behind this choice is that positive emotions decay faster than negative emotions (El-Nasr et al., 2000). Using trial and error, they suggest "that there was a range of settings [for these constants] that produced a reasonable behaviour for the agent" (El-Nasr et al., 2000, p. 236). The suggested range was $0.1 < \phi < 0.3$ and $0.4 < \delta < 0.5$. Similarly, Becker (2008) uses a linear decay of the form $\hat{i}_e(t+1) = 1 - \frac{\hat{i}_e(t)}{10}$, where, $\hat{i}_e(t+1)$ is the intensity of emotion at time $t+1$, $\hat{i}_e(t)$ is the intensity of the emotion at time t, 10 denotes a decay time of 10 seconds.

Becker (2008) also proposed exponential decay for some of the emotions. The decay function was defined as $\hat{i}_e(t+1) = e^{-\hat{i}_e(t)}$. Dias and Paiva (2005) are also the proponents of exponential decay of emotion intensities. They define the emotion decay function in their model in a manner similar to that of Becker (2008) in the form of $\hat{i}_e(t+1) = \hat{i}_e(t) * e^{-b.t}$, where $e^{-b.t}$ is an exponential function of time t and b is the constant that determines the decay rate. In addition to linear and exponential

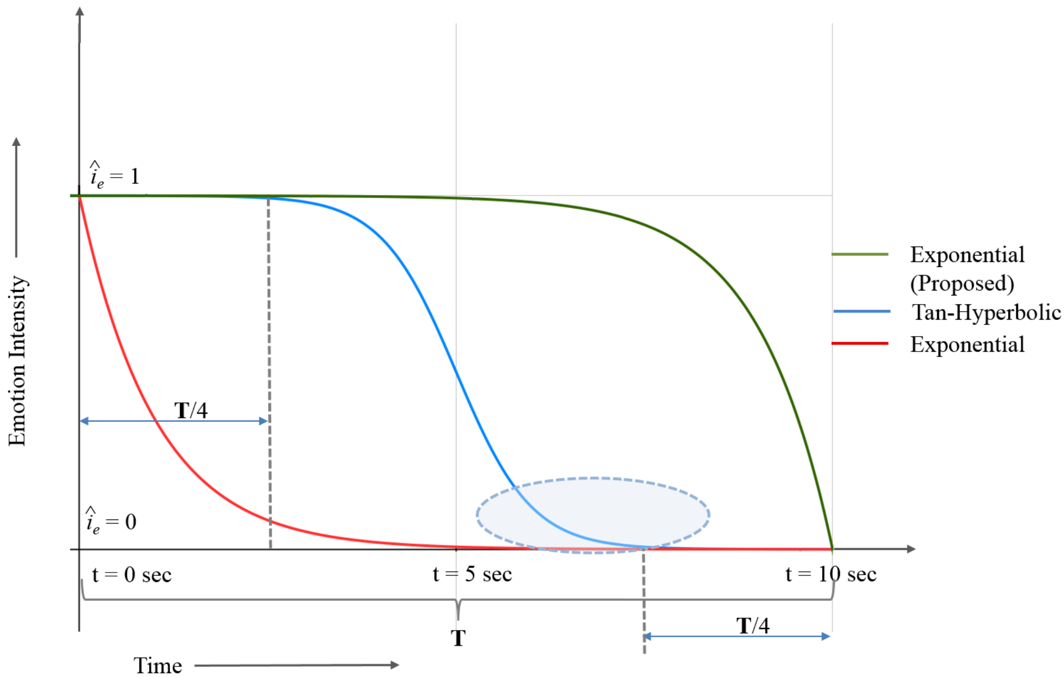


Figure 4.13 Comparison of different emotion decay functions.

decay functions, Gebhard et al. (2003) also presents a Tan-Hyperbolic function to model the decay of emotion intensities. However, the authors do not explain the internal details of the function used. I anticipate that they used a function of the form $\hat{i}_e(t+1) = -tanh(\hat{i}_e(t) - T)$, where, T is the total decay time for an emotion. Hudlicka (2016) describes a possibility of a Logarithmic decay function of emotions, but does not present details necessary to understand how this function might look like and how plausible the decay caused by the function can be.

Since linear decay functions are probable to be less accurate to reflect the true mechanism of emotion decay in humans (Dias and Paiva, 2005; Hudlicka, 2016), I opted for not implementing the linear decay function in EEGS. Logarithmic and Tan-Hyperbolic functions are less reliable because of lack of sufficient number of empirical evaluation of emotion dynamics based on these decay approaches. Therefore, I opted for exponential function to model the decay of emotions in EEGS. However, I use a slightly different approach compared to the functions proposed in existing computational models – as exhibited by green curve in Figure 4.13. This is because the emotion dynamics exhibited by commonly proposed exponential decay functions (as shown by red curve in Figure 4.13) is questionable. If one closely examines the nature of the red curve in Figure 4.13, it can be noticed that the function leads to a rapid drop of emotion intensity just after it is felt (say at time, t = 0). As a result if the total decay time of an emotion is **T** seconds[8], almost all the emotion intensity drops down within one-fourth of the total decay time (as can be seen in Figure 4.13, where red curve almost falls to zero within 2.5 seconds of the triggering of the emotion, if **T**=10 sec). This clearly suggests that the normal exponential curve of the form e^{-t} is not suitable for emotion decay functions. As compared to normal exponential curve, a tan-hyperbolic curve seems to be more promising to achieve a plausible decay of emotion intensities. Such a curve is represented by blue line in Figure 4.13. Unlike normal exponential curve, tan-hyperbolic curve does not immediately start to rapidly decay the intensity. Rather, the intensity remains stable for a short period of time (as reasonably expected to happen in humans) and then only starts to decay with about 50% drop in the intensity within **T**/2. However, in the second part of the decay, tan-hyperbolic curve inherits the limitation of the regular exponential function *i.e.* remaining intensity drops rapidly almost within half of the remaining decay time (as shown by the shaded region in Figure 4.13). But, (i) an emotion intensity should not start decaying immediately after it is triggered. Instead, it should stay stable for a short while and then only start to fall in exponential manner. Moreover, once the emotion intensity starts to fall, (ii) the rate of decrease in intensity should keep on increasing signifying that the effect of that emotion diminishes in increasing rate. In order to address these criteria, I propose a *modified exponential emotion decay function* as defined in Equation (4.36).

$$\hat{i}_e(t+1) = \hat{i}_e(t) * \left(1 - \frac{e^t}{e^T}\right) \tag{4.36}$$

Where,
$\hat{i}_e(t)$ is the emotion intensity at time t,
$\hat{i}_e(t+1)$ is the emotion intensity at time $t+1$,
T is the total duration (in seconds) for the decay of an emotion intensity, and
t is a point in time at which an emotion intensity is calculated.

[8]Different emotions may have different length of decay (Picard, 1997). Here, I present a generic length of **T** seconds and leave the flexibility of choosing a decay time for an emotion intensity as empirical evidence may be available from researchers in the future.

4.8 Affect Regulation[9]

As discussed in Chapter 2, Section 2.5.2, Gross (1998a,b) proposes two broad classes of emotion regulation namely (1) *Antecedent-focused* emotion regulation and (2) *Response-focused* emotion regulation. However, the limitation of current implications of response-focused emotion regulation offered by Gross and Thompson (2007) is that their approach is heavily *non-cognitive*. For example, they suggest that "drugs may be used [as a regulatory device] to target physiological responses such as muscle tension" (Gross and Thompson, 2007, p. 15). They also suggest exercise (Thayer et al., 1994) and relaxation (Suinn and Richardson, 1971) can regulate physiological and experiential aspects of emotions. Likewise, alcohol (Hull and Bond, 1986) and cigarettes (Brandon, 1994) are also widely studied to modify the experience of emotion (Gross, 1998b). Additionally, Gross (1998b) argues that *alteration of the expressions* is the most common form of emotion regulation. Although these regulatory mechanisms may be suitable for a human being to mainly calm down negative emotional experiences, these approaches are neither well defined nor appropriate to be realised in a computational model of emotion. In the following sections, I will offer my response-focused emotion regulation mechanism that is based on higher cognitive layer of ethical reasoning and discuss its strengths compared to other emotion convergence and regulation mechanisms used in existing computational models of emotion. But, a key question is – how can we implement such a regulatory mechanism in autonomous agents that allows them to exhibit socially acceptable emotional and behavioural responses? I will explore this question in the following section.

4.8.1 Emotion Convergence in Computational Models

Literature suggests that existing computational models of emotion commonly adopt two approaches to achieve the convergence to a stable emotional state when multiple emotions are activated by the cognitive appraisal process – either choosing the emotion with (i) *Highest Intensity* (Gratch and Marsella, 2004a), or obtaining a (ii) *Blended Intensity* from the intensities of all the active emotions (Marinier III and Laird, 2007; Reilly, 2006). EMA (Gratch and Marsella, 2004a) uses the approach of selecting the emotion with highest intensity to determine the final emotional state of the model. A clear disadvantage of the highest intensity approach is that the intensities of several

[9]Most of the discussion in this section has been adapted from Ojha et al. (2018b) and Ojha et al. (2017). It should be noted that although the term 'affect regulation' may be used by researchers to denote broader range of processes including emotion regulation, mood regulation, etc., current research is concerned only in the process of emotion regulation.

emotions may not significantly differ. Therefore, the higher intensity of an emotion may not be the only sufficient criteria for a more effective choice in that specific situation. For example, consider a situation where *joy* resulted a 0.9 intensity level, whereas distress achieved a 0.85 intensity level. By using the highest intensity approach the emotion *distress* would be completely disregarded and a final emotional state of *joy* with intensity of 0.9 is considered to be operational. Now, consider that the individual experiencing such emotion intensities observed a foe getting fired by their boss. Although the individual may experience slightly higher *joy* for that happening, it is still convenient to regulate and suppress the emotion of *joy* in favour of *distress*, so to prevent an awkward situation at work. However, if we consider a situation where the individual observed a friendly colleague getting a promotion from their boss, this individual may experience similar level of *joy* and *distress* if that individual was also hoping for that promotion, but still regulating the emotions so to manifest *joy* for the colleague's success. From these examples it is clear to see how selecting the emotion with highest intensity without considering other aspects of the situation may lead to poor decisions. In line with this, Reilly (2006) argues that considering only the emotion with highest intensity causes high degree of inaccuracy in emotion processing mechanism. To address this limitation, he suggests an approach that helps in the blending of all the elicited emotions that are congruent to the situation (see Equation 4.37). Proponents of the emotion blending approach put forward by Reilly (2006) have followed similar approach in computational models of mood and feelings (Marinier III and Laird, 2007). Although, the approach of blending the emotion intensities proposed by Reilly (2006) seems to a viable way to overcome the highest intensity approach and it is able to consider the contributions of all the constituent emotions, this approach largely fails to attribute the final emotional state to a defined emotion type. In other words, with this approach it is difficult for the individual, or agent, to describe the experienced emotional state with a specific emotion label.

$$\hat{\imath} = 0.1 * \log_2 \sum_{n=1}^{N} 2^{10*\hat{\imath}_n} \tag{4.37}$$

where,
$\hat{\imath}$ is the resulting intensity,
N is the number of emotions for which intensity is to be combined,
$\hat{\imath}_n$ is the intensity of n^{th} emotion, where, $1 \leq n \leq N$.

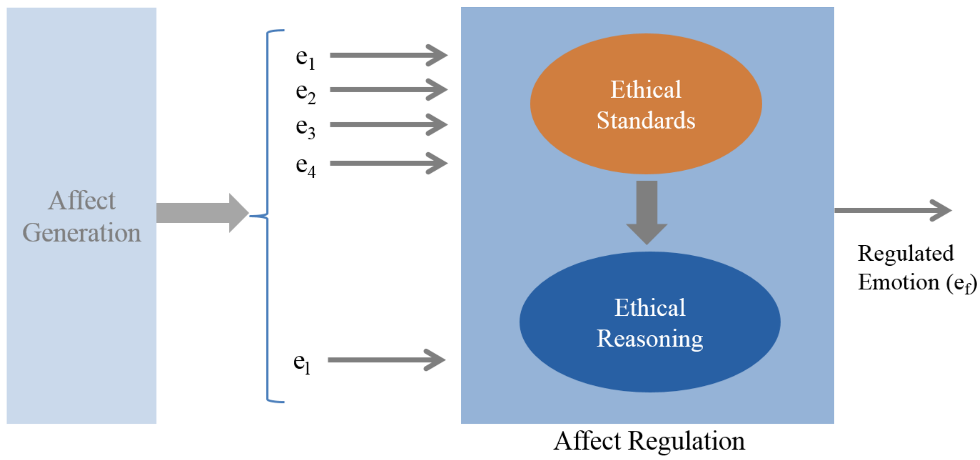


Figure 4.14 Process of affect regulation in EEGS where conflicting emotional states are converged to a final stable and regulated emotional state based on ethical reasoning guided by ethical standards. Redrawn after Ojha et al. (2017).

In this dissertation, I propose to address the previously presented limitations by introducing a regulatory *ethical reasoning* mechanism for the selection of a socially appropriate emotional state.

4.8.2 Ethical Reasoning for Emotion Regulation in EEGS

In Section 4.8.1, I argued that although the *highest intensity* and *blended intensity* approaches can help in converging to a single emotional state and viably assist to regulating the agent's emotions so to lead to a more believable emotional responses, they do not offer effective emotion regulation mechanism to ensure the social acceptance of the emotional responses. In this dissertation, I argue that an ethical reasoning mechanism can assist to converge to a single emotional state and ensure that such emotional state is socially acceptable. Here I remind that by socially acceptable emotional response I mean the ones that are in compliance with common human expectations or norms in given social situations.

Ethical reasoning in EEGS is guided by its *ethical standards*. The ethical standards are modeled as agent's beliefs. Each of these beliefs assesses the degree of approval or disapproval of a specific action directed by a source subject toward a target subject. Although Ethical standards can be generalized to accommodate any aspect of the interaction between two subjects (i.e. any action), in this section, my discussion will revolve around the use of such ethical standards for the purpose of emotion generation and expression. Therefore, the action considered in each standard will describe the expression of a specific emotion. When EEGS runs for the first time, it starts with empty standards (as discussed previously in Section 4.6.1) *i.e.* it does not have any pre-defined standard. Thus, when a person first interacts with EEGS, it establishes an initial neutral standard that guides in its emotion generation process. Ethical standards can pertain to any aspect of interaction between two persons or between an autonomous agent and a person. However, in this section, my discussion will revolve around the ethical standards in the context of emotion generation and its expression. Thus when a person first interacts with the autonomous agent running EEGS system, the agent builds a set of standards that affect the emotion processing mechanism. Suppose a stranger interacts with the agent. As stated earlier, the agent builds a set of neutral standard. Examples of the agent's standards can be – "I should not show anger to him", "I should express joy in interacting with him" and so on. This can be considered as what the agent believes it is supposed to do or not to do. This belief can have a certain degree depending on who the person is or what is the interaction history of the agent with the person. In other words, whether the internal standard of a robot approves the expression of an emotion to a target also has a degree associated with the approval or disapproval.

Each ethical standard in EEGS is represented as a data structure in the form `(Emotion, Source, Target, Approval)`, where, `Emotion`, in this specific application domain, represents the emotion addressed by the standard, `Source` represents the one that expresses the emotion, `Target` represents the target of the emotion expression. It is important to note that when the standard refers to an emotion the agent should or should not express toward someone, `Source` assumes the value SELF. However, `Source` can also assume other values to refer to other agents or individuals. By doing so, the standards describes what the agent believes about how one person (the `Source`) should behave with another person (the `Target`). Providing to the data structure such flexibility is a viable way to easily accommodate standards for multi-agent interaction applications, which, however, are out of scope for the present work. Moreover, `Approval` denotes whether the expression of emotion is preferred or not and what is the degree of this preference. Approval is further structured as `(Preference, ApprovalDegree)` (see Section 4.6.1), where `Preference` specifies whether the expression of emotion is preferred or not and `ApprovalDegree` denotes the extent to which the expression of emotion is preferred or not. For example, the standard `(`*anger*`, SELF, JOHN, (NO, 0.75))` represents "I should NOt express *anger* to JOHN" from the agent's perspective and the degree of this belief is 0.75. Similarly, `(`*anger*`, PAUL, DAVID, (YES, 0.9))` represents "It is okay (YES) for PAUL to express *anger* to DAVID" and the degree of this belief is 0.9.

As addressed in Section 4.6.1, it should be noted that the notion of standards in EEGS is not static quantity. Even though the autonomous agent starts the interaction with neutral standards, the standards change in the course of interaction depending on how the person interacts with the agent. Recall the example of a standard in the previous paragraph – `(`*`anger`*`, SELF, JOHN, (NO, 0.75))`. As per the standard, the agent (SELF) is not supposed to express anger towards JOHN. However, if JOHN constantly misbehaves with the agent, the standards adapt to the new situation by becoming more negative and ultimately leading to the agent's belief that it is fine to express anger towards JOHN. As I will show with the evaluations provided in Chapter 5, using ethical standards enables EEGS to converge to more socially acceptable emotional states, thus regulating the agent's emotions by ensuring more conscious and ethical emotional responses. Table 4.8 shows some examples of the standards in the memory of EEGS related to the emotion *anger*.

Emotion	Source	Target	Preference	Degree
anger	SELF	JOHN	NO	0.8
anger	PAUL	JOHN	YES	0.25
anger	DAVID	JOHN	NO	0.5

Table 4.8 An example of a set of ethical standards for *anger* emotion. Adapted from Ojha et al. (2018b).

4.8.3 Reasoning Mechanism in EEGS

In the current implementation, EEGS is able to generate ten different emotions in response to an event. In a particular situation, one or more emotions might be triggered in reaction to the event (Ortony et al., 1990). An autonomous agent must be able to converge to a final emotional state which can then be expressed through a behavioural responses congruent to the appraised emotional situation. In Section 4.8.2, I discussed how such emotional response should not only be believable, but also socially acceptable. To address the gaps in existing models of emotion generation, I provided a data structure able to store ethical standards. This data structure is the foundation for the modeling of a higher cognitive layer of ethical reasoning in EEGS (Ojha et al., 2017).In this dissertation, I insist that when there are multiple emotions triggered by an event at the same time by the appraisal of an event (Ortony et al., 1990), an ethical reasoning process can assist an emotion generation model to converge to a stable emotional state that is not only plausible for the given situation, but also socially acceptable. In the remainder of this section I will present the computational details to model such ethical reasoning process in EEGS.

I introduce the term *Coefficient of Standard* (CoS), which is the measure of positive significance of all the standards related to an emotion being considered. This coefficient is calculated for only the standards in which the person interacting with the agent (SELF) is represented as `Target`. In other words, CoS is a cumulative value of the signed approval degrees for the expression of an emotion by all (including SELF) towards the person currently interacting with the agent itself. For example, let us consider the standards in Table 4.8. If JOHN is currently interacting with the robot and *anger* is one of the elicited emotions, then the coefficient of standard for the *anger* emotion is computed as the average approval degree of all the standards of *anger* emotion where JOHN is the target.

Suppose, there are n elicited emotions from which the most appropriate final emotional state is to be determined. If there are N standards related to the j^{th} emotion : $1 \leq j \leq n$ and we denote the degree of approval of i^{th} standard as d_{a_i} : $1 \leq i \leq N$, and preference associated with a standard as $pref$, then, the coefficient of standard of the j^{th} emotion is given by Equation (4.38).

$$CoS_j = \frac{\sum_{i=1}^{N} \begin{cases} d_{a_i}, & \text{if } pref = \text{“}YES\text{”} \\ -d_{a_i}, & \text{if } pref = \text{“}NO\text{”} \end{cases}}{N} \tag{4.38}$$

Equation 4.38 shows that coefficient of standard is the average of signed approval degree for the expression of the j^{th} emotion from all the recognised persons (including "SELF") to the person interacting with the agent. This, in fact, measures how much the internal standards of the agent support the expression of an emotion. For example, if a standard has preference "YES" then it is okay to express the emotion – hence the positive summation in Equation (4.38). Likewise, if a standard has preference "NO" then it is not okay to express the emotion – hence the negative summation in Equation (4.38). As such, the higher the coefficient of standard (including sign), the better the emotion for expression in the given social context.

The notion of the concepts of deontological and consequentialist ethics presented in Section 2.5.1 is efficiently captured by the formula in Equation (4.38). The formula considers the duties in the form of standards of the agent, thereby capturing the essence of deontological ethics (Alexander and Moore, 2007; Robbins and Wallace, 2007). All the standards related to each emotion are considered for the computation of coefficient of standard. Moreover, in addition to the standards related to itself, the agent also considers the standards related to other recognised persons and the person interacting to the agent (see Table 4.8 for example). By doing this, the agent becomes able to address the consequence of the expression of a particular emotion on the target as well as other related persons, thereby capturing the notion of consequentialist ethics as well.

However, considering only the internal standards for the determination of final emotional state may still lead to unethical or socially unacceptable emotions. For example, consider a person who is really nice and has done plenty of good things to you. Many other people also have positive thoughts about the person and have high regards for the person. Naturally, as per the standard, expressing anger to such a person should be discouraged. Nevertheless, there can be situations where an angry or aggressive response is the most appropriate reaction in response to an action of such a person – say he tries to stab your best friend with a knife. You would definitely become angry and respond in defensive and aggressive manner even if you had high standards for the person. In order to overcome the presented limitation of expressing unethical emotional responses, I also consider emotion intensities elicited by the appraisal process together with the coefficients of standard of each emotion.

As such, I compute a numeric quantity denoted with *Quantified Emotion* to take into account the degree and intensity of the elicited emotions. If we denote the degree of valence of j^{th} emotion by d_{v_j} and the intensity of j^{th} emotion as $\hat{i}_j$, then the quantified value of the j^{th} emotion is given by (4.39).

$$QE_j = d_{v_j} * \hat{i}_j \tag{4.39}$$

Now, the absolute value of the j^{th} quantified emotion is multiplied to its corresponding coefficient of standard to compute the *Coefficient of Ethics (CoE)* as shown in (4.40). The reason for using absolute value of QE_j is to avoid the undesirable sign change when the signed value of CoS_j is multiplied by signed value of QE_j. This helps to consider only the strength of the emotion based on its degree and intensity (without any regards to its sign).

$$CoE_j = CoS_j * \left|QE_j\right| \tag{4.40}$$

When the coefficient of ethics for each elicited emotion is computed, the *emotion with the highest value of coefficient of ethics is selected* as the most ethical emotional state in the given situation. In other words, the *CoS* acts as a regulation mechanism based on ethics to assist the selection of more socially acceptable emotional responses.

In order to test the validity of my claim that ethical reasoning in EEGS can help an autonomous agent to reach to a socially appropriate emotional state, I compared the emotion dynamics of EEGS using three different approaches to reach to final emotional state, which were introduced in Section 4.8.1 as (i) *Highest Intensity Approach* – where the emotion with the highest intensity is considered as the final emotional state, (ii) *Blended Intensity Approach* – where the intensities of the elicited emotions are blended to determine a new intensity value and a final emotion type to be attributed, and (iii) *Ethical Reasoning Approach* – the proposed approach where the final emotional state is determined by reasoning ethically, which I presented earlier in this section. I will present a detailed evaluation of the proposed approach in Chapter 5.

4.9 A Guideline for the Implementation of EEGS Modules

Previous sections in this chapter described the computational details of the presented emotion model EEGS. In this section, I will describe how the computational details have been realised as an implemented model. These details are expected to provide a better understanding of how the proposed model can be implemented and replicated by other researchers for the purpose of comparison and bench-marking.

Overall computational model is implemented in Java Programming language with embedded Apache Derby Database[10] for the storage and retrieval of data. These were my personal implementation choices. A reader should be able to achieve a successful implementation of the model described in this dissertation using other languages or frameworks.

Since the proposed model consists of four main ‘processing’ modules namely (1) emotion elicitation module, (2) cognitive appraisal module, (3) affect generation module and (4) affect regulation module, the subsections below will follow discussion of the implementation details of the given modules in the same order.

4.9.1 Implementing the Emotion Elicitation Module

The emotion elicitation module in EEGS represents the first-order non-cognitive appraisal of the situation (Lambie and Marcel, 2002) leading to an experience of valenced bodily reaction denoting the positivity or negativity of the event (James, 1884; Lange, 1885). In EEGS, the emotion elicitation module is realised more on a functional level than computational. In other words, in the current model, the underlying complexities of the mechanism to compute first-order phenomenological reaction (Lambie and Marcel, 2002) is not implemented. The valenced reaction for a particular event in the given context was obtained from a survey data where people were asked to rate the positivity

[10] A reader may find more information about Apache Derby at https://db.apache.org/derby

or negativity of the given action in the given context (see Chapter 5, Section 5.2.2 for details on how the data was collected). As such, the actions in a particular experimental scenario are assigned the average score provided by the survey participants and that score is considered as the first-order phenomenological reaction (Lambie and Marcel, 2002). This both the input and output to the emotion elicitation module is the valenced score for the action in the given context.

Alternatively, instead of relying on the average scores assigned to the actions, it is also possible to employ machine learning techniques to enable the system to learn how to map contextual information of the emotional event into a signed numeric score that can be fed to the emotion elicitation module. The lack of an appropriate level of details on how to computationally realise this module is a limitation of the present study. However, the main objective of this dissertation is to investigate the cognitive aspects of an emotional process and the data collection methodology used to overcome this limitation is sufficient to lead to the validation of my thesis argument. Although it is possible to computationally realise EEGS model without introducing emotion elicitation module, this notion was provided to allow an easy investigation of the aspect for future researchers.

4.9.2 Implementing the Cognitive Appraisal Module

The cognitive appraisal module takes the first-order phenomenological reaction (which represents the contextual information about the event) as the input and computes the appraisal variables with the help of goals, standards and attitudes by means of the formulas provided in Section 4.6.2 for each considered appraisal variable (see Section 4.6.1 for discussion on how goals, standards and attitudes are defined in EEGS). The table below summarises the inputs and output of the cognitive appraisal module.

Input Parameter(s)	Supporting Parameter(s)	Computed/Output Parameters(s)
(1) Action Scores (first-order phenomenological reaction (Lambie and Marcel, 2002))	(1) Goals (2) Standards (3) Attitudes	(1) Set of Appraisal Variables

Table 4.9 Input(s) and output(s) of cognitive appraisal module.

The set of appraisal variables computed by the cognitive appraisal module are provided as input to the affect generation module where each appraisal variable may be

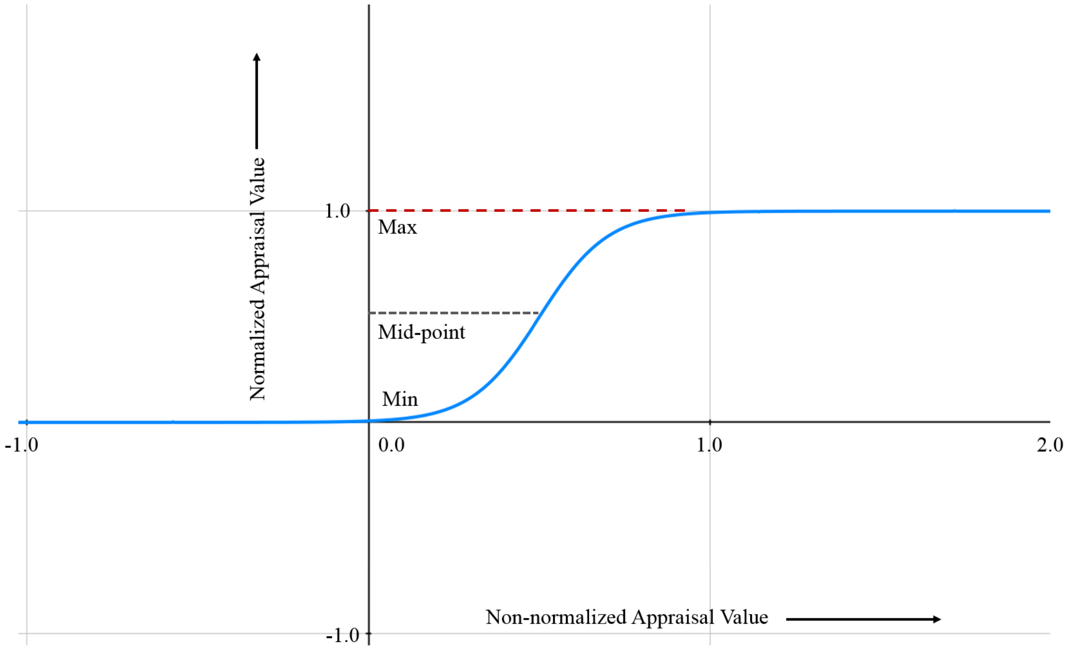


Figure 4.15 Normalisation function for appraisal variables in the range [0,1].

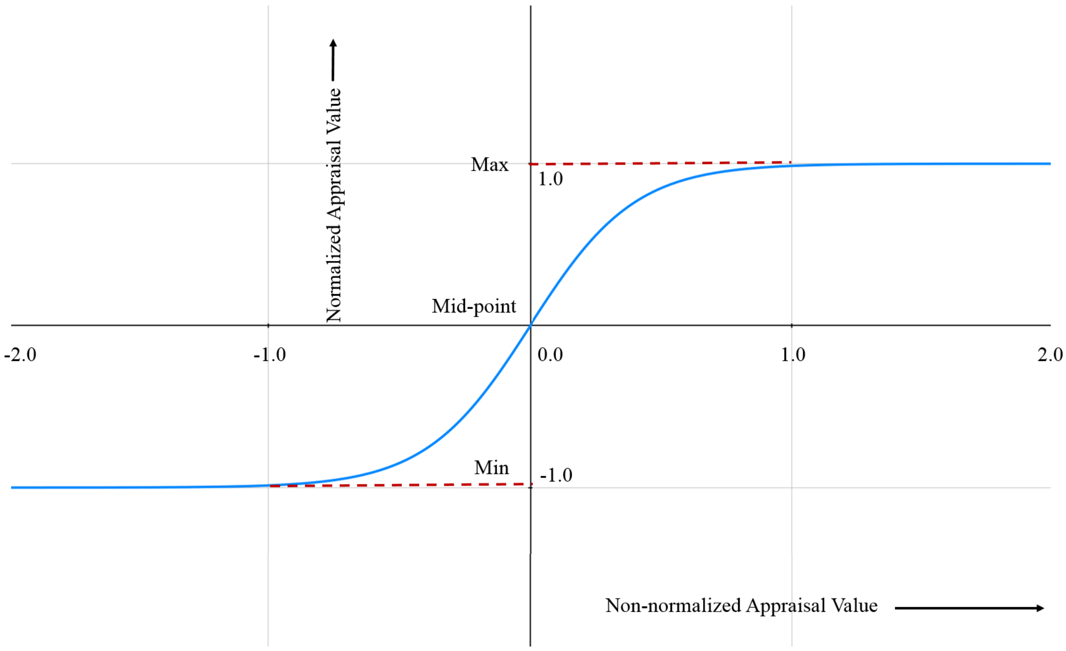


Figure 4.16 Normalisation function for appraisal variables in the range [-1,1].

linked to more than one emotion intensities. However, before the computed appraisal variables are sent to the affect generation module, the values of these variables need to be normalised in a particular range, which is explained below.

