UNIVERSITÉ DE GENÈVE
FACULTÉ DE PSYCHOLOGIE ET DES SCIENCES DE L'ÉDUCATION

Reinventing the Wheel: Emotional Awareness Enhancement in Computer-Mediated Collaboration with the Dynamic Emotion Wheel

Mémoire réalisé en vue de l'obtention de la Maitrise Universitaire en Sciences et Technologies de l'Apprentissage et de la Formation

Mattia A. FRITZ

DIRECTEUR DE MEMOIRE
Mireille BETRANCOURT

JURY
Gaëlle MOLINARI
Daniel K. SCHNEIDER

GENEVE, JUIN 2015

UNIVERSITE DE GENEVE
FACULTE DE PSYCHOLOGIE ET DES SCIENCES DE L'EDUCATION

Abstract

The Dynamic Emotion Wheel (DEW) is an Emotion Awareness Tool (EAT) developed as part of the EATMINT project, whose objective is to explore the relationship between emotion – defined accordingly to appraisal/componential theories – and computer-mediated collaboration. The DEW is inspired by the Geneva Emotion Wheel (GEW), a self-report instrument that arranges a set of modal emotions in a circle with two appraisal dimensions as axes: Valence and Control. The DEW is an adaptation of the GEW in the perspective of its co-existence with another computer-mediated task, which means that the EAT must occupy only a limited part of the screen and should not distract too much from the primary task. For this reason, the DEW does not show all modal emotions at the same time, but dynamically compute a subset of emotions according to two range-input values representing the same appraisal criteria of the GEW. The subset, thus, dynamically updates according to variations in appraisal criteria. Furthermore, the DEW provides graphical representation of the evolution of registered emotional episodes with respect to both the appraisal dimensions and the related subjective feeling. The DEW aims to enhance real time awareness of both self and other participant's emotional episodes in an ongoing collaborative task. This contribution illustrates the comprehensive development process of the DEW, which includes a usability test conducted in experimental settings where participants were engaged in a simulated collaborative task on a computer screen equipped with an eye-tracking device.

Keywords: Emotional Awareness, Computer-Mediated Collaboration, Affective Computing, Human-Computer Interaction, Emotion, Emotion Self-Report Tool, User-Experience, Test Driven Development, Rapid Application Development, Interaction Design, Usability Test, Eye Tracking.

Acknowledgments

I am grateful to many people for their help and support during the design, development and testing of the Dynamic Emotion Wheel. This project has a particular meaning to me, since it represented a turning point in my academic cursus: when I first heard about the prospect of developing an Emotion Awareness Tool, I was considering other opportunities about my future. This project contributed to clear my mind, renewing and corroborating my interest for research. Therefore, I would like to thank Professor Betrancourt, in the first place, for listening and understanding my doubts, and, secondly, for her acknowledged and thought-provoking support throughout the *mémoire*. I am also grateful to the EATMINT project members, in particular Gaëlle Molinari and Guillaume Chanel, for their open-minded approach and the freedom I was allowed during my internship.

The development and test of an application also requires substantial knowledge about the use of technologies. In this regard, I would like to thank Professor Schneider for his feedbacks about development methodologies, and Sandra Berney for her support with the eye-tracking equipment.

My deepest gratitude goes to my "TETRIS" fellow colleagues for their friendship and mutual support throughout the Master. In particular, I would like to thank Damien Gauthier for sharing his competence in usability and statistical issues.

I am grateful to my family for their love and support – and yes, my father was right: a university degree does change the perspective on things even at my age; to Roberto R. for sparkling my interest in the human mind; and to Matilde Legrand, without whom everything would be harder, less interesting and pleasant.

Contents

1 Introduction 1

1.1 Emotion Awareness Tool in Computer-Mediated Interaction 2

1.1.1 Theoretical Framework 3

1.1.2 Description of the First Version of the Emotion Awareness Tool 7

1.1.3 Requirements Analysis for the Second Version of the Tool 8

1.1.4 Theoretical Framework for the Development of an Emotion Awareness Tool 10

1.2 Emotion and Emotion Self-Report Tools 10

1.2.1 The Component Process Model 11

1.2.2 How to Measure Emotions 16

1.2.3 Emotion Self-Report Tools 16

1.2.4 The Geneva Emotion Wheel 18

1.2.5 Emotion as Interaction in Affective Computing 21

1.3 Awareness and Awareness Tools 22

1.3.1 General Assumptions about Awareness Useful for Designers 22

1.3.2 Displaying and Monitoring Awareness 23

1.3.3 Emotional Awareness in Computer-Mediated Environments 26

1.4 Summary 27

2 Objectives and Development Methodology 28

2.1 Interaction Design 29

2.2 Rapid Application Development 29

2.3 Test Driven Development 29

2.4 Measuring the User-Experience 30

3 Development of the Dynamic Emotion Wheel 31

3.1 Combining Dimensional and Discrete Emotions Approaches 31

3.1.1 The Interaction Problem with Sliders 33

3.1.2 Adding Semantic Meaning to a Button 33

3.1.3 Qualitative and Quantitative Data to Represent an Emotion 34

3.2 The Core of the Dynamic Emotion Wheel 35

3.2.1 The Dynamic Algorithm 35

3.2.2 Generation of a Subset of Adjacent Emotions 38

3.2.3 Unit Test of the Dynamic Algorithm 39

3.3 Description of the DEW's Interface 40

3.3.1 The Layout ... 40
3.3.2 Expected Interaction with the Tool ... 42

3.4 Architecture of the Application ... 46
3.4.1 Underlying Technologies ... 46
3.4.2 The Model-View-Controller Pattern ... 47

3.5 What the DEW Measures ... 48

3.6 Requirements Check ... 49
3.6.1 Theoretical Framework Requirement ... 50
3.6.2 Cognitive Load Requirement ... 50
3.6.3 Inter- and Intra-Personal Comparison Requirement ... 50
3.6.4 Adaptability and Portability Requirement ... 50

3.7 Summary ... 51

4 Testing the Dynamic Emotion Wheel ... 52

4.1 Objectives of the Test ... 52
4.1.1 Technical Objectives ... 52
4.1.2 Theoretical Objectives ... 53
4.1.3 Comprehensive Assessment of the Objectives ... 53

4.2 Method ... 53
4.2.1 Participants and Design ... 53
4.2.2 Experimental Task ... 54
4.2.3 Material ... 56
4.2.4 Procedure ... 59

4.3 Hypotheses ... 60
4.3.1 Test of the Displaying Function ... 61
4.3.2 Test of the Monitoring Function ... 61
4.3.3 Test of the Overall Function ... 62

4.4 Results ... 62
4.4.1 Assessment of the Displaying Function ... 64
4.4.2 Assessment of the Monitoring Function ... 70
4.4.3 Assessment of the Overall Function ... 73

4.5 Discussion ... 75
4.5.1 Did the Dynamic Mechanism Fulfill the Expectations? ... 75
4.5.2 Did Quantitative Data Add Useful Information to the Monitoring Function? ... 78
4.5.3 Usability Issues ... 79
4.5.4 Limits of the Usability Test ... 80

5 Conclusion ... 81

5.1 Limits of the Present Contribution ... 81
5.1.1 The Limits of Usability Perspective ... 81
5.1.2 An Excess of Cognition in Emotion Self-Reporting ... 81
5.1.3 Extent in Scope and Method ... 82

5.2 Future Developments ... 82
5.2.1 Development of a Version for Experimental Settings ... 82
5.2.2 Square the Circle: Modifications to the Core Algorithm of the DEW ... 83
5.2.3 Use of the DEW as a Self-Report Tool for Measuring Emotions ... 84

5.3 General Conclusion ... 84

6 Bibliography ... 86

7 Annexes ... 96

7.1 Survey ... 96

7.2 Outputs ... 100

1 Introduction

Contrary to popular wisdom, according to which emotions should be kept at bay when facing important situations, neuropsychological evidence suggests that emotions play an important role in high level functions like memory, attention, perception, and decision making (Brosch, Scherer, Grandjean, & Sander, 2013). For instance, the lack of emotions negatively affects decision-making in a way that can seriously compromise the quality of everyday life (Bechara, 2004). This evidence is consistent with the idea that emotions have adaptive functions: they allow the organism to better interact with the environment, mobilizing cognitive and physiological resources in order to adjust behavior to events (Frijda & Scherer, 2009). The importance of emotions, though, goes beyond the internal functioning of the organism, since they also play a significant role in social interactions (Van Kleef, 2009). As a consequence of emotions' influence in both individual and social regulation, in the last decades publications related to emotions have increased in many different fields (Scherer & Meuleman, 2013).

Among the fields that have shown particular interest in the mediating role of emotion in the last few years, Computer-Mediated Collaboration (CMC) has many reasons to consider the importance of emotions both at the individual and collective level. From an individual standpoint, collaboration requires the coordination of high-level functions, where emotion regulation is known to play an important role (Richards & Gross, 2000). At the social level, collaborative interactions between actors may lead to the emergence of socio-cognitive tensions (Andriessen, Baker, & Van der Puil, 2011). Negative emotions, elicited by unresolved conflicts, could undermine interpersonal relationships, and thus affect the collaborative outcomes (Barsade, 2002). The collaboration may therefore benefit from the acknowledgement of other participants' emotional state – that is, *emotional awareness* – since participants could adjust interactions accordingly (Eligio, Ainsworth, & Crook, 2012). However, CMC often implies that group members do not have face-to-face contact, and thus do not have access to verbal and para-verbal cues (Lund, Molinari, Séjourné, & Baker, 2007). For instance, emotional information cannot be shared via physical signals such as facial expression and voice (Bänziger, Mortillaro, & Scherer, 2012). Consequently, an alternative way is necessary for participants to (1) express their own emotions, and (2) perceive other members' emotional state. These two type of actions respectively relate to displaying and monitoring, which are the distinctive functions of awareness tools (Schmidt, 2002).

The *Emotion Awareness Tool for Computer-Mediated Interactions* (EATMINT) project (Chanel, Molinari, Cereghetti, Pun, & Bétrancourt, 2013; Molinari, Chanel, Bétrancourt, Pun, & Bozelle, 2013; Molinari, Bozelle, et al., 2013) aims at improving computer-mediated collaboration precisely through the implementation of Emotion Awareness Tools (EAT) allowing participants to share their emotional state. Preliminary experiments have shown that the use of an EAT during a collaborative task influences the perception of the quality of collaboration and group performance, and thus suggested the interest for further empirical research in the domain. In this regard, the EATMINT project members reckoned that the EAT used in pilot experiments showed some limitations, and therefore planned to implement a second version of the tool to be used in future studies. This contribution illustrates the comprehensive development process of an EAT, which started as part of a two-month internship in the project and successively evolved in a Master-thesis due to the complexity and interest of the topic. We named the tool

Dynamic Emotion Wheel[1] (DEW), for it is an adaptation of an existing emotion self-report tool – the Geneva Emotion Wheel (Scherer, Shuman, Fontaine, & Soriano, 2013). Its distinctive features refer to the combination of two major approaches in measuring emotions (i.e. the dimensional and discrete emotions approaches), and the integration of emotional awareness based on the Component Process Model (Scherer, 2009) definition of emotion.

In the first and introductive part of the contribution, we will put requirements for the new version in the perspective of the theoretical framework underlying the EATMINT project. At this stage, existing awareness and emotion self-report tools will be assessed as guidelines and inspiration. In the second part, we will define the objectives of the current contribution and specify the methodologies implicated to reach them. Thirdly, we will illustrate the technical features of the DEW, and depict how major design choices relate to the theoretical framework, the requirements, and human-computer interaction principles. Finally, we will show the result of a usability test conducted in experimental conditions implying a computer-mediated collaborative task. In conclusion, the results of the test will serve as a final assessment of the Dynamic Emotion Wheel and suggest future developments.

1.1 Emotion Awareness Tool in Computer-Mediated Interaction

The EATMINT project aims at studying the role of emotion awareness in Computer-Mediated Collaboration (CMC). It is a multidisciplinary project relating to affective science (Davidson, Scherer, & Goldsmith, 2003) and affective computing (Picard, 2000); cognitive science (Johnson-Laird, 1980); computer science and engineering, in particular Computer Supported Cooperative Work (Grudin, 1994); learning science, in particular Computer Supported Collaborative Learning (Stahl, Koschmann, & Suthers, 2006); human-computer interaction (Carroll, 1997) and interaction design (Rogers, Sharp, & Preece, 2011). The interest of the project is preeminently twofold:

1. Designing Emotion Awareness Tools, which allow either the explicit or the autonomic sharing of emotions among participants in a CMC task;
2. Studying the impact of emotional awareness on collaboration both at intra- and inter-personal levels, with respect to the emergence of strategies that improve collaboration such as monitoring, regulating and reflecting on emotions.

Ever since the EATMINT project began in mid 2011, few researches discussed the subject of emotional awareness during collaboration (e.g. Eligio et al., 2012; Feidakis, Daradoumis, & Caballé, 2011). Hence, the EATMINT project launched an exploratory experiment that studied the impact of emotional awareness with respect to three areas of interest: first, to what extent emotional awareness influences perceived emotions after collaboration and the perceived quality of interaction (Molinari, Chanel, et al., 2013); second, to what extent the perceived interaction during a collaborative task is linked to users' emotional traits, and how an emotional awareness tool can influence this relationship (Molinari, Bozelle, et al., 2013); and third, to what extent physiological and eye-movement coupling may predict situations that may benefit from emotional awareness information during the collaboration process (Chanel et al., 2013)

[1] For a working example see http://tecfaetu.unige.ch/perso/maltt/fritz0/memoire/dynamicemotionwheel/

Empirical results gathered to respond to these questions were obtained with the same experimental settings. Sixty participants, randomly assigned to form 30 same-sex dyads to avoid gender differences in the expression of emotions (e.g. Brody, 2000), executed a computer-mediated collaborative task, which consisted in co-authoring a slogan against bullying at school through the DREW environment (Lund et al., 2007). Participants disposed of audio connection, but not video, and had to jointly work on a concept map (Novak & Cañas, 2008) in order to assess mutual propositions and variations to the slogan. Half of the dyads disposed, in the right side of the screen, of a Graphical User Interface through which they could communicate their emotional state and watch, in the meantime, the emotional state of their partner (**Figure 1.1**). This interface corresponds to the first version of the Emotion Awareness Tool that this contribution aims to implement according to specific requirements. To put these requirements into perspective, we will first provide an overview of the EATMINT project theoretical framework and an accurate description of the first EAT.