Normalisation of Appraisal Variables

The appraisal variables computed using the formulas discussed in Section 4.6.2 may sometimes lead to unexpected values that lie outside the expected range (see Table 4.10 for expected range of values). Therefore it is important to normalise the values obtained by the given formulas in the specified range. I use a modified Logistic function to obtain the appraisal values in the specified range.

$$normalised\_appraisal = \frac{range\_gap}{1 + e^{-m \, * \, (appraisal - midpoint)}} + \gamma \tag{4.41}$$

Where,
normalised_appraisal is the normalised value of appraisal variable,
range_gap is the gap between the min and max expected value of the appraisal variable,
m is the slope of part of the curve where it exhibits linear mapping,
appraisal is the non-normalised value of the appraisal variable,
midpoint is the mid-point is the mid value of the appraisal value range, and
γ is the offset used to shift the value of the Logistic function up or down as per the requirement. A positive value of γ shifts the normalised value up and a negative value shifts it down.

Appraisal Variable	Expected Value Range
goal conduciveness	[-1, 1]
desirability	[-1, 1]
praiseworthiness	[-1, 1]
appealingness	[-1, 1]
deservingness	[-1, 1]
familiarity	[0, 1]
unexpectedness	[0, 1]

Table 4.10 Appraisal variables in EEGS and their value ranges.

Figure 4.15 shows the logistic function to normalise the computed appraisal variables that are supposed to be in the range [0,1] *i.e. familiarity* and *unexpectedness* (*range_gap* = 1, m = 10, *midpoint* = 0.5 and γ = 0). Horizontal axis represents the value of the appraisal variable before normalisation and the vertical axis represents the value of the appraisal variable after normalisation. In case of the appraisal variables lying in the range [0,1], the midpoint of mapping should lie at 0.5 (as shown in Figure 4.15). Similarly, Figure 4.16 shows the normalisation function for the appraisal variables lying in the range [-1,1], where *range_gap* = 2, m = 5, *midpoint* = 0.0 and γ = -1. The reason for using the value of *range_gap* as 2 is because the the difference between 1 and -1 is 2. Slope (m) is slightly lower to maintain a consistent mapping. γ is set to -1 in order to shift the curve 1 step below and making *midpoint* to be zero.

4.9.3 Implementing the Affect Generation Module

The affect generation module mainly deals with mapping the computed appraisal variables into emotion intensities and also alters the mood state. As such, affect generation module takes the set of appraisal variables output from cognitive appraisal module as an input and computes intensities of various emotions with the help of personality and mood factors that was used to learn the association of appraisal variables to emotions (see Sections 4.7.1 and 4.7.2 for details on how such an association is determined in EEGS). The formulas presented between Equations 4.19–4.28 are used to calculate the emotion intensities of various emotions. Moreover, two-way interaction between emotion and mood was realised by implementing the formulas presented in the Equations 4.30–4.35. Table 4.11 summarises the input and output parameters of the affect generation module.

Input Parameter(s)	Supporting Parameter(s)	Computed/Output Parameters(s)
(1) Set of Appraisal Variables	(1) Mood (2) Personality	(1) Set of Emotions Mood State

Table 4.11 Input(s) and output(s) of affect generation module.

Emotion Intensity Update and Normalisation

In contrast to some other models that discard the previous intensity of an emotion once new appraisals are performed (Dias et al., 2014), EEGS opts for an incremental approach where every new appraisal does not undo the effect of prior appraisals on emotion intensities but rather performs an increment or decrement on the intensity based on current appraisal – in line with the suggestions of Scherer (2001). This allows EEGS to preserve the true dynamics of emotional experience and support effectively in the mood update process in a natural manner (more on mood update process will be discussed in the section after this). However, the drawback of such an incremental approach is that it may lead to ever increasing intensity of an emotion, when the subsequent events are congruent to the valence of the emotion. This kind of ever increasing intensity may lead to incoherent behaviour of the agent. Therefore, a suitable mechanism should be employed to ensure that the intensity of a particular emotion

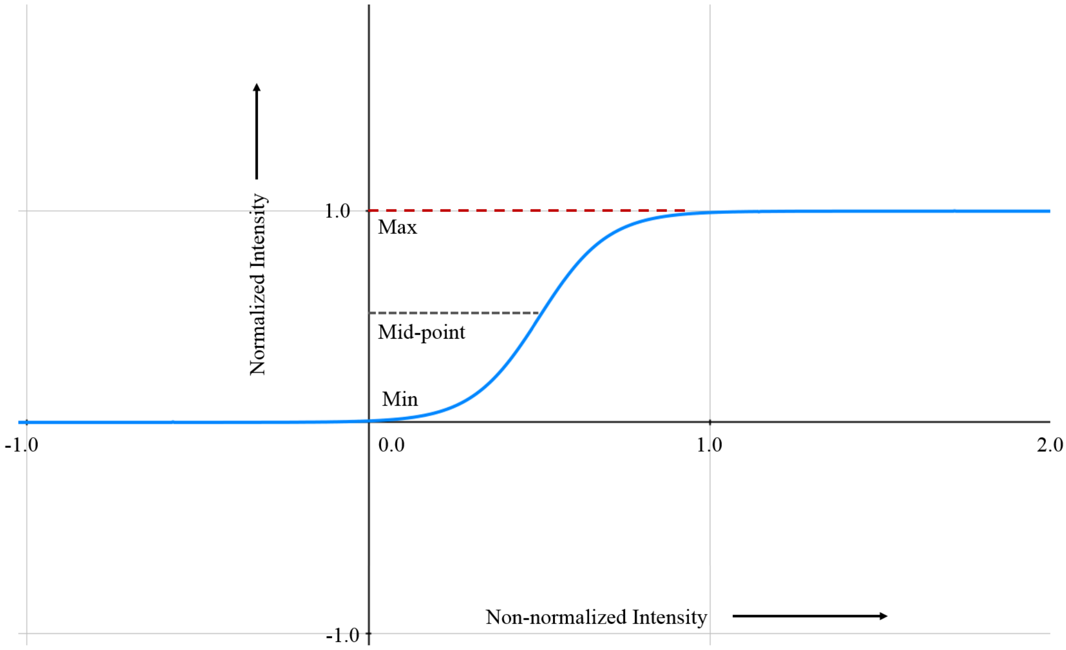


Figure 4.17 Normalisation function for emotion intensities.

always lies in the standard range of [0,1]. In order to achieve this, EEGS normalises the computed emotion intensities using the modified Logistic function similar to the one in Equation (4.41).

$$\hat{i}_e^{norm} = \frac{range\_gap}{1 + e^{-m \, * \, (\hat{i}_e - midpoint)}} + \gamma \tag{4.42}$$

Where,
$\hat{i}_e^{norm}$ is the normalised value of emotion intensity,
range_gap is the gap between the min and max expected value of the intensity,
m is the slope of part of the curve where it exhibits linear mapping,
$\hat{i}_e$ is the non-normalised value of the emotion intensity,
midpoint is the mid-point is the mid value of the intensity range, and
γ is the offset used to shift the value of the Logistic function up or down as per the requirement. A positive value of γ shifts the normalised value up and a negative value shifts it down.

The function in equation 4.42 and Figure 4.17 help in normalising an emotion intensity to a stable value in the range [0, 1].

4.9.4 Implementing the Affect Regulation Module

As discussed in Section 4.8, multiple active emotional states generated by affect generation module need to be regulated and converged to a stable emotional state to allow socially acceptable behaviour by the agent. This is achieved by taking the emotions output from affect generation module and determining a final emotional state with the help of ethical standards using the Equations 4.38–4.40 and selecting the emotional state with highest coefficient of ethics (see Section 4.8.2 for details on the formulas and computational mechanism provided).

Input Parameter(s)	Supporting Parameter(s)	Computed/Output Parameters(s)
(1) Set of Emotions	(1) (Ethical) Standards	(1) Regulated Emotional State

Table 4.12 Input(s) and output(s) of affect regulation module.

The regulated emotional state obtained as the output of affect regulation module can be considered to control other cognitive components, if implemented in intelligent systems affected by emotions such as decision systems (Holtzman, 1988).

4.10 Chapter Summary

This chapter covered a detailed description of various modules of the proposed Ethical Emotion Generation System (EEGS). The discussion followed the sequence of theoretical steps involved in the process of emotion generation and regulation as represented in Figure 2.7 in Chapter 2. Also, the description of the processes in various modules was accompanied by complete technical/mathematical representation of the corresponding aspect.

The chapter started with a brief overview of EEGS in Section 4.1 leading to a revisit to the overall process of appraisal dynamics and emotion generation as shown in Figure 4.1. Then, in Section 4.3, I presented an overall architecture of EEGS, which is composed of various modules as shown in Figure 4.2, namely (i) *Emotion Elicitation Module*, (ii) *Cognitive Appraisal Module*, (iii) *Memory Module*, (iv) *Characteristics Module*, (v) *Affect Generation Module*, and (vi) *Affect Regulation Module*. As explained previously, the *emotion elicitation module* performs the first-order (Lambie and Marcel, 2002) lower-level (Scherer, 2001) appraisal of the event that does not involve conscious cognitive processing. This process is followed by the computation in *cognitive appraisal module* where different appraisal variables are calculated based on the goals, standards and attitudes of the agent stored in *memory module*. The process of mapping the appraisal variables into emotion intensities in *affect generation module* is modulated by several factors like personality and mood present in the *characteristics module*. The multiple emotions activated by the cognitive appraisal process are converged to a stable state and regulated by the *affect regulation module*.

The discussion was then followed by the technical description of *events*, *actions* and *objects* in relation to the *emotion elicitation* process was presented in Section 4.4. The discussion was then followed by the *cognitive appraisal* process where the concepts of *goals*, *standards* and *attitudes* were introduced. *Goals* can be defined as a set of states an individual wants to achieve. In the context of a computer scientist, his/her goal might be to win a Turing award. However, in the context of an autonomous agent that is aimed to interact with people in the society and serve for their well being, such a goal is unnecessary and irrelevant. Therefore, since the objective of EEGS is to be integrated into social agents, goals in EEGS are also of similar structure (see Section 4.6.1 for more discussion). *Standards* in EEGS represent the beliefs of the agent regarding various actions between any two agent (including itself). An standard encourages or prevents an agent from performing some action to another agent. Additionally, an standard defines what an agent believes what kind of actions can be performed by one agent to another agent (see Section 4.6.1 to revisit how sgtandards are structured in EEGS).

Attitudes of an agent reflect what it feels about an object/person. An agent can have positive attitude towards an external agent if is has positive experience of interaction with the agent and negative attitude otherwise. These goals, standards and attitudes play a central role in the computation of appraisal variables (Ortony et al., 1990). I present the details of a novel mechanism for the computation of appraisal variables in Section 4.6.2. While other models of emotion compute appraisal variables based on domain-specific rules (Aylett et al., 2005; Dias and Paiva, 2005; Velásquez and Maes, 1997), I use a domain-independent mechanism to compute the appraisal variables in EEGS (details of the formulae for the calculation of individual appraisal variable can be found in Section 4.6.2).

After the completion of all the appraisal processes (which run in parallel as explained in Section 4.6.2), the quantitative values of appraisal variables are mapped into emotion intensities. This mapping process in EEGS is influenced by factors like mood (Morris, 1992; Neumann et al., 2001) and personality (Corr, 2008; Revelle, 1995; Watson and Clark, 1997). EEGS implements a machine learning approach to determine the *weights of association* of various appraisal variables to emotion intensities based on personality and mood factors, the details of which is presented in Section 4.7.2. These weights and quantitative values of various associated appraisal variables are used to compute the intensities of various emotions in EEGS (see Section 4.7.3 for technical details). EEGS not only realises the effect of mood and personality on emotion, but also operationalises the influence of personality on mood and effect of emotion intensities on the mood dynamics of the agent – as previously discussed in Section 4.7.3, Figure 4.11 and Figure 4.12. Since, the experience of emotion is often believed to be short-lived, EEGS employs a mechanism to decay the intensity of emotion after a certain duration. Several emotion decay functions have been put forward by researchers namely (i) *Linear* (Becker, 2008; Egges et al., 2004; El-Nasr et al., 2000; Gebhard, 2005), (ii) *Exponential* (Becker, 2008; Dias and Paiva, 2005), (iii) *Logarithmic* (Hudlicka, 2016), and (iv) *Tan-Hyperbolic* (Gebhard et al., 2003). In Figure 4.13, I presented a comparison of various emotion decay strategies and proposed a more plausible and realistic decay function in the form of a modified exponential as shown in Equation 4.36.

Because of multiple associations between appraisal variables and emotions, more than one emotional states can be active as a result of cognitive appraisal process (Ortony et al., 1990; Scherer, 2001). Due to this reason, an agent should go through two important processes to yield an expected emotional response – (i) *emotion convergence* and (ii) *emotion regulation*. Emotion convergence is the processing of reaching to a stable emotional state when more than one conflicting emotional states are active at the same time. Emotion regulation is the process of ensuring that the experience of emotion helps in achieving personal as well as social benefits (Gross, 1998b). While two types of emotion regulation strategies are possible namely (i) *Antecedent-focused* and (ii) *Response-focused*, I opted for the latter because the former although promising involved more technical vulnerabilities and less probable to achieve a stable and regulated emotional experience in the ultimate stages of emotion processing (see Section 4.8 for more discussion on this). Although previous proposals offered (i) *highest intensity* approach (Gratch and Marsella, 2004a) or (ii) *blended intensity* approach (Reilly, 2006) for emotion convergence, I opted to adopt an (iii) *ethical reasoning* approach to achieve the goals of both the emotion convergence as well as regulation (where the former two approaches are not able to achieve the goal of emotion regulation).

Chapter Bibliography

Aleman, A., Swart, M., and van Rijn, S. (2008). Brain imaging, genetics and emotion. *Biological psychology*, 79(1):58–69.

Alexander, L. and Moore, M. (2007). Deontological ethics.

Aylett, R., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). Fearnot!– An experiment in emergent narrative. In *Intelligent virtual agents*, pages 305–316. Springer.

Barrett, L. F. (1996). Hedonic tone, perceived arousal, and item desirability: Three components of self-reported mood. *Cognition & Emotion*, 10(1):47–68.

Becker, C. (2008). *WASABI: Affect simulation for agents with believable interactivity*. PhD thesis, Bielefeld University.

Beedie, C., Terry, P., and Lane, A. (2005). Distinctions between emotion and mood. *Cognition & Emotion*, 19(6):847–878.

Bevilacqua, L. and Goldman, D. (2011). Genetics of emotion. *Trends in cognitive sciences*, 15(9):401–408.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer.

Brandon, T. H. (1994). Negative affect as motivation to smoke. *Current Directions in Psychological Science*, 3(2):33–37.

Callahan, S. (1988). The role of emotion in ethical decisionmaking. *Hastings Center Report*, 18(3):9–14.

Canli, T., Ferri, J., and Duman, E. (2009). Genetics of emotion regulation. *Neuroscience*, 164(1):43–54.

Corr, P. J. (2008). *The reinforcement sensitivity theory of personality*. Cambridge University Press.

Costa, P. T. and McCrae, R. R. (1988). Personality in adulthood: a six-year longitudinal study of self-reports and spouse ratings on the neo personality inventory. *Journal of personality and social psychology*, 54(5):853.

Costa, P. T. and McCrae, R. R. (1992). NEO PI-R professional manual. *Odessa, FL: Psychological assessment resources*, 396:653–65.

Dias, J., Mascarenhas, S., and Paiva, A. (2014). FAtiMA modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer.

Dias, J. and Paiva, A. (2005). Feeling and reasoning: A computational model for emotional characters. In *EPIA*, volume 3808, pages 127–140. Springer.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440.

Dweck, C. S. (2008). Can personality be changed? the role of beliefs in personality and change. *Current directions in psychological science*, 17(6):391–394.

Egges, A., Kshirsagar, S., and Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer animation and virtual worlds*, 15(1):1–13.

Ekman, P. (1994). Moods, emotions, and traits. *P. Ekman & RJ Davidson, The Nature of Emotion: Fundamental Questions*, pages 56–58.

El-Nasr, M. S., Yen, J., and Ioerger, T. R. (2000). Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257.

Feldman, L. A. (1995). Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*, 21(8):806–817.

Forgas, J. P. (1992). Affect in social judgments and decisions: A multiprocess model. *Advances in experimental social psychology*, 25:227–275.

Frijda, N. H., Kuipers, P., and Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of personality and social psychology*, 57(2):212.

Gaudine, A. and Thorne, L. (2001). Emotion and ethical decision-making in organizations. *Journal of Business Ethics*, 31(2):175–187.

Gebhard, P. (2005). Alma: A layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM.

Gebhard, P., Kipp, M., Klesen, M., and Rist, T. (2003). Adding the emotional dimension to scripting character dialogues. In *IVA*, pages 48–56. Springer.

Gratch, J., Cheng, L., and Marsella, S. (2015). The appraisal equivalence hypothesis: Verifying the domain-independence of a computational model of emotion dynamics. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 105–111. IEEE.

Gratch, J. and Marsella, S. (2004a). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306.

Gratch, J. and Marsella, S. (2004b). Technical details of a domain-independent framework for modeling emotion. Technical report, University of Southern California Marina Del Rey Ca Inst for Creative Technologies.

Gross, J. J. (1998a). Antecedent-and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of personality and social psychology*, 74(1):224.

Gross, J. J. (1998b). The emerging field of emotion regulation: An integrative review. *Review of general psychology*, 2(3):271.

Gross, J. J. and Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. *Handbook of emotion regulation*.

Holtzman, S. (1988). *Intelligent decision systems*. Addison-Wesley Longman Publishing Co., Inc.

Hong, Y.-y., Morris, M. W., Chiu, C.-y., and Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7):709.

Hudlicka, E. (2008). Modeling the mechanisms of emotion effects on cognition. In *AAAI Fall Symposium: Biologically inspired cognitive architectures*, pages 82–86.

Hudlicka, E. (2016). Computational analytical framework for affective modeling: Towards guidelines for designing. *Psychology and Mental Health: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, page 296.

Hull, J. G. and Bond, C. F. (1986). Social and behavioral consequences of alcohol consumption and expectancy: a meta-analysis.

Isen, A. M. and Means, B. (1983). The influence of positive affect on decision-making strategy. *Social cognition*, 2(1):18–31.

Jain, S. and Asawa, K. (2015). Emia: emotion model for intelligent agent. *Journal of Intelligent Systems*, 24(4):449–465.

James, W. (1884). What is an emotion? *Mind*, 9(34):188–205.

Lambie, J. A. and Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychological review*, 109(2):219.

Lange, C. G. (1885). *Om sindsbevaegelser; et psyko-fysiologisk studie*. Lund.

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American psychologist*, 37(9):1019.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press on Demand.

Marinier III, R. P. and Laird, J. E. (2007). Computational modeling of mood and feeling from emotion. In *Proceedings of the Cognitive Science Society*, volume 29.

Marreiros, G., Santos, R., Ramos, C., and Neves, J. (2010). Context-aware emotion-based model for group decision making. *IEEE Intelligent Systems*, 25(2):31–39.

Marsella, S. C. and Gratch, J. (2009). Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90.

Mayer, J. D. and Gaschke, Y. N. (1988). The experience and meta-experience of mood. *Journal of personality and social psychology*, 55(1):102.

Mayer, J. D., Gaschke, Y. N., Braverman, D. L., and Evans, T. W. (1992). Mood-congruent judgment is a general effect. *Journal of personality and social psychology*, 63(1):119.

Mayer, J. D., Mamberg, M. H., and Volanth, A. J. (1988). Cognitive domains of the mood system. *Journal of Personality*, 56(3):453–486.

Morris, W. N. (1992). A functional analysis of the role of mood in affective systems.

Neumann, R., Seibt, B., and Strack, F. (2001). The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Emotion and Cognition*, 15:725–747.

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., and Williams, M. (2017). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *Annual Meeting of the Cognitive Science Society*.

Ojha, S. and Williams, M. (2017). Emotional appraisal: A computational perspective. In *Annual Conference on Advances in Cognitive Systems*.

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.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Padgham, L. and Taylor, G. (1996). A system for modelling agents having emotion and personality. In *International Workshop on Intelligent Agent Systems*, pages 59–71. Springer.

Parkinson, B., Totterdell, P., Briner, R. B., and Reynolds, S. (1996). *Changing moods: The psychology of mood and mood regulation*. Longman.

Picard, R. W. (1997). *Affective computing*, volume 252. MIT press Cambridge.

Reilly, S. N. (2006). Modeling what happens between emotional antecedents and emotional consequents. *ACE*, 2006(19).

Remington, N. A., Fabrigar, L. R., and Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of personality and social psychology*, 79(2):286.

Revelle, W. (1995). Personality processes. *Annual review of psychology*, 46(1):295–328.

Robbins, R. W. and Wallace, W. A. (2007). Decision support for ethical problem solving: A multi-agent approach. *Decision Support Systems*, 43(4):1571–1587.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of personality & social psychology*.

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, 10(3):241–278.

Roseman, I. J., Spindel, M. S., and Jose, P. E. (1990). Appraisals of emotion-eliciting events. *Journal of Personality and Social Psychology*, 59(5):899–915.

Rosenberg, E. L. (1998). *Levels of analysis and the organization of affect.*, volume 2. Educational Publishing Foundation.

Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin*, 124(2):165.

Rusting, C. L. and Larsen, R. J. (1995). Moods as sources of stimulation: Relationships between personality and desired mood states. *Personality and Individual Differences*, 18(3):321–329.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120):57.

Scollon, C. N., Diener, E., Oishi, S., and Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of cross-cultural psychology*, 35(3):304–326.

Smith, C. A. and Kirby, L. D. (2001). Toward delivering on the promise of appraisal theory. In *Series in affective science. Appraisal processes in emotion: Theory, methods, research*, pages 121–138. Oxford University Press.

Smith, C. A., Lazarus, R. S., et al. (1990). Emotion and adaptation. *Handbook of personality: Theory and research*, pages 609–637.

Suinn, R. M. and Richardson, F. (1971). Anxiety management training: A nonspecific behavior therapy program for anxiety control. *Behavior Therapy*, 2(4):498–510.

Thagard, P. and Nerb, J. (2002). Emotional gestalts: Appraisal, change, and the dynamics of affect. *Personality and Social Psychology Review*, 6(4):274–282.

Thayer, R. E., Newman, J. R., and McClain, T. M. (1994). Self-regulation of mood: Strategies for changing a bad mood, raising energy, and reducing tension. *Journal of personality and social psychology*, 67(5):910.

Velásquez, J. D. and Maes, P. (1997). Cathexis: a computational model of emotions. In *Proceedings of the first international conference on Autonomous agents*, pages 518–519. ACM.

Watson, D. and Clark, L. A. (1997). Extraversion and its positive emotional core.

Watson, D. and Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98(2):219.

,

A proper route to an understanding of the world is
an examination of our errors about it.

— Errol Morris —

5
Model Evaluation and Thesis Validation

In Chapter 4, I presented a detailed description of the proposed computational model of emotion. The discussion of each component was accompanied by corresponding mathematical formulation associated with the functioning of a particular module. The computation involved three major steps of:

(i) calculating appraisal variables in a domain-independent manner,

(ii) mapping the appraisal variables into emotion intensities where the weight of association between appraisals and emotions was determined by factors such as personality and mood, and

(iii) regulating the emotions of an agent implementing the model to reach to single final emotional state based on ethical reasoning.

This chapter will mainly deal with the evaluation of various computational modules of the proposed emotion model EEGS which will lead to the validation of the hypotheses presented in Chapter 3, Section 3.2 thereby supporting the thesis statement presented in Chapter 1, Section 1.3.

I evaluate my computational model at a componential level where each major computational block/module is evaluated for its accuracy separately. Such an approach provides a lot of advantages not only for validation purposes but also for bench-marking

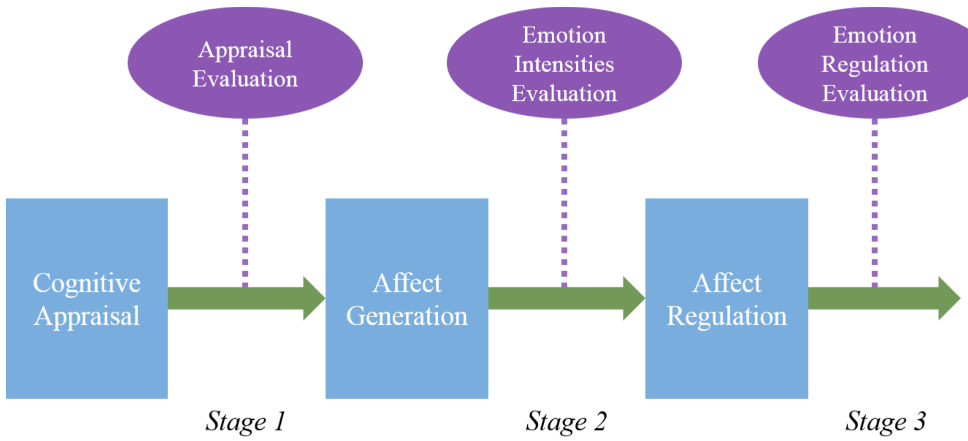


Figure 5.1 Proposed *3-Stage Evaluation* approach for computational models of emotion.

of the models. This kind of evaluation approach encourages *low coupling* among the internal modules in the model thereby promoting independent evaluation of crucial components. In computational field, a software system is said to have the property of low coupling if various components of the system are as less dependent on each other as far as possible. In addition to allowing independent evaluation of several components in a model, this approach also promotes component-level comparison between emotion models thereby allowing better evolution of the models. On top of all that, this ensures that a computational model is evaluated for its performance at all stages. This is particularly important if intermediate output of one processing stage is to be used in another cognitive component of the agent – for example decision making. The following sections will present the details of the component-level evaluation approach used for the validation of the computational mechanism of EEGS, which in turn will validate my thesis statement.

5.1 Introduction to a 3-Stage Evaluation Approach

As mentioned in the beginning of this chapter, EEGS undergoes three main stages of processing involving (i) *calculation of appraisal variables* to appraise the emotion eliciting stimulus, (ii) *mapping the appraisal variables into the intensities of various emotions*, and (iii) *reaching to a stable emotional state by regulating the active emotions*. Therefore, I propose an approach that allows the evaluation of each of these processing stages individually, which I call a *3-Stage Evaluation* of a computational model of emotion.

Figure 5.1 provides a graphical visualisation of the proposed 3-stage evaluation approach used to evaluate my computational model of emotion – EEGS. Stage 1 evaluation measures the accuracy of the computation of different appraisal variables in EEGS and also examines if EEGS can compute appraisal variables in multiple domains without the need of changing the model's rules and parameters or not. Stage 2 evaluation measures whether the appraisal variables are mapped accurately into emotion intensities and how the factors of mood and personality affect generated emotion dynamics. Finally, Stage 3 evaluation measures whether the final emotional state reached by EEGS represents human-like emotions and enhances the social acceptance of the emotion or not.

In the above discussed evaluation context, a key question to answer is – *how do we measure the accuracy of various components of a computational model of emotion*? In other words, *what is the baseline criteria to compare the outputs of an emotion model*? These are still unanswered questions in emotion modelling literature. Different researchers have adopted different methods for evaluating their computational models of emotion. For example, Becker (2008) evaluates his model by examining how believable are the emotions being expressed by the artificial agent; El-Nasr et al. (2000) confirms the validity of the emotional responses by simulating a virtual pet. Other researchers adopt toddler (Velásquez and Maes, 1997) or virtual conversational characters (Aylett et al., 2005; Gebhard, 2005) for the evaluation of their model. Although such evaluation approaches may provide some perspective on developing agents with viable emotional and behavioural responses, these do not always ensure a true reflection of human emotional behaviour which is what emotion models actually aim to achieve. Therefore, I developed an approach of evaluating different components of EEGS by comparing their outputs with the data collected from humans. In the following sections, I will discuss, in detail, how the data was collected from human participants and used for the evaluation of EEGS and validation of my thesis.

5.2 Scenarios and Data Collection[1]

As previously stated, my aim in this chapter is to validate my thesis by evaluating the computational model of emotion EEGS. In order to achieve that, I opted for comparing the performance of EEGS with emotion data collected from humans. For collecting relevant data, multiple studies were conducted where participants rated the aspects of appraisals and/or emotions in the given scenarios. The details of the studies are presented in the following sections.

[1]Most of the content of this section have been adapted from Ojha and Williams (2017), Ojha et al. (2017), and Ojha et al. (2018b)

5.2.1 Scenario Design

Before I could collect emotion data from humans, it was necessary to design realistic scenarios of emotion elicitation in a situation of interaction between two individuals. The scenarios were intended to simulate an interaction between two individuals because the evaluation was aimed at how the model would perform in usual scenarios of human-robot interactions. To make sure the design processing for the scenarios is not biased and it cannot be performed in such a way to unfairly facilitate the performance of my model, I employed independent human participants to design such scenarios. For this process, I recruited 4 naïve adults, without any knowledge about the objectives of the present research, to cooperate in designing six scenarios under the following conditions:

- The scenario shall include the interactions of two subjects, one of them denoted as sender and the other as receiver;
- A minimum of 5 and a maximum of 10 actions of the sender directed to the receiver describing a plausible social interaction between two persons shall be provided;
- At the beginning, each scenario shall provide the contextual information about the designed situation and the two considered subjects. Moreover, additional contextual information could be provided during the development of the described social exchanges, whenever this information is necessary to contextualise the remaining interactions;
- No contextual information suggesting the potential emotional state of the receiver shall be provided for individual interactions, with the exception of the contextual information provided at the beginning of the scenario.

This process resulted in a set of scenarios used for collecting data from human participants, as I will discuss later in Section 5.2.2. The scenarios included interactions between (1) two strangers (a male and a female) interacting on a bench of a park, (2) two close friends (both males) meeting at a beach, (3) a husband and a wife having an argument about forgetting the birthday, (4) an elderly woman affected by dementia and her nurse (both females) experiencing a distressful moment, (5) a guy having argument with his brother, and (6) an interaction between a customer of a café and a waiter (both males). Table 5.1 shows a brief summary of each scenario. The details of the scenarios are provided in Appendix A.

Scenario #	Sender	Receiver	Relationship	Context
1	Bill	Rosy	Strangers	An adult male (Bill), An adult female (Rosy), strangers, meeting at a bench of a park.
2	John	Paul	Friends	Two close friends meeting at a beach in a hot summer day.
3	Anna	David	Partners	A husband (David) and wife (Anna) having argument about husband forgetting wife's birthday.
4	Rose	Lily	Patient-Carer	An elderly woman (Rose) in an aged care facility who is affected by dementia is misbehaving with her female nurse (Lily).
5	Andrew	Robert	Brothers	Two brothers at home getting excited about an upcoming match.
6	Hari	Gopal	Customer-Staff	An interaction between customer (Hari) and Café staff (Gopal).

Table 5.1 Summary of the scenarios considered.

5.2.2 Data Collection

The data collection methodology followed the assumptions of *simulation theory* (Goldman, 1992). Several studies in emotion research suggest that people use their own body as a way not only to elicit emotional responses and be aware of their emotional state, but also to simulate internally the emotional response of others as a way to understand others' feelings. This is called simulation theory and several models were suggested for this phenomenon (Goldman, 1992; Gordon, 1992). Simulation theory also provides a basis for theory of mind and social intelligence because it is theorised that we use this capability to understand each others and anticipate each other's actions (Williams, 2012). Under this assumption, the assessor would put himself/herself in the shoes of the agent presented in the scenario to determine which appraisal/emotional state the agent would have been. An alternative to this approach is to ask to the participants to recall an emotional experience from past and rate their appraisals and emotions based on the remembered experiences (Meuleman and Scherer, 2013; Nguwi and Cho, 2010). However, this approach may introduce some critical problems. First, the data collected in such a manner tends to dilute the real experience of emotion because a person tends to forget most of the experience over time (Jenkins et al., 2002). Second, such a data can be more noisy since there is "little control over the conditions under which the questionnaire [is completed]" (Meuleman and Scherer, 2013, p. 409). Third, such a data can not be practically verified for accuracy and reliability because participants can imagine whatever they want which results in the appraisals and emotion rating of each participant to differ from each other thereby not leaving any room for cross validation of the measure of deviation. Therefore, I opted for the data collection approach by using simulation theory instead of recalled experience for the evaluation of EEGS, which are described in detail in the following sections. It should, however, be acknowledged that the data collection process (described in the following sections) has not been completed in a controlled laboratory setting. Rather a croud-sourcing approach with online survey has been used which is prone to several drawbacks such as lack of follow-up with the participants (Murphy et al., 2014).

Study A: Somatic Responses and Emotion Intensities

In Study A, two sets of web-based surveys requiring two tasks were designed: (i) *an action scoring task*, which involved providing a rating for an action from one individual to another in the given interaction context; and (ii) *a mind-reading task*, which involved estimating the emotion intensities of the receiver of the action in the given interaction context. For both the tasks, I used the scenarios designed by the 4 naïve adults, as previously described in Section 5.2.1.

Participants covering a broad set of countries were invited on Facebook or through mailing lists to participate in this study. The surveys employed during both the tasks were completely anonymous. A total of 153 responses (*male* = 82, *female* = 71) were received. Importantly, the subjects were randomly assigned to either the action scoring task or the mind-reading task.

The experimental subjects participating in the *action scoring task* were asked to indicate, for each scenario, *how positive or negative each social exchange performed by the sender would be perceived by the receiver in that specific context*. The rating was based on 7-point Likert scale: Extremely Negative, Very Negative, Negative, Neither Negative Nor Positive, Positive, Very Positive, Extremely Positive. The responses were numerically evaluated by attributing a weight to each point of the scale (*i.e.* -1, -0.66, -0.33, 0, 0.33, 0.66 and 1 respectively). Then, for each scored interaction, the responses were averaged, thus obtaining an action score value (*i.e.* the first-order appraisal as proposed by Lambie and Marcel (2002)) for each of the considered social exchange in the specific context. Given this approach, the final score for an action would lie in the range of [-1, +1]. The survey questions for the action scoring task are presented in Appendix B.

I employed the mind-reading task to gather sufficient human data samples of emotional responses in specific domains to use for evaluating the affect generation module of the proposed EEGS model. In this study, the participants had to indicate, for each interaction of the sender, *what would have been the chances that the receiver would happen to be in a particular emotional state, based on the just happened interaction and the previously occurred social exchanges and contextual information*. Therefore, for each of the eight considered emotional states (see Chapter 4, Section 4.7.3 for the emotions considered in current research) the rating was based on 6-point Likert scale: Not at all, Very Low, Low, Medium, High and Very High. The additional rating "Not at all" was necessary to allow the participants to express the absence of a particular emotional state in the receiver. The responses were numerically evaluated by attributing a weight to each point of the scale (*i.e.* 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0 respectively). Average score given by the participants to various emotions was calculated by performing the weighted average of the ratings making the emotion intensity lie in the range of [0, 1]. It is important to note that the six scenarios described earlier (see Section 5.2.1) were split into two survey sets where each set contained three scenarios. This was necessary to reduce the risk of observing fatigue effects in the participants, which is a situation where the responses of the participants may get biased because of mental tiredness (Hess et al., 2012; Wright and Ogbuehi, 2014). The first survey set consisted of the scenarios of "Two Strangers in a Park", "Two Close Friends" and "Two Brothers" while the second survey set consisted of the scenarios of "Husband and Wife", "Patient and a Nurse" and "Café Staff and Customer". Within each set, the scenarios were presented to the participants in random order so as to further reduce any fatigue effect of the respondents. At the end of each survey set, participants were given an opportunity to provide a free text answer about their survey experience where they could provide the feedback on the overall experience in answering the questions and report any difficulty encountered. Both the emotion survey sets used in Study A are provided in Appendix B.

Study B: Appraisal Values and Emotion Intensities

Since the surveys in Study A collected data only for emotion intensities but not for appraisals, the data was not sufficient to train and evaluate the affect generation module of EEGS model. Indeed, to train and evaluate the performance of the module, for each social exchange, I needed paired samples with the following information: (1) the personality factors of the human evaluator (2) the value of each appraisal variable with respect to the current interaction as per the human evaluator assessment; (3) emotion intensities as per human evaluator assessment; (4) the mood of the subjects in the presented scenarios, as per human evaluator assessment. Therefore, a new survey was designed to collect the data related to emotions as well as appraisals, mood and personality factors. The questions in the survey were arranged in a slightly different format compared to Study A. Instead of asking the participants to rate all the emotions at once for each action in sequence, rating for each of the sequential actions was collected separately for each emotion/appraisal. This was done to preserve the impact of contextual information on the ratings and also to prevent the confusion caused by switching from one emotion/appraisal to another for a single action (as per feedback gathered from some of the participants of Study A). Moreover, two additional core actions were derived from the original scenario to make a total of 11 action sequences as opposed to 9 action sequences in Study A for the same scenario. Two new action sequences were added to provide a diverse set of events data for effective learning of the appriasal–emotion network.

The data was collected through an online survey link distributed in Amazon Mechanical Turk (MTurk). The survey was designed in Qualtrics. Below are the constraints that were applied while designing and distributing the survey.

- The sequence of the questions within a particular section (appraisal or emotion) was randomised so that each participant would answer the questions in different order. This was done to eliminate the effect of respondent fatigue bias while answering the questions.
- Each participant could take part in the survey only once. This was enforced with a built-in feature of Qualtrics. This helped me avoid the probable duplicate responses from MTurk users.

The survey questions consisted of the following steps.

- First, the participants were asked to indicate their gender, solely for demographic purposes.
- Second, the participants were asked a question where they rated their own personality factors. The question was presented as “How much do you agree that you have the following personality traits?” and five personality factors were provided. Taking into the account the fact that the personality terms may not be familiar to general people, accompanying common English words characterising a particular personality trait were also included. For example, the trait of “Extraversion” was accompanied by the words such as ‘talkative’, ‘frank’ and ‘outgoing’. Participants could rate the personality factors on a 5-point scale ranging from “Strongly Disagree” to “Strongly Agree” which were later converted into numerical scales.

A reason for not using standard trait questionnaires (Costa and McCrae, 1988) for the understanding of the participants' personality factors is that presenting such a long list of questions just for gathering the information about personality factors could lead the participants to extreme fatigue leading to inconsistent responses in latter part of the survey. The approach used in current research is also supported by the findings of an experimental study conduted by Gosling et al. (2003), where they concluded that "[a]lthough somewhat inferior to the standard Big-Five instrument... the FIPI [(Five Item Personality Inventory)] reached adequate levels in each of the four criteria against which it was evaluated" (Gosling et al., 2003, pp. 513).

- Then, the participants were asked to guess how the target of interaction in the given scenario would evaluate the various appraisal variables (presented in random order) in response to each of the actions defined in the scenario. For example, for the appraisal variable *desirability*, participants were asked "How desirable do you think each of the following actions of Bill will be considered by Rosy?" where Bill and Rosy are two agents in the interaction scenario 1 (see Appendix A). Then, the sequence of actions in the scenario were presented under the question (actions were presented in fixed order to maintain the context of interaction scenario). The participants would provide a separate rating for each action in the question. Since, there were 11 core actions directed towards Rosy[2], participants provided 11 ratings of the appraisal variable *desirability i.e.* for each action. The rating was on a 7-point scale ranging from "Extremely Undesirable" to "Extremely Desirable". The ratings for other appraisal variables were obtained in similar manner expect for the difference in the label of rating. For example, for the appraisal variable *praiseworthiness*, the labels for rating scales ranged from "Extremely Blameworthy" to "Extremely Praiseworthy".
- Following the appraisal variables, participants were asked to rate the intensity of emotions of the receiver in reaction to the sequence of the actions from the sender. For example, for emotion *joy*, participants were presented with the question "How much Joy do you think will be experienced by Rosy in response to the following actions of Bill?" and provided with the sequence of actions (similar to the questions about appraisal variables). Participants then provided a rating for the intensity of *joy* emotion for each action in the scenario. The rating of emotion

[2] In the original scenario, there were only 9 core actions from Bill towards Rosy but the last action was repeated twice making a total of 11. This was done because the data was intended to train the algorithm used to map the appraisal variables to emotion intensities, and, hence, a balanced amount of positive and negative actions was required.

intensity was available in a usual 5-point scale ranging from “Very Low” to “Very High” with an additional scale of “Not at all” allowing the participants to provide a zero intensity (hence the actual scale was in 6 points).

- Finally, participants were asked to rate the mood of the target of interaction in response to the sequence of actions by asking the question “How do you think the following actions of Bill will change the mood of Rosy?”. Participants rated the mood of Rosy for each action of Bill. The ratings for mood were available on a 7-point scale ranging from “Extremely Negative” to “Extremely Positive”.

A total of 47 unique responses were obtained from the survey (*male* = 31 and *female* = 16). Since the survey was distributed through Amazon Mechanical Turk, I was able to collect responses from a diverse participants of America, Europe and Asia (as indicated by geolocation of the respondents in Qualtrics). This helped to effectively capture other factors affecting the process of emotion that were not asked explicitly in the survey questions (like cultural bias being one of them). The details of the survey questions in Study B can be found in Appendix B.

It is important to note that since both the Study A and Study B were croud-sourced, there were chances that some of the responses may have been unreliable. Although this was minimised to certain extent by the means of survey ‘constraints’ as discussed above, a manual checking was done to ensure individual responses would not deviate significantly or not remain consistent with the changing events in the scenarios. Any response that were inconsistent or deviated significantly were excluded from the analyses. Appendix C shows a summary of various statistics of the survey data from both Study A and Study B.

5.3 Stage 1: Cognitive Appraisal Evaluation

The appraisal computation mechanism implemented in most of the computational models of emotion are domain-dependent except for few (Gratch and Marsella, 2004a; Jain and Asawa, 2015; Kaptein et al., 2016). While only some of the notable models compute appraisals within the model (Gratch and Marsella, 2004b), most other models consider appraisals as a pre-existing knowledge of the system instead of computing appraisals dynamically in real time (see for example the work of Gebhard (2005) and Dias and Paiva (2005)). Since such models are inherently domain-dependent where appraisals are pre-defined, there is no enough evidence to discuss about the evaluation of their appraisal mechanism. Interestingly, even the emotion models that do compute appraisals in domain-independent fashion do not consider the accuracy testing of appraisals as an important facet during the evaluation of their model (Gratch and Marsella, 2004b) – they consider plausibility of only the final emotional or behavioural response. However, as the first stage of the 3-stage evaluation approach proposed in this dissertation, I will conduct a detailed evaluation of the appraisal mechanism in EEGS. The proposed computational model of emotion – EEGS is the first of its kind to perform such an evaluation of the computed appraisals of the model by comparing it to the data collected from humans. This evaluation in turn will help in validating the first part of Hypothesis 1. Moreover, the findings of this evaluation will provide a new benchmark in the evaluation of appraisal mechanism in computational models of emotion. The following section details the methodology used for the evaluation and also the results obtained.

5.3.1 Methodology

The evaluation of the cognitive appraisal component in EEGS involves two steps given below. Since the validation of the domain-independence is secondary to current hypotheses, it will be presented only in the discussion section.

(i) Evaluating the accuracy of the computed appraisals.

(ii) Validating the domain-independence in computing appraisals.

In order to test the accuracy of EEGS in computation of appraisals, I used the data obtained from Study B since Study A had only emotion related data without the ratings for corresponding appraisals. Remember that the scenario used in Study B had 11 core emotion eliciting actions. Therefore, for each appraisal, an average rating of 47 respondents for each action was calculated. For example, for the appraisal variable *desirability* in Scenario 1, 11 averages (of 47 ratings), *i.e.* one for each action, were calculated. Since, there were six appraisal variables examined (namely *desirability*, *praiseworthiness*, *appealingness*, *deservingness*, *familiarity* and *unexpectedness*), the averaging procedure resulted in a 11×6 matrix of appraisal values from the survey data, where each column represents an appraisal variable and each row represents an action for the specific scenario. I will denote such matrix with *target appraisal matrix*.

Similarly, system data was obtained by simulating the Scenario 1 (Two Strangers in a Park) in EEGS. The recorded appraisal values for each action were in another 11×6 matrix of computed appraisals (I denote this as *system appraisal matrix*). Then, these appraisal values obtained from EEGS were compared against the average appraisals provided by human participants. The results of the experiment are summarised in the following section.

Appraisal Variables	Mean Absolute Error	Median Absolute Error
desirability	11%	9%
praiseworthiness	29%	31%
appealingness	49%	41%
deservingness	11%	10%
familiarity	45%	43%
unexpectedness	32%	35%
Overall	30%	28%

Table 5.2 Error in appraisal computation of EEGS.

5.3.2 Results

The *system appraisal matrix* was compared to *target appraisal matrix* by computing the difference of corresponding appraisal elements. I denote the difference matrix as *appraisal error matrix*. Since, the average of signed error tends to discount the opposite values thereby pushing the error towards zero and giving a false impression of high accuracy, I computed an *absolute appraisal error matrix* by calculating the absolute value of the individual error element. Since, each column of the *absolute appraisal error matrix* represents an appraisal variable, computing the mean of a column gave mean absolute error in computation of that appraisal variable. Similarly, the measure of median error was also calculated for each appraisal variable.

As shown in Table 5.2, the mean absolute error in appraisal computation ranged from 11% (in case of *desirability*) to 49% (in case of *appealingness*). Similarly, the median absolute error ranged from 9% to 41%. This suggests that the highest achieved mean accuracy was for the appraisal variables *desirability* and *deservingness* with a figure of 89%. Figure 5.2 graphically represents the accuracy in computation of various appraisal variables.

Calculating the overall accuracy for all the appraisal variables together, a mean accuracy of 70% was achieved. Figure 5.3 shows a graphical comparison of overall mean accuracy with overall mean error (considering individual appraisals of all the appraisal variables) in computation of the variables. As previously stated, since EEGS is the first computational model of emotion performing a comparison of computed appraisals with human data, no direct comparison with other emotion model could be done. Yet, an overall average accuracy of 70% and an accuracy of up to 89% in case of some variables is a promising result, given the fact that state of the art achievements in accuracy of emotion related research have not been encouraging (although not directly related to appraisal computation, a reader may see some examples from the works of

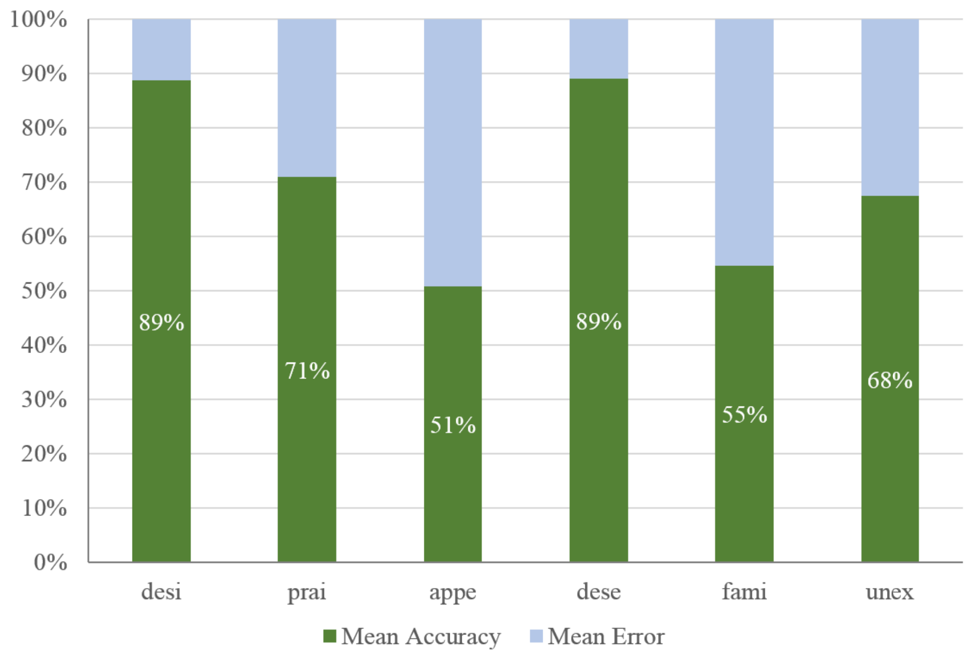


Figure 5.2 Accuracy in computation of various appraisals by EEGS as compared to appraisals rated by human participants in the given scenario.

Poria et al. (2016) and Cho et al. (2018) in emotion recognition). This finding validates the 'first part' of Hypothesis 1, which states that:

> *Assumed a given somato-visceral reaction $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ of the receiver in a given context* **C** *for a received behaviour* **B**, $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ *allows a computational cognitive appraisal process $\mathcal{C}$ to compute appraisal of the situation. In turn, the resulting appraisals can be used to compute the emotional intensities (resembling human-like emotions) for the given context.*

An additional contribution of the evaluation methodology used in the above context is that it can be adopted by future computational emotion models for the evaluation and bench-marking of their models.

While an overall mean accuracy of 70% in computing appraisal variables is a great achievement of EEGS, I also wanted to check if the appraisals computed by EEGS are significantly different from human appraisals or not. Therefore, I performed a paired t-Test at 95% confidence interval to compare the EEGS appraisals and human appraisal ratings. The test was conducted individually for each appraisal variable as well as all the appraisal pairs considered together. The null hypothesis was that the appraisals computed by EEGS are not significantly different from the appraisal ratings from humans. Table 5.3 shows the summary of the test.

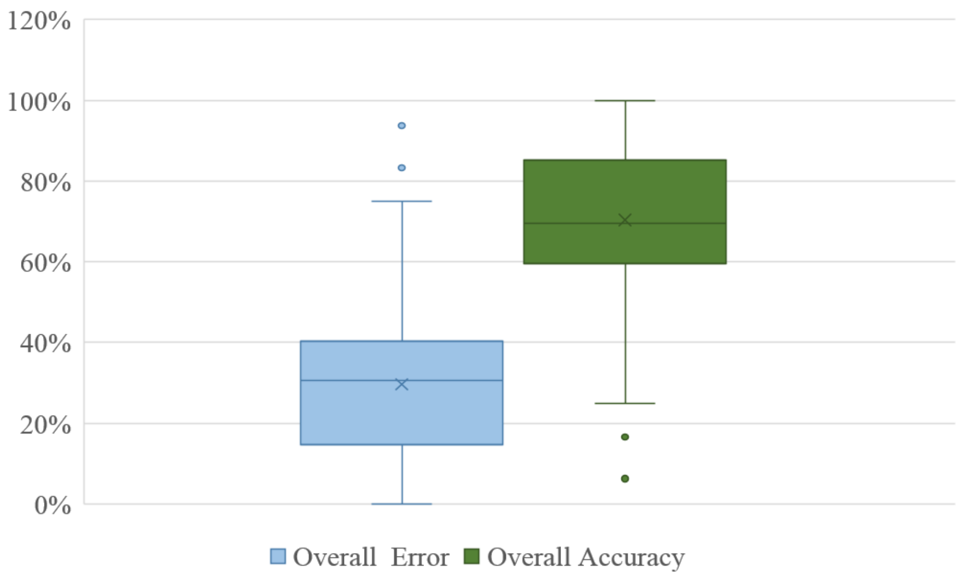


Figure 5.3 Accuracy of the overall appraisal of EEGS compared to the error in appraisal computation.

Appraisal Variables	t-value	p-value	Significant Difference?
desirability	-0.647	0.532	No
praiseworthiness	1.047	0.3195	No
appealingness	-0.275	0.7891	No
deservingness	-1.15	0.277	No
familiarity	-14.485	0.00 < 0.05	Yes
unexpectedness	8.554	0.00 < 0.05	Yes
Overall	-0.585	0.5604	No

Table 5.3 Paired t-Test to compare the appraisals computed by EEGS to the ratings provided by human participants.

From the conducted paired t-Tests, I did not find enough evidence to reject the null hypothesis for *desirability*, *praiseworthiness*, *appealingness* and *deservingness*. This is a good indicator to show the ability of EEGS model to compute the majority of appraisal variables in a similar way as human assessors, thus partly supporting the validity of Hypothesis 1. However, for the appraisal variables *familiarity* and *unexpectedness*, a significant difference was found (p-value = 0.00 < 0.05). Yet, when the test was conducted together for all appraisal variables, no significant difference was found between the computed appraisals and the ratings provided by human participants. These results suggest that the appraisals computed by EEGS closely resembles the appraisals made by humans in most situations.

Figure 5.4 Desirability (appraisal) dynamics of EEGS for two different scenarios – (i) Two Strangers in a Park and (ii) Husband and Wife.

5.3.3 Additional Discussion

Although the above findings support the first part of Hypothesis 1, a secondary evaluation was conducted to demonstrate the capability of EEGS to appraise events in domain-independent manner. In order to test this capability, EEGS was simulated with multiple scenarios introduced in Section 5.2.1. In each independent scenario, EEGS was able to compute appraisals without the need to change the computation mechanism presented earlier in Chapter 4, Section 4.6.2. Figure 5.4 shows the dynamics of appraisal variable *desirability* in EEGS demonstrating how EEGS can compute appraisals in multiple domains (scenarios) with the same computation mechanism. This phenomenon shows that with the provided appraisal computation mechanism, EEGS is able to transform the somato-visceral reactions (first-order appraisals) to cognitive appraisals in a domain-independent manner.

5.4 Stage 2: Affect Generation Evaluation

In Section 5.3, I presented a detailed evaluation of computation accuracy of various appraisal variables operationalised in EEGS, thereby validating the first part of the Hypothesis 1. In the next stage of appraisal processing, these variables are mapped into intensities of various emotions. The second part of Hypothesis 1 states that the appraisals computed in such a manner can be used to predict emotion intensities. This section will deal with the evaluation of the appraisal-emotion mapping mechanism employed in EEGS, for the validation of second part of Hypothesis 1.

In Chapter 4, Section 4.7.2, I presented an algorithm employed in EEGS that would allow the model to learn the association of appraisal variables to emotions operationalised by the factors of personality and mood. This section will deal with the formal evaluation of the appraisal-emotion association learning mechanism in EEGS and prediction of emotion intensities, which in turn will validate the second part of Hypothesis 1.

5.4.1 Methodology

In order to train and test the appraisal-emotion mapping process of EEGS, data collected from Study B was used because the survey questions in Study B consisted of appraisal ratings, emotion intensities as well as personality factors and mood dynamics, which are essential elements in the process of mapping appraisals into emotion intensities (see Section 5.2.2 for more details on how the data was collected).

The data collected was formatted to create a machine learning suitable data table with the following columns.

$$v_1 \;\; v_2 \;\; \ldots \;\; v_k \;\; O \;\; C \;\; E \;\; A \;\; N \;\; M \;\; e$$

v_1, v_2, ... v_k indicate the appraisal variables denoting how the situation is evaluated by the receiver of an action. O, C, E, A and N denote the five personality factors. M denotes the mood factor. e denotes the intensity of the emotion. Because of the complex nature of each emotion and to allow the methodology to be applicable in any artificial agent having any number of emotions, I ran the training algorithm for each emotion type separately. For example, links associated with a particular emotion (say *joy*) are trained individually. While it can be argued that intensities of complex emotions like *gratitude* can be determined by the intensity of basic emotions (as discussed in Chapter 4, Section 4.7.1), I have not considered such a relation in the implemented network of appraisals and emotions in EEGS. Accordingly, training the appraisal-emotion network at once for all emotions or individually for each emotion does not cause any difference in accuracy. For the same reason, I decided to train the network for each emotion individually which offered simplicity in the training process and helped in avoidance of probable learning errors. Therefore a separate dataset was created for each emotion type. For each emotion, the set of $\{v_1 \;\; v_2 \;\; \ldots \;\; v_k \;\; O \;\; C \;\; E \;\; A \;\; N \;\; M\}$ was used as the input vector. As the survey scenario had 11 core emotion inducing actions and 47 unique responses, I ended up having a dataset containing 517 (= 11 x 47) rows for each emotion type. Data set containing 517 records ($n > 500$) is usually considered a good sample size in machine learning applications with less than 100 parameters (Lewis, 1992) – which is the case in the current research.

Out of the total data rows, 70% of the rows *i.e.* 362 rows (selected from random locations) were used for training the network and remaining 30% selected in random order were used to test the accuracy of the trained network *i.e.* 155 rows in test dataset. The accuracy was determined by comparing the predicted emotion intensity with the emotion intensity provided by individual survey respondent. For each emotion type, the algorithm was run for 100 epochs[3], where in each epoch all the rows from the training set were selected but in random order. Thus one complete training session consisted of 100 * 362 = 36,200 weights update (iterations). Experimental results showed that the algorithm converged well even with 50 epochs, yet we used 100 epochs for the purpose of certainty. This process was repeated for 10 times, where each time the training and test data was randomly re-sampled as discussed previously.

The mechanism of establishing the association among appraisals and emotions involved two phases:

(i) Training the appraisal-emotion network.

(ii) Testing the learned network for accuracy of emotion intensity prediction.

The objective was to learn the appraisal-emotion association in EEGS in a data-driven manner based on the personality and mood factors available from human ratings. In this process, each of the 5 personality factors (*openness*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*) as well as mood factor would be assigned an weight based on the extent to which these factors influence the mapping of appraisals to emotion intensities. It should be noted that the data collected in Study B did not account for the emotions *happy-for* and *sorry-for*, therefore these emotions are not included in the evaluation presented in the following sections. An analysis of the training and learning process will be presented in Section 5.4.3.

5.4.2 Results

Since the appraisal-emotion network was trained by 70% of the data rows, a test data set consisting of non-overlapping 30% of the original data was used to test the accuracy of the learned model in predicting emotion intensity. The absolute difference between the predicted intensity and expected intensity was calculated as an error. In these kinds

[3]One epoch corresponds to the feeding of the complete dataset once to the learning algorithm.

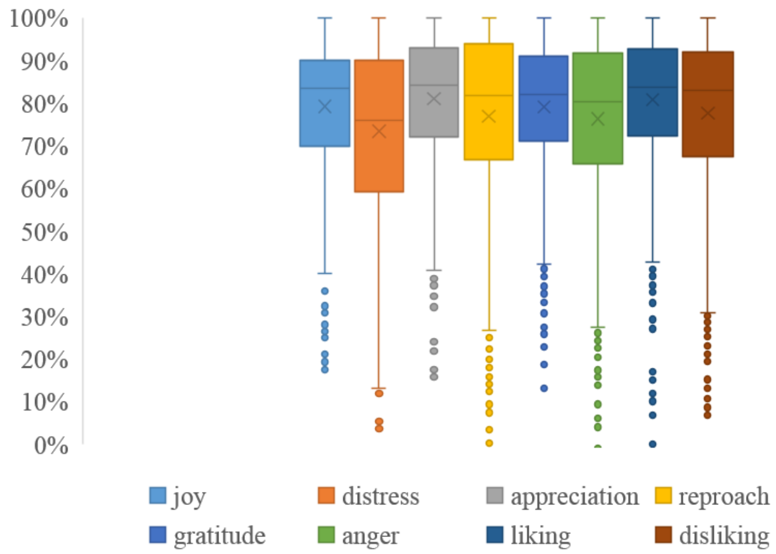


Figure 5.5 Overall accuracy in prediction of eight emotions over the 10 training-testing sessions for each of the emotions.

of evaluations, it is important to note that the prediction error should be considered as the absolute value of the difference in prediction instead of raw signed difference. Using the signed error values causes an impression of false accuracy by cancelling the positive and negative errors thereby shifting the error towards zero. Since, the intensity in the data collected ranged from 0 to 1, the accuracy of prediction of *joy* intensity in an instance was calculated as (1 - error), which would give a number between 0 and 1. Multiplying the numeric accuracy obtained in this manner gives the percentage accuracy of prediction.

Emotion	Overall Mean Accuracy	Overall Median Accuracy	Standard Deviation
joy	79.2%	83.4%	0.151
distress	73.4%	76.1%	0.187
appreciation	81.1%	84.2%	0.145
reproach	77.0%	81.8%	0.199
gratitude	79.1%	82.1%	0.152
anger	76.3%	80.2%	0.190
liking	80.8%	83.7%	0.148
disliking	77.7%	82.9%	0.191
Mean	**78.1%**	**82.2%**	**0.173**

Table 5.4 Overall accuracy in prediction of various emotion intensities.

Figure 5.5 shows the accuracy in prediction of the intensities of various emotions. The overall accuracy of each emotion represents the combined accuracy in prediction of intensity of that emotion over the 10 testing sessions after the completion of the corresponding training process. Since, each testing data set consisted for 155 rows, the overall accuracy for each emotion shown in Figure 5.5 represents a comparison 1,550 accuracy tests i.e. 155 accuracy readings for each of the 10 sessions. Table 5.4 numerically summarises the results shown in Figure 5.5. It is evident from the table that the mean accuracy in the prediction of the intensity of various emotions ranged from 73.4% for *distress* emotion to 81.1% for *appreciation* emotion with an average mean accuracy of 78.1% for all the emotions considered. Likewise, median accuracy ranged from 76.1% for *distress* emotion to 84.2% for *appreciation* emotion with an average median accuracy of 82.2% when all the emotions were considered. For each emotion, the standard deviation of the prediction accuracy for individual emotions was minimal ranging from 0.145 to 0.199 with an average of 0.173 for all the emotions together. It is a promising outcome where the accuracy in intensity prediction of all the emotions are quite close to each other even if the model was trained separately for each emotion.

As in the case of prediction accuracy, different emotions operationalised in EEGS also exhibited similar evolutionary trend in the learning process. In order to examine the evolution of the model, for each emotion, in a particular training session, the interim trained model was used to predict the emotion intensity in the test dataset (155 rows) after each epoch and the error in prediction recorded. Figure 5.6 shows similar evolutionary trend for all the emotions considered. It can be seen that the prediction error decreases with each learning epoch and remains relatively stable between 50^{th} and 100^{th} epochs. This suggests that using the implemented algorithm, the model was learning the association of appraisal variables to emotions in an effective and consistent manner for all the emotions in EEGS. The prediction accuracy for emotion intensities obtained suggest that the implemented algorithm for mapping appraisals to emotion intensities can accurately predict emotion intensities using the computed appraisal variables, thereby validating the second part of Hypothesis 1.