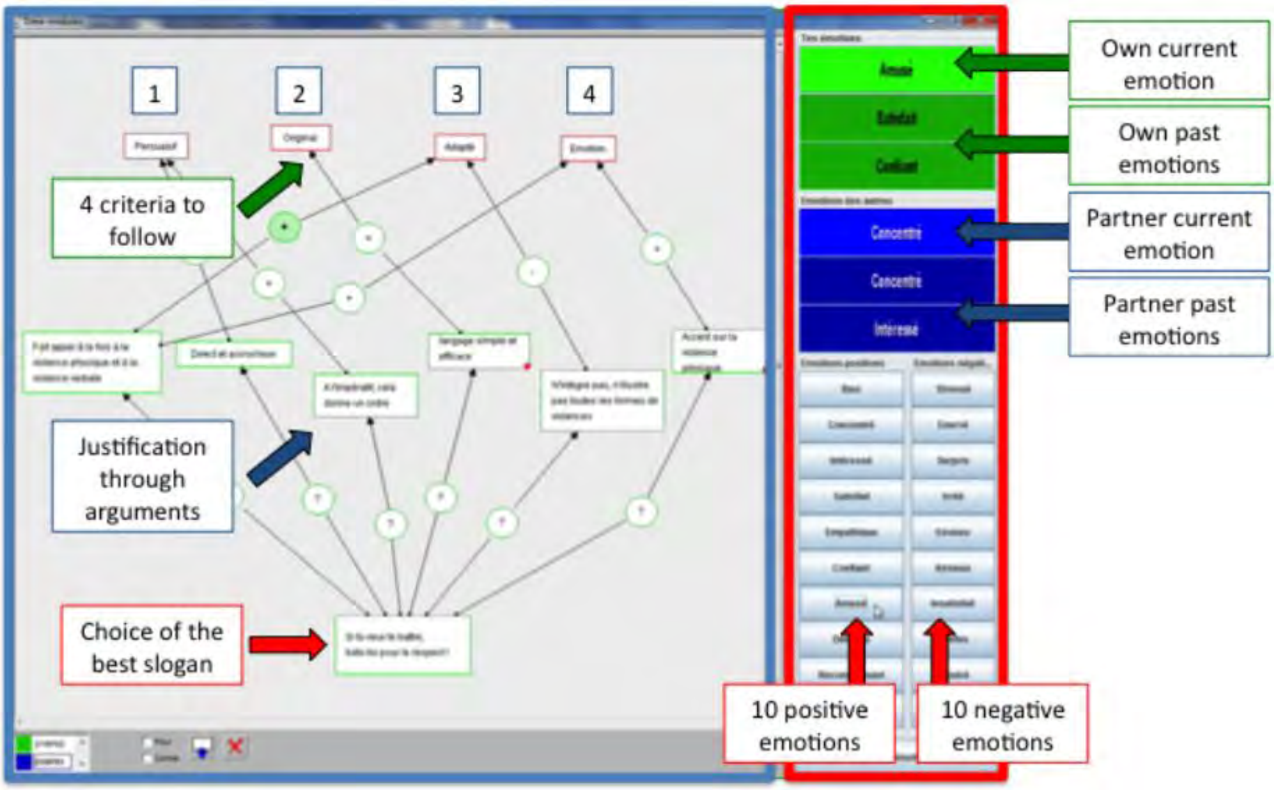


Figure 1.1. The DREW concept map (left), and the first version of the Emotion Awareness Tool used during exploratory EATMINT experiment (Molinari, Chanel, et al., 2013, p. 339 Figure 1)

1.1.1 Theoretical Framework

The EATMINT project studies the interaction of three theoretical constructs: (1) computer-mediated collaboration, (2) awareness, and (3) emotion. Each of these concepts is widely studied in many fields of social sciences, and thus possesses extensive – and sometimes contradicting – theoretical and empirical references (e.g. Dillenbourg, 1999; Frijda & Scherer, 2009; T. Gross, Stary, & Totter, 2005). It is beyond the scope of this contribution to present a comprehensive review of the literature about these concepts for two interrelated reasons. First, this contribution focuses on the design and development of a tool within the ongoing project, and must therefore adhere to its theoretical framework. Second, this theoretical framework has been extensively reviewed in a recent contribution (Cereghetti, 2013) and, to our knowledge, no evidence suggests the need to revise or update it. A brief overview of the project background is nevertheless necessary to put this contribution into perspective.

1.1.1.1 Awareness Tools: Beyond the Limits of Computer-Mediated Collaboration

One of the most challenging issues with respect to computer-mediated activities such as Computer-Supported Cooperative Work (Grudin, 1994), Computer-Supported Collaborative Learning (Dillenbourg, Järvelä, & Fischer, 2009), or even distance and blended learning (Jézégou, 2010) is the lack of a holistic perception of other people taking part in the same activity (Molinari, 2011). This shortcoming is especially problematic in the field of distance learning, where students may tend to feel abandoned on their own, which negatively affects their motivation and perseverance (Jacquinot, 1993). Consequently, shared workspaces accessed through electronic devices integrated tools to counter this phenomenon by displaying information about other people's presence and activities: a process generally referred to as awareness (Dourish & Bellotti, 1992).

Initially, awareness primarily provided spatial-like information such as which user is online and what is he working on (Buder, 2011). In other words, awareness utility was shaped upon the standard of face-to-face communication (Jonassen, Davidson, Collins, Campbell, & Haag, 1995). However, even in the case of seamless audio/video connection, users have access to a limited set of verbal and para-verbal cues available in a co-located situation (Baltes, Dickson, Sherman, Bauer, & LaGanke, 2002; Lund et al., 2007). More recently, thus, awareness tools moved from a limited visual-spatial perspective, to the integration of information and cues that are not physically observable (Buder, 2011). For example, *knowledge awareness* (Ogata & Yano, 2000) refers to the "awareness of a group member regarding the knowledge of collaboration partners" (Engelmann, Dehler, Bodemer, & Buder, 2009, p. 950). Knowing what someone else knows is not an easy inferential process even face-to-face, often leading to under- or over-estimation (Nickerson, 1999, 2001) and therefore affecting the collaboration. In this regard, Sangin and collaborators (2011) provided participants to a remote collaborative learning task with cues about their peer's level of prior knowledge. This information was available through a knowledge awareness tool (KAT) only to half of the participants. Results suggested that the use of the KAT positively influenced learning outcomes allowing participants "to be more accurate in estimating their partner's actual knowledge and by extension, [enhance] their learning gain" (Sangin et al., 2011, p. 1064).

Especially in Computer-Supported Collaborative Learning, thus, awareness tools have progressively replaced – or integrated – spatial information with cognitive and social cues (Buder, 2011). In this way, awareness is not limited to "fill-in" the shortcomings of remote collaboration compared to co-located conditions, but rather enhance collaboration providing elements that participants can use to regulate and improve their interactions.

1.1.1.2 The Cognitive and Social Influence of Emotions

Even though emotion "may be one of the fuzziest concepts in all of the sciences" (Frijda & Scherer, 2009, p. 142), the number of publications in social sciences containing the term "emotion" as a keyword has multiplied by a factor of 1.13 per year between 1990 and 2011 (Scherer & Meuleman, 2013). The interest for emotion, especially in psychology and cognitive sciences, relates, among other things, to the fact that emotion is no more considered as a nuisance to cognitive processing (Bechara, 2004; Damasio, 2006). On the contrary, according to some theories that have gained widespread consensus in the last few years – most notably appraisal (Roseman & Smith, 2001; Schorr, 2001; Siemer, Mauss, & Gross, 2007; Smith & Ellsworth, 1985) and componential (Scherer, 1982, 2001, 2009, 2013) theories – emotion and cognition are interrelated in a functional manner (Leventhal & Scherer, 1987).

With respect to the objectives of the EATMINT project, we may identify at least three significant ways by which emotion and cognition intertwine. First, a cognitive evaluation of a situation seems to be both sufficient and necessary for eliciting an emotion (Siemer et al., 2007). Second, since the evaluation of the situation determines what kind of emotion in particular is to be felt (e.g. joy, anger, sadness, etc.), emotion regulation is possible through the re-appraisal of the situation (Lazarus, 1966). Empirical research suggests that a cognitive re-evaluation of an event requires less mental effort than trying to suppress the consequences of efferent manifestations such as blushing, crying or laughing (J. J. Gross, 2002; Richards & Gross, 2000). Third, if cognition influences emotion, the same happens the other way around: psychological and neuroscientific evidences suggest that emotions impact high-level cognitive functions such as perception, attention, memory and decision-making (Brosch et al., 2013). Emotions, thus, play an important role in self-regulation since they contribute to determine the allocation and orientation of resources in order to adapt to changes in the environment that are considered of major concern for the organism (Scherer, 2005).

The importance of emotion is nevertheless not limited to the internal functioning of a human being: emotions are also social phenomena that influence the way we interact with other people (Parkinson, 1996; Rime, 2009). We can perceive and understand the emotional state of a person through efferent manifestations such as facial expressions, body posture and voice (Bänziger et al., 2012; Ekman, 1992; Johnstone & Scherer, 2000) and use this information to regulate our behavior. In this regard, the *Emotion As Social Information* (EASI) model (Van Kleef, 2009, 2010) identifies “two processes through which observers’ behavior may be influenced: inferential processes and affective reactions” (Van Kleef, 2009, p. 184). Inferential processes assume that an emotional episode has an intelligible cause (Lin, Keysar, & Epley, 2010). The observer may therefore infer that his very behavior could be the cause of the emotion elicitation and, thus, modify or maintain the behavior according to the outcomes. Affective reactions, on the other hand, imply that someone’s manifested emotion can itself represent an emotion-eliciting event for the observer. This may occur through spontaneous contagion (e.g., being sad if a beloved person is also sad), or through modification of impressions and interpersonal linking (e.g., a person which often reacts in anger may be despised for her irritability). Both processes depend on two modulating factors: information processing and social-relational factors. On the one hand, inferential processes and affective reactions depend upon the ability and motivation of the observer to take emotional cues into account – that is, reflecting upon their causes and influences. On the other hand, emotional expression and perception are also regulated by social norms that define, for instance, if an emotion is appropriate for a given social situation (Ekman, Sorenson, & Friesen, 1969).

To sum up, emotion is a pivotal phenomenon that influences both intrapersonal and interpersonal mechanisms. At the intrapersonal level, it plays a significant role in what a person perceives, decides, and pays attention to. In the meantime, at the interpersonal level, it serves as a window for other to infer, among other things, the very same internal processes: knowing that a person is angry or happy changes the perspective about what she may perceive, decide or pay attention to. Furthermore, emotions also serve a social function, since they influence the way people interact.

1.1.1.3 Why Emotional Awareness Can Improve Computer-Mediated Collaboration

In a collaborative task, participant needs to build and maintain some degree of mutual understanding (Roschelle & Teasley, 1995) – a process identified as *grounding* (Baker, Hansen, Joiner, & Traum, 1999) – which implies that each participant builds and constantly updates a mental representation of his own and his colleagues' understanding of the topic at hand (Molinari, Sangin, Dillenbourg, & Nüssli, 2009). According to Andriessen and collaborators (2011) this process implies two interrelated adjustments: one concerning the *cognitive aspects* (e.g. agreement and disagreement on procedures or expected results), and the other concerning *socio-relational aspects* (e.g. build interpersonal relationships that enhance motivation, persistence, authenticity, etc.). Considering that emotions play an important role both at cognitive and socio-relational levels, they may contribute in the regulation of socio-cognitive tensions arising during collaboration.

In this regard, it is worth noting that efficient emotion regulation does not necessarily correspond to the suppression of all negative emotions. For instance, the expression of anger – which is often considered inacceptable in social relations – can in some circumstances enhance creativity in problem solving and decision making (Van Kleef, Anastasopoulou, & Nijstad, 2010), from which the collaboration may take advantage. In other words, a collaborative task will not exclusively benefit from positive emotions and suffers from negative ones. For example, if an excess of negative emotions may certainly undermine the inter-personal relationship and compromise the collaboration (Barsade, 2002), an excess of positive emotions may lead participants to tolerate behaviors such as social loafing (Karau & Williams, 1993).

Emotional awareness utility is thus unrelated – or at least not exclusively related – to the elicitation of positive emotions. On the contrary, EATs must provide authentic, meaningful and trustworthy emotional information, for which a double process is necessary. On the one hand, the EAT must collect information that accurately represent what users "really feel". On the other hand, this information must be conveyed and presented in a way that allows other participants to understand, not only what a user feels, but also the general context, possible causes and consequences of emotions – that is, stimulate metacognition about affective experiences (Flavell, 1979). If these premises are maintained, Emotion Awareness Tools provide participants with valuable information. More specifically, emotional awareness may:

1. Contribute to build a holistic representation of collaborators, especially in the absence of audio-video connection.
2. Facilitate the expression of emotions, and decrease the risk of dysfunctional suppression strategies. For instance, without a tool that directly asks participants to express emotions, they could be reluctant to voluntary share them.
3. Stimulate emotion understanding in terms of eliciting causes and consequences, both at intra- and inter-personal level (Eligio et al., 2012)
4. Provide feedbacks – even alerts in some cases – useful to regulate emotions between members of the same group (Järvenoja & Järvelä, 2009), instead of ignoring them.

5. Assess a situation in real-time, limiting the risks of misunderstandings and second thoughts about the collaboration.

1.1.2 Description of the First Version of the Emotion Awareness Tool

In EATMINT's experimental settings, emotional awareness was conveyed through a customized Graphical User Interface that allowed participants to share their emotions during the collaborative task (see **Figure 1.1**, p. 3). The Emotion Awareness Tool, located at the right edge, occupied the whole height and about a fifth of the width of the screen. The tool vertically divided in two areas of almost the same size.

In the lower part of the tool, users could press one of the 20 buttons with a label representing an emotion adjective (e.g. happy, amused, anxious, etc.). The list of emotions were defined mixing a previous study in the field (D'Mello & Graesser, 2012) and 2 pre-experiments that aimed to identify the most frequent and intense emotions felt during a situation of collaboration, real or imagined. The 20 emotions were equally divided in positive and negative labelled groups, but did not follow any particular order within each list (see **Table 1.1**). At the bottom-end of the screen, a "no-emotion" button spanning across the two columns was also available.

In the upper side of the tool, participants could see the last three emotions they had expressed and the last three emotions manifested by the partner. Their own emotions appeared in the upper side as a vertical list of text-centered cases with a green background. The very last emotion had a lighter background and resulted top of the list. This case also served as a text input where the participant himself could type an emotion not in the list. Their partner's emotions appeared just below, in exactly the same way, except for a blue background and the first light-blue case not editable.

Each participant could express the emotion he felt in one of three ways: (1) clicking one of the 20 emotion labelled buttons; (2) typing an emotion not in the list directly in the first light-green case; or (3) by clicking the "no-emotion" button at the bottom-end of the screen. The expressed emotion would then appear in the upper light-green case of his screen, and, respectively, in the upper light-blue case of his partner's screen. A pop-up window reminded participants to express their emotion after five minutes from the last emotion submitted if they had not provided another emotion in the meantime.

Table 1.1. List of the 20 emotion adjectives used as buttons in the EAT. Our translation from the original French word in italic – partially based on a document by the Geneva Emotion Research Group (1988) – is limited to the contribution's purpose. The order in the table reflects the order of the buttons in the interface.

#	Positive emotions	Negative emotions
1	Delighted – *Ravi*	Stressed – *Stressé*
2	Attentive – *Concentré*	Annoyed – *Enervé*
3	Interested – *Intéréssé*	Surprised – *Surpris*
4	Satisfied – *Satisfait*	Irritated – *Irrité*
5	Empathic – *Empathique*	Envious – *Envieux*

#	Positive emotions	Negative emotions
6	Confident – *Confiant*	Anxious – *Anxieux*
7	Amused – *Amusé*	Disappointed – *Insatisfait*
8	Relaxed – *Détendu*	Confused – *Confus*
9	Grateful – *Reconnaissant*	Frustrated – *Frustré*
10	Relieved – *Soulagé*	Bored – *Lassé*

1.1.3 Requirements Analysis for the Second Version of the Tool

Even though the experimental setting of the EATMINT contributed to significant results in three published papers (Chanel et al., 2013; Molinari, Chanel, et al., 2013; Molinari, Bozelle, et al., 2013), the EATMINT project members assessed that the EAT could be implemented with respect both to theoretical and technical aspects. They agreed to develop a second version of the tool for experimental settings where participants voluntary share their emotions during an ongoing computer-mediated collaborative task. A requirement analysis has thus been conducted to identify what the new version of the EAT "should do or how it should perform" (Rogers et al., 2011, p. 355). Requirement statements were not supposed to be very specific at this stage of the process. On the contrary, the EATMINT project members wished to consider alternatives and were thus open even to significant modifications of the tool. Nevertheless, a set of directives and suggestions emerged from the analysis. The following four points put the main requirements and suggestions into a technical and/or theoretical perspective.

1.1.3.1 The Control-Value Theory of Achievement Emotions

Research in social sciences often suffers from the fact that a different conceptual meaning is given to the term "emotion" and, consequently, contributions are difficult to compare and integrate (Scherer, 2005). For this reason, the EATMINT project members planned to strengthen the link between the Emotion Awareness Tool and an underlying theoretical framework. Since collaboration is generally bond to the perspective of obtaining a result that is satisfactory for the people involved, *achievement emotions theory* (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011; Pekrun, 2006) seemed a suitable reference. Achievement emotions refer to "emotions tied directly to achievement activities or achievement outcomes" (Pekrun, 2006, p. 317), and the associated theory provides a comprehensive framework to analyze emotions experienced in achievement and academic settings.

This framework is consistent with the EATMINT perspective preeminently for three reasons. First, the theory is in line with the concept of emotion as a multi-component process driven by appraisal, and thus shares the same theoretical framework of appraisal/componential theories adopted by the EATMINT project. Second, the theory relates to the field of education, in which the use of computer-mediated collaboration – especially Computer-Supported Collaborative Learning – is widely adopted (Stahl et al., 2006). Third, the theory posits that achievement emotions result from the evaluation of a situation according to two criteria: control and value.

> The control-value theory […] posits that two groups of appraisals are of specific relevance for achievement emotions: (1) subjective control over achievement activities and their outcomes (e.g., expectations that persistence at studying can be enacted, and that it will lead to success); and (2) the subjective values of these activities and outcomes (e.g., the perceived importance of success) (Pekrun, 2006, p. 317).

In this regard, one of the suggestion for modifying the EAT concerned the possibility to replace the set of 20 emotion-labelled buttons with dimensional measurement of emotion-related appraisal criteria. In other words, participants would not click on a word that best represents their emotional state, but rather assess it on a continuous scale, such as if they were more or less in control of the activity, or if they gave more or less importance to it.

1.1.3.2 Cognitive Load: Limiting the Number of Emotions on Screen

Another reason to consider the switch to a dimensional measure related to the number of emotions – and thus choices – present on the interface. Due to the limited capacity of working memory (Baddeley, 2003), the number of coexistent choices available in the meantime on the interface was considered as a potential threat for cognitive overload (Gevins et al., 1998; Sweller, 1988, 1994). Participants, in fact, had to perform a seek-and-evaluate task (Rouet & Tricot, 1995) every time they wanted to express an emotion. Even though the 20 emotions appeared in positive and negative groups, within the list they did not follow any particular order. This made finding the right button even harder, especially since participants were meanwhile engaged in the collaboration, creating thus a situation of dual-task, which is known to represent a source of impairment for cognitive efficiency (Pashler, 1994).

1.1.3.3 Intra- and Inter-Personal Comparison of the Emotional State Evolution

The presence of 20 buttons on the interface also reduced the space at disposal for showing the emotions shared through the EAT. Only the last three emotions for each participant appeared on the interface without any further information such as the time when they occurred or in which order they appeared considering the dyad as a whole. In other words, it was possible to establish in which order each participant had expressed his last emotions, but not if the very last emotion of the dyad belonged to the first or second participant. It was also impossible to tell, from the interface alone, if the order of the emotion followed a one-each turn, or a participant expressed three emotions in a row. Consequently, for the new version of the tool, the EATMINT project members wanted to enhance the possibility for participants to (1) dispose of the evolution of their own emotional state during time, and (2) compare this evolution with their partner's one.

1.1.3.4 Adaptation of the Tool to Different Settings and Tasks

A last requirement, of more general scope, concerned the flexibility of the tool with respect to both theoretical and technical features. On the one hand, the new version of the tool should provide experiments with maximum choice about the underlying theoretical framework, for example with respect to the number and labels of the emotions. On the technical level, on the other hand, the interface of the tool could also be considered as a potential factor in the experimental settings. For example, a group of participants could dispose of an interface that particularly highlights inter-personal comparison, whereas the other group an interface that provides more insight on the intra-personal evolution over time. Finally, the tool should also provide a way to set up emotional awareness in experimental settings without specific technical knowledge, and dispose in the meantime of collected data in a practical form for exportation and subsequent statistical treatment.

1.1.4 Theoretical Framework for the Development of an Emotion Awareness Tool

The requirement analysis suggests a procedural approach to the development of an Emotion Awareness Tool. First, it is necessary to define what an emotion is, what an emotion means to the person herself, and how the person can share her emotional state with others. This first step relates to the field of affective sciences, and more specifically to the psychology of emotion. Consequently, the next section of this introduction will provide a framework to define an emotion, and illustrate how emotions are measured, in particular with the use of emotion self-report tools used in computer-mediated environments.

Once determined the piece of information that users will share during the collaboration, it is necessary to establish how to convey this information, for users to perceive, process, and integrate it during the collaborative task. This second step refers to the concept of awareness, and more specifically awareness in computer-mediated environments, since the only specific condition about the collaboration is that it will take place using a computer. We will therefore provide some general assumptions about the design of awareness tools.

1.2 Emotion and Emotion Self-Report Tools

What is an emotion? A common reaction to this question consists in an immediate perception of a self-evident answer and a subsequent difficulty to put it into words. It is unlikely to obtain two times the same answer to this question, even from researchers in the growing interdisciplinary field of affective sciences (Davidson et al., 2003). This assumption may seem unrealistic considering the widespread use of emotion-related terms such as anger, fear or joy in everyday speech (Douglas-Cowie, Campbell, Cowie, & Roach, 2003). These terms often elicit to the mind past experiences or potential situations when these emotions occur, and bodily symptoms or expressions that accompany them. Despite this intuitive representation of emotion as a familiar experience, though, there is not yet a consensual definition of emotions as unambiguous objects of research (Scherer, 2005). The lack of a consensual definition is due to the coexistence of different underlying theoretical frameworks that conceptualize emotion in different ways. Most of the theories originated over the years belong to one of three families, which “differ with respect to the definition of the basic units of emotional experience, the underlying mechanism that is assumed, and the role of the verbal emotion labels In ordinary language in relation to the psychobiological emotion process” (Scherer, 2013, p. 10).

1. *Basic (or discrete) emotion theories* (Ekman et al., 1969; Ekman, 1992) identify a direct link between specific types of events and a limited number of discrete emotions elicited by those events. According to these theories, thus, each discrete emotion disposes of an automated affect program that triggers a specific physiological and expressive response depending on the event. This response is universal and can be identified by a specific word in the language.
2. *Dimensional/constructivist emotion theories* (Osgood, 1952; Russell, 1980) posit that affect phenomena in general organize in low-dimensional spaces, represented by one or more factors. For

instance, Russell (1980) uses a two-dimensional circumflex composed by the arousal and valence factors. Every affect phenomenon can be identified by its arousal value (i.e., if it is more or less activating) and valence value (i.e., if it is more or less pleasant). The person then names the corresponding affect according to cultural and situational factors.

3. *Appraisal emotion theories* (Schorr, 2001; Smith & Ellsworth, 1985) postulate that it is the evaluation a person makes of an event that determines what kind of emotional response is triggered. Since the evaluation is subjective and depends upon the combination of some criteria, a potentially infinite number of emotions exist. However, since human beings often face similar situations during their lives, they regrouped the most frequent emotional responses under similar labelling.

A recent comprehensive and multicultural study analyzed emotion with the GRID instrument (Fontaine, Scherer, & Soriano, 2013), in which 24 emotion terms are displayed as rows of a table, and 142 emotion features – pertaining to all the three emotion theory families – as columns. The overall results of the GRID analysis suggest that, while the three families "may disagree as to the mechanism underlying the emotion process, they are complementary when it comes to the meaning of emotion terms" (Fontaine & Scherer, 2013, p. 125). In this regard, appraisal theories present elements that belong to both families, even if not precisely in the same way. On the one hand, appraisal theories consider that emotion differentiation results from appraisal criteria, which can be considered similar to low-dimensional factors. On the other hand, appraisal theories posit that emotions, even if not limited to a strict number, are regrouped in modal emotion families. Since the EATMINT project adopts the appraisal theories of emotion approach, the next section will present a comprehensive framework providing a thorough description of what an emotion is.

1.2.1 The Component Process Model

The Component Process Model (CPM) (Scherer, 1982, 1993, 2001, 2005, 2009) implements appraisal emotion theories' perspective into a comprehensive theoretical framework "based on the idea that during evolution, emotion replaced instincts in order to allow for more flexible response to events in a complex environment" (Scherer, 2009, p. 93). According to this adaptability perspective, emotion represents a latency time during which the organism mobilizes cognitive and physical resources in order to optimize the subsequent response to an event. This mobilization is anticipatory of a change in behavior that is likely to take place as a result of the modified internal and/or external situation of the organism. More specifically, the CPM suggests that the mobilization consists in the coordination of five components (depicted in **Table 1.2**), each of them performing a particular function through specific organismic subsystems:

1. The *cognitive or appraisal component* is responsible for the ongoing evaluation of the environment in search of elements that may represent a particular concern for the organism.
2. The *neurophysiological component* relates to alterations in the internal body functions like respiration or heart rate.
3. The *motivational component* facilitates the preparation of the organism to the possibility of an abrupt change in behavior.
4. The *motor expression component* manifests, especially through facial expression, the emotional episode the organism is going through.

5. The *subjective feeling component* gathers information from the other components, for the organism to be aware of the undergoing phenomena.

Table 1.2. Relationships between organismic subsystems and the functions and components of emotion, adapted from Scherer (2005, p. 698)

#	Emotion component	Organismic subsystem and major substrata	Emotion function
1	Cognitive component (appraisal)	Information processing (CNS)	Evaluation of objects and events
2	Neurophysiological component (bodily symptoms)	Support (CNS, NES, ANS)	System regulation
3	Motivational component (action tendencies)	Executive (CNS)	Preparation and direction action
4	Motor expression component (facial and vocal expression)	Action (SNS)	Communication of reaction and behavioral intention
5	Subjective feeling component (emotional experience)	Monitor (CNS)	Monitoring of internal state and organism-environment interaction

Note: CNS = central nervous system; NEW = neuro-endocrine system; ANS = autonomic nervous system; SNS = somatic nervous system.

The five components intertwine so that a variation in one of them may lead to modifications in the others. The resulting theoretical concept of emotion is thus defined as "*an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism*" (Scherer, 2005, p. 697). The synchronization of the subsystems starts with an evaluation of the situation, which subsequently influences the other components in a chain reaction:

> Briefly put, the CPM suggests that the event and its consequences are appraised with a set of criteria on multiple levels of processing, producing a motivational effect or action tendency that often changes or at least modifies the status quo. Specifically, the appraisal results and the concomitant motivational changes will produce efferent effects in the autonomic nervous system (in the form of somatovisceral changes), and motor expressions are centrally represented and constantly fused in a multimodal integration area (with continuous updating as events and appraisals change). Parts of this central integrated representation may then become conscious and subject to assignment to fuzzy emotion categories, as well as being labeled with emotion words, expressions, or metaphors" (Scherer, 2013, p. 13).