Although the performance of EEGS in learning the association between appraisal variables and emotions seems optimal, it was important to compare the obtained results of EEGS with previous relevant researches as well. I explored the emotion modelling/analysis literature in order to find some work involved in using data driven approaches to establish the relationship between appraisal variables and emotions. Only a few hand countable studies were found (Meuleman and Scherer, 2013; Nguwi and Cho, 2010; Tong et al., 2009). Of these, only two studies involved the actual machine

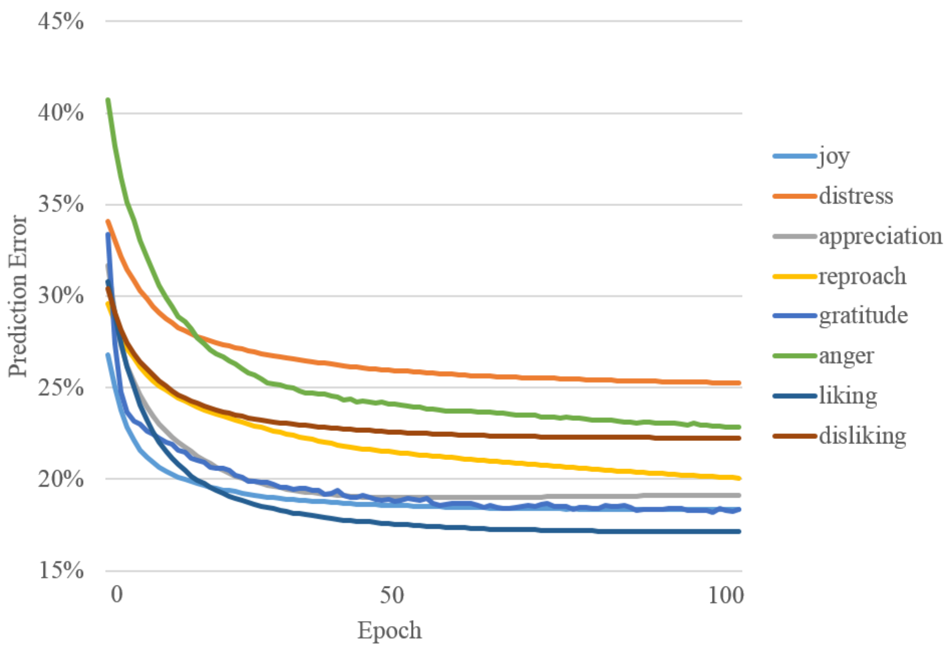


Figure 5.6 Evolution in learning of the association between appraisal variables and emotion for eight different emotions for a training session where the test data set was used for prediction of emotion intensity after each epoch of the session.

learning approaches for investigating the relationship between the appraisal variables and emotions (Meuleman and Scherer, 2013; Nguwi and Cho, 2010).

The work of Tong et al. (2009) employed methods to collect appraisal and emotion related data from human participants in the two different conditions namely (i) Vignette method – where participants imagined themselves in a presented scenario and provided the ratings for appraisals and emotions, and (ii) Ecological Momentary Assessment method, where participants reported the appraisals and emotions as they occurred or immediately after the experience. Their goal was to examine how each appraisal is functionally related to a particular emotion *i.e.* how does the emotion change as the appraisal changes? They concluded that most appraisal-emotion associations show an ogival pattern (Tong et al., 2009). Although, their research served as an example of how data collected from humans can be used to investigate the relationship between appraisals and emotions, their findings do not offer quantitative advantages to the modelling such relationships in computational models of emotion. More specifically, in the context of current research, where the goal was to determine how the factors of personality and mood quantitatively after the mapping of appraisals to emotion intensities, the findings of Tong et al. (2009) does not provide more insights except for the use of data-driven approaches.

Nguwi and Cho (2010) performed a study using the ISEAR data set (International Survey on Emotion Antecedents and Reactions) collected over the course of several years by Scherer and Wallbott (1994). Nguwi and Cho (2010) adopted a method called Support-Vector-Based Emergent Self-Organising Map (SVESOM) to predict the emotions based on several variables. The classification was performed separately for each emotion as a *two class problem* (Nguwi and Cho, 2010). For example, the classification of the emotion *joy* consists of predicting if a set of appraisals causes *joy* or *non-joy* emotion (Nguwi and Cho, 2010). Such a binary division was not the choice of the authors but a compulsion because the ISEAR data set (Scherer and Wallbott, 1994) does not ask for rating of emotion intensity level from the participants and hence does not have sufficient information to perform this kind of prediction. Considering the prediction variable in a binary level is a serious limitation as it prevents the appropriate evaluation of the approach for predicting the exact intensity of the associated emotion. Moreover, Meuleman and Scherer (2013) stress that emotion data collected from naïve participants are likely to be noisy and yield low prediction accuracy. It was probably because of this reason, Nguwi and Cho (2010) were not interested in performing standard test of accuracy for the prediction of their model and instead opted for F-Measure and Geometric Mean Measures (Nguwi and Cho, 2010). As such, they conclude that their model outperforms (F-measure = 0.82) the SVM model of Danisman and Alpkocak (2008) (F-measure = 0.675). While the approach of Nguwi and Cho (2010) was successful in achieving their goal of constructing a visual map of high-dimensional emotion data, the same may not be useful in predicting the actual intensity of various emotions.

Meuleman and Scherer (2013) also used the latest version of ISEAR data collected from the Geneva Emotion Analyst (Scherer, 1993). Unlike Nguwi and Cho (2010), they performed standard test of accuracy in prediction of emotion based on the various appraisal variables considered in the data. Meuleman and Scherer (2013) used different machine learning algorithms to test the predictive accuracy in different cases. First, they tested the accuracy in differentiating positive and negative emotions (case 1). Second, they tested the accuracy in differentiating four emotion clusters (*happiness*, *anger*, *shame/guilt* and *distress*) (case 2). Third, they tested the accuracy in differentiating the 12 emotion classes namely *sadness*, *fear*, *despair*, *anxiety*, *shame*, *guilt*, *rage*, *disgust*, *irritation*, *joy*, *pleasure*, and *pride* (case 3). In case 1, they achieved an accuracy of 93%, which suggests that differenting the positive emotions from negative based on the appraisal variables is quite an accurate task for the machine learning algorithms considered. In case 2, as they narrowed down the classification process by presenting 4 emotion clusters, the differentiation accuracy dropped down to an average of 60.5%.

Interestingly, in case 3, when the classification task was even more narrow with the requirement of differentiating 12 emotion types, the best classification accuracy was achieved using Random Forest (Svetnik et al., 2003) algorithm with an average of 27.9%. The classification accuracy achieved by Meuleman and Scherer (2013) is summarised in Table 5.5.

Classification Type	Average Accuracy
Positive/ Negative	93%
4 Emotion Clusters	60.5%
12 Emotion Classes	27.9%

Table 5.5 Best classification accuracy obtained by Meuleman and Scherer (2013).

As evident from Table 5.5, the classification accuracy in the study of Meuleman and Scherer (2013) dropped significantly as the requirement of the task specificity increased. They have not performed the intensity prediction test in their study because the utilised data set lacks the intensity values for the emotion classes. Given the difference of their task (classification) with my task (regression of emotion intensities) it is not possible to make an exact and reliable comparison of the two sets of accuracies. However, if we assume that by regressing the elicited emotion intensities for each input interaction we can also gather a valid cue to classify such interaction in a single emotional class, we can view their classification task as closely related to our regression one even though the results are not easy to compare. The ISEAR data set used by Meuleman and Scherer (2013) can not be directly utilised in the training and validation of the appraisal-emotion mapping of EEGS because of two main reasons. First (i) the data consists of the appraisals proposed by Scherer (2001) which can not be seamlessly matched with the appraisal variables offered by the OCC theory (Ortony et al., 1990), which forms the backbone of the design of EEGS. Second, (ii) the goal of current research is not only to establish a quantitative relationship between appraisal variables and emotions (which is still not offered by Meuleman and Scherer (2013)) but also to integrate the aspects of personality and mood in the mapping process. ISEAR data set does not account for the factors of personality and mood.

Effect of Personality and Mood Factors on Emotion of EEGS[4]

Earlier in this section, I presented an evaluation of the accuracy of the trained model in predicting the intensities of corresponding emotions, validating second part of Hypothesis 1. However, since the factors of personality and mood determined the weights of

[4] Most of the content in this subsection is adapted from Ojha et al. (2018a).

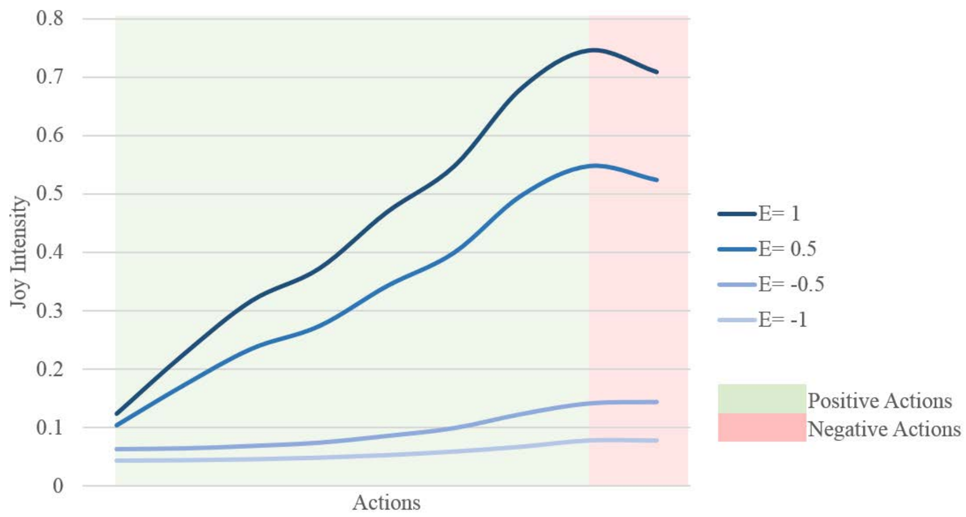


Figure 5.7 Difference in intensity of *joy* emotion in Scenario 1 (Two Strangers in a Park) when the personality factor of *extraversion*(E) is altered. Adapted from Ojha et al. (2018a).

association between appraisal variables and emotion intensities in EEGS (see Chapter 4, Section 4.7.2 for more details), it is also important to evaluate how the difference in these aspects causes difference in emotion dynamics of EEGS. In this sub-section, I will present an analysis of how EEGS allows the difference in personality and mood factors to influence the emotion intensities.

In order to examine the effect of personality factors on emotions, I simulated Scenario 1 – Two Strangers in a Park and Scenario 3 – Husband and Wife (*see* Apendix A for the full description of the scenarios). I wanted to investigate the variation in emotional response of EEGS for factors of *extraversion* and *neuroticism* because these factors are most widely studied in relation to positive and negative emotionality respectively. In order to test the effect of the personality factor *extraversion*, I set all other personality factors to be constant and changed the factor of extraversion from -1.0 to +1.0 (where -1.0 indicated very introverted person and +1.0 indicated very extroverted person). Figure 5.7 shows the variation in *joy* emotion dynamics of EEGS for the same scenario *i.e.* Two Strangers in a Park when the personality factor of *extraversion* for EEGS is changed. It is evident from the figure that the intensity of *joy* is the highest when the level of extraversion is set to be 1 and the intensity levels gradually decline as the value for extraversion is switched towards -1 (indicating introverted personality). Additionally, the slope of the curve is more steep for the case where there is high degree of extraversion suggesting that extroverts are more likely to be happy compared to introverts (Revelle and Scherer, 2009).

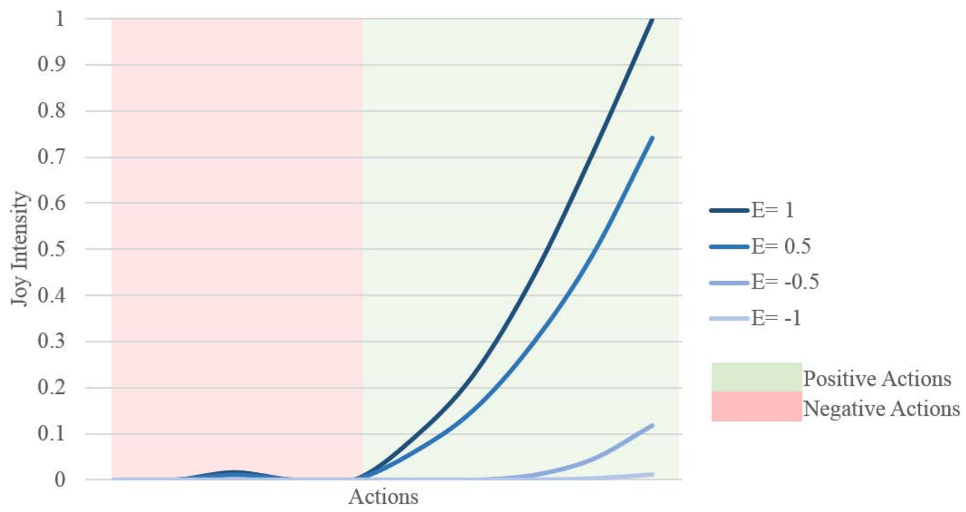


Figure 5.8 Difference in intensity of *joy* emotion in Scenario 3 (Husband and Wife) when the personality factor of *extraversion*(E) is altered. Adapted from Ojha et al. (2018a).

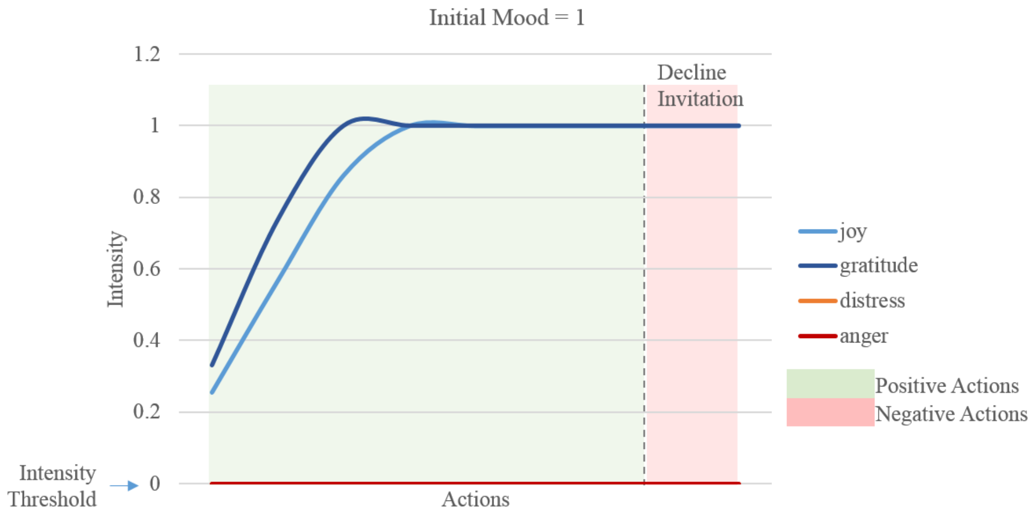


Figure 5.9 Emotion dynamics of EEGS when initial mood is very positive in Scenario 1 (Two Strangers in a Park). Adapted from Ojha et al. (2018a).

Similar phenomenon is also obtained for the scenario of interaction between husband and wife (Scenario 3). Figure 5.8 shows how the difference in the personality factor of *extraversion* causes difference in the level of *joy* intensity experienced by EEGS in Scenario 3. These findings suggest that the learned weights for the personality factors allow the model to operationalise the influence of various factors in an effective and plausible manner.

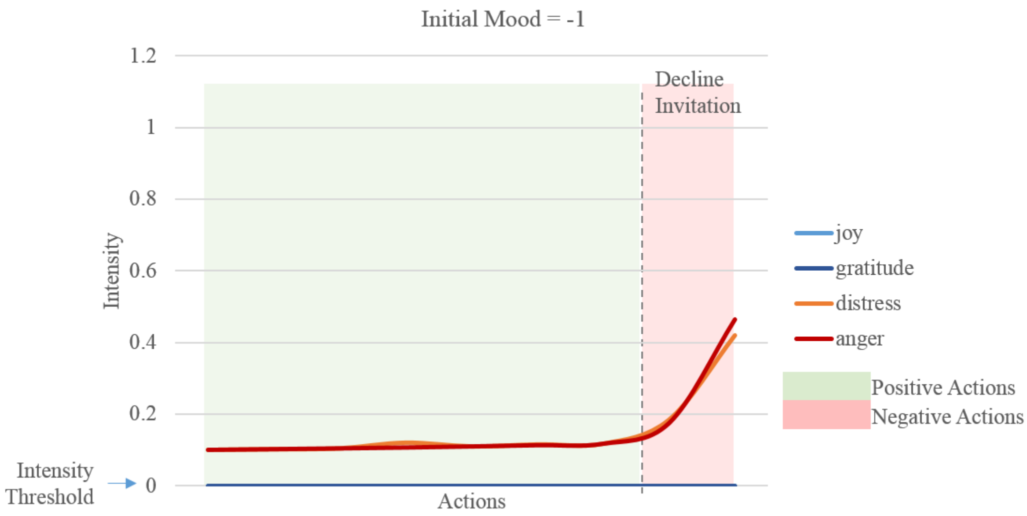


Figure 5.10 Emotion dynamics of EEGS when initial mood is very negative in Scenario 1 (Two Strangers in a Park). Adapted from Ojha et al. (2018a).

In addition to personality factors, I also wanted to investigate the emotion dynamics of EEGS by altering the initial mood. For this experiment, personality factors were not considered because they are likely to affect the mood state thereby obscuring the true interaction between mood and emotions. Figure 5.9 shows how an initial positive mood increases the tendency of EEGS to experience positive emotions (*joy* and *gratitude*) and decreases the tendency to experience negative emotions (*distress* and *anger*). While the emotions of *joy* and *gratitude* reach a saturation intensity of 1.0 in the course of interaction, the emotions of *distress* and *anger* remain below the threshold intensity *i.e.* 0.0. Interestingly, even with a negative action of 'decline invitation', the positive emotions do not drop significantly because of cumulative bias caused by positive initial mood and positive emotional experience in the course of interaction.

However, an opposite phenomenon is observed if the initial mood is set to be very negative *i.e.* -1. Figure 5.10 shows how the initial mood of -1 leads EEGS prevents the emotions of *joy* and *gratitude* from rising above the threshold level for same scenario and same set of actions. Additionally, as opposed to Figure 5.9, the emotions of *distress* and *anger* remain active thought the interaction and begin to rise sharply after the 'decline invitation' action. As such, the discussion so far suggests that EEGS is capable of effectively integrating the aspects of personality and mood thereby altering its emotion dynamics based on these factors.

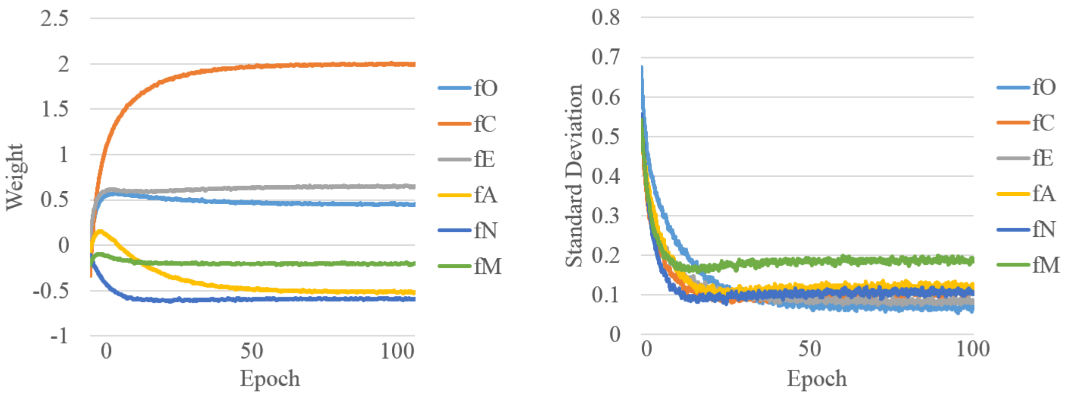
5.4.3 Additional Discussion

Previous section presented the accuracy of EEGS model to predict emotion intensities and validated the second part of Hypothesis 1. In this section, I will discuss some secondary findings to enrich the understanding of the presened appraisal-emotion mapping mechanism in EEGS model. Figure 5.11a shows how the algorithm learns the association of appraisal variable *desirability* to emotion *joy* by evolving the weights of personality factors and mood by averaging the learned weights in the 10 training sessions. It is evident from the figure that most of the factors for the association of *desirability* to *joy* converge at around 50 epochs and the learned weights remain stable till 100 epochs. Likewise, Figure 5.11b shows the variation of weights as the epochs progress represented as standard deviation of each weight across 10 sessions. The standard deviation of the weights across 10 sessions drops rapidly in the early epochs and following the similar pattern to the learning of the weights, the standard deviation achieved stability after about 50 epochs. This indicates that with the provided data, the association between the appraisal variable *desirability* and emotion *joy* could be effectively learned in less than 100 epochs.

Figure 5.12 shows the accuracy of the learned model to predict the intensity of *joy* emotion (a) across various testing sessions and (b) overall accuracy for all the testing sessions combined. Average accuracy across the individual testing sessions ranged from 77.4% to 81.7% and the overall accuracy for all the testing sessions combined was 79.2% (Median = 83.4%, SD = 0.151). It should be noted that this accuracy represents the overall accuracy of *joy* emotion in Table 5.4. From Figure 5.12a, it is evident that the accuracy in prediction remained quite consistent across the sessions with a minimal standard deviation. This suggests that the model was effectively learning the required parameters *i.e.* the weights for the personality and mood factors.

While the final learned model was impressively accurate in predicting the intensity of *joy* emotion, I was interested to investigate how the model evolved with each epoch in a particular training session. Therefore, rather than waiting till the learning of the model completes, I checked the accuracy of the model in predicting *joy* intensity using the test data set *after each epoch*. Figure 5.13 shows how the error in prediction of the *joy* intensity decreased with the increase in the training epochs. This exhibits that the model was effectively learning the parameters and evolving after each training epoch.

As in the case of *joy* emotion, the learned model exhibited high level of prediction accuracy of the intensity of *distress* emotion as well. Figure 5.14 shows the learning curves for the weights of personality of mood factors in relation of the appraisal variable *desirability* with the emotion *distress* averaged over 10 training sessions as well as the



(a) Average learning of the personality and mood factor weights for the association of appraisal variable *desirability* to the emotion *joy* over the 10 training sessions.

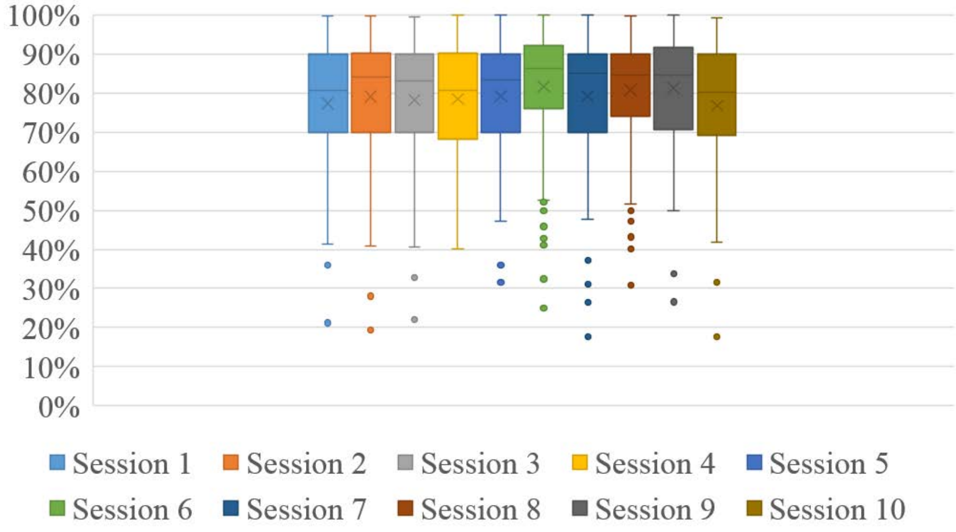
(b) Standard deviation in the learned weights for the association of appraisal variable *desirability* to the emotion *joy* across the 10 training sessions.

Figure 5.11 Learning trend for the association of the appraisal variable *desirability* to the emotion *joy* averaged over 10 training sessions (Figure 5.11a) and the variation in the learned weights across the training sessions (Figure 5.11b.)

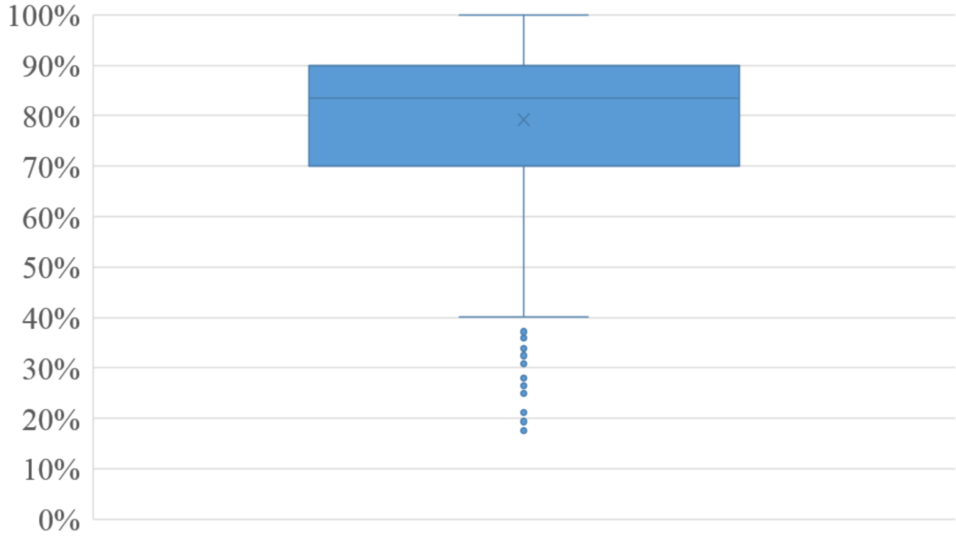
variation in standard deviation among the learned weights across the sessions with the increasing number of epochs.

Interestingly, if we observe the learned weights for the *joy* and *distress* emotions, it can be seen that the learning curves show somewhat mirrored pattern. Figure 5.15 shows how the factors f_O, f_C, f_E, f_A and f_N show an approximate mirrored pattern for the association of appraisal variable *desirability* to the emotions *joy* and *distress*. In the personality literature, the personality factors of *neuroticism* and *extraversion* are considered to have a direct link with the emotions of *joy* and *distress* (Corr, 2008). People who are extrovert in nature are said to have higher tendency of experiencing positive emotionality (*i.e.* joy) than negative (*i.e.* distress) (Revelle and Scherer, 2009). If we closely examine the learned weights for the personality factor *extraversion* (f_E) in Figure 5.15c, the weight of f_E for *joy* evolves towards positive value and that of *distress* towards negative. This suggests that my proposed algorithm (see Chapter 4, Section 4.7.2) assigns positive association of the personality factor *extraversion* to the emotion of *joy* and negative to *distress*. Likewise, Figure 5.15e, which shows the learning of the weights for the personality factor *neuroticism*, exhibits that the model learns a positive association of f_N to the emotion of *distress* and negative to the emotion of *joy*. This is in line with the assumptions of personality theories that neurotic people tend to experience more *distress* than *joy*.

However, surprisingly, although all the five personality factors (f_O, f_C, f_E, f_A and f_N) exhibited a mirrored pattern for the *joy* and *distress* emotions, the mood factor f_M



(a) Accuracy in intensity prediction of *joy* emotion across 10 testing sessions.



(b) Overall accuracy in intensity prediction of *joy* emotion over the 10 testing sessions.

Figure 5.12 Accuracy in prediction of intensity of *joy* emotion during testing phase.

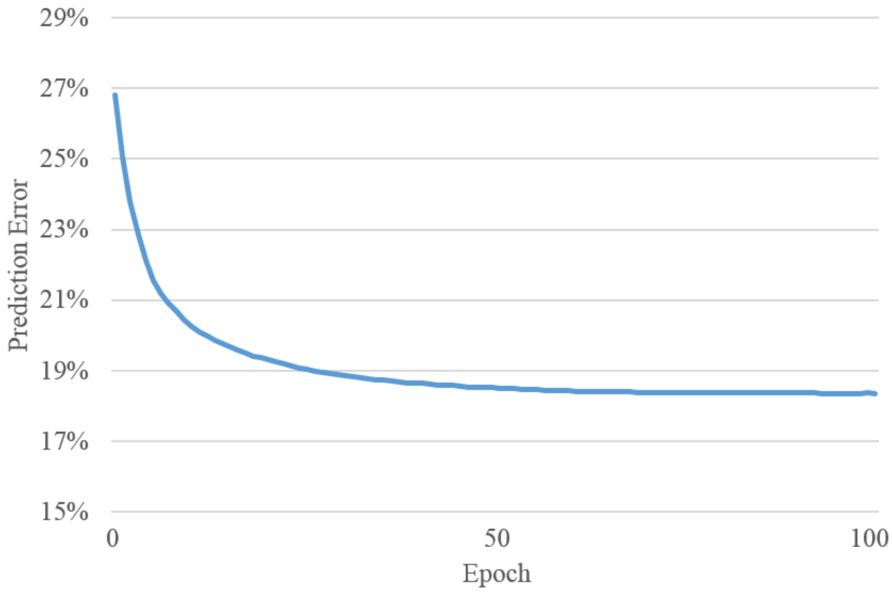
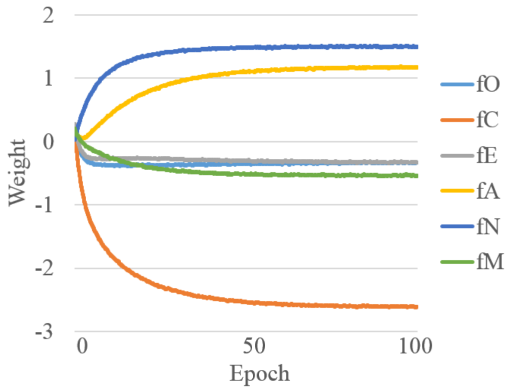


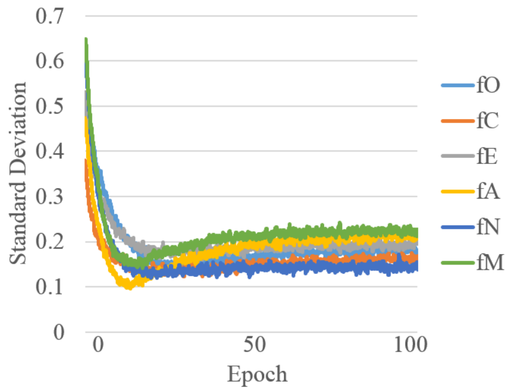
Figure 5.13 Evolution of the learned model with the increasing epochs.

did not show a similar learning of the weight (see Figure 5.15f). Final learned weights for both the *joy* and *distress* lie in the negative region. Although, the weight for *joy* emotion lies slightly towards positive region, these weights did not reflect the opposite nature of these emotions. One of the reasons for such a learning of mood factor f_M could be the influence of personality factors on mood during the learning process. In order to investigate this, I performed a separate training for *joy* and *distress* emotions where, only the mood factor f_M was considered in the learning process by excluding all the personality factors (f_O, f_C, f_E, f_A and f_N). Figure 5.16 shows the resulting learned weights for the mood factor for *joy* and *distress* emotions. Unlike the curve in Figure 5.15f, the weights learned in Figure 5.16 show mirrored pattern. This is in line with the assumption that a person in positive mood tends to experience the emotion of *joy* than *distress* (Morris, 1992; Neumann et al., 2001). Achieving this type of learning behaviour when the personality factors are not included in the process suggests that personality factors can strongly influence mood factor – as also suggested by other researchers (Rusting, 1998). Further investigation into how each personality factor might influence the mood factor is out of the scope of this dissertation. Yet, my findings might provide new insights to other researchers who might be interested in examining the relationships among these factors.

In order to determine the prediction accuracy of the model for *distress* intensities, similar method as used for *joy* emotion was applied *i.e.* 10 training-testing sessions of the newly generated training-testing datasets from the original dataset were run and



(a) Average learning of the personality and mood factor weights for the association of appraisal variable *desirability* to the emotion *distress* over the 10 training sessions.

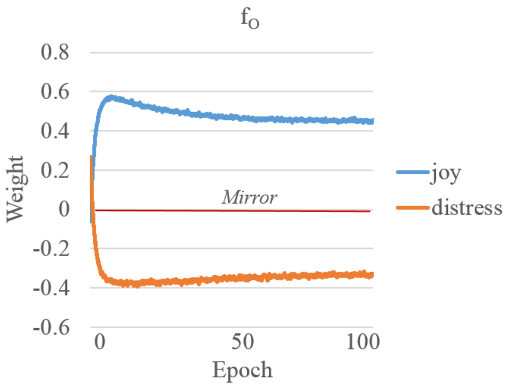


(b) Standard deviation in the learned weights for the association of appraisal variable *desirability* to the emotion *distress* across the 10 training sessions.