The CPM conceptualization of emotion is particularly adapted to empirical research because it posits that emotion possesses specific features, which help to discriminate it from other affective phenomena such as feelings (which, according to CPM, is just a component of the emotion), preferences, attitudes, or moods. According to Scherer (2005), the multi-component nature of emotion results in seven distinctive features. (a) *Event focus*: emotions are elicited by an internal or external event. (b) *Appraisal driven*: the event must be evaluated as relevant for the organism. (c) *Response synchronization*: in order to maximize adaptability, the mobilization of cognitive and physical resources must follow a synchronized pattern. (d) *Rapidity of change*: the minimization of latency time facilitates the adaptation to sudden modifications. (e) *Behavioral impact*: since the appraised event represents a major concern to the organism, it is likely that the organism must modify its behavior to adapt to the new circumstances. (f) *Intensity*: to capture full access to consciousness, the emotional episode must outrun co-occurring stimulations. (g) *Duration*: in order to enhance flexibility and limit the effort of resource mobilization, emotion must be relatively short in time.

Since the function of the EAT is to voluntary express emotions through a Graphical User Interface, we can illustrate the CPM conceptualization of emotion in a hypothetical input-output communication workflow that consists in 3 phases:

1. *Elicitation*, including the appraisal component;
2. *Patterning*, including motor expression, autonomic physiology and action tendencies components;
3. *Recognition*, including the subjective feeling component.

If we posit that an emotion is elicited by an external event that is part of the perceptive environment of the person, and that the person wishes to communicate the resulting emotional episode to someone else, then we may consider that the appraisal processes and the subjective feeling are gateways between the inter- and intra-individual levels (see **Figure 1.2**). The motor expression could also be considered as an output to the inter-individual level, since efferent manifestations of an emotion such as facial expression, voice and posture communicate the emotional state of a person. Nevertheless, this information is missing in a computer-mediated situation where audio-video connection is not provided. Consequently, on the inter-individual level, the appraisal and subjective feeling components are the most interesting ones.

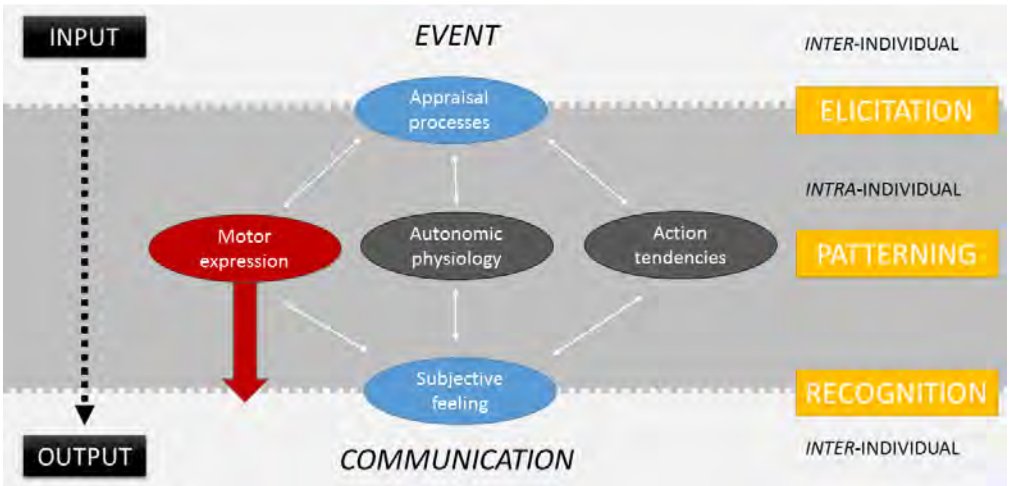


Figure 1.2. Input-Output communication workflow of an emotion with respect to the CPM five components. Adapted from Scherer (2010, p. 50 Figure 2.1.1).

1.2.1.1 Appraisal Processes: How Emotions Are Elicited

The core principle of appraisal theories consists in the assumption that "it is the appraisal of a situation, not the situation per se, that determines the quality and the intensity of an emotional response" (Siemer et al., 2007, p. 592). This hypothesis explains (1) why two people's emotional reaction to the same event may diverge, and (2) why the very same person experiences different emotions facing the same event over time. Imagine you just realized that the presentation file you are supposed to use in tomorrow's meeting has vanished from your hard disk. Possible emotional reactions to this event may include (a) anger, since you spent a lot of time creating the presentation and your efforts turn up to be useless; (b) shame, imagining yourself in front of the audience without anything to show; (c) relief, when you realize that you sent a copy of the presentation to a colleague. The three emotional reactions are not mutually exclusive. On the contrary, you may feel them in a row depending on the criterion you focus on: (a) the useless effort in case of anger, (b) the mocking image of yourself in front of others in the case of shame, and (c) a solution to avoid the catastrophe in the case of relief. With respect to the Component Process Model illustrated above, each evaluation will lead to a particular emotional pattern – that is, synchronization of subsystems – eventually resulting in the conscious labelling of the emotion as anger, shame or relief. The main interest of appraisal is thus to determine if specific patterns of evaluation lead to specific discrete emotional responses (Roseman & Smith, 2001; Scherer, 1993; Smith & Ellsworth, 1985). In other words, do anger and sadness – or any other emotion – present (1) a distinct appraisal pattern with respect to each other; and (2) a consistent appraisal pattern in every elicitation?

As an integrated corollary to the Component Process Model, Scherer (2001) posits that the role of the cognitive component is an ongoing evaluation of event related to specific and chronologically ordered criteria: the Stimulus Evaluation Checks (SECs) – this idea, though, is not shared among all appraisal theories, which rather prefer more flexible combination of criteria (Roseman & Smith, 2001). The SECs represent "the minimal set of dimensions or criteria that are considered necessary to account for the differentiation of the major families of emotional states" (Scherer, 2001, p. 94). Scherer (2001) identifies four appraisal objectives:

1. *Relevance*: How relevant is this event for me? Does it directly affect me or my social reference group? Corresponding checks include *novelty*, *intrinsic pleasantness*, and *goal relevance*.
2. *Implications*: What are the implications or consequences of this event and how do these affect my well-being and my immediate or long-term goals? Corresponding checks include *causal attribution*, *outcome probability, discrepancy from expectations*, *goal/need conduciveness*, and *urgency*.
3. *Coping potential*: How well can I cope with or adjust to these consequences? Corresponding checks include *control*, *power*, and *adjustment*.
4. *Normative significance*: What is the significance of this event with respect to my self-concept and to social norms and values? Corresponding checks include *internal* and *external standards*.

1.2.1.2 Subjective Feeling: What Emotions Mean to Oneself and the Others

The subjective feeling (feeling hereafter) may be considered the most important component with respect to what an emotion *means* for the person herself – for this reason, it is sometimes confused with the whole emotional episode (Ellsworth, 1994). In fact, the feeling is "a holistic cognitive representation that integrates the temporarily coordinated changes of the other components […], allowing the individual to reach awareness of his/her state and label it – stating that he/she 'has' or 'feels' a particular emotion" (Scherer, Shuman, et al., 2013, p. 281).

Grandjean, Sander & Scherer (2008, p. 487) illustrate this process with a Venn diagram where three different aspects of feeling are represented by circles (**Figure 1.3**):

A. The first circle represents the holistic cognitive representation of an emotional episode as *raw* psychophysiological data, and is therefore unconscious.
B. The second circle refers to the part of this information of which the person becomes aware, and it is thus the conscious representation of what a feeling "really is".
C. The third circle includes all the verbal and communication elements that a person possesses to express the consciousness of the emotional episode.

The three circles, though, only partially overlap, which – particularly in the cases of circles B and C – "draws attention to the fact that the use of linguistic labels or expressions to describe the conscious part of feeling will not cover all of what is conscious, partially because of the absence of appropriate verbal concepts" (Grandjean et al., 2008, p. 487).

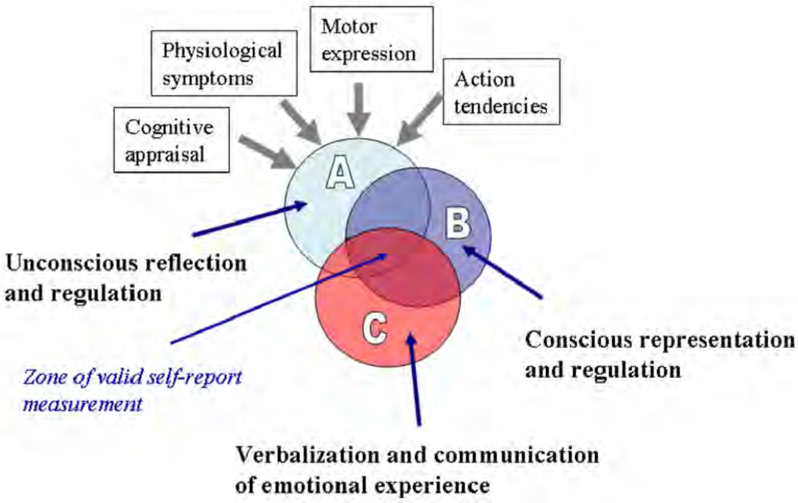


Figure 1.3. Venn diagram of the three hypothetical types of central representation of the component process (Grandjean et al., 2008, p. 487 Fig. 3)

Despite this shortcoming, verbalization remains the most accessible way to express an emotion to someone else. It would be impossible – or at least highly impractical – to communicate all the amount of psychophysiological data that accounts for the multi-componential emotional episode just to tell how we feel. Furthermore, an open-online study conducted with close to 6'000 participants confirmed that "it is possible to predict the emotion words that people will use to describe their emotional experiences on the basis of self-reported appraisal" (Scherer & Meuleman, 2013, p. 2). This roughly means that an emotion label also conveys the elicitation profile that accounts for emotion differentiation. In the meantime, though, the study suggested that participants often require more than one word to describe their emotional state, corroborating the idea that emotion labels can only partially account for the complexity of the conscious emotional experience.

1.2.2 How to Measure Emotions

The complexity of the emotional episode also affects its study, “sometimes generating the implicit assumption that emotions cannot be scientifically studied in the sense of lawful causation” (Scherer & Meuleman, 2013, p. 1). The measurement of emotional episode presents, in fact, many technical and situational challenges. On the technical level, coherently with the CPM definition of emotion, the measure of emotion should gather data from all the five interrelated components, which requires specific equipment (e.g. Valstar, Mehu, Pantic, & Scherer, 2012). On the situational level, in order to measure an emotion, that emotion must first be elicited – and this can be difficult in “unnatural” experimental settings (Scherer & Ceschi, 1997). The quickest and less intrusive way to measure an emotion remains by questioning participants about how they feel – that is, asking their subjective feeling by introspection. This is generally done in one of two ways (Scherer, 2005): (1) *open choices*, where participants can freely express how they feel with whatever figure of speech comes to their minds; and (2) *forced choices*, where experimenters provide a pre-determined set of emotion representations among which the participants can choose the most accurate. Both methods have advantages and disadvantages. On the one hand, the open choice method overcomes the limitation of single labels that hardly account for the complexity of emotional episode (Grandjean et al., 2008). On the other hand, though, this method produces data that is difficult to compare in statistical tests, and also suffers from the differences in participants’ “emotional vocabulary” (Shaver, Schwartz, Kirson, & O’Connor, 1987).

Forced choices, on the contrary, allow for standardization within the same experiment, but provide participants with a limited set of options that may impose an unspontaneous choice. Furthermore, since every experiment often adapts forced choices to the specific needs of the contribution, data is difficult to compare across studies. Two types of forced choices are used in self-reports (Scherer, 2005): (1) the *discrete emotions approach*, and (2) *the dimensional approach*, which are clearly linked with the respective aforementioned theories. The first approach is the one adopted by the first version of the EAT, which provides participant with 20-emotion buttons (even if the tool also allows for an open choice approach, since participants can type an emotion not in the list). The second approach often implies participants to rate how they feel on a continuous scale characterized by opposite poles: pleasant vs. unpleasant, calm vs. excited, etc. With respect to gathered data, thus, the discrete emotions approach provides qualitative observations, whereas the dimensional approach produces quantitative data.

1.2.3 Emotion Self-Report Tools

Contributions that directly aim at studying emotions usually measure them with questionnaires (e.g. Doherty, 1997; Pekrun et al., 2011; Scherer, 2001), which are often customized for the objectives of the research. Over the years, though, some self-report tools have been created for computers (e.g. Caicedo & van Beuzekom, 2006) and even mobile devices (Isomursu, Tähti, Väinämö, & Kuutti, 2007). To our knowledge, prior to the first version of the EATMINT’s tool, no other emotion self-report tool has been developed with the clear intention to convey emotional awareness in different experimental settings. In this section, we will briefly present some of the self-report tools intended for a generalized use whenever affect phenomena are taken into account.

1.2.3.1 Self-Assessment Manikin

The Self-Assessment Manikin (SAM, Bradley & Lang, 1994) is a picture-oriented self-report tool that exists both in paper-pencil and computer software versions. The SAM adopts a dimensional approach to emotions, identifying an emotional episode with respect to three affective dimensions: (1) pleasure, (2) arousal, and (3) dominance (Osgood, 1952).

> [The paper-pencil version of] SAM ranges from a smiling, happy figure to a frowning, unhappy figure when representing the pleasure dimension, and ranges from an excited, wide-eyed figure to a relaxed, sleepy figure for the arousal dimension. The dominance dimension represents changes in control with changes in the size of SAM: a large figure indicates maximum control in the situation (Bradley & Lang, 1994, pp. 50–51).

Users manifest their emotional state by placing a cross inside a figure on each row (**Figure 1.4**), or between two figures if they consider that their true feeling is somewhere in between. The computer version, on the other hand, dynamically updated the figure according to a 20-point scale for each dimension.