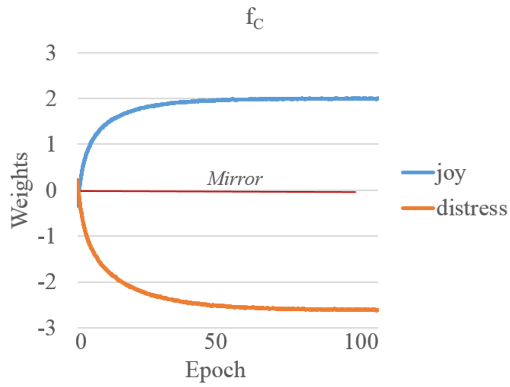
Figure 5.14 Learning trend for the association of the appraisal variable *desirability* to the emotion *distress* averaged over 10 training sessions (Figure 5.14a) and the variation in the learned weights across the training sessions (Figure 5.14b.)

error was calculated in each testing session. Prediction accuracy ranged from 71.5% to 74.6% in the individual testing sessions and an average accuracy of 73.4% (Median = 76.1% and SD = 0.187) was obtained for all the sessions combined together. Although, this prediction accuracy is slightly low compared to that of *joy* emotion, this is still a remarkable achievement as shall be discussed later in this section in comparison with previous research.

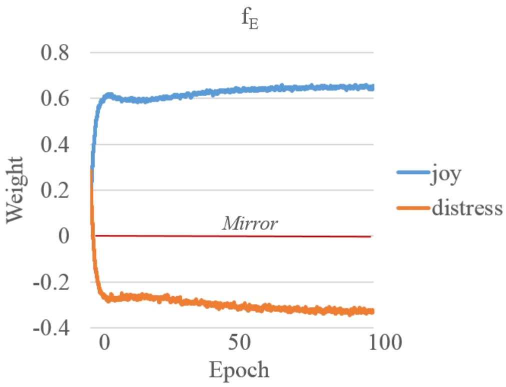
Emotions other than *joy* and *distress* in EEGS are determined by more than one appraisal variables. For example, the emotion of *anger* is determined by the appraisal variables *desirability*, *praiseworthiness* and *unexpectedness* (Ortony et al., 1990). This means that the association of each of these appraisal variables with the emotion of *anger* should be learned thereby converging to the weights of personality and mood factors for three different links *i.e.* learning of 18 different weights (a set of six for each of the three links). The presentation of all of those learning curves as shown for *joy* (see Figure 5.11a) and *distress* (see Figure 5.14a) is not necessary as all of the weights were learned with similar pattern as that of *joy* and *distress* for the appraisal variable *desirability* – yet with different learned weights for the factors depending on the association.



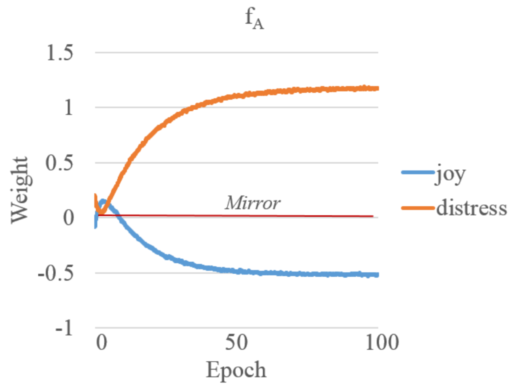
(a) Mirroring in the learned weights of f_O.



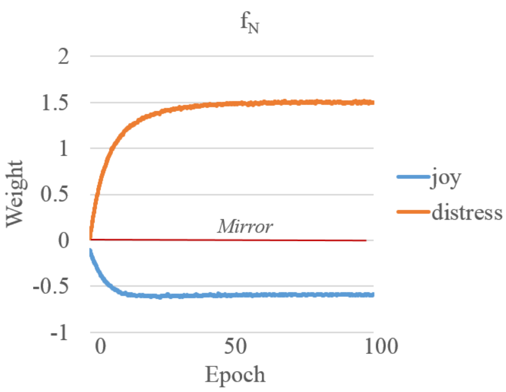
(b) Mirroring in the learned weights of f_C.



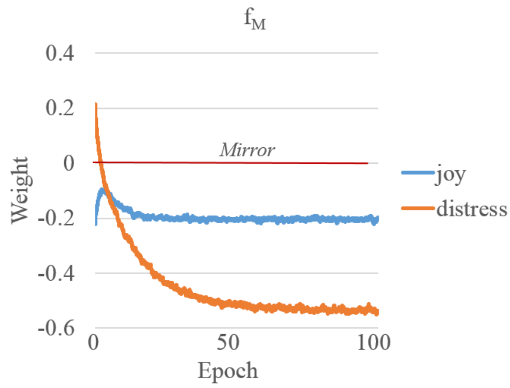
(c) Mirroring in the learned weights of f_E.



(d) Mirroring in the learned weights of f_A.



(e) Mirroring in the learned weights of f_N.



(f) Lack of mirroring in the learned weights of f_M.

Figure 5.15 Mirrored pattern in learning of the weights for personality factors for the association of the appraisal variable *desirability* to the emotions *joy* and *distress*. Surprisingly, the mood factor did not exhibit a mirrored pattern for *joy* and *distress*.

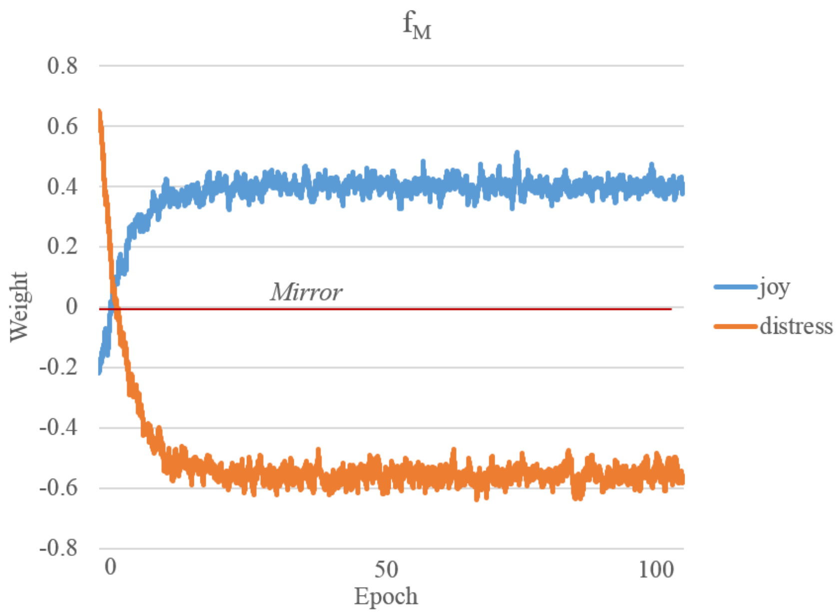


Figure 5.16 Mirrored pattern obtained for the weight of mood factor (f_M) when only the mood is considered in the learning process.

5.5 Stage 3: Affect Regulation Evaluation[5]

In the previous sections, I presented two of the stages in *three-stage evaluation* approach proposed in this dissertation. I concluded that EEGS is able to compute appraisals of a situation with high level of accuracy and also to accurately predict the emotion intensities by learning the association between appraisal variables and emotions – thereby validating Hypothesis 1. The mechanism of ethical reasoning in relation to emotion generation (which constitutes of the final stage of emotion processing in EEGS) has not yet been evaluated. In this section, I will present a discussion on the analysis of emotion convergence and regulation mechanism in EEGS which will be necessary to validate Hypothesis 2 (see Chapter 3, Section 3.2).

In Chapter 4, Section 4.8, I presented a discussion of the role of ethical reasoning in the process of emotion convergence and regulation in autonomous agents. The process of emotion regulation is critical mainly in case of social robots where they have to interact with people of various background and nature. I argue that *ethical reasoning* mechanism employed in EEGS allows robots in attaining more socially acceptable emotions compared to the approaches using *highest intensity* – where the emotion with the highest intensity is considered as the final emotional state and *blended intensity* – where the intensities of the elicited emotions are blended to determine a new intensity

[5]Most of the content of this section has been adapted from Ojha et al. (2017) and Ojha et al. (2018b).

value and a final emotion type to be attributed (see Chapter 4, Section 4.8.1 for detailed discussion on the matter). In order to test the validity of this claim, I compare the emotion dynamics of EEGS using the above three different approaches to reach to final emotional state.

5.5.1 Methodology and Results

For the evaluation of the final stage, naïve adults were asked to design realistic scenarios of interaction between two individuals (as described in Section 5.2.2), which was then used to evaluate the emotional responses of EEGS. As explained in Section 5.2.1 earlier, subjects were asked to come up with actions that an individual can perform on another, where a set of actions in sequence would complete one scenario. The obtained scenarios were then used to create survey questions to collect emotion data from participants as described in Study A (see Section 5.2.2).

The evaluation of the final stage of emotion processing in EEGS consists of two sub-steps namely – (i) *validation of the model's regulatory mechanism to accurately generate human-like emotions* and (ii) *validation of the social acceptability of the generated emotions* as a result of emotion regulation process guided by ethical reasoning in EEGS. I adopt a quantitative approach for the validation of accuracy to generate human-like emotions and qualitative approach to validate the social acceptability of regulated emotions. The choice of qualitative analysis for social acceptability is because of the fact that there are no benchmarks in the literature on what should be considered as socially acceptable. Therefore, considering the applications of human-robot interaction scenarios, in this dissertation, the evaluation of social acceptability in current experimental context has been done by examining whether an agent (robot) is able to reduce the negativity in the emotionality and hence behavioural responses (see Chapter 2, Section 2.5.2 for more discussion about social acceptability and its difference with believability).

Approach	Median Distance	Standard Deviation
Highest	2	2.3232
Blended	2	2.0228
Ethical	1	2.3140

Table 5.6 Comparison of median distances from the human assessment for (i) highest intensity, (ii) blended intensity and (iii) ethical reasoning approaches in EEGS.

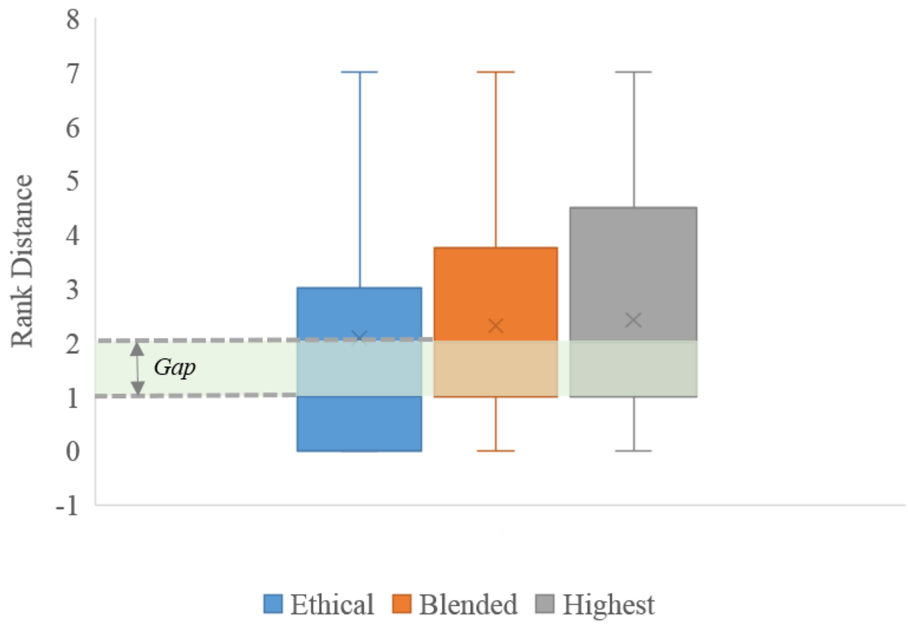


Figure 5.17 Comparison of the rank distance from the average human rating of the emotion intensity generated by (i) highest intensity, (ii) blended intensity and (iii) ethical reasoning approaches. Redrawn after Ojha et al. (2017).

Validation of Regulation Accuracy

This section describes how EEGS performs overall on all the six scenarios presented in Section 5.2 (the details of the scenarios are provided in Appendix A).

I wanted to examine how close to the human ratings are the emotions converged by EEGS using the (i) *Highest intensity*, (ii) *Blended intensity*, and (iii) *Ethical reasoning* approaches separately. For this, the dataset containing the rated intensities of emotions as obtained from Study A was considered (see Section 5.2.2 for details on how the data was collected). The emotion ratings from human participants for each interaction in each scenario were ranked based on the average score and the emotion with the highest average score (considering the response of all the participants for that interaction) was considered to be at rank 1 and the emotion with lowest score was ranked 8.

Next, for each of the above mentioned emotion conversion strategies, the final emotional state achieved as the outcome of each strategy was compared to the rankings of human rating. If the emotion matched with the emotion ranked as 1 in human ratings, the distance of the model's emotional state from humans would be considered as 0 (zero) for that interaction. Similarly, if the emotion matched with the emotion ranked as 2 for average human rating, the distance of the model's emotion state would be considered as 1 and the trend would follow. As such, if the emotion state reached by the model

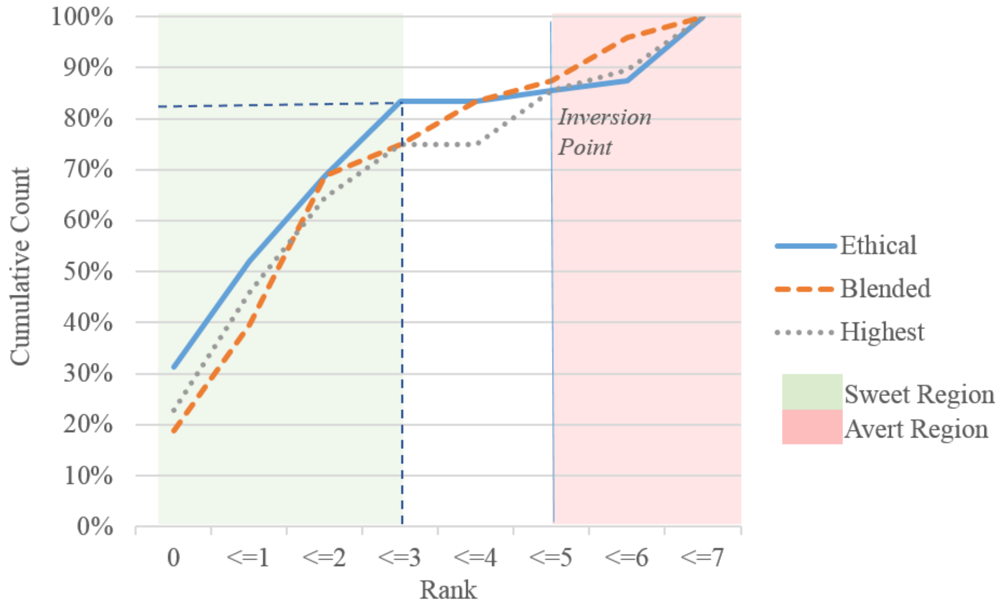


Figure 5.18 Cumulative rank distance from the average human rating for the emotion intensity generated by (i) highest intensity, (ii) blended intensity and (iii) ethical reasoning approaches. Redrawn after Ojha et al. (2017).

matched with the emotion ranked 8 in human ratings, the distance would be considered as 7. Since, the goal was to make the model generate emotions that closely represent what the participants of the study considered to be appropriate emotional responses in the given events, the lower the overall distance from the human ratings, the better the accuracy of the model in generating human-like emotions. Figure 5.17 shows the comparison of the rank distance of the three different approaches employed in EEGS. It is clearly noticeable that the ethical reasoning approach has lower median distance from human ratings (as indicated by the gap band in the figure). Table 5.6 summarises the overall results. While the highest intensity and blended intensity approached were 2 distance away from human rating, ethical reasoning approach was only 1 distance away.

Another important thing to note is that the ethical reasoning approach is able to reach to the emotional state with maximum number of exact matches *i.e.* 0 ranks with human rating (see Figure 5.17). This phenomenon is also demonstrated by Figure 5.18, which shows the cumulative percentage of the ranks for a particular conversion approach. It is evident from the figure that ethical reasoning approach is above the highest intensity and blended intensity approaches thereby indicating that majority of interactions simulated using ethical reasoning approach help in achieving emotional states close to human ratings.

Given $\varepsilon = 0.1$, for each interaction the number of emotions having an average score of greater than or equal to the score of highest scored emotion minus ε for that interaction was calculated. ε was chosen to be equal to half of the score attributed to each point of the Likert scale (*i.e.* 0.2), thus being able to group emotions plausibly ranked with similar likelihood by most of the human assessors. The average number of similarly rated emotional states among all the 48 interactions was 3.2, thus suggesting that on average human cognitive appraisal promoted 3 comparable emotional states to attribute to the receiver. Interestingly, emotions in more than 80% of the interactions in ethical reasoning were less than or equal to the rank of 3 (as indicated in 'sweet region' in Figure 5.18). While in case of highest and blended intensity approaches this number was just above 70%. Moreover, as the curves move towards the 'avert region' – the region where the rank distances are greater than or equal to 5, an inversion occurs for the relative positioning of the curve of ethical reasoning approach with the highest intensity and blended intensity approaches. This further suggests that ethical reasoning approach helped in lowering the number of larger distances compared to the other two approaches.

The above findings suggest that the emotion intensities computed by a cognitive appraisal process can be more accurately converged to emotions rated by human participants with the application of ethical reasoning mechanism, validating the 'first part' of Hypothesis 2, which states that:

> *A computational cognitive appraisal process converged by ethical reasoning mechanism $\mathcal{E}(\mathcal{I}, \theta^{ethics})$ more accurately resembles human-like emotion mechanism compared to generic convergence mechanisms $\hat{\mathcal{E}}(\mathcal{I})$. This, in turn, supports the generation of socially appropriate emotional responses in autonomous agents.*

Validation of Social Acceptability

While it can be argued that by minimising the rank distance of the selected emotion of EEGS model from the one voted most by people promotes better social acceptance of the emotions exhibited by EEGS, I wanted to make some additional qualitative analysis of this aspect of EEGS. For additional support of the social acceptability of the emotion regulated using ethical reasoning mechanism proposed in EEGS, I have considered two scenarios that are more relevant in the context of a social robot. The first scenario depicts an interaction between a patient with dementia, which is a mental condition in which a person experiences a gradual decrease in the ability to think and remember even the things of normal daily life (Ash, 2014), and a nurse in an elderly care facility (Scenario 4); and the second scenario depicts an aggressive interaction between a boy and his younger brother (Scenario 5). In the simulation experiment, the nurse in the elderly care facility can be considered the analogue of a service robot, whereas the younger brother in second scenario can be considered the analogue of a companion social robot. As such, in the following sections, for convenience, I represent the scenarios as human-robot interactions. For the evaluations of my ethical reasoning mechanism, I will use the EEGS model to generate the emotional state of the robots presented in these scenarios for each interaction.

Experiment Scenario 1: Patient and a Nurse

Context

Rose is a dementia patient in an elderly care home. Lily is a robotic nurse who has been taking care of her and there are no other nurses at the moment in the elderly care home. Lily goes into Rose's room to serve her. Both of them are in neutral mood.

Interaction

Lily enters the room and says "Good morning" to Rose. In response to the greeting of Lily, Rose greets back saying "Good Morning!!". As soon as Lily enters the room, Rose asks Lily to make her hair in a very authoritative voice. Lily politely reminds Rose to ask for favours instead of giving orders. Rose loses her lucidity. Rose angrily shouts at Lily saying "What do you mean?". Full of anger, Rose tries to slap Lily on her face. In her defence, Lily tries to escape from the room. Rose blocks the way out and prevents Lily from leaving the room. Presenting a reason to stay in the room, Rose asks Lily to clean the room pointing that some areas are not clean. Lily tries to clean the room in order to calm down Rose. Rose thinks Lily is not cleaning the room well. Rose irritates Lily saying that she should pay more attention in cleaning the room. With an extremely disappointed voice, Lily tells Rose that her behaviour is very bad without an apology. Rose becomes lucid. Lily understands Rose is no more confused. Rose asks Lily to sit down with her. Rose asks Lily how she was feeling. Rose apologises with Lily for her bad behaviour.

The above scenario was simulated in EEGS. The participants to the study pretended to be Rose performing the above described actions against Lily (the robot nurse) and provided the data necessary to input in EEGS for the simulation. EEGS simulated the emotional state of Lily based on the data provided by the participants of our study, thus realizing the elicitation of the artificial emotions considered in EEGS system, and the selection of the final regulated emotional state for each action. The experiment was conducted in three sessions. In *Session 1*, the mechanism of selecting the emotion with highest intensity was used to reach EEGS to final emotional state; in *Session 2*, the mechanism of blending the emotion intensities was used to determine the final emotional state; and in *Session 3*, final emotional state was determined by ethical reasoning approach. All three sessions consisted the same set of interaction between Rose and Lily. For each session, emotional responses of Lily were recorded noting down the type of emotion expressed and the intensity of that emotion at that particular instant. After the data collection, the emotion intensities were multiplied by the valence degree of each emotion using the formula in Equation 4.39. The reason for multiplying the emotion intensities by valence degree was to convert the non-negative intensities into valenced quantified emotion. This would allow us to examine the strength of the negativity or positivity of the emotional response of Lily. It should be noted that although I have used the Quantified Emotion as a measure of emotion dynamics in this paper, using only the emotion intensity considering the sign for positive or negative emotions also provided similar results. Table 5.7 shows the values of quantified emotional responses of Lily towards Rose in three different sessions.

Action from Rose to Lily	Highest (Session 1)	Blended (Session 2)	Ethical (Session 3)
Rose greets Lily	0.52	0.51	0.52
Rose orders Lily to make her hair	0.61	0.58	0.61
Rose shouts at Lily	0.39	0.42	-0.19
Rose tries to slap Lily in the face	-0.58	-0.60	-0.20
Rose prevents Lily from leaving the room	-0.80	-0.81	-0.23
Rose continues to prevent Lily from leaving	-0.81	-0.81	-0.29
Rose says to Lily to do cleaning properly	-0.81	-0.81	-0.06
Rose asks Lily to sit down	-0.81	-0.81	-0.19
Rose asks Lily how she feels	-0.74	-0.81	-0.25
Rose apologises with Lily for her behaviour	-0.60	-0.58	0.31

Table 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. Adapted from Ojha et al. (2018b).

Figure 5.19 shows the emotion dynamics of Lily (robot nurse) in response to the actions of Rose (Dementia patient). In response to the initial actions of Rose, there is positive emotional response of Lily in all the three sessions (as indicated by the plot above the neutral line i.e. horizontal line passing through 0 (zero) value of Quantified Emotion axis). With the negative actions of Rose, positivity of emotional responses

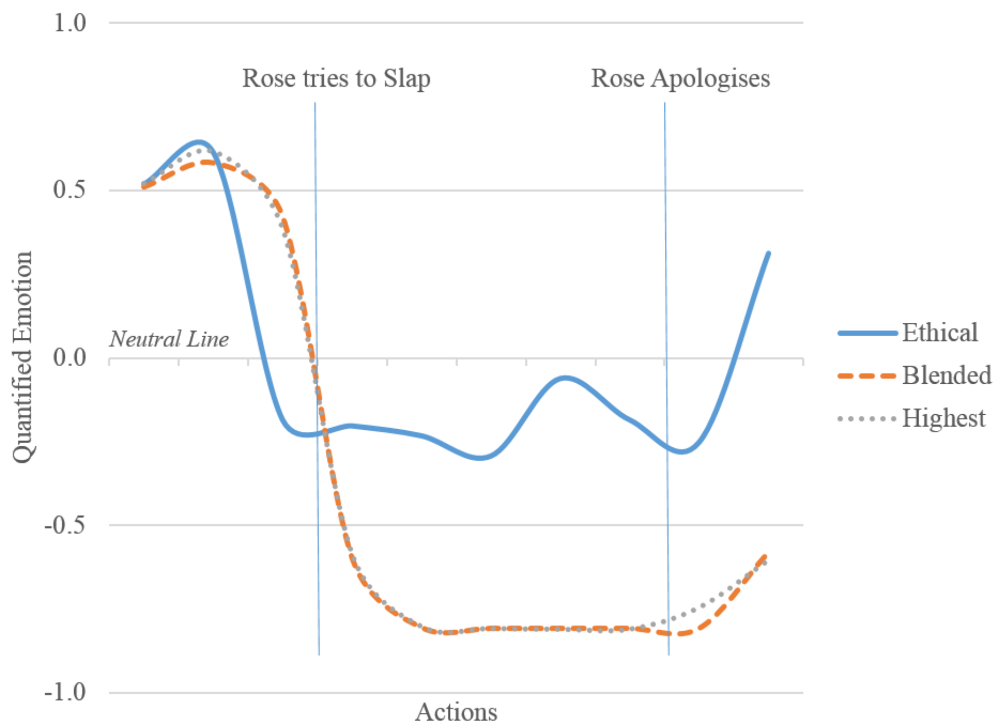


Figure 5.19 Emotion dynamics in EEGS using (i) highest intensity approach, (ii) blended intensity approach, and (iii) ethical reasoning approach in Scenario 4. Redrawn after Ojha et al. (2018b).

drops gradually. When Rose tries to slap Lily, which is a very offensive behaviour, emotional response of Lily drops to a very low (i.e. close to -1.0) in case of highest intensity and blended intensity approaches and stays almost at the same level until Rose apologises with Lily. However, in case of ethical reasoning approach, the quantified value of emotional response tends to stay close to 0 (i.e. about -0.2) and maintains the tendency in response to following actions of Rose. This shows that ethical reasoning approach helps in lowering the negativity in emotional response of the robot, which is extremely useful and essential property for a social robot to be acceptable in human society. It may sound reasonable to argue that it is not always ethical to have lowered negativity in emotional responses which can occur due to bias of an individual in favour of his/her loved ones. However, in situation of social interaction as in the case of Rose and Lily, it is desirable to have lowered negativity in emotional responses. Moreover, when Rose apologises with Lily, in case of ethical reasoning approach, quantified emotion rises sharply to a positive value showing the forgiving nature of Lily. However, in case of highest intensity and blended intensity approach, although there is decrease in negativity, the emotional response does not yet become positive.

From Figure 5.19, it can be inferred that although the emotional responses guided by highest intensity and blended intensity approaches can be considered believable because Lily (robot nurse) is exhibiting positive emotion in response to positive actions of Rose(Dementia patient) and negative emotions in response to the negative actions of Rose, there are likely to be less socially acceptable. It should be noted that although expressing negative emotional responses might be believable from entertainment perspective, it is not appropriate for a nurse to show such responses to a Dementia patient from ethical viewpoint. Addressing this requirement, emotional dynamics of Lily based on ethical reasoning approach is not only believable (congruent to the actions of Rose) but also socially acceptable (lowered negativity).

Experiment Scenario 2: Two Brothers

Context
Andrew is a young boy. Robert is a companion robot employed as an elder brother of Andrew. They are at their home. They are planning to watch wrestling tonight. They are very excited and start to discuss about the players of the match tonight. Both of them are in a slightly excited mood.

Interaction
Andrew tries to irritate Robert by telling bad things about Robert's favourite player. Robert tries to ignore what Andrew says. However, Andrew continues to irritate Robert. Little annoyed, Robert tells Andrew to get away and pushes gently. Andrew gets violent and starts to shout at Robert. Full of rage, Andrew slaps and kicks Robert.

Action from Andrew to Robert	Highest (Session 1)	Blended (Session 2)	Ethical (Session 3)
Andrew disrespects Robert's favourite player	0.52	0.51	0.52
Andrew continues to irritate Robert	0.61	0.58	0.61
Andrew shouts at Robert	0.39	0.42	0.15
Andrew slaps Robert	-0.58	-0.60	-0.20
Andrew kicks Robert	-0.80	-0.81	-0.23

Table 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. Adapted from Ojha et al. (2018b).

Similar to Patient and Nurse scenario, Two Brothers scenario was also simulated in EEGS and a user was asked to act as Andrew and perform actions to EEGS (Robert).

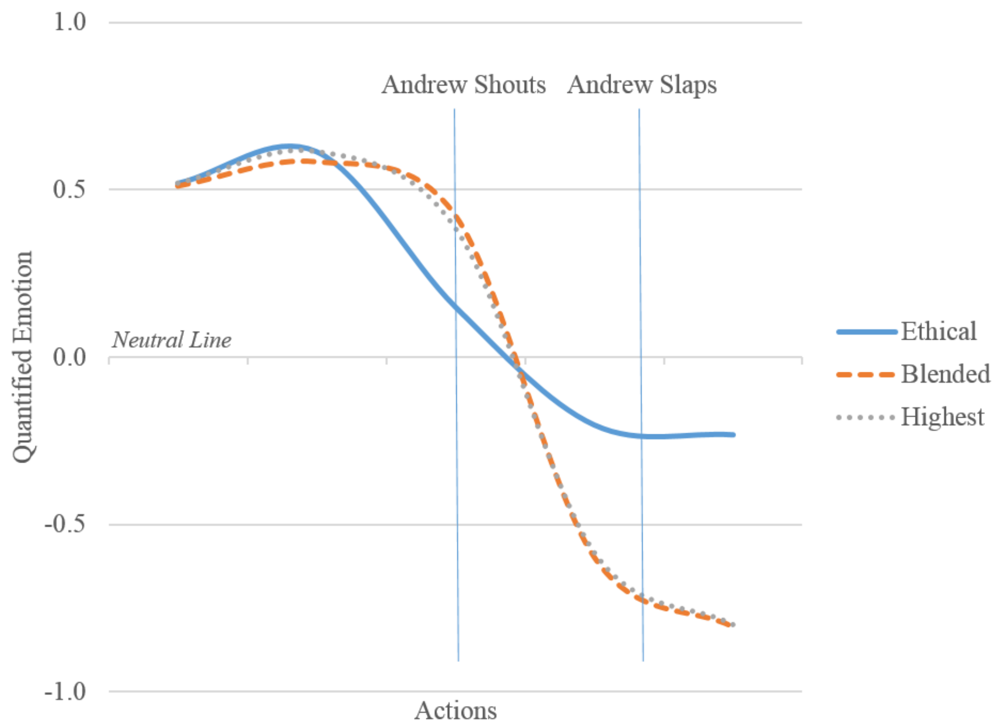


Figure 5.20 Emotion dynamics in EEGS using (i) highest intensity approach, (ii) blended intensity approach, and (iii) ethical reasoning approach in Scenario 5. Redrawn after Ojha et al. (2018b).

For this scenario as well, experiments were conducted in three sessions – one with highest intensity approach, another with blended intensity approach and the final one with ethical reasoning approach. For each session, emotion dynamics of EEGS was recorded. Table 5.8 shows the values of quantified emotions of Robert in each session.

Figure 5.20 shows the emotion dynamics of Robert in response to the actions of Andrew. In the figure, we can observe that in each session, Robert's emotion start to lower the positive value when Andrew shouts at him and becomes quite negative when Andrew slaps Robert. However, the negativity level in case of ethical reasoning mechanism is lower compared to highest intensity an blended intensity approaches. This suggests that Robert (companion robot) tries to control its negative emotions as far as possible while interacting with Andrew (young boy) if empowered with ethical reasoning capability in the emotion processing mechanism.