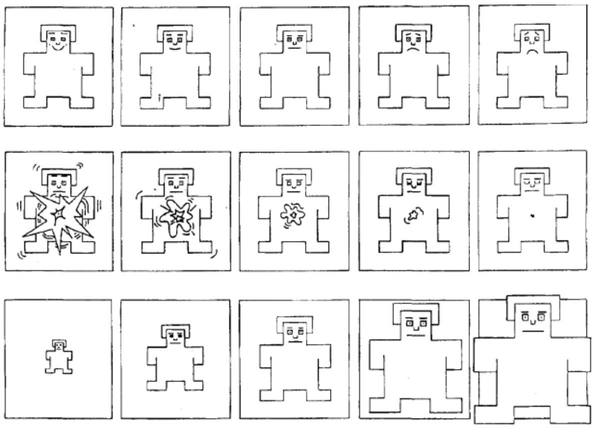


Figure 1.4. The Self-Assessment Manikin (SAM) used to rate the affective dimensions of valence (top panel), arousal (middle panel), and dominance (bottom panel) (Bradley & Lang, 1994, p. 51 Figure 1)

1.2.3.2 AffectButton

The AffectButton [*sic*] (Broekens & Brinkman, 2013) adopts the same dimensional approach and the very same dimensions of the SAM, but integrates them in a single graphical representation (**Figure 1.5**). The tool uses in fact a face that changes the expression according to the direction and the distance of the mouse from the center of the image. The horizontal movement determines the pleasure dimensions, whereas the vertical movement the dominance dimension. The arousal dimension is calculated according to the distance of the mouse from the central point of the image. Each dimensions ranges from -1 to 1; at the central point of the image, each dimension equals 0. Moving the mouse on the left, pleasure decreases to negative values; on the contrary, it increases to positive values moving on the right. Dominance becomes positive moving up, and negative moving down. Arousal is negative if the mouse remains at less than half the distance from the center point to the edges of the image, and becomes positive after the halfway has passed.

Figure 1.5. The AffectButton figure changes according to mouse-over movements with respect to the center of the image (Broekens & Brinkman, 2013, p. 11 Figure 2).

1.2.3.3 Product Emotion Measurement

The Product Emotion Measurement (PrEmo) instrument (Desmet, 2005) adopts a discrete emotions approach to self-report. The tool also maintains, though, a graphical representation of the emotional episode (**Figure 1.6**).

> PrEmo is a non-verbal self-report instrument that measures 14 emotions that are often elicited by product design. Of these 14 emotions, seven are pleasant (i.e. desire, pleasant surprise, inspiration, amusement, admiration, satisfaction, fascination), and seven are unpleasant (i.e. indignation, contempt, disgust, unpleasant surprise, dissatisfaction, disappointment, and boredom). Instead of relying on the use of words, respondents can report their emotions with the use of expressive cartoon animations (Desmet, 2005, p. 5).

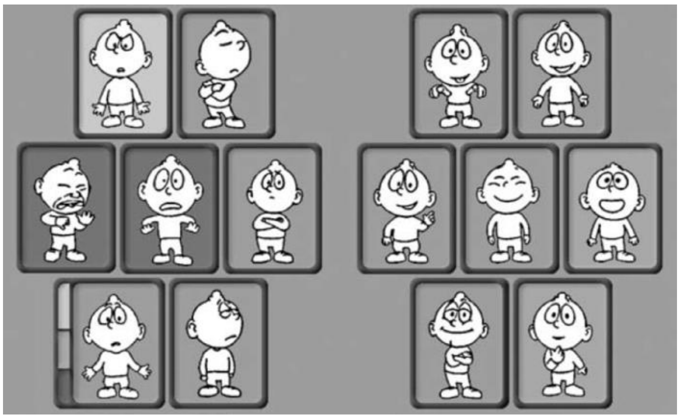


Figure 1.6. The Product Emotion Measurement instrument uses 14 discrete emotions represented by 14 different animations (Desmet, 2005, p. 5 Figure 1).

1.2.4 The Geneva Emotion Wheel

The Geneva Emotion Wheel[2] (GEW) (Scherer, Shuman, et al., 2013; Scherer, 2005) deserves a more comprehensive assessment since it combines the dimensional and discrete emotions approaches, in order to overcome the respective shortcomings. The GEW emerged on the premise that "it might be worth investigating in the development of an instrument capable of combining the advantages of the precise differentiation provided by natural language labels with the simple organizational structure afforded by a two-dimensional space" (Scherer, Shuman, et al., 2013, p. 283). The GEW integrates the two approaches by

---

[2] Official material available at http://www.affective-sciences.org/gew

the disposition of discrete emotions on a circumflex, and by determining the position of each emotion according to a combination of two underlying dimensions. The main features of the tool are the following:

1. Emotions are disposed with respect to the Valence and Control/Power (Control hereafter) dimensions. Valence may be defined as if "the situation is experienced as (un)pleasant and enjoyable (disagreeable) and/or is likely to have positive and desired (negative and undesired) consequences for the person", whereas Control determines if "the person believes that he/she can (cannot) influence the situation to maintain or improve it (if desired)" (Scherer, Shuman, et al., 2013, p. 289);
2. Discrete emotions correspond to natural language words;
3. The position of each emotion is determined by the underlying dimensions;
4. Users can express the intensity of an emotion;
5. The graphical user interface provides users with affordances, allowing them to understand the architecture of the tool and its use.

1.2.4.1 The Valence and Control Dimensions as Appraisal Checks

The choice of the two underlying dimensions was made on the appraisal emotion theories premise that "[i]f emotions are indeed elicited and differentiated by appraisal patterns, the structure of the emotional space should be largely determined by the major appraisal dimensions" (Scherer, Shuman, et al., 2013, p. 285). In this regard, the authors of the GEW identify a close relationship between the Valence dimension and the appraisal checks *intrinsic (un)pleasantness*, *goal (in)consistency*, and *goal conduciveness/obstructiveness*, representing the *Relevance* and *Implications* SECs. The Control dimension, on the other hand, refers to the *Coping potential* SEC, including thus the degree of control a person has on the situation, and the possibility to express this control. The authors suggest that numerous studies corroborated the fact that these two dimensions "have the strongest impact on emotion differentiation" (Scherer, Shuman, et al., 2013, p. 285).

1.2.4.2 The Choice of Modal (Discrete) Emotions and Their Disposition

For the third and most updated version of the tool at present, the authors used the GRID instrument (Fontaine et al., 2013; Scherer, Fontaine, & Soriano, 2013) in order to determine the position of each emotion with respect to the underlying dimensions of Valence and Control. The tool comprises 20 emotions (4 emotions have been re-rated specifically for the GEW, since they were not part of the 24 emotions used in the original GRID study): admiration, amusement, anger, compassion, contempt, contentment, disappointment, disgust, fear, guilt, hate, interest, joy, love, pleasure, pride, regret, relief, sadness, shame. The 20 emotions are arranged on a circle, which can be divided in four quadrants according to the combination of the underlying dimensions (**Figure 1.7**). The first quadrant clockwise includes emotions positive in Valence, and high (positive) in Control; the second, emotions positive in Valence and low (negative) in Control; the third, emotions negative/low in both Valence and Control; and the fourth, emotions negative in Valence and high (positive) in control.

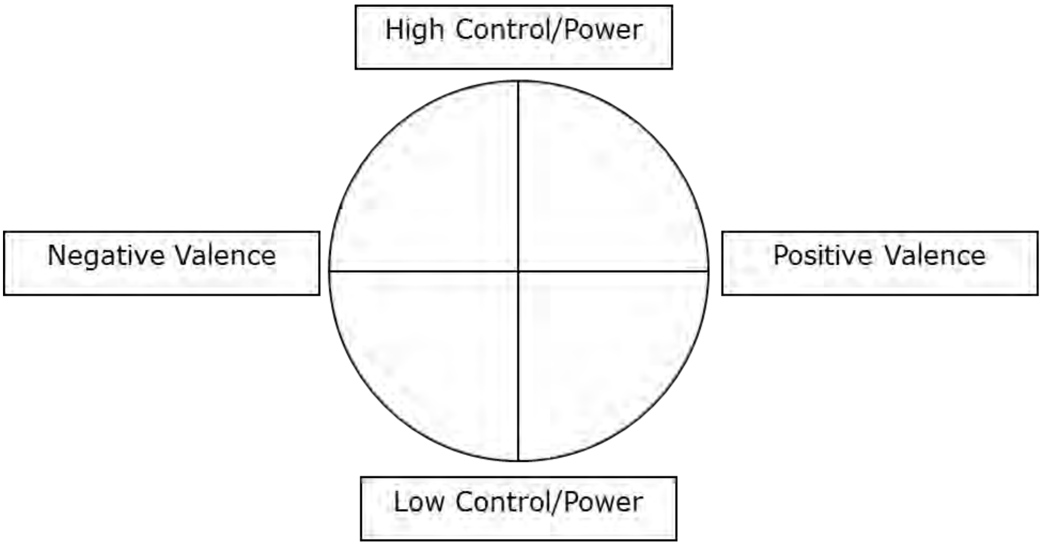


Figure 1.7. The four quadrants resulting from the combination of the two underlying dimensions Valence and Control/Power (from the official material).

1.2.4.3 Graphical User Interface of the GEW

The resulting interface of the GEW 3.0 is illustrated in **Figure 1.8**. Users can manifest an emotion by clicking to the circle that represents the intensity of the corresponding subjective feeling. Users can also provide an emotion not in list in the "other" space, or do not provide any emotion at all (with the "none" space).

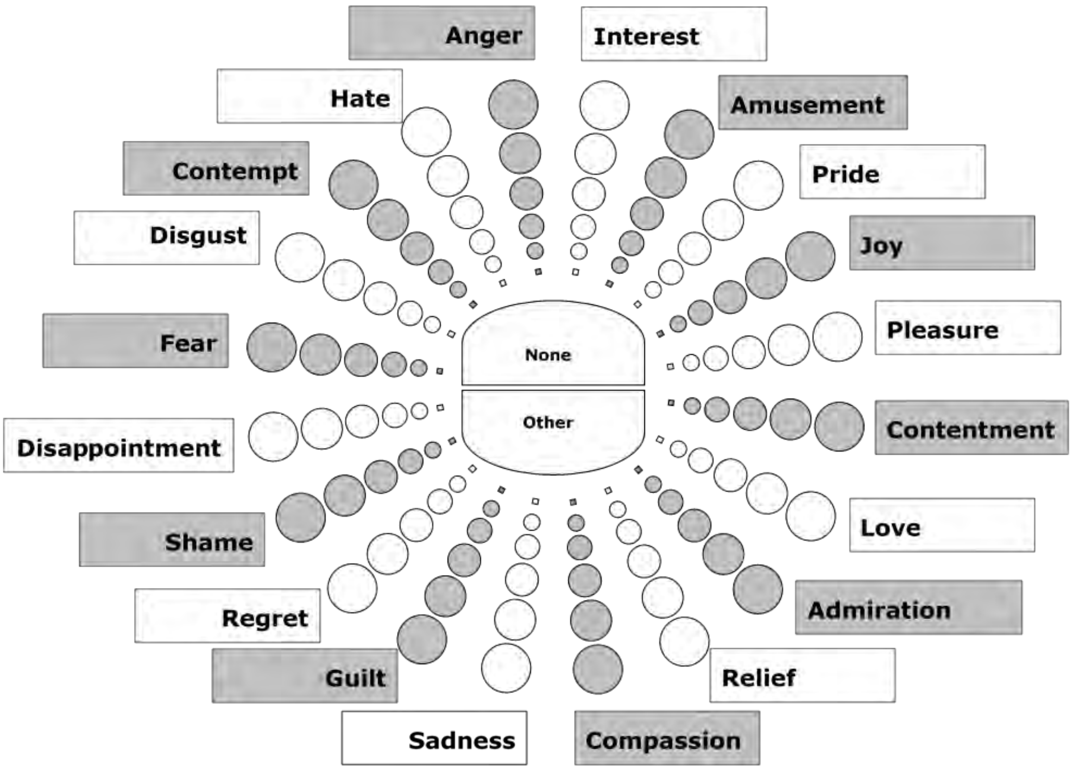


Figure 1.8. Graphical User Interface of the GEW 3.0 (from the official material).

1.2.4.4 Validity and Limitations of the GEW

The GEW is an emotion self-report tool deeply rooted in the CPM theoretical framework and based upon empirical evidence. For this reason, it has been adopted in many studies in different fields and with different empirical settings. In particular, Tran and collaborators (2012) used the GEW in a collective decision-making process, and they observed that members of a team used the tool to discuss about their emotions and keep track of changes in the emotional climate. Furthermore, the usability of the GEW was also compared in a comprehensive test with most of the other tools presented in the previous section (Caicedo & van Beuzekom, 2006). Results suggested that the GEW has a more user-friendly approach.

On the other hand, studies conducted on the disposition of emotions on the two-dimensional space highlighted two problems (Sacharin, Schlegel, & Scherer, 2012). First, the two dimensions correlated, which is inconsistent with the idea of emotion differentiation over the entire circumflex. Second, the disposition of emotions was more problematic for the Control/Power dimension, to the extent that "it might be useful to use a different approach to validate the position of emotion words in VALENCE x POWER [Control] structure, abandoning the exclusive reliance on coping potential appraisals in favor of a more comprehensive representation of POWER" (Scherer, Shuman, et al., 2013, p. 298).

1.2.5 Emotion as Interaction in Affective Computing

The GEW is an example of *affective computing*, a term first introduced by Picard (2000, p. 3) as "computing that relates to, arises from, or deliberately influences emotions". Picard (2009) identifies four research areas for affective computing. (1) Technology for sensing, recognizing, modelling and predicting affective state. (2) Methods for computers to respond intelligently and respectfully to handle perceived emotions. (3) Technology for displaying emotional information or mediating the expression or communication of emotion. (4) Computational mechanisms that simulate internal emotions or implement their regulatory and biasing functions.

Emotional Awareness clearly refers to the third area of research. In this regard, though, Boehner and collaborators (2007) posit that researches in affective computing primarily focused on the other areas, which consider emotion as an internal and individual phenomenon. They identify this approach to emotion as *informational*, which is "rooted in a longer laboratory-science tradition of studying emotion in which subjective experience is considered suspect, to be replaced by objective measures" (Boehner et al., 2007, p. 276). The authors propose instead an *interactional* approach, which (1) focus on how computers can help users understand their emotions, rather than how computers can understand users' emotions; and (2) sees emotions as a cultural, dynamic and social phenomena, constructed in action and interaction. The interactional approach, thus, implies a form of inter-personal communication that goes beyond the liability of the coding and encoding mechanism. The mediating tool's utility is not limited to transfer information from a user to another, but should rather create the conditions for the information to be meaningful and useful. This assumption is consistent with the concept of awareness, defined in the following section. To sum up:

> The role of affective systems is not to transmit pre-existing emotional units, but to provide a resource for emotional meaning-making. Success of such a system is measured by whether users find the system's responses useful for interpreting, reflecting on, and experiencing their

> emotions. Evaluation does not aim at finding a user's original, 'true' emotions, but in tracking how emotions are constructed and interpreted over time, and correlating these dynamics with aspects of system design (Boehner et al., 2007, p. 287).