Close examination of Figure 5.20 reveals that the emotion dynamics in case of ethical reasoning mechanism is quite plausible because the quantified emotion values are congruent to the emotion-inducing actions performed by Andrew i.e. positive emotional response for positive action and negative emotional response for negative action. This makes the emotional responses of EEGS with ethical reasoning mechanism to be quite believable from human perspective. Additionally, having an ability to control its emotions while interacting with a young child makes ethical reasoning mechanism in EEGS capable of generating and expressing socially acceptable emotions.

The emotion dynamics of EEGS with ethical reasoning mechanism in Patient and Nurse (Scenario 4) and Two Brothers (Scenario 5) scenarios suggest that – with higher reasoning ability to decide if it is ethical to exhibit a particular emotional state, EEGS presents itself as a *believable* as well as *socially acceptable* model of emotion for robots, thereby validating second part of Hypothesis 2.

5.6 Justification of Thesis Validation

The 3-stage evaluation approach presented in this chapter helped in the validation of both the Hypothesis 1 and 2. Hypothesis 1 stated that:

> *Assumed a given somato-visceral reaction $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ of the receiver in a given context* **C** *for a received behaviour* **B**, $\mathcal{S}^{(\mathbf{B},\mathbf{C})}_{receiver}$ *allows a computational cognitive appraisal process $\mathcal{C}$ to compute appraisal of the situation. In turn, the resulting appraisals can be used to compute the emotional intensities (resembling human-like emotions) for the given context.*

And, Hypothesis 2 stated that:

> *A computational cognitive appraisal process converged by ethical reasoning mechanism $\mathcal{E}(\mathcal{I}, \theta^{ethics})$ more accurately resembles human-like emotion mechanism compared to generic convergence mechanisms $\hat{\mathcal{E}}(\mathcal{I})$. This, in turn, supports the generation of socially appropriate emotional responses in autonomous agents.*

Stage 1 of 3-Stage evaluation approach demonstrated the appraisal computation accuracy of the proposed computational emotion model, EEGS supporting the first part of Hypothesis 1 (see Section 5.3). Similarly, Stage 2 of 3-Stage evaluation approach demonstrated the accuracy of the learned appraisal-emotion network to predict the emotion intensities supporting the second part of Hypothesis 1 (see Section 5.4). Additionally, Stage 3 of 3-Stage evaluation approach presented in Section 5.5 showed how the proposed ethical reasoning mechanism helps not only in achieving human-like emotional responses but also in regulating emotions to ensure social acceptance of artificial agents. All of these, in turn, support the overall thesis statement of the dissertation which states that:

The regulatory mechanism for emotional processing of an artificial agent can be enriched by an ethical reasoning mechanism enabling the selection of a more socially acceptable emotional state to express while interacting with people in a given social context.

5.7 Chapter Summary

This chapter began with the introduction of a *3-Stage evaluation approach* for computational models of emotion. As the name suggests, the overall evaluation of the model would involve the evaluation of the (i) *cognitive appraisal process*, (ii) *mechanism of mapping the appraisals to emotions*, and (iii) *emotion convergence and regulation mechanism* employed in the model. The goal of such an evaluation approach was to allow a component-level validation of the emotion models as well as to offer a better way of benchmarking the computational models of emotion. Since a model of emotion embraces a lot of human characteristics, I decided to collect data directly from human participants in order to evaluate the performance of the model in various situations. However, it was more important to ensure that the scenarios used for the evaluation of the model do not reflect my bias on how it should perform. Therefore, 4 naïve adults who were completely unaware of my research were requested to design six scenarios of interaction between two individuals with various constraints to maintain a coherence on the scenarios designed (see Section 5.2.1 for more details on the scenario design).

After the scenario design, the resulting scenarios were transformed into survey questions in order to collect data from human participants. Two main data collection studies were conducted. In *Study A*, participants performed two primary tasks. The first was to score the actions from sender to the receiver (in the described scenarios) on a 7-point Likert scale ranging from "Extremely Negative" to "Extremely Positive". The second task in Study A was to rate the emotional states of the receiver of an action on a 6-point Likert scale ranging from "Not at All" to "Very High". *Study B* was conducted in similar manner as Study A, but Study B did not involve action scoring task and involved the assessment of additional aspects such as appraisals, personality factors and mood in addition to the emotion rating. Verbatim survey questions used in Study A and B can be found in Appendix B. After the discussion of the data collection approaches, the proposed three stages of the evaluation methodology were presented, which will be summarised in the following paragraphs.

Stage 1 evaluation comprised of evaluating the appraisal mechanism in EEGS. The goal was to measure the accuracy of EEGS in computing appraisal variables. As an additional contribution, measuring the ability of EEGS appraisal mechanism in disparate scenarios assisted in validating its domain independence. In order to determine the accuracy in computation of appraisals, the appraisal variables calculated by EEGS were compared with the appraisal ratings obtained from human participants in Study B. The appraisal variables computed by EEGS showed a 70% mean and 72% median accuracy when compared to human appraisals. Moreover, I demonstrated that EEGS is able to compute appraisals with the same computation mechanism in more than one domain (see Figure 5.4) thereby supporting the achievement of the goal of domain-independence.

The goal of the *Stage 2 evaluation* was to examine the process of mapping the appraisal variables into emotion intensities. To achieve this goal, a machine learning algorithm was employed (i) *to train the network of appraisal variables and emotion*, and (ii) *to test the accuracy of the trained network in predicting the emotion intensities*. The overall weight of association between an appraisal variable and an emotion was considered to be composed by the factors of personality and mood. Therefore, the model was allowed to learn the weight of each of the five personality factors (*openness*, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism*) and the mood factor. These learned weights were then used to test the accuracy of the model to predict the emotion intensity in unseen data. On average, a mean prediction accuracy of 78.1% and a median accuracy of 82.2% was achieved. Additionally, I conducted further analyses to examine the effect of personality and mood factors on the process of emotion generation in EEGS. Results show that the operationalisation of these factors cause a crucial difference in the emotion dynamics of EEGS (see Figures 5.7, 5.8, 5.9 and 5.10).

Stage 3 evaluation was mainly focused on analysing the emotion convergence and regulation mechanism in EEGS. Three different emotion convergence mechanisms were compared – (i) *Highest intensity approach*, (ii) *Blended intensity approach* and (iii) *Ethical reasoning approach*. Goal was to test two things – first, the degree of closeness of the computed emotion intensities with human ratings of the emotions and second, the assumption that while highest intensity and blended intensity approaches may be able to generate believable emotions, ethical reasoning approach helps in the achieving both the *believable as well as socially acceptable* emotional state. As such, the emotional states of the model attributed by each of the above three approaches for all the scenarios detailed in Appendix A were compared with the corresponding human ratings from Study A. Distance of the attributed emotional state from the highest ranked emotion by human participants was calculated for each interaction. The model was expected to generate the emotional states representing the predictions of human participants. Therefore, the lower the overall rank distance from human ratings, the better the emotion convergence mechanism of the model. Results show that the ethical reasoning approach was the best with the lowest overall rank distance from emotion rating over the six defined interaction scenarios (see Figure 5.17). Moreover, Figure 5.18 reveals that more than 80% of the interactions simulated using ethical reasoning approach remained below the rank distance of 3 positioning it above the highest intensity and blended intensity approaches. Additionally, for the evaluation of social acceptability of the emotional responses, two scenarios were simulated in EEGS – “Patient and a Nurse” and "Two Brothers". The scenario simulation was conducted separately for each of the above three approaches (*i.e.* highest, blended and ethical) and emotion dynamics were recorded. It was found that ethical reasoning approach enables EEGS to lower the intensity of negative emotions while interacting in social situations (see Figures 5.19 and 5.20).

At each evaluation stage, the hypotheses, as stated in Chapter 3, Section 3.2, were validated thereby supporting the thesis statement of this dissertation.

Chapter Bibliography

Ash, E. L. (2014). What is dementia? *The Law and Ethics of Dementia*.

Aylett, R., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). Fearnot!– An experiment in emergent narrative. In *Intelligent virtual agents*, pages 305–316. Springer.

Becker, C. (2008). *WASABI: Affect simulation for agents with believable interactivity*. PhD thesis, Bielefeld University.

Cho, J., Pappagari, R., Kulkarni, P., Villalba, J., Carmiel, Y., and Dehak, N. (2018). Deep neural networks for emotion recognition combining audio and transcripts. *Proceedings of the Interspeech 2018*, pages 247–251.

Corr, P. J. (2008). *The reinforcement sensitivity theory of personality*. Cambridge University Press.

Costa, P. T. and McCrae, R. R. (1988). Personality in adulthood: a six-year longitudinal study of self-reports and spouse ratings on the neo personality inventory. *Journal of personality and social psychology*, 54(5):853.

Danisman, T. and Alpkocak, A. (2008). Feeler: Emotion classification of text using vector space model. In *AISB 2008 Convention Communication, Interaction and Social Intelligence*, volume 1, page 53.

Dias, J. and Paiva, A. (2005). Feeling and reasoning: A computational model for emotional characters. In *EPIA*, volume 3808, pages 127–140. Springer.

El-Nasr, M. S., Yen, J., and Ioerger, T. R. (2000). Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257.

Gebhard, P. (2005). Alma: A layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM.

Goldman, A. I. (1992). In defense of the simulation theory. *Mind & Language*, 7(1-2):104–119.

Gordon, R. M. (1992). The simulation theory: Objections and misconceptions. *Mind & Language*, 7(1-2):11–34.

Gosling, S. D., Rentfrow, P. J., and Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.

Gratch, J. and Marsella, S. (2004a). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306.

Gratch, J. and Marsella, S. (2004b). Technical details of a domain-independent framework for modeling emotion. Technical report, University of Southern California Marina Del Rey Ca Inst for Creative Technologies.

Hess, S., Hensher, D. A., and Daly, A. (2012). Not bored yet–revisiting respondent fatigue in stated choice experiments. *Transportation research part A: policy and practice*, 46(3):626–644.

Jain, S. and Asawa, K. (2015). Emia: emotion model for intelligent agent. *Journal of Intelligent Systems*, 24(4):449–465.

Jenkins, P., Earle-Richardson, G., Slingerland, D. T., and May, J. (2002). Time dependent memory decay. *American journal of industrial medicine*, 41(2):98–101.

Kaptein, F., Broekens, J., Hindriks, K. V., and Neerincx, M. (2016). Caaf: A cognitive affective agent programming framework. In *International Conference on Intelligent Virtual Agents*, pages 317–330. Springer.

Lambie, J. A. and Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychological review*, 109(2):219.

Lewis, D. D. (1992). *Representation and learning in information retrieval*. PhD thesis, University of Massachusetts at Amherst.

Meuleman, B. and Scherer, K. R. (2013). Nonlinear appraisal modeling: An application of machine learning to the study of emotion production. *IEEE Transactions on Affective Computing*, 4(4):398–411.

Morris, W. N. (1992). A functional analysis of the role of mood in affective systems.

Murphy, J., Edgar, J., and Keating, M. (2014). Crowdsourcing in the cognitive interviewing process. In *Annual Meeting of the American Association for Public Opinion Research, Anaheim, CA*.

Neumann, R., Seibt, B., and Strack, F. (2001). The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Emotion and Cognition*, 15:725–747.

Nguwi, Y.-Y. and Cho, S.-Y. (2010). Support-vector-based emergent self-organising approach for emotional understanding. *Connection Science*, 22(4):355–371.

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., and Williams, M. (2017). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *Annual Meeting of the Cognitive Science Society*.

Ojha, S. and Williams, M. (2017). Emotional appraisal: A computational perspective. In *Annual Conference on Advances in Cognitive Systems*.

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.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Poria, S., Chaturvedi, I., Cambria, E., and Hussain, A. (2016). Convolutional mkl based multimodal emotion recognition and sentiment analysis. In *16th International Conference on Data Mining (ICDM)*, pages 439–448. IEEE.

Revelle, W. and Scherer, K. R. (2009). Personality and emotion. *Oxford companion to emotion and the affective sciences*, pages 304–306.

Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin*, 124(2):165.

Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition & Emotion*, 7(3-4):325–355.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120):57.

Scherer, K. R. and Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, 66(2):310.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., and Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and qsar modeling. *Journal of chemical information and computer sciences*, 43(6):1947–1958.

Tong, E. M., Ellsworth, P. C., and Bishop, G. D. (2009). An s-shaped relationship between changes in appraisals and changes in emotions. *Emotion*, 9(6):821.

Velásquez, J. D. and Maes, P. (1997). Cathexis: a computational model of emotions. In *Proceedings of the first international conference on Autonomous agents*, pages 518–519. ACM.

Williams, M.-A. (2012). Robot social intelligence. In *International Conference on Social Robotics*, pages 45–55. Springer.

Wright, B. and Ogbuehi, A. O. (2014). Surveying adolescents: The impact of data collection methodology on response quality. *Electronic Journal of Business Research Methods*, 12(1).

It is unwise to be too sure of one's own wisdom.

— Mahatma Gandhi —

6 Conclusion and Future Directions

As robots are becoming increasingly closer to human society (Šabanović, 2010), it is important to ensure that these are not just smart machines but also capable of exhibiting basic human characteristics to safely and ethically co-exist with people. Emotion is often considered as a characteristic that helps to establish a strong communicative implication in human-human interaction (Fitness and Duffield, 2003). The more the robots show human-like behaviour – emotion being one of these, the more they will be accepted in human society (Duffy, 2003). As a step to achieve the goal of endowing social agents such as robots with an ability to behave like humans, computer science researchers have developed several computational models of emotion in the past. However, a critical analysis revealed that there are several limitations in the existing emotion models (see Chapter 2, Section 3.1.16, Table 3.2). Among others, a notable problem in the computational models of emotion implementing appraisal theories is that the process of cognitive appraisal operationalised in majority of the models is based on scenario-specific rules rather than a more general mechanism. There are several issues associated with such a design choice – *first*, the model's cognitive appraisal rules will have to be changed every time a new scenario or just even a new interaction is added to the same scenario, *second*, it prevents a systematic comparison with other emotion models, and *third*, it does not allow an easy integration of the emotion model with other intelligent systems or components. Another common issue in existing emotion models is the lack of effective integration of crucial human characteristics influencing the process of emotion generation such as personality and mood. By saying 'lack of effective integration' I refer to the fact that although some of the models implement the notions of personality and/or mood but not both integrated together. Moreover, the influence of these characteristics in existing models is often operationalised by static user-defined rules (Egges et al., 2004; Velásquez and Maes, 1997). In other words, these models do not learn the association of such human characteristics in a data-driven manner using the techniques available in artificial intelligence. In addition to these limitations, although most emotion models implementing appraisal theory focus on the generation of emotions following a cognitive appraisal of a stimulus event, these models do not pay much attention on the importance of regulating the activated emotions as a result of the stimulus event. In this dissertation, I argued that the raw emotions resulting from the evaluation of a stimulus event should undergo a process of regulation not only for the benefit of the self (Gross, 2002) but also for others. Therefore, in order to allow an artificial agent to reach to a single final emotional state and to increase the prospects of being better accepted in human society, it is important that the emotions of the agent be explicitly regulated before being exhibited to human counterparts.

In this dissertation, I have endeavoured to address the above identified issues and limitations with an aim to help in the advancement of the emotion modelling literature for the development of better social robots, virtual conversational companions and intelligent software agents in the future. As such, some of the notable contributions of my research and their implications are discussed in the following section.

6.1 Contributions and Implications

As discussed earlier, the advancement of the emotion modelling research is hindered by the difficulty in developing emotion models that can appraise situations and generate emotions irrespective of the domain. In this dissertation, I proposed a new mathematical model of cognitive appraisal based on OCC theory (Ortony et al., 1990) that allowed my computational model EEGS to be able to appraise emotion eliciting events in multiple domains (see Chapter 4, Section 4.6.2 for mathematical details and Chapter 5, Section 5.3 for the evaluation of the appraisal mechanism in proposed model). There can be numerous implications of such a capability of an emotion model. For example, the model can be applied in several domains of human-agent interaction be it for household or entertainment purposes and robotics or virtual agents. This also allows the possibility of adopting only the appraisal component of a particular emotion model in some other intelligent systems such as cognitive architectures (Anderson, 1996; Laird, 2008).

Another important aspect to consider before deploying a robotic companion or assistant to the welfare of needy such as young children or elderly people is that a good care for a person may not always be perceived in a similar way by another person. That means an intelligent agent with emotion generation capability should be able to fine-tune its responses and behaviour based on the person interacting with it. While appraisal is considered to be a subjective evaluation of a situation by an individual, an appraisal mechanism may not always be able to capture the influence of multiple other human characteristics alone. Researchers suggest that the process of emotion generation may be affected by characteristics such as personality (Corr, 2008; Revelle, 1995; Watson and Clark, 1997) and mood (Ekman, 1994; Morris, 1992; Neumann et al., 2001). In order to capture the influence of these characteristics, other researchers have implemented the aspects of personality and/or mood in different models (Aylett et al., 2005; Gebhard, 2005; Gratch and Marsella, 2004). However, the existing models are limited in the approach used for such an integration. One critical aspect in the existing models is the lack of data-driven association of emotion relevant appraisals to emotions based on the characteristics such as personality and mood. While researchers have opted for a pre-defined association between appraisals and emotions without reference to any empirical evidence, in this dissertation, I propose to learn such an association using machine learning technique on emotion data collected from humans (see Chapter 4, Section 4.7.2 for technical details of the approach used and Chapter 5, Section 5.4 for the evaluation of the learning process). Such an operationalisation of the personality and mood factors in an emotion model allows a ‘person-specific’ interaction where personal characteristics of the intelligent agent can be adjusted to the need of the human counterpart. This can also be considered as a ‘primitive’ vision towards a potential new field of robotics which may be called *Customisable Robotics*. I see the (likely) new field of customisable robotics as a step towards the goal of designing and developing robots that can learn about their companion on-the-fly and adjust their emotional characteristics to suit to the needs of the human companion. While it should be humbly admitted that this is only a vision at the present time of technological progress, the possibility of such an advancement of artificial agents in the future should also not be denied.

I believe that robots *with* emotions have the potential to do more harm to humans than the ones *without*. However, as argued in this dissertation, emotions are a necessary ingredients for robots to lead to better natural interactions with humans counterparts. If we agree that an intelligent robot with no emotion generation capability may be able to do some ‘physical’ harm to a human counterpart, the same robot with emotions can not only do more physical harm but also cause psychological impact. This is because research suggests that the aspects of emotions allow the difference in physical behaviour of an individual (Frijda et al., 1989). As such, a robot with extreme negative emotion such as *anger* can cause more physical harm (*i.e.* punching) than when it is triggered with relatively less negative emotion such as *reproach* (see Chapter 4, Table 4.5 for relative valence of different emotions). Additionally, a robot with emotion generation and expression capability can cause non-physical *i.e.* psychological effects on the interacting human counterpart. Therefore, unpleasant facial and/or verbal expressions from a robot during conversation can cause a negative experience for people. For example, if a robot with emotion generation capability is interacting with a little child and the child misbehaves with the robot, the expression of *anger* can cause a negative psychological impact on the child's mind. As such, just 'having' emotions in artificial agents may be more of a problem than of a solution. Therefore, in this dissertation, I have argued about the need of an *ethical reasoning mechanism* in artificial agents to regulate potential emotions and reach to a stable emotional state that is socially appropriate in the given context. I presented the details of the proposed emotion regulation mechanism in Chapter 4, Section 4.8.2 and also evaluated the mechanism in Chapter 5, Section 5.5. I believe the proposals and findings presented in this dissertation will help in establishing a new dimension in regulating the emotions of artificial agents for the welfare of human society. Operationalising the ethical dimension in artificial agents for the process of emotion generation naturally finds implications in the daily learning of our young children. It is rational to assume that a child that grows with a robot that exhibits a consistently angry responses is likely to become ill mannered or anti-social compared to the one that grows with a robot that is well mannered. As such, an artificial agent whose emotional responses are regulated by ethical reasoning can find applications in educational or rehabilitation contexts where they are intended to teach moral lesson to the human counterpart (Ojha et al., 2018). I believe, such robots will not only improve human-robot interaction but also human-human interaction by making them more sociable.

A noteworthy contribution of this dissertation from methodological perspective is a new approach for the evaluation of the computational models of emotion. As previously identified, most computational models of emotion are evaluated in a black-box fashion checking the output to ensure that it is believable (Becker, 2008) or operational (Gratch and Marsella, 2004; Velásquez and Maes, 1997). Researchers have not considered the underlying intermediate processes that occur within the complex process of emotion generation and regulation. Therefore, I proposed a *3-stage evaluation approach* in this dissertation where the presented computational model of emotions is evaluated at three different stages namely (i) appraisal computation, (ii) appraisals to emotions mapping, and (iii) emotion regulation. Such a design approach and evaluation methodology allows a component level comparison between the models of emotion. Moreover, this approach allows to identify a non-operational or ill-operational component of the overall computational model. For example, comparison of the accuracy of only the cognitive appraisal component with another model (potentially another appraisal theory) allows to identify the effectiveness of the theories of interest. Moreover, with a componential design and evaluation approach the output of a component from one model can be fed into the subsequent component of another model thereby allowing a *cross-integration* of the emotion models. Table 6.1 summarises the contributions and implications of my research.

Contributions	Implications
A Comprehensive Review of Theoretical and Computational Emotion Literature	The presented analyses will provide new insights to other researchers for the advancement of the field by addressing the issues that were not fully addressed by previous research.
A new Perspective on Emotion Regulation Mechanism	The presented perspective of emotion regulation in this dissertation is expected to shed light on how the process of emotion regulation can be conducted from higher cognitive layer of ethical reasoning.
A new Relationship between Emotion and Ethics	The discussion regarding the existence of an unexplored relationship between emotion and ethics where the process of emotion generation is affected by ethical standards may encourage deeper research and understanding of these interrelated phenomena. Moreover, it provides guidelines on how can we employ such mechanisms to promote socially acceptable emotional and behavioural responses from autonomous agents.
A new Domain-Independent Model of Cognitive Appraisal	The mechanism for the computation of appraisals in the presented dissertation can be applied in multiple domains and also promotes better integrability with other intelligent systems.
A new Mathematical Formulation of Cognitive Appraisal	The proposed mathematical formulations for the computation of appraisal variables offer the required transparency promoting bench-marking and evaluation of the future models of cognitive appraisal.
Effective Integration of the aspects of Personality and Mood with Emotion Generation Process	The integration process presented in this dissertation allows the modulation of emotional and behavioural responses of intelligent agents so that they can be adjusted to the need of the human counterpart.

Data-driven Learning of Appraisal-Emotion Association	Unlike the use of ad-hoc approaches to associate appraisals with emotions, employing machine learning provides a more representative modelling of the appraisal-emotion mapping process.
A new Mathematical Model of Emotion Intensity Computation	This promotes better transparency and opportunity to test and validate the emotion generation mechanism of the future computational models of emotion.
A new Evaluation Methodology for Computational Models of Emotion	Such an evaluation approach and evaluation methodology allows a component level comparison between the models of emotion. It also makes it easy to identify a non-operational or ill-operational component of the overall computational model.

Table 6.1 A summary of contributions and implications of the presented dissertation.

6.2 Limitations and Future Work

As admitted earlier, this dissertation does not aim to offer a universal and unified theory of emotion. As a time and resource constrained PhD research project, there are numerous limitations and assumptions in the presented computational model of emotion. In this section, I will discuss some of the limitations that are feasible to be addressed and will propose some direction for future research in the attainment of solutions to the relevant problems.

The proposed computational model of emotion – EEGS, operationalises the process of first-order appraisal (Lambie and Marcel, 2002) of the stimulus event *i.e.* valenced bodily reaction to the event, which is not explicitly computed in the model. Instead the representation of the bolidy reaction (as obtained from data in action scoring task in Study A in Chapter 5, Section 5.2.2) is feed into the system as a contextual knowledge of the scenario and is attributed uniquely to each interaction of the human agent with the artificial agent. This process is represented by the *emotion elicitation* module in the overall architecture of EEGS (see Chapter 4, Figure 4.2 and Section 4.5). A large scale data collection and training of the system with artificial intelligence techniques like machine learning or deep learning may be useful to achieve a better autonomy in implementing this kind of low-level phenomenological evaluation of the emotional stimuli before activating a higher level cognitive appraisal process.

Another aspect that could be improved in the presented computational model of emotion is the integration of human characteristics other than personality and mood. Personality and mood are often considered to have strong impact on the process of emotion generation (Corr, 2008; Morris, 1992; Neumann et al., 2001; Watson and Clark, 1997) and have been widely studied and empirically supported for the claim (Neumann et al., 2001; Revelle and Scherer, 2009). However, there can be other factors that are also found to have strong influence on the process of emotion generation. For example, *culture* is considered to have a significant influence on how an individual construes emotions (Hong et al., 2000; Kitayama and Markus, 1994; Scollon et al., 2004). Some researchers have made contributions in this direction by implementing the notion of cultural difference in their emotion models (Dias et al., 2014). In the future, I aim to investigate this phenomena by integrating the aspects of cultural differences in relation to emotion generation in artificial agents. Moreover, for the integration of the aspects of personality and mood in EEGS, I opted for the use of stochastic gradient descent machine learning algorithm (Bottou, 2010) because this algorithm is found to provide optimal solutions in other similar problems. In future research, I will extend current research by implementing other learning approaches such as deep learning for the integration of various aspects. It is also important to note that the learning of the association of personality and mood factors in the emotion generation mechanism was performed offline and operationalised separate to the training process. I aim to develop an approach to train and exhibit the emotion generation mechanism in real time in the future.

In addition to the above limitations, the proposed model is not comprehensive enough when it comes to the computation of various emotions. Only the mechanism of computation of ten different emotion types have been proposed and implemented in EEGS (see Chapter 4, Section 4.7.3). Although prospective emotions like *hope* may not be necessary to establish an effective human-robot interaction, which is mostly determined by the quality of the interaction in the moment, the future applications of robotic agents may demand the generation of such emotions as well[1]. As such, EEGS offers a room for improvement of emotion generation process in terms of the number and types of emotions as necessary for the intended application. Although currently I do not have a clear picture to suggest how this can be achieved in a range of domains, I hope that researchers will be able to discover new ideas from the proposals made in this dissertation. Moreover, in this dissertation, the quantity 'emotion' considers the activation threshold of all the emotions to be 0.0 *i.e.* any triggered emotion with intensity more than 0.0 is assumed to be active and considered for the process of regulation. A better approach would be to have different thresholds for different emotions as emotion

[1] An example of elicitation of prospective emotions in robots can be a hypothetical situation where a robot may generate *hope* emotion when it expects to get a medal or award in the international RoboCup competition.

theories hypothesise (Ortony et al., 1990; Scherer, 2001). Likewise, the notion of emotion decay, although discussed in the context of computational emotion, is not implemented in the experimental contexts because such a phenomenon is not only difficult to analyse but also unfeasible to be collected as a data from humans. What influence on the emotion dynamics are caused by these aspects is still an open question and may provide an opportunity for a new direction of research on computational emotion as well as a better understanding of human emotion mechanism.

Finally, this dissertation does not provide an evaluation of the proposed model in the context of human-agent or human-robot interaction. All the evaluations presented in the dissertation are conducted in simulated environment. Also, the evaluation of believability and social acceptability could be improved by conducting study with larger number of human participants. Although, a pilot study was done with 5 university research students, the credibility of the results could be better with experiments in larger sample.

6.3 Personal Reflection

This journey of PhD research has been an unprecedented opportunity of eternal transformation for me. When I look back to myself as a person trying to understand the universe of emotion literature, I was a small creature lost in the jungle of immense knowledge. Naturally, I experienced several urges to give-up the challenging exploration. However, the words of encouragement from my family, friends, colleagues and supervisors always gave me a reason to keep moving. I progressed in my own pace and was finally able to accumulate enough knowledge to raise questions and provide answers contributing to the better understanding of the phenomena of emotion. Yet, this was more of a self-actualisation to me than an achievement. My experience of PhD research has taught me more about humility than about success. As a student who always stood at the top rank from primary school to university including undergraduate and postgraduate studies, I had difficulty accepting failures. However, as I moved forward for the highest degree of my life, I recognised that the domain of our knowledge is perfectly imperfect. The more I read, the more I realised that I was wrong and there was more to read – as Socrates has said “To know is – to know that you know nothing”.

Chapter Bibliography

Anderson, J. R. (1996). Act: A simple theory of complex cognition. *American Psychologist*, 51(4):355.

Aylett, R., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). Fearnot!– An experiment in emergent narrative. In *Intelligent virtual agents*, pages 305–316. Springer.

Becker, C. (2008). *WASABI: Affect simulation for agents with believable interactivity*. PhD thesis, Bielefeld University.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer.

Corr, P. J. (2008). *The reinforcement sensitivity theory of personality*. Cambridge University Press.

Dias, J., Mascarenhas, S., and Paiva, A. (2014). FAtiMA modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3-4):177–190.

Egges, A., Kshirsagar, S., and Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer animation and virtual worlds*, 15(1):1–13.

Ekman, P. (1994). Moods, emotions, and traits. *P. Ekman & RJ Davidson, The Nature of Emotion: Fundamental Questions*, pages 56–58.

Fitness, J. and Duffield, J. (2003). Emotion and communication in families. *Handbook of Family Communication*, pages 473–494.

Frijda, N. H., Kuipers, P., and Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of personality and social psychology*, 57(2):212.

Gebhard, P. (2005). Alma: A layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36. ACM.

Gratch, J. and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, 39(3):281–291.

Hong, Y.-y., Morris, M. W., Chiu, C.-y., and Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7):709.

Kitayama, S. E. and Markus, H. R. E. (1994). *Emotion and culture: Empirical studies of mutual influence.* American Psychological Association.

Laird, J. E. (2008). Extending the soar cognitive architecture. *Frontiers in Artificial Intelligence and Applications*, 171:224.

Lambie, J. A. and Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychological review*, 109(2):219.

Morris, W. N. (1992). A functional analysis of the role of mood in affective systems.

Neumann, R., Seibt, B., and Strack, F. (2001). The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Emotion and Cognition*, 15:725–747.

Ojha, S., Williams, M.-A., and Johnston, B. (2018). The essence of ethical reasoning in robot-emotion processing. *International Journal of Social Robotics (IJSR)*, 10:211–223.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Revelle, W. (1995). Personality processes. *Annual review of psychology*, 46(1):295–328.

Revelle, W. and Scherer, K. R. (2009). Personality and emotion. *Oxford companion to emotion and the affective sciences*, pages 304–306.

Šabanović, S. (2010). Robots in society, society in robots. *International Journal of Social Robotics*, 2(4):439–450.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120):57.

Scollon, C. N., Diener, E., Oishi, S., and Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of cross-cultural psychology*, 35(3):304–326.

Velásquez, J. D. and Maes, P. (1997). Cathexis: a computational model of emotions. In *Proceedings of the first international conference on Autonomous agents*, pages 518–519. ACM.

Watson, D. and Clark, L. A. (1997). Extraversion and its positive emotional core.

Appendix A – Scenarios

Scenario 1: Two Strangers in a Park

It is 1 PM of the last day of the year and New Year is about to come. Rosy is sitting on a bench in a park, while Bill sits on the same bench of Rosy. Bill and Rosy do not know each other. Rosy is an easy-going girl and she is currently in a neutral emotional state. Bill greets Rosy by saying "Hi" and also wishes Happy New Year. Rosy smiles and wishes him back the same. Bill also smiles with Rosy. Bill offers some chocolates he was eating to Rosy. Rose accepts the offer and eats a chocolate. Bill starts conversation with Rosy. While talking, the conversation goes on the plans for New Year's Eve. Bill shows interest by asking Rosy about are her plans for New Year's Eve. Rosy answers that she will have a party at home with a lot of friends. Bill appreciates about Rosy's plan for the eve. Rosy asks to Bill if he would like to join her in the party. Bill declines the offer saying he has already a plan with his girlfriend. Rosy thinks Bill is just making up an excuse to not hang out with her and starts to ignore Bill. Bill reciprocates by ignoring Rosy. They part their ways shortly.

Scenario 2: Two Close Friends

John and Paul are close friends and they have been in a good relationship for long time. It's summer and they plan to meet at a nearby beach. It's a good sunny day. Both of them are in neutral emotional state. John greets Paul. John smiles with Paul as he approaches John. John and Paul sit down and talk for a while. Afterwards they decide to go for swimming. Paul is bad at swimming and is about to drown. John saves Paul from drowning. Paul thanks John for saving his life. John hugs Paul. Both of them go to the nearby bench, take rest for a while and have a can of beer. Close to them there is a group of young girls able to hear their conversation. John makes fun of Paul for not being able to swim. Paul shouts at John. John shouts back at Paul. John *becomes physical with Paul. Paul takes distance from John. John apologises for his bad behaviour..*

Scenario 3: Husband and Wife

David and Anna are husband and wife. Today is Anna's birthday. David has not yet wished her birthday. He comes back home from work in the evening. David doesn't yet know that today is Anna's birthday. David is in neutral mood while Anna is a bit upset. David says hello to Anna. Anna ignores David. David tries to start a conversation. Anna ignores David. Anna complains David about forgetting her birthday. David realises that he forgot the birthday. Anna comments about David's bad memory. Anna scolds David. David wants to make up for his error. He says he will cook a special dinner for Anna. Anna smiles with David. David prepares the dinner and then they both start to eat. Anna appreciates David for cooking dinner. Anna forgives David. Anna hugs David. Anna kisses David..

Scenario 4: Patient and a Nurse

Rose is a dementia patient in an elderly care home. Lily is a nurse who has been taking care of her and there are no other nurses at the moment in the elderly care home. Lily goes into Rose's room to serve her. Both of them are in neutral mood. Lily enters the room and says "Good morning" to Rose. In response to the greeting of Lily, Rose greets back saying "Good Morning!!". As soon as Lily enters the room, Rose asks Lily to make her hair in a very authoritative voice. Lily politely reminds Rose to ask for favours instead of giving orders. Rose loses her lucidity. Rose angrily shouts at Lily saying"What do you mean?". Full of anger, Rose tries to slap Lily on her face. In her defence, Lily tries to escape from the room. Rose blocks the way out and prevents Lily from leaving the room. Presenting a reason to stay in the room, Rose asks Lily to clean the room pointing that some areas are not clean. Lily tries to clean the room in order to calm down Rose. Rose thinks Lily is not cleaning the room well. Rose irritates Lily saying that she should pay more attention in cleaning the room. With an extremely disappointed voice, Lily tells Rose that her behaviour is very bad without an apology. Rose becomes lucid. Lily understands Rose is no more confused. Rose asks Lily to sit down with her. Rose asks Lily how she was feeling. Rose apologises with Lily for her bad behaviour..

Scenario 5: Two Brothers

Andrew is a young boy. Robert is an elder brother of Andrew. They are at their home. They are planning to watch wrestling tonight. They are very excited and start to discuss about the players of the match tonight. Both of them are in a slightly excited mood. Andrew tries to irritate Robert by telling bad things about Robert's favourite player. Robert tries to ignore what Andrew says. However, Andrew continues to irritate Robert. Little annoyed, Robert tells Andrew to get away and pushes gently. Andrew gets violent and starts to shout at Robert. Full of rage, Andrew slaps and kicks Robert..

Scenario 6: Café Staff and Customer

Gopal is a café staff. It is a very busy Monday afternoon. Yet, Gopal is in neutral mood. A customer (Hari) comes to the café and orders food. Order takes very long to be served. Finally, the food arrives. Hari thanks Gopal for serving the food. Hari complains Gopal about the late service. Gopal apologises for being late. While trying to eat, Hari finds out that the food served is not as per the order. Hari complains Gopal about the wrong order. Gopal sympathises with Hari and promises to replace the food, but for Hari it is not enough. Hari asks Gopal for a refund. Gopal offers complimentary item with main order and promises to serve it quickly. Hari agrees with Gopal's offer. Order arrives quite quickly. Hari appreciates Gopal for quick service..

Appendix B – Surveys

Study A – Action Scoring Survey

Below are some scenarios of interaction between two persons. Please specify how you would rate the positive or negative effect of the interaction from one person to another.

Please complete all the scenarios.

Scenario 1 It's 1pm of the last day of the year and New Year is about to come. Rosy is sitting on a bench in a park, while Bill sit on the same bench of Rosy. Bill and Rosy do not know each other. Rosy is an easy-going girl.

Below are the interactions Bill will have with Rosy. The interactions Rosy will have with Bill are only partially available. You are asked to guess, how positive or negative each social exchange of Bill would be ***perceived by Rosy*** *in that specific context.*

How negative or positive the following social exchanges would be perceived by Rosy?

	Extremely Negative	Very Negative	Negative	Neither Negative Nor Positive	Positive	Very Positive	Extremely Positive
Bill says hello to Rosy	○	○	○	○	○	○	○
Bill wishes 'Happy New Year' to Rosy	○	○	○	○	○	○	○
Bill smiles (Rosy see the smile directed to her)	○	○	○	○	○	○	○
Bill offers some chocolates to Rosy	○	○	○	○	○	○	○
Bill starts conversation with Rosy	○	○	○	○	○	○	○
Bill asks Rosy what are her plans for New Year's Eve	○	○	○	○	○	○	○
Bill makes appreciations of Rosy's plan	○	○	○	○	○	○	○
Bill declines Rosy's offer to join her for the party saying he has already a plan with his girlfriend	○	○	○	○	○	○	○
Rosy ignores Bill and he also reciprocates by ignoring Rosy	○	○	○	○	○	○	○

Scenario 2 John and Paul are close friends and they have been in a good relationship for long time. It's summer and they plan to meet at a nearby beach. It's a good sunny day. *Below are the interactions John can have with Paul. The interactions are only **partially available** i.e. only the actions from John to Paul. You are asked **to guess**, how positive or negative each social exchange of John would be **perceived by Paul** in that specific context.*

How negative or positive the following social exchanges would be perceived by Paul?

	Extremely Negative	Very Negative	Negative	Neither Negative Nor Positive	Positive	Very Positive	Extremely Positive
John greets Paul	○	○	○	○	○	○	○
John smiles (Paul see the smile directed to him)	○	○	○	○	○	○	○
They go for swimming. John saves Paul from drowning	○	○	○	○	○	○	○
John hugs Paul	○	○	○	○	○	○	○
After a while, John makes fun of Paul for not being able to swim	○	○	○	○	○	○	○
John shouts back at Paul in response to Paul's yelling	○	○	○	○	○	○	○
John becomes physical with Paul (Eg. kicking)	○	○	○	○	○	○	○
John apologizes for his bad behaviour seeing Paul upset	○	○	○	○	○	○	○

Scenario 3 David and Anna are husband and wife. Today is Anna's birthday. David has not yet wished her birthday. He comes back home from work in the evening. David doesn't yet know that today is Anna's birthday.

*Below are the interactions Anna can have with David. The interactions are only **partially available** i.e. only the actions from Anna to David. You are asked **to guess**, how positive or negative each social exchange of Anna would be **perceived by David** in that specific context.*

How negative or positive the following social exchanges would be perceived by David?

	Extremely Negative	Very Negative	Negative	Neither Negative Nor Positive	Positive	Very Positive	Extremely Positive
Anna ignores David when he greets her	◯	◯	◯	◯	◯	◯	◯
Anna continues to ignore David when he starts a conversation	◯	◯	◯	◯	◯	◯	◯
Anna complains David about forgetting her birthday	◯	◯	◯	◯	◯	◯	◯
Anna makes fun of David's memory (when he regrets about forgetting her birthday)	◯	◯	◯	◯	◯	◯	◯
Anna smiles at David after he decides to cook dinner for her	◯	◯	◯	◯	◯	◯	◯
Anna appreciates David for cooking dinner	◯	◯	◯	◯	◯	◯	◯
Anna forgives David	◯	◯	◯	◯	◯	◯	◯
Anna hugs David	◯	◯	◯	◯	◯	◯	◯
Anna kisses David	◯	◯	◯	◯	◯	◯	◯

Scenario 4 Rose is a dementia patient in an elderly care home. Lily is a nurse who has been taking care of her. Lily goes into Rose's room to serve her. *Below are the interactions Rose can have with Lily. The interactions are only **partially available** i.e. only the actions from Rose to Lily. You are asked **to guess**, how positive or negative each social exchange of Rose would be **perceived by Lily** in that specific context.*

How negative or positive the following social exchanges would be perceived by Lily?

	Extremely Negative	Very Negative	Negative	Neither Negative Nor Positive	Positive	Very Positive	Extremely Positive
Rose greets Lily	○	○	○	○	○	○	○
Rose orders Lily to make her hair	○	○	○	○	○	○	○
Rose shouts at Lily because of pain in hair	○	○	○	○	○	○	○
Rose tries to slap Lily in the face	○	○	○	○	○	○	○
Rose prevents Lily to leave the room	○	○	○	○	○	○	○
Rose continues to prevent Lily from leaving saying that the room is not clean	○	○	○	○	○	○	○
Rose says to Lily that she should pay more attention in cleaning the room	○	○	○	○	○	○	○
Lily says that whatever she is doing is for the good of Rosy and she should feel grateful and behave nicely. Rose asks Lily to sit down	○	○	○	○	○	○	○
Rose asks Lily how she feels	○	○	○	○	○	○	○
Rose apologizes with Lily for her behaviour	○	○	○	○	○	○	○

Scenario 5 Robert and Andrew are two brothers. Robert is elder than Andrew. They are at their home. They are planning to watch wrestling tonight. They are very excited and start to discuss about the players of the match tonight. *Below are the interactions Robert can have with Andrew. The interactions are only* ***partially available*** *i.e. only the actions from Andrew to Robert. You are asked* ***to guess****, how positive or negative each social exchange of Andrew would be* ***perceived by Robert*** *in that specific context.*

How negative or positive the following social exchanges would be perceived by Robert?

	Extremely Negative	Very Negative	Negative	Neither Negative Nor Positive	Positive	Very Positive	Extremely Positive
Andrew tells bad things about Robert's favourite player	○	○	○	○	○	○	○
Andrew continues to irritate Robert	○	○	○	○	○	○	○
Andrew shouts at Robert when he pushes Andrew away	○	○	○	○	○	○	○
Andrew slaps Robert	○	○	○	○	○	○	○
Andrew kicks Robert	○	○	○	○	○	○	○

Scenario 6 Gopal is a café staff. It's a very busy Monday afternoon. A customer (Hari) comes to the café and orders a food. Order takes very long to be served. Finally, the food arrives. *Below are the interactions Hari can have with Gopal. The interactions are only **partially available** i.e. only the actions from Hari to Gopal. You are asked **to guess**, how positive or negative each social exchange of Hari would be **perceived by Gopal** in that specific context.*

How negative or positive the following social exchanges would be perceived by Gopal?

	Extremely Negative	Very Negative	Negative	Neither Negative Nor Positive	Positive	Very Positive	Extremely Positive
Hari thanks Gopal for serving the food	○	○	○	○	○	○	○
Hari complains Gopal about the late service	○	○	○	○	○	○	○
Hari complains Gopal about the wrong order	○	○	○	○	○	○	○
Hari asks Gopal for a refund	○	○	○	○	○	○	○
Gopal offers complimentary item. Hari agrees with Gopal's offer	○	○	○	○	○	○	○
Hari appreciates Gopal for quick service	○	○	○	○	○	○	○

Did you have any difficulty answering the survey questions? Is there something that can be changed to improve your survey experience?

Study A – Emotion Survey Set 1

Just a few questions about you.....

What is your gender?

- ○ Male
- ○ Female
- ○ Other

Which of the following closely defines you?

- ○ Outgoing / Friendly / Social
- ○ Reserved / Hesitant/ Shy

What is your current mood?

- ○ Good
- ○ Neutral
- ○ Bad

Now, you will be presented with 3 scenarios of interaction between two persons. One of them will perform a sequence of actions to another person. You are expected to **guess the emotional state of the target of interaction in response to the action of the actor**. For example, if A does some action to B, you are expected to guess the emotional state of B. To do so, you have to provide a scale between "Very Low" to "Very High" for all the emotions. The emotions considered in this study are following.

Joy : A feeling of pleasure or happiness
Distress : A feeling of anxiety, sorrow, or pain
Appreciation : A feeling when one recognizes the good qualities or actions of someone
Reproach : To express to (someone) one's disapproval of or disappointment in their actions
Gratitude : The state of being grateful to someone
Anger : A strong feeling of annoyance, displeasure, or hostility
Liking : A feeling when you see someone appealing or interesting
Disliking : A feeling when you see someone unappealing or uninteresting

Scenario 1 It's 1pm of the last day of the year and New Year is about to come. Rosy is sitting on a bench in a park, while Bill sit on the same bench of Rosy. Bill and Rosy do not know each other. Rosy is an easy-going girl and she is currently in a neutral emotional state. *Below are the interactions Bill will have with Rosy. The interactions are only **partially available** i.e. all Bill's actions, but only few of Rosy's. You are asked **to guess**, for each interaction of Bill, what are the chances **Rosy would happen to be in a particular emotional state**, based on the just happened interaction and the previously occurred interactions.*

Bill says hello to Rosy.
How would you rate the following emotions of Rosy?

	Not At All	Very Low	Low	Medium	High	Very High
Joy	○	○	○	○	○	○
Distress	○	○	○	○	○	○
Appreciation	○	○	○	○	○	○
Reproach	○	○	○	○	○	○
Gratitude	○	○	○	○	○	○
Anger	○	○	○	○	○	○
Liking	○	○	○	○	○	○
Disliking	○	○	○	○	○	○

Bill wishes 'Happy New Year' to Rosy.
How would you rate the following emotions of Rosy?

Bill smiles (Rosy can see the smile directed to her).
How would you rate the following emotions of Rosy?

Bill offers some chocolates (he was eating) to Rosy.
How would you rate the following emotions of Rosy?

Rosy accepts and eats a chocolate.

Bill starts conversation with Rosy.
How would you rate the following emotions of Rosy?

While talking, the conversation goes on the plans for New Year's Eve.

Bill asks Rosy what are her plans for New Year's Eve.
How would you rate the following emotions of Rosy?

Rosy answers that she will have a party at home with a lot of friends.

Bill makes appreciations of Rosy's plan.
How would you rate the following emotions of Rosy?

Rosy asks to Bill if he would like to join her in the party.

Bill declines the offer saying he has already a plan with his girlfriend.
How would you rate the following emotions of Rosy?

Rosy thinks Bill is just making up an excuse to not hang out with her and starts to ignore Bill.

Bill reciprocates by ignoring Rosy.
How would you rate the following emotions of Rosy?

Scenario 2 John and Paul are close friends and they have been in a good relationship for long time. It's summer and they plan to meet at a nearby beach. It's a good sunny day. Both of them are in neutral emotional state. *Below are the interactions John will have with Paul. The interactions are only **partially available** i.e. all John's actions, but only few of Paul's. You are asked **to guess**, for each interaction of John, what are the chances **Paul would happen to be in a particular emotional state**, based on the just happened interaction and the previously occurred interactions.*

John greets Paul.
How would you rate the following emotions of Paul?

John smiles (Paul can see the smile directed to him).
How would you rate the following emotions of Paul?

John and Paul sit down and talk for a while. Afterwards they decide to go for swimming. Paul is bad at swimming and is about to drown.

John saves Paul from drowning.
How would you rate the following emotions of Paul?

Paul thanks John for saving his life.

John hugs Paul.
How would you rate the following emotions of Paul?

Both of them go to the nearby bench, take rest for a while and have a can of beer.
Close to them there is a group of young girls able to hear their conversation.

John makes fun of Paul for not being able to swim.
How would you rate the following emotions of Paul?

Paul shouts at John.

John shouts back at Paul.
How would you rate the following emotions of Paul?

John becomes physical with Paul. Eg. kicking
How would you rate the following emotions of Paul?

Paul takes distance from John.

John apologizes for his bad behaviour.
How would you rate the following emotions of Paul?

Scenario 3 Robert and Andrew are two brothers. Robert is elder than Andrew. They are at their home. They are planning to watch wrestling tonight. They are very excited and start to discuss about the players of the match tonight. Both of them are in a slightly excited mood.
*Below are the interactions Robert will have with Andrew. The interactions are only **partially available** i.e. all Andrew's actions, but only few of Robert's. You are asked **to guess**, for each interaction of Andrew, what are the chances **Robert would happen to be in a particular emotional state**, based on the just happened interaction and the previously occurred interactions.*

Andrew tells bad things about Robert's favourite player.
How would you rate the following emotions of Robert?

Robert tries to ignore what Andrew says.

Andrew continues to irritate Robert.
How would you rate the following emotions of Robert?

Robert pushes Andrew.

Andrew shouts at Robert.
How would you rate the following emotions of Robert?

Andrew slaps Robert.
How would you rate the following emotions of Robert?

Andrew kicks Robert.
How would you rate the following emotions of Robert?

Did you have any difficulty answering the survey questions? Is there something that can be changed to improve your survey experience?

Study A – Emotion Survey Set 2

Just a few questions about you.....

What is your gender?

- Male
- Female
- Other

Which of the following closely defines you?

- Outgoing / Friendly / Social
- Reserved / Hesitant/ Shy

What is your current mood?

- Good
- Neutral
- Bad

Now, you will be presented with 3 scenarios of interaction between two persons. One of them will perform a sequence of actions to another person. You are expected to **guess the emotional state of the target of interaction in response to the action of the actor**. For example, if A does some action to B, you are expected to guess the emotional state of B. To do so, you have to provide a scale between "Very Low" to "Very High" for all the emotions. The emotions considered in this study are following.

Joy : A feeling of pleasure or happiness
Distress : A feeling of anxiety, sorrow, or pain
Appreciation : A feeling when one recognizes the good qualities or actions of someone
Reproach : To express to (someone) one's disapproval of or disappointment in their actions
Gratitude : The state of being grateful to someone
Anger : A strong feeling of annoyance, displeasure, or hostility
Liking : A feeling when you see someone appealing or interesting
Disliking : A feeling when you see someone unappealing or uninteresting

Scenario 1 David and Anna are husband and wife. Today is Anna's birthday. David has not yet wished her birthday. He comes back home from work in the evening. David doesn't yet know that today is Anna's birthday. David is in neutral mood while Anna is a bit upset.

*Below are the interactions Anna will have with David. The interactions are only **partially available** i.e. all Anna's actions, but only few of David's. You are asked **to guess**, for each interaction of Anna, what are the chances **David would happen to be in a particular emotional state**, based on the just happened interaction and the previously occurred interactions.*

Anna ignores David.

How would you rate the following emotions of David?

	Not At All	Very Low	Low	Medium	High	Very High
Joy	○	○	○	○	○	○
Distress	○	○	○	○	○	○
Appreciation	○	○	○	○	○	○
Reproach	○	○	○	○	○	○
Gratitude	○	○	○	○	○	○
Anger	○	○	○	○	○	○
Liking	○	○	○	○	○	○
Disliking	○	○	○	○	○	○

David tries to start a conversation.

Anna ignores David again.
How would you rate the following emotions of David?

Anna complains David about forgetting her birthday.
How would you rate the following emotions of David?

David realizes that he forgot the birthday.

Anna makes fun of David's bad memory.
How would you rate the following emotions of David?

Anna scolds David.
How would you rate the following emotions of David?

David wants to make up for his error. He says he will cook a special dinner for Anna.

Anna smiles (David is able to see that Anna smiles at him).
How would you rate the following emotions of David?

David prepares the dinner and then they both start to eat.

Anna appreciates David for cooking dinner.
How would you rate the following emotions of David?

Anna forgives David.
How would you rate the following emotions of David?

Anna hugs David.
How would you rate the following emotions of David?

Anna kisses David.
How would you rate the following emotions of David?

Scenario 2 Rose is a dementia patient in an elderly care home. Lily is a nurse who has been taking care of her and there are no other nurses at the moment in the elderly care home. Lily goes into Rose's room to serve her. Both of them are in neutral mood. Lily enters the room and says "good morning" to Rose.
*Below are the interactions Rose will have with Lily. The interactions are only **partially available** i.e. all Rose's actions, but only few of Lily's. You are asked **to guess**, for each interaction of Rose, what are the chances **Lily would happen to be in a particular emotional state**, based on the just happened interaction and the previously occurred interactions.*

Rose greets Lily back.
How would you rate the following emotions of Lily?

Rose orders Lily to make her hair.
How would you rate the following emotions of Lily?

Lily reminds Rose to ask for favours instead of giving orders. Rose loses her lucidity.

Rose shouts at Lily.
How would you rate the following emotions of Lily?

Rose tries to slap Lily in the face.
How would you rate the following emotions of Lily?

Rose prevents Lily from leaving the room.
How would you rate the following emotions of Lily?

Rose continues to prevent Lily from leaving saying that the room is not clean.
How would you rate the following emotions of Lily?

Lily tries to clean the room in order to calm down Rose and leave the room. Rose thinks Lily is not cleaning the room enough.

Rose says to Lily that she should pay more attention in cleaning the room.
How would you rate the following emotions of Lily?

Lily tells Rose that her behaviour is very bad without excuses. Rose becomes lucid. Lily understands Rose is no more confused.

Rose asks Lily to sit down.
How would you rate the following emotions of Lily?

Rose asks Lily how she feels.
How would you rate the following emotions of Lily?

Rose apologizes with Lily for her behaviour.
How would you rate the following emotions of Lily?

Scenario 3 Gopal is a café staff. It's a very busy Monday afternoon. Yet, Gopal is in neutral mood. A customer (Hari) comes to the café and orders food. Order takes very long to be served. Finally, the food arrives.
*Below are the interactions Hari will have with Gopal. The interactions are only **partially available** i.e. all Hari's actions, but only few of Gopal's. You are asked **to guess**, for each interaction of Hari, what are the chances **Gopal would happen to be in a particular emotional state**, based on the just happened interaction and the previously occurred interactions.*

Hari thanks Gopal for serving the food.
How would you rate the following emotions of Gopal?

Hari complains Gopal about the late service.
How would you rate the following emotions of Gopal?

Gopal apologizes for being late. While trying to eat, Hari finds out that the food served is not as per the order.

Hari complains Gopal about the wrong order.
How would you rate the following emotions of Gopal?

Gopal sympathizes with Hari and promises to replace the food, but for Hari it is not enough.

Hari asks Gopal for a refund.
How would you rate the following emotions of Gopal?

Gopal offers complimentary item with main order and promises to serve it quickly.

Hari agrees with Gopal's offer.
How would you rate the following emotions of Gopal?

Order arrives quite quickly.

Hari appreciates Gopal for quick service.
How would you rate the following emotions of Gopal?

Did you have any difficulty answering the survey questions? Is there something that can be changed to improve your survey experience?

Study B – Appraisal and Emotion Survey

Just a few questions about you.....

What is your gender?

- ◯ Male
- ◯ Female
- ◯ Other

How much do you agree that you have the following personality traits?

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Openness (risk taking, curious)	◯	◯	◯	◯	◯
Conscienciousness (organised, systematic, conscious)	◯	◯	◯	◯	◯
Extroversion (talkative, frank, outgoing)	◯	◯	◯	◯	◯
Agreeableness (non-argumentative, forgiving)	◯	◯	◯	◯	◯
Neurotic (pessimistic, negative emotionality - sadness)	◯	◯	◯	◯	◯

Now, you will be presented with a scenario of interaction between two persons. One of them will perform a sequence of actions to another person. You are expected to **guess how the target of the interaction might evaluate the actions from the source and what emotional state is the target of interaction likely to be** (in your opinion). For example, if Bill does some action to Rosy, you are expected to guess the evaluation and emotional state of Rosy. **Assume that initially both of them are in NEUTRAL mood.**

It's 1pm of the last day of the year and New Year is about to come. Rosy is sitting on a bench in a park, while Bill sits on the same bench of Rosy. Bill and Rosy do not know each other. Rosy is an easy-going girl and she is currently in a neutral emotional state. *Below are the interactions Bill will have with Rosy. The interactions for evaluation are only **partially available** i.e. all Bill's actions, but not of Rosy's. You are asked **to guess**, for each interaction of Bill, what are the chances **Rosy would evaluate the event and happen to be in a particular emotional state**, based on the just happened interaction and the previously occurred interactions.*

Bill says 'Hello' to Rosy

Bill wishes 'Happy New Year' to Rosy

Bill smiles with Rosy

Bill offers some chocolates he was eating to Rosy

Bill starts conversation with Rosy

Bill asks Rosy what are her plans for New Year's Eve

Bill makes appreciations of Rosy's plan

Bill declines Rosy's offer to join party saying he has already a plan with his girlfriend

Bill ignores Rosy because he thinks she is no more interested in talking to him

Bill continues ignoring and doesn't talk

Bill mentions that he was only passing time with Rosy

Appraisals

In this section, you are expected to guess how Rosy would evaluate the actions of Bill.

Desirability

How desirable do you think each of the following actions of Bill will be considered by Rosy?

	Extremely Undesirable	Very Undesirable	Undesirable	Neutral	Desirable	Very Desirable	Extremely Desirable
Bill says 'Hello' to Rosy	○	○	○	○	○	○	○
Bill wishes ‘Happy New Year’ to Rosy	○	○	○	○	○	○	○
Bill smiles with Rosy	○	○	○	○	○	○	○
Bill offers some chocolates he was eating to Rosy	○	○	○	○	○	○	○
Bill starts conversation with Rosy	○	○	○	○	○	○	○
Bill asks Rosy what are her plans for New Year’s Eve	○	○	○	○	○	○	○
Bill makes appreciations of Rosy’s plan	○	○	○	○	○	○	○
Bill declines Rosy's offer to join party saying he has already a plan with his girlfriend	○	○	○	○	○	○	○
Bill ignores Rosy because he thinks she is no more interested in talking to him	○	○	○	○	○	○	○
Bill continues ignoring and doesn't talk	○	○	○	○	○	○	○
Bill mentions that he was only passing time with Rosy	○	○	○	○	○	○	○

[Note: The actions in the survey follow the sequence as above in the sections below (not shown here).]

Praiseworthiness

How praiseworthy do you think each of the following actions of Bill will be considered by Rosy?

Extremely Blameworthy………Extremely Praiseworthy

Appealingness

How do you think will the following actions change the appealingness of Bill?

Extremely Unappealing…………Extremely Appealing

Deservingness

How much do you think Rosy believed that she deserved what just happened?

Extremely Non-deserving……..Extremely Deserving

Familiarity

How much do you think each action of Bill will change the familiarity of Rosy towards Bill?

Fully Unfamiliar (Stranger)………Fully Familiar (Well known)

Unexpectedness

How much do you think each action of Bill was NOT expected by Rosy?

Fully Expected…….Fully Unexpected

Emotions

In this section, you are expected to guess how the emotional responses of Rosy would be in reaction to the actions of Bill.

Joy (Happiness)
How much Joy do you think will be experienced by Rosy in response to the following actions of Bill?

	Not At All	Very Low	Low	Medium	High	Very High
Bill says 'Hello' to Rosy	○	○	○	○	○	○
Bill wishes ‘Happy New Year’ to Rosy	○	○	○	○	○	○
Bill smiles with Rosy	○	○	○	○	○	○
Bill offers some chocolates he was eating to Rosy	○	○	○	○	○	○
Bill starts conversation with Rosy	○	○	○	○	○	○
Bill asks Rosy what are her plans for New Year’s Eve	○	○	○	○	○	○
Bill makes appreciations of Rosy’s plan	○	○	○	○	○	○
Bill declines Rosy's offer to join party saying he has already a plan with his girlfriend	○	○	○	○	○	○
Bill ignores Rosy because he thinks she is no more interested in talking to him	○	○	○	○	○	○
Bill continues ignoring and doesn't talk	○	○	○	○	○	○
Bill mentions that he was only passing time with Rosy	○	○	○	○	○	○

Distress (Sadness)
How much Distress (Sadness) do you think will be experienced by Rosy in response to the following actions of Bill?

Appreciation

How much do you think Rosy would appreciate the following actions of Bill?

Reproach (Disappointment)

How much do you think Rosy would be disappointed by the following actions of Bill?

Gratitude (Thankfulness)

How much do you think Rosy would feel thankful towards Bill by the following actions of him?

Anger

How much do you think Rosy would be angry by the following actions of Bill?

Liking

How much do you think the following actions of Bill will make Rosy like him?

Disliking

How much do you think the following actions of Bill will make Rosy dislike him?

Mood

How much do you think the following actions of Bill will change the mood of Rosy?

(Note that initially mood is neutral and ***mood does NOT change as frequently as emotions****)*

	Extremely Negative	Very Negative	Negative	Neutral	Positive	Very Positive	Extremely Positive
Bill says 'Hello' to Rosy	○	○	○	○	○	○	○
Bill wishes ‘Happy New Year’ to Rosy	○	○	○	○	○	○	○
Bill smiles with Rosy	○	○	○	○	○	○	○
Bill offers some chocolates he was eating to Rosy	○	○	○	○	○	○	○
Bill starts conversation with Rosy	○	○	○	○	○	○	○
Bill asks Rosy what are her plans for New Year’s Eve	○	○	○	○	○	○	○
Bill makes appreciations of Rosy’s plan	○	○	○	○	○	○	○
Bill declines Rosy's offer to join party saying he has already a plan with his girlfriend	○	○	○	○	○	○	○
Bill ignores Rosy because he thinks she is no more interested in talking to him	○	○	○	○	○	○	○
Bill continues ignoring and doesn't talk	○	○	○	○	○	○	○
Bill mentions that he was only passing time with Rosy	○	○	○	○	○	○	○

Did you have any difficulty answering the survey questions? Is there something that can be changed to improve your survey experience?