1.3 Awareness and Awareness Tools

Staying aware of other people's actions and reactions in order to adapt our behavior accordingly is a mechanism that emerges from the early stages of human development (Feinman, 1982) and is therefore "something that we take for granted in the everyday world" (Gutwin & Greenberg, 2002, p. 411). In remote computer-mediated collaboration, though, people do not have access to the same amount and quality of verbal and para-verbal cues available in a face-to-face, co-located situation (Baltes et al., 2002; Lund et al., 2007). Furthermore, the possibility to share information inside a system is mediated by the electronic device used, and therefore faces challenges such as "how to capture, process, and present information about others or context" (T. Gross et al., 2005, p. 324). Since "being able to stay aware of others plays an important role in the fluidity and naturalness of collaboration" (Gutwin & Greenberg, 2002, p. 411), computer-mediated environments faced the challenge to provide information about other people's presence and activities, which is generally referred to as *awareness* (Dourish & Bellotti, 1992) or *groupware* (Ellis, Gibbs, & Rein, 1991). Dourish and Belotti (1992, p. 1) define awareness as "an *understanding of the activities of others*, which provides a *context for your own activity*". Consequently, awareness is of particular interest in the context of collaboration, where it can "ensure that individual contributions are relevant to the group's activity as a whole, and to evaluate individual actions with respect to group goals and progress" (*ibid.*).

In this section, we will first provide general assumptions about awareness, which are useful for designers. Second, we will provide a description of the two distinctive functions of awareness tools, the *displaying* and *monitoring* functions (Schmidt, 2002). Finally, we will illustrate to what extent emotions relate to awareness in computer-mediated environments.

1.3.1 General Assumptions about Awareness Useful for Designers

Considering the important role played by awareness in collaboration, Gutwin & Greenberg (2002) propose a framework to help designers understand awareness and its objectives. The authors suggest that awareness comprises four basic characteristics (Gutwin & Greenberg, 2002, p. 416):

1. Awareness is knowledge about the state of an environment bounded in time and space.
2. Environments change over time, so awareness is knowledge that must be maintained and kept up to date.
3. People interact with and explore the environment, and the maintenance of awareness is accomplished through this interaction.
4. Awareness is a secondary goal in the task – that is, the overall goal is not simply to maintain awareness but to complete some task in the environment.

Furthermore, awareness should help to answer to basic questions such as: (a) *who* executed a particular activity; (b) *what* was the result of the activity; (c) *where* did the activity occur in the environment; (d) *when* did the activity occur; and (e) *how* the activity affects the environment. In this regard, designers should evaluate each questions with respect to two factors: "[1] the degree of interaction between the participants in the activity indicates how specific or general the information in the interface should be; [2] the dynamism of the element – how often the information changes – indicates how often the interface will need to be updated" (Gutwin & Greenberg, 2002, p. 421).

Finally, efficiently conveyed awareness may contribute to enhance collaboration in five ways. (1) Participants will organize their exchanges around awareness information, which would help them to distinguish between moments that are favorable to discussion and collective assessment, and periods where colleagues are engaged in individual work. (2) Awareness contributes to simplify the communication with respect to the *least collaborative effort* (Clark & Brennan, 1991), since being acquainted with what others are doing provide "shortcuts" for mutual understanding. (3) Awareness facilitate coordination of action through communication or visual perception about what activity has been accomplished, which is still undergoing, and what is pending. (4) Awareness maximize the likelihood that participants can anticipate future behaviors, according to past or present activities of which they became aware. (5) Awareness fosters mutual help, since it highlights if participants are struggling with a particular task, and incite colleagues to intervene.

1.3.2 Displaying and Monitoring Awareness

In his article *The Problem with 'Awareness'*, Schmidt (2002) identifies a series of controversies and different theoretical approaches to awereness, starting from the assumption that the very term "is one of those highly elastic English words that can be used to mean a host of different things" (p. 287). The author identifies that one of the main issues with awareness in collaborative settings concerns "how actors so effortlessly pick up what is going on around them and make practical sense of it" (Schmidt, 2002, p. 291). In fact, actors shall strike a balance between the need to use information about the others, and the need to maintain the attention focused on their own activity. The integration of the two aspects results from distinct but mutually dependent functions: *displaying* and *monitoring*.

Displaying refers to the act of making something available to others: "actors display those aspects of their activities that may be of relevance to their colleagues" (Schmidt, 2002, p. 291). Awareness, thus, emerges as the result of someone's activity, which is consistent with the idea that awareness is intrinsically linked to action. Monitoring, on the other hand, relates to the act of perceiving the generated information about others' activities: "[a]ctors obviously somehow 'monitor' the activities of their colleagues in the setting – by observing or listening – so as to ascertain the state, progress, direction, etc. of these activities" (Schmidt, 2002, p. 291). Displaying and monitoring are therefore two integrated functions that account for creating and retrieving awareness information. In other words, "[c]oordination between collaborators can be regarded as an ongoing cycle between displaying and monitoring activities" (Buder, 2011, p. 1115).

Buder (2011) adopts Schmidt's distinction between displaying and monitoring functions to assess how to convey awareness in Computer-Supported Collaborative Learning, which, compared to Computer-Supported Cooperative Work, "has a much stronger emphasis on cognitive or social categories that are not directly observable" (Buder, 2011, p. 1114). According to the author, in fact, CSCW interprets awareness primarily as a mean to reproduce "spatial" information that is normally available in a face-to-face, co-located situation, such as who is around, what is he doing, etc. CSCL, on the other hand, does not consider face-to-face the aspiring model to reproduce in virtual environments, and focus more on cognitive and social aspects that would be difficult or impossible to detect even with seamless audio/video connection. In this regard, Buder (2011) identifies eight issues that designers face in the development of awareness tools, four concerning the displaying function, and four the monitoring function (**Table 1.3**).

Table 1.3. Empirical issues in research on group awareness tools, from Buder (2011, p. 1117 Table 1)

#	Process	Issue
1	Displaying	Explicit (ratings) vs. implicit feedback
2	Displaying	Dynamic display vs. repeated display vs. static display
3	Displaying	Voluntary vs. enforced display activities
4	Displaying	Closed vs. open display format
5	Monitoring	Obtrusive vs. non-obtrusive monitoring
6	Monitoring	Interpersonal comparability of performances
7	Monitoring	Normative pressure vs. evaluation of apprehension
8	Monitoring	Guidance and directivity of tools

1.3.2.1 Methods for Displaying Awareness

According to Buder (2011) the displaying function of awareness may be distinguished with respect to 4 aspects. First, users may provide information about their activity in an explicit or implicit way. The explicit manifestation of awareness implies that users know they are displaying information, they do it voluntary, and are acquainted with the use that would be made of it. The implicit method, on the other hand, does not require any particular action from the user, since the system automatically figures out what kind of information conveying. On the one hand, implicit mechanism are less obtrusive and may be considered unbiased, since awareness derives from the collaborative activity itself. On the other hand, even if explicit feedbacks require additional effort since the displaying action is accessory to the collaborative task, explicit actions may better adhere to the idea of sharing information that is not necessarily visible.

Second, the displaying function varies with respect to its frequency over time. In case of dynamically generated awareness, the process is seamless, for users to dispose at any minute of the most updated information. This process is difficult, though, in case of explicit feedbacks, where a balance between immediacy and effort to maintain the information updated is necessary. On the other hand, a static representation of awareness does not provide any information about the evolution in time, and is therefore limited to a general assumption on a previous state. It is also possible to update information at specific moments in time, either planned by the system or determined by users. This last option, combined with explicit feedback, adds semantic meaning to the entry, considering that users expressed the information at this particular time for a specific reason.

Third, awareness tools may force users to display information, for example at given moments during the collaboration, or let them free to choose the moment they see fit. The issue at stake, in this case, is an equilibrium between the quantity of information at disposal, and the nuisance of constantly interrupting users in what they are doing, which may lead to reject the presence of the tool altogether.

Fourth, awareness information may be collected in more or less open or closed formats. On the one hand, open formats (such as text fields) allow for authentic and detailed information. On the other hand, closed formats (such as choices between a list of predefined items) reduce the workload and are more easily comparable.

> In sum, awareness tool designers for learning have to decide whether to use explicit or implicit feedback; they should implement measures in order to ensure that awareness information does not become outdated through ongoing interaction dynamics; they might consider introducing display activities as a requirement for learners; and they should choose a display format that learners are most likely to use (Buder, 2011, p. 1116).

1.3.2.2 Methods for Monitoring Awareness

Similarly, monitoring function faces 4 issues as well. First, monitoring may be more or less obtrusive. The degree of obtrusiveness refers to the distance between the collaborative activity and the activity required to display awareness information. The more the two activities diverge, more likely the tool would create a dual-task situation (Pashler, 1994), where users must split their attention over two different and potentially unrelated actions.

Second, depending on the type of information available through the awareness tool, information may be more or less easy to compare. The comparison does not simply refer to the open/closed format of the displaying function, but also concerns the effects of comparing one user to the others, specifically oneself to others. This may lead to the emergence of collective dynamics that are known to influence the perception of oneself inside a group (Doise, Deschamps, & Mugny, 2010).

Third, and tightly related to the second issue, awareness tools convey information about individual's contribution to the collective task, which means that individuals may fear the evaluation of their peers. On the other hand, knowing that someone has contributed less than others may push this person to increase her involvement. Awareness tools may therefore be more or less intrusive of users' privacy, determining what other will know about one's activity.

Fourth, awareness tools require to some extent that the system direct users' attention, for them to notice awareness related information, and therefore adapt subsequent actions accordingly. This means that awereness tools integrate the primary task in a way that can more or less affects its "normal" workflow – that is the same task without awareness monitoring.

> In sum, designers of group awareness tools for learning must try to find a balance between primary tasks (collaboration) and secondary tasks (monitoring); they should consider potentials and drawbacks of providing learners with information that will foster social comparison; they should take issues like evaluation apprehension into account; and they should find ways to provide guidance without diminishing learner autonomy (Buder, 2011, p. 1117).

1.3.3 Emotional Awareness in Computer-Mediated Environments

As suggested by Schmidt (2002), awareness is a term that may refer to different concepts, even within the field of CSCW and CSCL. In this regard, Gutwin, Greenberg and Roseman (1996, p. 5) reckon that "[p]eople are aware of many different things when they work in groups, some of which relate to the group, and some to the task or situation more generally". More specifically, the authors identify four types of awareness proposed in CSCW related researches, which are normally available in a face-to-face situation: informal awareness, social awareness, group-structural awareness, and workspace awareness. Emotions are mentioned in particular with respect to:

> Social awareness is the information that a person maintains about others in a social or conversational context: things like whether another person is paying attention, their emotional state, or their level of interest. Social awareness is maintained through conversational cues such as back-channel feedback, and through non-verbal cues like eye contact, facial expression, and body language (Gutwin et al., 1996, p. 6).

On the other hand, Gross, Stary and Totter (2005) suggest that CSCW should implement awareness with respect to a more general scope, referring to the concept of awareness as defined in social sciences such as sociology, cognitive and social psychology. In this regard, they propose operational definitions of different types of awareness, for developers to take full advantage of interdisciplinary transfer of knowledge. Among these types of awareness, one directly mentions emotions:

> Group awareness is related closely to emotions, the perception of others' properties, and the provision of collective leads. There might be patterns of interaction or group roles giving direction to social developments in the group or workspace or both. In the context of CSCW, this type of awareness strongly affects the collaboration aspect when work is supported by computers, in terms of how group members perceive one another and handle this information when cooperating via information and communication technology in work tasks (T. Gross et al., 2005, p. 341).

These examples corroborate the idea that emotions play an important role in conveying meaningful information, even in a computer-mediated environment that does not necessarily provide participants with seamless audio-video connection. In this regard, Derks, Fischer and Bos (2008) conducted a review of studies in order to assess if it is more difficult to share emotions in computer-mediated communication (CMC) compared to face-to-face (F2F) communication. Results infirmed the general assumption that emotional communication is limited in CMC due to the lack of non-verbal cues; on the contrary "people have found ways to cope with the restrictions of CMC, for example by the use of emoticons, or by verbalizing emotions in a more explicit way" (Derks et al., 2008, p. 780). Furthermore, Cheshin, Rafaeli and Bos (2011) demonstrated that emotional contagion occurs in computer-mediated environments where the communication is exclusively text-based.

On the other hand, though, Eligio and collaborators (2012) showed – in one of the few studies investigating the relationship between emotions and computer-mediated collaboration – that without explicit instructions about emotion sharing, participants to collaborative tasks have a poor understanding of their partners' emotions. More specifically, participants tend to attribute to other a reflection of their own emotions. On the other hand, if participants received precise instructions to share their emotional state at key moments during the task, both performance and emotion understanding improved. These findings suggest that emotional awareness may not be automatically taken into account without explicit instructions, or, precisely, tools that are dedicated to this task.

1.4 Summary

In this introduction, we defined the general and more specific scope of this contribution. The general scope refers to the EATMINT's project objective to study the relationship between emotions and computer-mediated collaboration. The specific scope relates to the development of an Emotion Awareness Tool that will allow participants to share their emotional episodes during an ongoing collaborative task. To put the development of an EAT into perspective, we provided an overview of the EATMINT's project theoretical framework, which suggests that collaboration in a computer-mediated environment would benefit from emotional awareness. We then defined what an emotion is, and how it can be measured using emotion self-report tools. In particular, we illustrated the Geneva Emotion Wheel, an emotion self-report tool deeply rooted in the Component Process Model theoretical framework, which combines the two major approaches to emotion measurement: the dimensional and discrete emotions approaches. Finally, we proposed some general assumptions about awareness that developers should consider in order to provide users with efficient tools, allowing them to take information about others into account for their own task. In this regard, we defined the displaying and monitoring functions of awareness, suggesting how developers face major choices about the best way to split attention between awareness information and the primary collaborative task. In the next chapter, we will specify the objectives of this contribution and the methodology adopted to reach them.

2 Objectives and Development Methodology

The specific purpose of this contribution consists in the development of an Emotion Awareness Tool, which would be used preeminently in experimental settings. This implies that the tool must serve up to two distinct but mutually dependent functions:

1. It must provide participants with an intuitive and accessible Graphical User Interface through which they can share their emotional state and take this information into account as meaningful indications in the perspective of the undergoing collaborative task. In the meantime, activities related to emotional awareness should not diverge participant's attention from the primary collaborative task they are performing (Pashler, 1994). The fourth chapter of this contribution will more clearly define these objectives in the perspective of an empirical test.
2. It must provide experimenters with a trustworthy tool for measuring the phenomenon they are interested to register, which, in this case, consists in voluntary shared emotions. Consequently, experimenters should be able to easily integrate the tool in their experimental design and obtain measurements of interest for their analysis (Rosenbaum, 2009). In this regard, the next chapter depicting the application architecture will define which measures the tool gathers. These measures would also be integrated the empirical test illustrated in the fourth chapter.

Since participants to experimental settings usually represent a heterogeneous population and people do not participate in experiments on a daily basis, it would have been difficult to strictly follow an interaction design approach (Cooper, Reinman, & Cronin, 2007; Rogers et al., 2011) and assess directly from participants which features they would expect in an Emotion Awareness Tool. In fact, interaction design aims at "designing interactive products to support the way people communicate and interact in their everyday and working lives" (Rogers et al., 2011, p. 9). Even though this objective is premature with respect to an Emotion Awareness Tool, we nevertheless considered the process of interaction design suitable as a guideline for the development of the tool. The process consists in four stages: (1) establishing requirements; (2) designing alternatives; (3) prototyping; and (4) evaluating. We therefore adapted this paradigm to the specific context of experimental material purposes, integrating both users and experimenters' perspectives.

Furthermore, since the requirements analysis did not provide specific features, but rather suggested possibilities to explore, the Rapid Application Development (Gottesdiener, 1995) method provided the best procedural paradigm to follow. In fact, we expected the tool to emerge as the integrated result of individual features. In the meantime, we were aware of the importance of accuracy in experimental settings, and therefore decided to adopt a Test Driven Development (Pančur & Ciglarič, 2011) approach from the early stages of coding.

Finally, as a way to figure out whether the tool fulfilled the expectations both in terms of the experimenter and participant's perspective, we decided to conduct the usability test in experimental settings implicating a computer-mediated collaborative task. The test would therefore be the occasion to: (1) assess the utility of the tool with respect to the gathered data; and (2) measure the User Experience (Tullis & Albert, 2013) resulting from the use of the tool in similar conditions to the EATMINT experiment. In the rest of this chapter, we will provide a brief overview of each methodology adopted in this contribution.

2.1 Interaction Design

Interaction Design (ID) refers to an approach to developing products based upon the user's perspective, and therefore involves the user in the designing process from an early stage. Cooper et al. (2007) proposes the concept of *Goal-Directed Design* to emphasize the importance ID should give to a fundamental question: "[w]hy is a user performing an activity, task, action, or operation in the first place?" (Cooper et al., 2007, p. 15). In order to provide an answer to this question, and subsequently develop a product that allows the user to achieve his goals, ID suggests a developing process consisting in 4 phases (Rogers et al., 2011):

1. *Establishing requirements*: during this phase, developers must identify what the users want to obtain by using the product. This phase is sometimes problematic, especially when an unfamiliar technology is introduced, since users do not yet dispose of enough elements to determine if and how they would benefit from the use of a product.
2. *Designing alternatives*: ID posits that there is not a unique way to achieve a result, but not all forms of the product are equivalent. Therefore, design choices affect the interaction with a product, and alternatives extend the likelihood to meet user's goals.
3. *Prototyping*: prototypes are representation of the products that give a sense of the interaction they allow. ID emphasizes the use of prototyping at different level, starting from low-fidelity wireframes, which can be easily created and modified.
4. *Evaluating*: evaluation determines if the product fulfills user's goals, and more generally assess the perception of the product by the user (also see 2.4)

2.2 Rapid Application Development

Rapid Application Development (RAD) refers to an "integrated set of techniques, guidelines and tools that facilitate deploying a customer's software needs within a short period of time" (Gottesdiener, 1995, p. 28). RAD emerged as an alternative to the waterfall method, which implies a structured and sequential cycle of development. In contrast, "RAD is based on the premise that software development is a discovery process" (Gottesdiener, 1995, p. 38). In other words, the development of the software affects the result, since the development itself provides helpful cues about what the software should do and how it should perform.

RAD implies a continuous implementation of customer's feedback for the software to adapt accordingly. This process allows customers to consider modifications they did not necessarily foresee at the beginning, but rather emerged, for instance, as the result of a prototype test. RAD is also a highly iterative process by which the software is not necessarily delivered as a whole, but rather in functional parts that can be tested, and serve as basis for further functionalities added in a subsequent cycle of development.

2.3 Test Driven Development

The Test Driven Development (TDD) approach consists in the general assumption that a software is the integrated result of functionalities. Thus, "TDD practice starts with thoughts on how to test the required functionality" (George & Williams, 2004, p. 338). Functionalities mainly refer to object, methods and functions that form the computational logic underneath the software.

> The main idea behind TDD concept is that the programmer writes unit tests before writing any new code. When writing the test code, programmer in fact specifies how exactly the program would need to work in order to pass the test. The test necessarily fails, since there is no code yet. After that, the programmer writes the code needed to pass the test. When the test passes (and other previous tests pass too), the programmer looks at the code and if needed, improves its design (Pančur & Ciglarič, 2011, p. 557).

Empirical research suggests that TDD improves the quality of software and decrease development time (George & Williams, 2004; Pančur & Ciglarič, 2011) since it allows for small development cycles that are highly iterative and limits the possibility that more complex functionalities are built upon dysfunctional code. At the same time, it breaks up the logic of the overall software in small units, and increases the likelihood of reusing and combining functions in different parts of the application.

2.4 Measuring the User-Experience

The term *User Experience* (UX) emerged in the field of human-computer interaction as a comprehensive term referring to "the experience the product creates for the people who use it in the real world" (Garrett, 2011, p. 6).

> *Usability* is usually considered the ability of the user to use the thing to carry out a task successfully, whereas *user experience* takes a broader view, looking at the individual's entire interaction with the thing, as well as the thoughts, feelings, and perceptions that result from that interaction (Tullis & Albert, 2013, p. 5).

UX requires three elements: (1) a user, (2) an interaction with "something" that has an interface; and (3) an interest in observing and measuring the subjective or objective outcomes of the interaction in a direct or indirect way (Tullis & Albert, 2013). Different methods and techniques exist to measure UX such as tests, scales, or audits. Usability tests often imply different users to execute the same tasks so that the measures – such as task success, errors, execution time, etc. – are comparable. Measuring the user experience provides useful cues in the perspective of an iterative process, since it indicates what part of the products can be improved.

3 Development of the Dynamic Emotion Wheel

The development of the new version of the EAT must consider at least three interrelated layers. First, on a macro-level, general assumptions about awareness and emotions, provided by theoretical and empirical evidence. Second, on a meso-level, the specific requirements defined by the EATMINT project members. Third, on a micro-level, human-computer interaction principles and development paradigms as guidelines to meet these requirements. In this section, we describe how the three layers intertwined in the development of the Dynamic Emotion Wheel distinctive feature: the dynamic generation of a subset of discrete emotions based on the appraisal of the situation on two-dimensional criteria. Since the DEW was developed in the perspective of the usability test as part of an iterative process, the focus of development was preeminently to dispose of a functional tool to test, but neither in a final nor stable version. It is therefore difficult to outline the features of the tool independently from the design process, because those very features emerged during the development itself. Consequently, first, we describe how the idea of combining the discrete emotions and the dimensional approaches to emotion self-reporting emerged to solve a usability issue during the design of alternatives. Second, we define the core principle of the Dynamic Emotion Wheel, in particular the *dynamic algorithm* that "adapts" the Geneva Emotion Wheel to emotional awareness purposes. In this regard, we took advantage from the validity of the GEW and maintained the same underlying dimensions (i.e., Valence and Control) and the same set of 20 discrete emotions on the circumflex as guidelines for development and testing. Third, we provide a general outlook of the DEW's architecture and illustrate the specific measures recorded through its use. Finally, we compare the result of the development process with the requirements established in the introduction. Whenever possible, we also illustrate some limits of the tool that derive from conceptual or technical shortcomings – that is, structural problems independent from the test conducted and illustrated later in the contribution.

To facilitate the reading of this technical section, references will often be made to the two functions of awareness as mentioned by Schmidt (2002): the displaying and the monitoring functions. Displaying refers to the information that goes from the user to the tool (i.e. inserting data), whereas monitoring to information that goes from the tool to the user (i.e. retrieving and presenting data). In addition, the term emotion will sometimes be used, for its immediacy, interchangeably with the subjective feeling component of the CPM. Although this is inconsistent with the theoretical framework adopted, we think that the reading will benefit from this "infraction". Finally, given that this section will be rather conceptual, the reading of the comprehensive development process could benefit from the overview of a working example of the DEW, available on the TECFA web space dedicated to students[3].

3.1 Combining Dimensional and Discrete Emotions Approaches

The first version of the EAT adopted the discrete emotions approach both for displaying and monitoring emotional information, since users clicked on a button with an emotion-adjective on it, and the same adjective then appeared top of the list in the three last emotions cases of the interface (see 1.1.2). Following the Interaction Design process, we designed an alternative tool adopting a pure dimensional approach (**Figure 3.1**), in which users would rather assess their emotional state using two *sliders* – that is, inputs to express numerical values between a predefined range by moving a cursor to the left or to the

[3] http://tecfaetu.unige.ch/perso/maltt/fritz0/memoire/dynamicemotionwheel/

right on a continuous scale. Each slider would determine the user's position proportionally to minimal and maximal values. Using the Valence dimension of the GEW as example, participants would determine to what extent they evaluate the situation as pleasant/unpleasant, in a range comprised between *very unpleasant* and *very pleasant*.

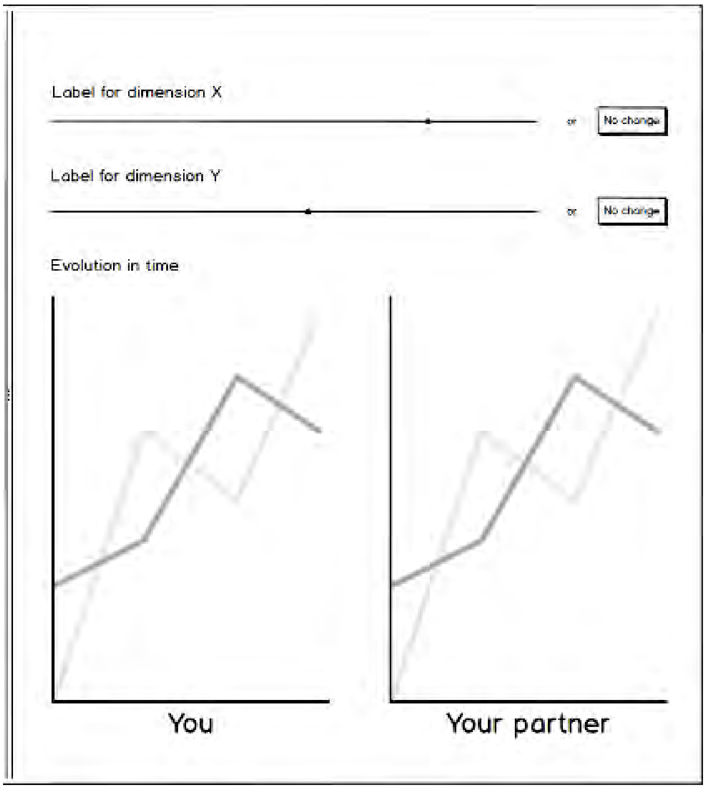


Figure 3.1. Wireframe of an alternative design, based on a bi-dimensional approach to emotion displaying, and a quantitative representation through line charts for monitoring.

A pure dimensional approach, though, presents at least two major drawbacks. First, it is insufficient to express and perceive all the nuances of emotional episodes. In other words, telling the interface (or knowing from the interface) that someone feels minus 50 on dimension X and 60 on dimension Y does not provide a meaningful representation of the emotional episode for participants to regulate and reflect on their emotions. Second, the quantitative data gathered is difficult to "translate" in a bounded emotional episode. The use of iconic representations, whose features change with respect to the dimensional values (such as in the AffectButton), is subject to an intrinsic polysemy (Darras, 1998; Eco, 2013). Furthermore, icons are not suitable for the monitoring function over time: users should assess every icon in a row, and then compare it with the previous one to perceive the changes. In this regard, charts are certainly more suitable, but they also present drawbacks. First, reading a graph is often considered a difficult task (Pinker, 1990). Second, even though the evolution in time of emotional episodes was one of the requirements, the visual representation of an emotion must nevertheless be consistent with its distinctive features as affective phenomenon. In particular, an emotion is limited in time, and considered a major change in behavior. Representing emotions exclusively through, say, a line chart risks to convey a misleading perception of continuity and modulation. In spite of these major drawbacks, though, the idea of integrating the dimensional and discrete emotions approaches emerged for a usability issue concerning the sliders.

3.1.1 The Interaction Problem with Sliders

A major change from the one-click button of the first EAT version to the two-dimensional sliders of the alternative wireframe consists in spreading an observation over time. In fact, handling two inputs instead of the single button causes a latency time between the manipulations (**Figure 3.2**). Saving all manipulations automatically is not an option: participants could just modify a previous value that they judged incorrect, or provide sequential entries over a short period of time, leading to the impossibility to distinguish a bounded episode.