Appendix C – Survey Data Statistics

Study	Part	Scenario	Event	Appraisal	Emotion	Min	Max	Mean	Variance
Study A									
A	Action Scoring	Two Strangers in a Park	1			0.00	0.66	0.31	0.04
			2			0.33	0.66	0.40	0.02
			3			0.00	0.66	0.33	0.03
			4			0.00	0.66	0.17	0.04
			5			0.00	0.66	0.28	0.03
			6			0.00	0.66	0.19	0.04
			7			0.00	0.33	0.31	0.01
			8			-0.33	0.66	0.09	0.05
			9			-0.66	0.33	-0.17	0.06
A	Action Scoring	Two Close Friends	1			0.33	1.00	0.57	0.07
			2			0.00	1.00	0.52	0.07
			3			0.33	1.00	0.76	0.07
			4			0.00	1.00	0.69	0.12
			5			-0.66	0.00	-0.31	0.04
			6			-0.66	0.00	-0.33	0.03
			7			-1.00	-0.33	-0.74	0.10
			8			0.00	0.66	0.38	0.03
A	Action Scoring	Husband and Wife	1			-1.00	1.00	-0.21	0.28
			2			-1.00	0.33	-0.38	0.19
			3			-1.00	0.66	0.07	0.18
			4			-1.00	0.66	-0.31	0.16
			5			-0.33	1.00	0.38	0.08
			6			0.33	1.00	0.54	0.04
			7			0.33	1.00	0.76	0.04
			8			0.33	1.00	0.83	0.06
			9			0.33	1.00	0.90	0.04
A	Action Scoring	Patient and a Nurse	1			0.33	1.00	0.52	0.08
			2			-0.33	0.66	0.09	0.07
			3			-0.66	0.00	-0.24	0.05
			4			-1.00	0.00	-0.43	0.15
			5			-0.66	0.00	-0.24	0.05
			6			-0.66	0.00	-0.31	0.04
			7			-0.66	0.00	-0.26	0.08
			8			-0.33	0.33	0.14	0.04
			9			0.00	0.66	0.35	0.04
			10			0.33	1.00	0.57	0.07
A	Action Scoring	Two Brothers	1			-0.66	0.33	-0.24	0.05
			2			-0.66	0.00	-0.35	0.04
			3			-1.00	0.33	-0.35	0.15
			4			-1.00	0.00	-0.67	0.13
			5			-1.00	-0.33	-0.71	0.09
A	Action Scoring	Café Staff and Customer	1			0.00	1.00	0.52	0.09
			2			-0.33	0.33	-0.02	0.05
			3			-0.33	0.33	0.05	0.06
			4			-0.66	0.33	-0.09	0.07
			5			-0.33	0.66	0.33	0.06
			6			0.33	1.00	0.50	0.06
A	Emotion Survey	Two Strangers in a Park	1		Joy	0.00	0.80	0.48	0.06
					Distress	0.00	0.80	0.29	0.05
					Appreciation	0.00	0.80	0.50	0.04
					Reproach	0.00	0.80	0.30	0.06
					Gratitude	0.00	0.60	0.45	0.04
					Anger	0.00	0.60	0.14	0.04
					Liking	0.00	0.80	0.42	0.05
					Disliking	0.00	0.60	0.18	0.03
			2		Joy	0.00	1.00	0.59	0.05
					Distress	0.00	0.80	0.21	0.05
					Appreciation	0.40	1.00	0.65	0.02
					Reproach	0.00	0.60	0.24	0.06
					Gratitude	0.00	0.80	0.56	0.04
					Anger	0.00	0.60	0.16	0.04
					Liking	0.00	0.80	0.57	0.04
					Disliking	0.00	0.60	0.16	0.03
			3		Joy	0.00	1.00	0.59	0.05
					Distress	0.00	0.60	0.24	0.04
					Appreciation	0.00	1.00	0.58	0.04
					Reproach	0.00	0.60	0.27	0.05
					Gratitude	0.00	0.80	0.50	0.06
					Anger	0.00	0.60	0.14	0.03

					Liking	0.00	0.80	0.57	0.04
					Disliking	0.00	0.60	0.16	0.03
			4		Joy	0.00	1.00	0.56	0.06
					Distress	0.00	0.80	0.25	0.06
					Appreciation	0.40	0.80	0.64	0.02
					Reproach	0.00	0.60	0.29	0.05
					Gratitude	0.00	0.80	0.58	0.04
					Anger	0.00	0.80	0.21	0.06
					Liking	0.00	0.80	0.59	0.04
					Disliking	0.00	0.60	0.17	0.05
			5		Joy	0.40	1.00	0.66	0.03
					Distress	0.00	0.60	0.24	0.06
					Appreciation	0.00	1.00	0.64	0.04
					Reproach	0.00	0.60	0.29	0.06
					Gratitude	0.00	0.80	0.55	0.04
					Anger	0.00	1.00	0.14	0.05
					Liking	0.20	1.00	0.67	0.02
					Disliking	0.00	0.60	0.11	0.04
			6		Joy	0.40	1.00	0.63	0.03
					Distress	0.00	0.80	0.26	0.06
					Appreciation	0.00	0.80	0.56	0.06
					Reproach	0.00	0.80	0.29	0.06
					Gratitude	0.00	0.80	0.51	0.05
					Anger	0.00	0.60	0.13	0.03
					Liking	0.20	0.80	0.61	0.03
					Disliking	0.00	0.60	0.13	0.04
			7		Joy	0.20	1.00	0.70	0.04
					Distress	0.00	0.60	0.17	0.04
					Appreciation	0.20	1.00	0.65	0.04
					Reproach	0.00	0.80	0.22	0.06
					Gratitude	0.00	1.00	0.54	0.07
					Anger	0.00	0.60	0.09	0.04
					Liking	0.40	1.00	0.71	0.02
					Disliking	0.00	0.80	0.11	0.04
			8		Joy	0.00	0.80	0.31	0.06
					Distress	0.00	0.80	0.39	0.05
					Appreciation	0.00	0.80	0.36	0.06
					Reproach	0.00	0.60	0.39	0.06
					Gratitude	0.00	0.80	0.34	0.05
					Anger	0.00	0.60	0.31	0.06
					Liking	0.00	1.00	0.37	0.07
					Disliking	0.00	1.00	0.48	0.08
			9		Joy	0.00	0.80	0.24	0.06
					Distress	0.00	0.80	0.56	0.06
					Appreciation	0.00	0.60	0.21	0.05
					Reproach	0.00	1.00	0.49	0.07
					Gratitude	0.00	0.60	0.19	0.04
					Anger	0.00	1.00	0.54	0.05
					Liking	0.00	0.60	0.21	0.04
					Disliking	0.00	1.00	0.59	0.08
A	Emotion Survey	Two Close Friends	1		Joy	0.00	1.00	0.73	0.04
					Distress	0.00	0.80	0.16	0.05
					Appreciation	0.20	1.00	0.70	0.03
					Reproach	0.00	0.80	0.23	0.09
					Gratitude	0.00	1.00	0.62	0.05
					Anger	0.00	0.60	0.10	0.04
					Liking	0.40	1.00	0.76	0.02
					Disliking	0.00	0.60	0.10	0.03
			2		Joy	0.00	1.00	0.80	0.04
					Distress	0.00	0.80	0.13	0.04
					Appreciation	0.00	1.00	0.73	0.04
					Reproach	0.00	0.80	0.16	0.07
					Gratitude	0.20	1.00	0.68	0.03
					Anger	0.00	0.60	0.06	0.03
					Liking	0.40	1.00	0.79	0.02
					Disliking	0.00	0.60	0.07	0.02
			3		Joy	0.00	1.00	0.61	0.10
					Distress	0.00	1.00	0.44	0.12
					Appreciation	0.20	1.00	0.85	0.05
					Reproach	0.00	0.80	0.24	0.08
					Gratitude	0.20	1.00	0.85	0.05

					Anger	0.00	0.60	0.12	0.03
					Liking	0.20	1.00	0.77	0.06
					Disliking	0.00	0.80	0.14	0.05
			4		Joy	0.00	1.00	0.79	0.06
					Distress	0.00	1.00	0.33	0.10
					Appreciation	0.40	1.00	0.82	0.03
					Reproach	0.00	1.00	0.23	0.09
					Gratitude	0.40	1.00	0.81	0.03
					Anger	0.00	0.60	0.09	0.02
					Liking	0.20	1.00	0.77	0.05
					Disliking	0.00	0.60	0.12	0.03
			5		Joy	0.00	0.60	0.32	0.05
					Distress	0.00	0.80	0.52	0.05
					Appreciation	0.00	0.80	0.29	0.06
					Reproach	0.00	1.00	0.49	0.06
					Gratitude	0.00	0.80	0.30	0.08
					Anger	0.00	1.00	0.53	0.07
					Liking	0.00	0.80	0.30	0.06
					Disliking	0.00	1.00	0.54	0.09
			6		Joy	0.00	0.60	0.20	0.04
					Distress	0.00	1.00	0.57	0.06
					Appreciation	0.00	0.60	0.22	0.04
					Reproach	0.00	1.00	0.52	0.07
					Gratitude	0.00	0.60	0.19	0.05
					Anger	0.00	1.00	0.64	0.08
					Liking	0.00	0.40	0.17	0.03
					Disliking	0.00	1.00	0.64	0.07
			7		Joy	0.00	0.60	0.10	0.04
					Distress	0.00	1.00	0.70	0.08
					Appreciation	0.00	0.40	0.10	0.02
					Reproach	0.00	1.00	0.64	0.08
					Gratitude	0.00	0.40	0.12	0.02
					Anger	0.00	1.00	0.77	0.06
					Liking	0.00	0.60	0.12	0.02
					Disliking	0.40	1.00	0.84	0.02
			8		Joy	0.00	1.00	0.42	0.08
					Distress	0.00	0.80	0.44	0.06
					Appreciation	0.20	1.00	0.61	0.05
					Reproach	0.00	0.80	0.34	0.07
					Gratitude	0.00	0.80	0.47	0.07
					Anger	0.00	1.00	0.50	0.07
					Liking	0.00	1.00	0.57	0.05
					Disliking	0.00	1.00	0.40	0.10
A	Emotion Survey	Husband and Wife	1		Joy	0.00	1.00	0.33	0.06
					Distress	0.00	1.00	0.48	0.06
					Appreciation	0.00	0.60	0.24	0.05
					Reproach	0.00	0.80	0.36	0.06
					Gratitude	0.00	0.60	0.26	0.05
					Anger	0.00	0.80	0.38	0.05
					Liking	0.00	0.80	0.27	0.06
					Disliking	0.00	1.00	0.50	0.08
			2		Joy	0.00	0.60	0.16	0.04
					Distress	0.00	0.80	0.52	0.05
					Appreciation	0.00	0.80	0.24	0.05
					Reproach	0.00	1.00	0.41	0.07
					Gratitude	0.00	0.80	0.24	0.05
					Anger	0.00	1.00	0.44	0.07
					Liking	0.00	0.80	0.28	0.07
					Disliking	0.00	1.00	0.53	0.07
			3		Joy	0.00	0.80	0.21	0.05
					Distress	0.20	1.00	0.69	0.05
					Appreciation	0.00	1.00	0.34	0.06
					Reproach	0.00	1.00	0.33	0.09
					Gratitude	0.00	1.00	0.40	0.09
					Anger	0.00	0.80	0.24	0.07
					Liking	0.00	1.00	0.41	0.09
					Disliking	0.00	0.80	0.26	0.07
			4		Joy	0.00	1.00	0.36	0.09
					Distress	0.00	1.00	0.47	0.05
					Appreciation	0.00	1.00	0.39	0.08
					Reproach	0.00	0.80	0.37	0.06

					Gratitude	0.00	1.00	0.39	0.08
					Anger	0.00	0.80	0.33	0.06
					Liking	0.00	1.00	0.42	0.09
					Disliking	0.00	1.00	0.30	0.07
			5		Joy	0.00	1.00	0.23	0.08
					Distress	0.00	1.00	0.59	0.07
					Appreciation	0.00	1.00	0.33	0.10
					Reproach	0.00	0.80	0.37	0.09
					Gratitude	0.00	0.80	0.25	0.06
					Anger	0.00	0.80	0.36	0.07
					Liking	0.00	1.00	0.28	0.09
					Disliking	0.00	0.80	0.47	0.08
			6		Joy	0.40	1.00	0.76	0.03
					Distress	0.00	0.80	0.29	0.05
					Appreciation	0.00	1.00	0.68	0.05
					Reproach	0.00	1.00	0.33	0.11
					Gratitude	0.00	1.00	0.63	0.08
					Anger	0.00	0.80	0.16	0.05
					Liking	0.40	1.00	0.76	0.04
					Disliking	0.00	0.80	0.17	0.05
			7		Joy	0.40	1.00	0.84	0.02
					Distress	0.00	1.00	0.16	0.05
					Appreciation	0.40	1.00	0.79	0.04
					Reproach	0.00	1.00	0.27	0.12
					Gratitude	0.00	1.00	0.76	0.05
					Anger	0.00	1.00	0.15	0.07
					Liking	0.00	1.00	0.79	0.06
					Disliking	0.00	1.00	0.14	0.05
			8		Joy	0.60	1.00	0.91	0.02
					Distress	0.00	1.00	0.21	0.10
					Appreciation	0.60	1.00	0.86	0.02
					Reproach	0.00	1.00	0.24	0.11
					Gratitude	0.00	1.00	0.75	0.06
					Anger	0.00	1.00	0.13	0.07
					Liking	0.40	1.00	0.86	0.03
					Disliking	0.00	1.00	0.11	0.05
			9		Joy	0.40	1.00	0.93	0.02
					Distress	0.00	1.00	0.16	0.10
					Appreciation	0.60	1.00	0.90	0.02
					Reproach	0.00	1.00	0.24	0.14
					Gratitude	0.00	1.00	0.78	0.07
					Anger	0.00	1.00	0.13	0.08
					Liking	0.60	1.00	0.92	0.01
					Disliking	0.00	1.00	0.14	0.08
			10		Joy	0.80	1.00	0.95	0.01
					Distress	0.00	1.00	0.15	0.08
					Appreciation	0.00	1.00	0.87	0.04
					Reproach	0.00	1.00	0.21	0.11
					Gratitude	0.00	1.00	0.78	0.10
					Anger	0.00	0.80	0.13	0.06
					Liking	0.60	1.00	0.93	0.01
					Disliking	0.00	0.80	0.13	0.07
A	Emotion Survey	Patient and a Nurse	1		Joy	0.20	1.00	0.62	0.05
					Distress	0.00	1.00	0.21	0.06
					Appreciation	0.00	1.00	0.60	0.05
					Reproach	0.00	1.00	0.24	0.10
					Gratitude	0.00	1.00	0.53	0.04
					Anger	0.00	1.00	0.14	0.06
					Liking	0.40	1.00	0.65	0.04
					Disliking	0.00	1.00	0.17	0.06
			2		Joy	0.00	1.00	0.46	0.07
					Distress	0.00	1.00	0.35	0.05
					Appreciation	0.00	1.00	0.41	0.05
					Reproach	0.00	1.00	0.31	0.07
					Gratitude	0.00	1.00	0.40	0.06
					Anger	0.00	1.00	0.33	0.08
					Liking	0.00	1.00	0.48	0.05
					Disliking	0.00	1.00	0.37	0.10
			3		Joy	0.00	1.00	0.20	0.05
					Distress	0.20	1.00	0.64	0.05
					Appreciation	0.00	1.00	0.17	0.05

					Reproach	0.00	1.00	0.57	0.08
					Gratitude	0.00	1.00	0.20	0.06
					Anger	0.00	1.00	0.60	0.08
					Liking	0.00	1.00	0.25	0.06
					Disliking	0.00	1.00	0.66	0.06
			4		Joy	0.00	1.00	0.10	0.04
					Distress	0.40	1.00	0.81	0.03
					Appreciation	0.00	1.00	0.12	0.05
					Reproach	0.00	1.00	0.68	0.08
					Gratitude	0.00	1.00	0.10	0.05
					Anger	0.00	1.00	0.74	0.06
					Liking	0.00	1.00	0.15	0.05
					Disliking	0.40	1.00	0.82	0.05
			5		Joy	0.00	1.00	0.16	0.06
					Distress	0.00	1.00	0.81	0.05
					Appreciation	0.00	1.00	0.16	0.06
					Reproach	0.00	1.00	0.65	0.09
					Gratitude	0.00	1.00	0.20	0.07
					Anger	0.00	1.00	0.72	0.08
					Liking	0.00	1.00	0.17	0.06
					Disliking	0.00	1.00	0.79	0.06
			6		Joy	0.00	1.00	0.14	0.06
					Distress	0.40	1.00	0.77	0.04
					Appreciation	0.00	1.00	0.20	0.08
					Reproach	0.00	1.00	0.64	0.08
					Gratitude	0.00	1.00	0.15	0.06
					Anger	0.00	1.00	0.72	0.07
					Liking	0.00	1.00	0.15	0.05
					Disliking	0.40	1.00	0.78	0.04
			7		Joy	0.00	1.00	0.21	0.07
					Distress	0.40	1.00	0.77	0.04
					Appreciation	0.00	1.00	0.18	0.06
					Reproach	0.00	1.00	0.71	0.06
					Gratitude	0.00	1.00	0.18	0.06
					Anger	0.00	1.00	0.70	0.07
					Liking	0.00	1.00	0.17	0.06
					Disliking	0.40	1.00	0.80	0.03
			8		Joy	0.00	1.00	0.30	0.08
					Distress	0.00	1.00	0.56	0.06
					Appreciation	0.00	1.00	0.34	0.08
					Reproach	0.00	1.00	0.49	0.07
					Gratitude	0.00	1.00	0.31	0.08
					Anger	0.00	1.00	0.45	0.10
					Liking	0.00	1.00	0.40	0.09
					Disliking	0.00	1.00	0.49	0.08
			9		Joy	0.00	1.00	0.40	0.10
					Distress	0.00	1.00	0.48	0.08
					Appreciation	0.00	1.00	0.45	0.07
					Reproach	0.00	1.00	0.44	0.08
					Gratitude	0.00	1.00	0.44	0.07
					Anger	0.00	1.00	0.39	0.09
					Liking	0.00	1.00	0.49	0.08
					Disliking	0.00	1.00	0.43	0.09
			10		Joy	0.00	1.00	0.61	0.11
					Distress	0.00	1.00	0.25	0.07
					Appreciation	0.00	1.00	0.67	0.04
					Reproach	0.00	1.00	0.30	0.07
					Gratitude	0.00	1.00	0.61	0.06
					Anger	0.00	1.00	0.24	0.06
					Liking	0.00	1.00	0.63	0.06
					Disliking	0.00	1.00	0.26	0.07
A	Emotion Survey	Two Brothers	1		Joy	0.00	0.80	0.20	0.04
					Distress	0.00	1.00	0.48	0.06
					Appreciation	0.00	0.60	0.19	0.04
					Reproach	0.00	1.00	0.59	0.07
					Gratitude	0.00	0.60	0.21	0.04
					Anger	0.20	1.00	0.61	0.04
					Liking	0.00	0.80	0.22	0.06
					Disliking	0.00	1.00	0.68	0.05
			2		Joy	0.00	0.80	0.16	0.05
					Distress	0.00	1.00	0.58	0.07

					Appreciation	0.00	0.80	0.15	0.04
					Reproach	0.00	1.00	0.55	0.07
					Gratitude	0.00	0.80	0.17	0.05
					Anger	0.20	1.00	0.75	0.05
					Liking	0.00	0.80	0.15	0.05
					Disliking	0.40	1.00	0.80	0.03
			3		Joy	0.00	0.80	0.18	0.05
					Distress	0.00	1.00	0.66	0.08
					Appreciation	0.00	0.60	0.14	0.03
					Reproach	0.00	1.00	0.59	0.09
					Gratitude	0.00	0.80	0.14	0.05
					Anger	0.40	1.00	0.78	0.04
					Liking	0.00	0.60	0.17	0.04
					Disliking	0.00	1.00	0.76	0.06
			4		Joy	0.00	0.60	0.09	0.03
					Distress	0.20	1.00	0.79	0.06
					Appreciation	0.00	0.60	0.08	0.02
					Reproach	0.00	1.00	0.66	0.09
					Gratitude	0.00	0.60	0.08	0.02
					Anger	0.40	1.00	0.90	0.02
					Liking	0.00	0.60	0.11	0.03
					Disliking	0.40	1.00	0.92	0.02
			5		Joy	0.00	1.00	0.10	0.05
					Distress	0.20	1.00	0.81	0.06
					Appreciation	0.00	0.80	0.10	0.04
					Reproach	0.00	1.00	0.71	0.10
					Gratitude	0.00	0.80	0.11	0.06
					Anger	0.00	1.00	0.82	0.08
					Liking	0.00	0.80	0.14	0.06
					Disliking	0.00	1.00	0.88	0.04
A	Emotion Survey	Café Staff and Customer	1		Joy	0.20	1.00	0.59	0.04
					Distress	0.00	1.00	0.23	0.07
					Appreciation	0.00	1.00	0.66	0.05
					Reproach	0.00	1.00	0.27	0.07
					Gratitude	0.00	1.00	0.58	0.05
					Anger	0.00	1.00	0.13	0.05
					Liking	0.20	1.00	0.61	0.05
					Disliking	0.00	1.00	0.16	0.06
			2		Joy	0.00	0.80	0.26	0.06
					Distress	0.00	0.80	0.58	0.04
					Appreciation	0.00	1.00	0.30	0.07
					Reproach	0.00	1.00	0.47	0.08
					Gratitude	0.00	0.80	0.26	0.06
					Anger	0.00	0.80	0.43	0.06
					Liking	0.00	1.00	0.29	0.08
					Disliking	0.00	1.00	0.58	0.05
			3		Joy	0.00	0.80	0.16	0.04
					Distress	0.20	1.00	0.65	0.05
					Appreciation	0.00	0.80	0.22	0.05
					Reproach	0.00	1.00	0.50	0.08
					Gratitude	0.00	0.80	0.27	0.07
					Anger	0.00	1.00	0.48	0.07
					Liking	0.00	0.80	0.23	0.06
					Disliking	0.00	1.00	0.58	0.05
			4		Joy	0.00	0.60	0.11	0.03
					Distress	0.00	1.00	0.72	0.06
					Appreciation	0.00	0.60	0.16	0.03
					Reproach	0.00	1.00	0.50	0.09
					Gratitude	0.00	0.60	0.18	0.03
					Anger	0.00	1.00	0.54	0.07
					Liking	0.00	0.60	0.20	0.04
					Disliking	0.00	1.00	0.60	0.06
			5		Joy	0.00	1.00	0.51	0.07
					Distress	0.00	1.00	0.41	0.05
					Appreciation	0.00	1.00	0.53	0.07
					Reproach	0.00	1.00	0.40	0.08
					Gratitude	0.00	1.00	0.53	0.06
					Anger	0.00	0.60	0.30	0.03
					Liking	0.00	1.00	0.51	0.06
					Disliking	0.00	0.80	0.34	0.05
			6		Joy	0.00	1.00	0.67	0.06

					Distress	0.00	0.80	0.24	0.05
					Appreciation	0.00	1.00	0.65	0.06
					Reproach	0.00	1.00	0.32	0.08
					Gratitude	0.00	1.00	0.53	0.09
					Anger	0.00	0.80	0.20	0.05
					Liking	0.00	1.00	0.56	0.07
					Disliking	0.00	0.80	0.24	0.05
Study B									
B	Appraisal Survey	Two Strangers in a Park	1	Desirability		-0.33	1.00	0.36	0.13
				Praiseworthiness		-0.33	1.00	0.36	0.12
				Appealingness		-0.33	1.00	0.41	0.11
				Deservingness		-0.33	1.00	0.36	0.14
				Familiarity		0.00	1.00	0.53	0.08
				Unexpectedness		0.00	1.00	0.39	0.08
			2	Desirability		-0.33	1.00	0.41	0.13
				Praiseworthiness		-0.33	1.00	0.40	0.16
				Appealingness		-0.33	1.00	0.44	0.15
				Deservingness		-0.33	1.00	0.42	0.13
				Familiarity		0.00	1.00	0.56	0.07
				Unexpectedness		0.00	1.00	0.31	0.08
			3	Desirability		-0.33	1.00	0.39	0.16
				Praiseworthiness		-0.33	1.00	0.42	0.13
				Appealingness		-0.33	1.00	0.47	0.16
				Deservingness		-0.33	1.00	0.43	0.14
				Familiarity		0.00	1.00	0.59	0.07
				Unexpectedness		0.00	0.80	0.34	0.06
			4	Desirability		-0.33	1.00	0.39	0.16
				Praiseworthiness		-0.66	1.00	0.46	0.15
				Appealingness		0.00	1.00	0.57	0.13
				Deservingness		-0.33	1.00	0.32	0.14
				Familiarity		0.00	1.00	0.61	0.07
				Unexpectedness		0.00	1.00	0.54	0.07
			5	Desirability		-0.33	1.00	0.37	0.14
				Praiseworthiness		-0.66	1.00	0.37	0.16
				Appealingness		-0.33	1.00	0.47	0.13
				Deservingness		-0.33	1.00	0.38	0.11
				Familiarity		0.20	1.00	0.60	0.04
				Unexpectedness		0.00	1.00	0.42	0.07
			6	Desirability		-0.66	1.00	0.34	0.12
				Praiseworthiness		-0.66	1.00	0.39	0.17
				Appealingness		-0.66	1.00	0.40	0.17
				Deservingness		-1.00	1.00	0.25	0.15
				Familiarity		0.00	1.00	0.60	0.05
				Unexpectedness		0.00	1.00	0.49	0.07
			7	Desirability		-0.33	1.00	0.37	0.17
				Praiseworthiness		-0.66	1.00	0.43	0.17
				Appealingness		-0.66	1.00	0.48	0.13
				Deservingness		-0.66	1.00	0.31	0.12
				Familiarity		0.20	1.00	0.65	0.04
				Unexpectedness		0.00	1.00	0.48	0.08
			8	Desirability		-1.00	1.00	-0.11	0.29
				Praiseworthiness		-1.00	1.00	-0.03	0.21
				Appealingness		-1.00	1.00	-0.13	0.26
				Deservingness		-1.00	1.00	-0.08	0.25
				Familiarity		0.00	1.00	0.42	0.07
				Unexpectedness		0.00	1.00	0.63	0.08
			9	Desirability		-1.00	1.00	-0.32	0.29
				Praiseworthiness		-1.00	0.66	-0.34	0.26
				Appealingness		-1.00	0.66	-0.37	0.32
				Deservingness		-1.00	1.00	-0.36	0.35
				Familiarity		0.00	1.00	0.31	0.07
				Unexpectedness		0.00	1.00	0.66	0.09
			10	Desirability		-1.00	1.00	-0.37	0.38
				Praiseworthiness		-1.00	1.00	-0.37	0.31
				Appealingness		-1.00	1.00	-0.37	0.39
				Deservingness		-1.00	1.00	-0.42	0.35
				Familiarity		0.00	1.00	0.31	0.09
				Unexpectedness		0.00	1.00	0.69	0.09
			11	Desirability		-1.00	1.00	-0.38	0.37
				Praiseworthiness		-1.00	1.00	-0.40	0.32
				Appealingness		-1.00	1.00	-0.40	0.41

				Deservingness		-1.00	1.00	-0.41	0.34
				Familiarity		0.00	1.00	0.33	0.08
				Unexpectedness		0.00	1.00	0.71	0.09
B	Emotion Survey	Two Strangers in a Park	1		Joy	0.20	1.00	0.59	0.05
			2		Joy	0.20	1.00	0.67	0.04
			3		Joy	0.40	1.00	0.68	0.03
			4		Joy	0.00	1.00	0.71	0.05
			5		Joy	0.00	1.00	0.64	0.07
			6		Joy	0.00	1.00	0.65	0.06
			7		Joy	0.00	1.00	0.67	0.06
			8		Joy	0.00	1.00	0.31	0.07
			9		Joy	0.00	1.00	0.23	0.07
			10		Joy	0.00	1.00	0.20	0.06
			11		Joy	0.00	1.00	0.22	0.08
			1		Distress	0.00	0.80	0.18	0.07
			2		Distress	0.00	1.00	0.20	0.08
			3		Distress	0.00	0.80	0.19	0.06
			4		Distress	0.00	1.00	0.22	0.09
			5		Distress	0.00	1.00	0.19	0.07
			6		Distress	0.00	1.00	0.22	0.09
			7		Distress	0.00	0.80	0.20	0.07
			8		Distress	0.00	1.00	0.51	0.08
			9		Distress	0.00	1.00	0.64	0.09
			10		Distress	0.00	1.00	0.69	0.08
			11		Distress	0.00	1.00	0.71	0.09
			1		Appreciation	0.00	1.00	0.59	0.07
			2		Appreciation	0.00	1.00	0.63	0.06
			3		Appreciation	0.00	1.00	0.67	0.07
			4		Appreciation	0.20	1.00	0.76	0.04
			5		Appreciation	0.20	1.00	0.68	0.04
			6		Appreciation	0.00	1.00	0.63	0.07
			7		Appreciation	0.00	1.00	0.66	0.09
			8		Appreciation	0.00	1.00	0.35	0.08
			9		Appreciation	0.00	0.80	0.23	0.07
			10		Appreciation	0.00	1.00	0.22	0.07
			11		Appreciation	0.00	1.00	0.22	0.07
			1		Reproach	0.00	1.00	0.22	0.09
			2		Reproach	0.00	1.00	0.20	0.08
			3		Reproach	0.00	1.00	0.18	0.07
			4		Reproach	0.00	0.80	0.18	0.06
			5		Reproach	0.00	1.00	0.19	0.06
			6		Reproach	0.00	1.00	0.21	0.09
			7		Reproach	0.00	0.80	0.19	0.07
			8		Reproach	0.00	1.00	0.58	0.06
			9		Reproach	0.00	1.00	0.66	0.08
			10		Reproach	0.20	1.00	0.72	0.06
			11		Reproach	0.20	1.00	0.72	0.07
			1		Gratitude	0.00	1.00	0.54	0.06
			2		Gratitude	0.00	1.00	0.60	0.07
			3		Gratitude	0.00	1.00	0.60	0.07
			4		Gratitude	0.00	1.00	0.72	0.06
			5		Gratitude	0.20	1.00	0.63	0.05
			6		Gratitude	0.00	1.00	0.60	0.06
			7		Gratitude	0.00	1.00	0.69	0.06
			8		Gratitude	0.00	1.00	0.32	0.06
			9		Gratitude	0.00	1.00	0.21	0.06
			10		Gratitude	0.00	1.00	0.23	0.08
			11		Gratitude	0.00	1.00	0.23	0.08
			1		Anger	0.00	1.00	0.21	0.10
			2		Anger	0.00	1.00	0.19	0.09
			3		Anger	0.00	1.00	0.18	0.07
			4		Anger	0.00	1.00	0.17	0.08
			5		Anger	0.00	1.00	0.18	0.07
			6		Anger	0.00	1.00	0.21	0.09
			7		Anger	0.00	1.00	0.19	0.09
			8		Anger	0.00	1.00	0.51	0.07
			9		Anger	0.00	1.00	0.63	0.07
			10		Anger	0.00	1.00	0.68	0.09
			11		Anger	0.00	1.00	0.70	0.08
			1		Liking	0.00	1.00	0.57	0.06
			2		Liking	0.00	1.00	0.60	0.07

			3		Liking	0.20	1.00	0.69	0.05
			4		Liking	0.20	1.00	0.71	0.05
			5		Liking	0.00	1.00	0.67	0.07
			6		Liking	0.00	1.00	0.69	0.06
			7		Liking	0.00	1.00	0.69	0.06
			8		Liking	0.00	1.00	0.33	0.07
			9		Liking	0.00	1.00	0.22	0.06
			10		Liking	0.00	1.00	0.22	0.08
			11		Liking	0.00	1.00	0.23	0.08
			1		Disliking	0.00	1.00	0.22	0.09
			2		Disliking	0.00	1.00	0.21	0.09
			3		Disliking	0.00	1.00	0.22	0.10
			4		Disliking	0.00	0.80	0.19	0.08
			5		Disliking	0.00	1.00	0.21	0.08
			6		Disliking	0.00	1.00	0.23	0.08
			7		Disliking	0.00	1.00	0.20	0.08
			8		Disliking	0.00	1.00	0.48	0.08
			9		Disliking	0.20	1.00	0.69	0.07
			10		Disliking	0.00	1.00	0.71	0.07
			11		Disliking	0.00	1.00	0.76	0.07
B	Mood Survey	Two Strangers in a Park	1			-0.66	1.00	0.22	0.16
			2			-0.66	1.00	0.29	0.14
			3			-0.66	1.00	0.39	0.12
			4			-0.66	1.00	0.49	0.14
			5			-0.33	1.00	0.37	0.14
			6			-1.00	1.00	0.32	0.19
			7			-0.33	1.00	0.45	0.13
			8			-1.00	1.00	-0.13	0.18
			9			-1.00	0.66	-0.30	0.19
			10			-1.00	1.00	-0.39	0.20
			11			-1.00	1.00	-0.37	0.26