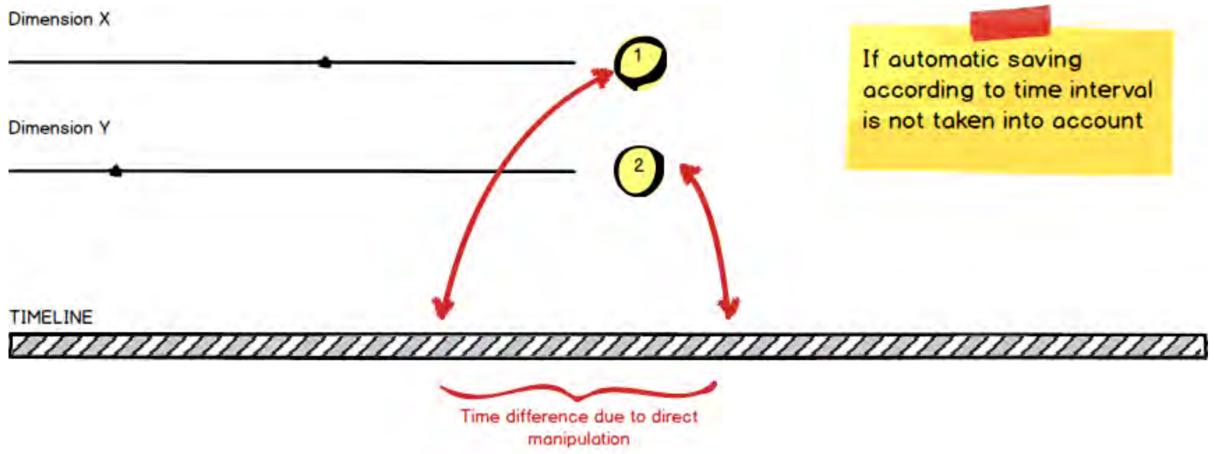


Figure 3.2. Latency time between the two manipulations of the sliders could result in confusion if participants provide sequential entries in a short time.

A simple technical solution to this problem consists in adding a button, whose only purpose is to save both values as a single observation, as it is often the case with online web-forms. Adopting this solution, though, implies three actions to manifest an emotion, against a single action in the case of emotion-labelled buttons. Moreover, the third action (i.e. clicking a button to send data) is a system-related action that does not convey any semantic information.

3.1.2 Adding Semantic Meaning to a Button

Since the necessity of a third manipulation seems inevitable in order to provide two values at the same time, we considered that at least this action should convey meaningful information. As emerged from the theoretical framework, emotion words remain the most suitable way to synthetize the subjective feeling and thereafter communicate semantic emotional information to others. We therefore imagined that an emotion-labelled button similar to those used in the first EATMINT version could replace the synchronizing "send data" button. The resulting interface would then imply (1) *two dimensional sliders* whose role is the cognitive evaluation of the event/situation, and (2) *a discrete emotion button* that represent the subjective feeling. In other words, the interface would present the two components of the CPM that, in a communication perspective, are linked to the inter-individual level. Since the appraisal of the situation accounts for emotional differentiation, we speculated that the emotion-label could directly depend on the combination of the two-dimensional values. However, considering that two dimensions are not sufficient for an accurate discrimination (Fontaine, Scherer, Roesch, & Ellsworth, 2007), it was not conceivable to obtain a perfect match on a single button. Moreover, a single button would not add any meaningful information to the dimension values, since the two representations would be equivalent. On the contrary, the interface should present more emotion-labelled buttons among which users choose the most appropriate. That would add valuable information to the data entry, without incrementing the number of manipulations (**Figure 3.3**a).

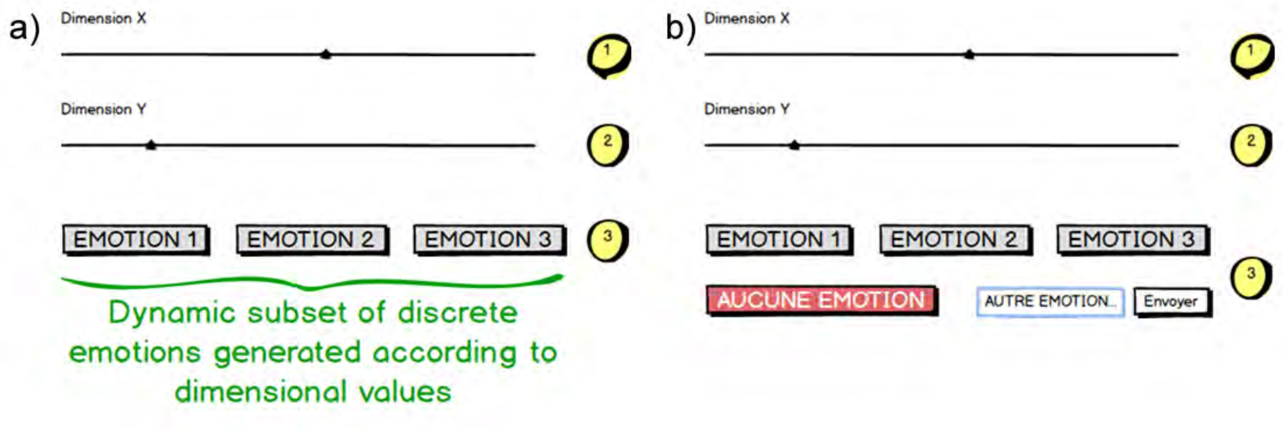


Figure 3.3. (a) Replacing the "send data" button with a series of emotion-labelled buttons integrates semantic meaning to the manipulation. The emotions should be determined with respect to the appraisal dimensions. (b) The interface can be implemented with the additional features proposed by the first EATMINT version and the GEW: typing an emotion not listed and clicking on a "no emotion" button.

We then implemented the wireframe with the additional self-reporting features of the first EAT version and the GEW: the possibility to type an emotion not in the list, and a “no emotion” button (**Figure 3.3**b). The interface would then respond to the requirements concerning (1) the integration of a qualitative measure, suitable for statistical analysis; and (2) reducing the number of emotions displayed on the interface.

3.1.3 Qualitative and Quantitative Data to Represent an Emotion

The addition of the subjective feeling component produces both quantitative (two-dimensional sliders) and qualitative (emotion-labelled button) data. This widen the possibilities of representing an emotion in an adequate form. On the one hand, the two-dimensional values do not represent the bounded emotional episode, but rather the cognitive evaluation of the situation. Consequently, a graphical representation of the appraisal dimensions through a line chart would not generate confusion with the temporal features of an emotional episode. In fact, a chart that shows appraisal dimensions over time only accounts for variations in the evaluation of the situation. In the meantime, the holistic subjective feeling, as a single word, is more clearly delimited as a distinct occurrence.

Furthermore, both appraisal dimensions and subjective feeling relates to the inter-individual level. We therefore assumed that, in the absence of motor expression cues when audio-video connection is not provided, both components would be useful to convey meaningful information about the undergoing emotional episode. The cognitive evaluation of the situation, on the one hand, provides a hint about the eliciting causes of an emotion. For instance, if the user assesses an event very high in Valence and very high in Control, the partner could infer that something positive has happened as a result of the user’s direct action. The subjective feeling, on the other hand, provides nuances of emotional meaning that quantities cannot provide, and therefore maximizes the possibility that both partners share the same mental representation of the emotional state, including its causes and consequences (Lin et al., 2010).

Furthermore, reflecting the same information provided through the interface may facilitate a process of *reverse engineering,* through which users reconstitute the episode backwards. A conveyed feeling of admiration by our partner, corroborated by a positive Valence and negative Control appraisal, may be interpreted as “he is admiring because what I did is good [Valence], and he could not or did not think to do it [Control]”. Limiting the information to the dimensions may generate, for instance, an inference closer to relief: “he was overwhelmed by the situation [Control], and I bailed him out [Valence]”. On the other hand, restricting the information to the feeling could result in confusion about what the partner is admiring – his own work (on which he has Control) or his partner’s work (on which he has not)?

3.2 The Core of the Dynamic Emotion Wheel

Since the Geneva Emotion Wheel combines dimensional and discrete emotions approaches, we thought of adapting it to emotional awareness, because neither its size, nor the lack of a dedicated monitoring function fitted with the purpose. In the GEW, the underlying dimensions serve as coordinates to position the emotions in an order that reflects their extent both on the Valence and Control scales. We therefore approached this principle from a slightly different angle. Users of the GEW do not directly interact with the underlying dimensions, which serve as “placeholder guidelines”. What if, on the contrary, users could assess the Valence and Control dimensions on two quantitative scales, and obtain a list of emotions fitting with the appraised criteria?

This idea is the core principle of the Dynamic Emotion Wheel, which precisely refer to the dynamic computation of a subset of discrete emotions according to user’s evaluation of an event with respect to two-dimensional criteria. Since we are aware that this mechanism does not represent an accurate reproduction of the emergent synchronized pattern of the multi-componential emotional phenomenon, we considered the dynamic mechanism *ex negativo*. The mechanism should exclude those emotions that are less likely to occur given a specific evaluation of the situation. Using the GEW as an example, if the user appraises the situation as positive both in Valence and Control, the odds are that he is not feeling one of the emotions whose position is determined by at least one negative dimension. That leaves the emotions positioned in the top-right quadrant of the GEW as potential candidates for the corresponding subjective feeling – namely interest, amusement, pride, joy, or pleasure.

3.2.1 The Dynamic Algorithm

What we needed to put this principle into practice, then, was an algorithm computing the position of an emotion in a circumflex given two values on a continuous range. We imagined superposing to the interface of the GEW a Cartesian plan where the vertical axis represents the Control dimension, whereas the horizontal axis refers to the Valence. We then considered the circumflex as a circle of 360°, where each emotion refers to an angle of the radius with respect to the y-axis. For instance, the GEW includes 20 emotions, so we divided the round angle by 20, obtaining a mean radial distance of 18° between each emotion. Since no emotion on the GEW is placed exactly on the edge of the four intersections between the axis and the circumflex, we moved all emotions by half of the average radial distance. Consequently, the *interest* feeling (the first label clockwise) received an angle of 9°, *amusement* of 27°, and so on until *anger*, associated with an angle of 351°. The complete list is illustrated in **Table 3.1**.

Table 3.1. Angle and expected value (positive or negative) for the Valence and Control dimensions for the 20 emotion-labels used in the GEW 3.0

#	Label	Angle	Valence	Control	Quadrant
1	Interest	9°	Positive	Positive	I
2	Amusement	27°	Positive	Positive	I
3	Pride	45°	Positive	Positive	I
4	Joy	63°	Positive	Positive	I
5	Pleasure	81°	Positive	Positive	I
6	Contentment	99°	Positive	Negative	II
7	Admiration	117°	Positive	Negative	II
8	Love	135°	Positive	Negative	II
9	Relief	153°	Positive	Negative	II
10	Compassion	171°	Positive	Negative	II
11	Sadness	189°	Negative	Negative	III
12	Guilt	207°	Negative	Negative	III
13	Regret	225°	Negative	Negative	III
14	Shame	243°	Negative	Negative	III
15	Disappointment	261°	Negative	Negative	III
16	Fear	279°	Negative	Positive	IV
17	Disgust	297°	Negative	Positive	IV
18	Contempt	315°	Negative	Positive	IV
19	Hate	333°	Negative	Positive	IV
20	Anger	351°	Negative	Positive	IV

At this point, we disposed of an identifier for each emotion that was related to the origin of the Cartesian plan. Each emotion, in fact, could be “hit” by a hypothetical radius sharing the same angle of the emotion. Applying a confidence interval on both sides of this radius creates a circular sector that, according to its width, includes a number of adjacent emotions (see **Figure 3.4**). What was needed, then, was an algorithm to pass from two-dimensional values to an angle.

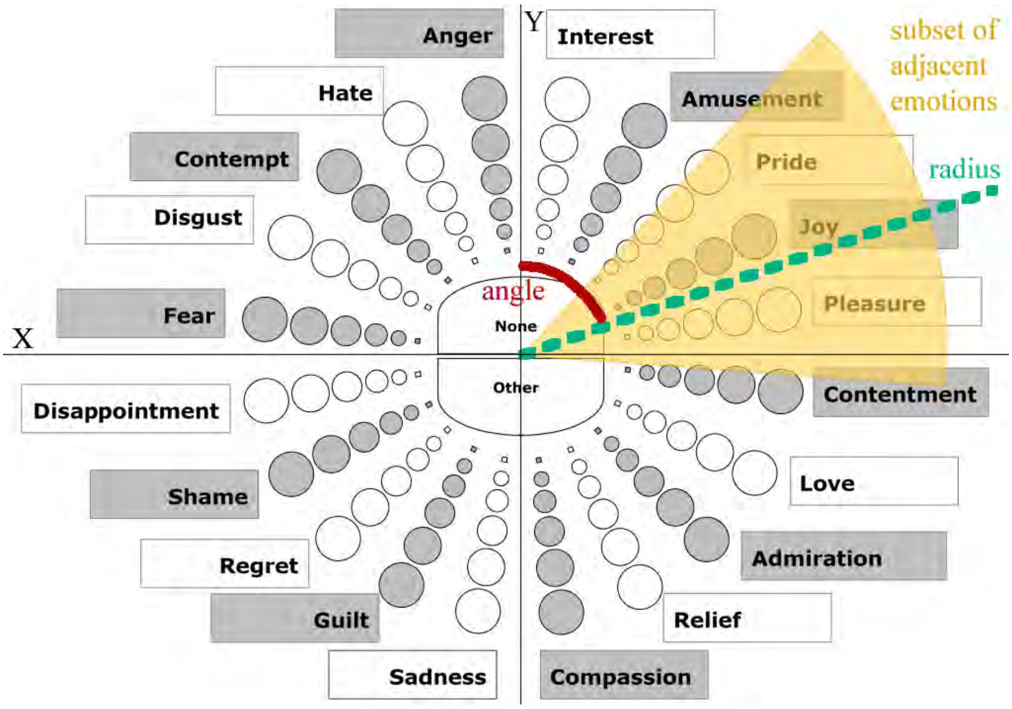


Figure 3.4. Circular sector including a number of adjacent emotions based on their positions on the circumflex (using GEW 3.0 as underlying model).

3.2.1.1 Expected Function of the Algorithm in TDD Perspective

Following the Test Driven Development approach[4], we started by planning the expected result of the algorithm: depending on the combined values of Valence and Control, a hypothetical radius starting from the origin of the Cartesian plan should cross the circumflex in one of the four quadrants. More specifically, given any two numbers representing the Valence and Control sliders, the radius should:

1. Cross the first quadrant if both numbers are positive;
2. Cross the second quadrant if the Valence is positive and Control is negative;
3. Cross the third quadrant if both numbers are negatives;
4. Cross the fourth quadrant if the Valence is negative and the Control is positive.

We built the expected behavior defining the quadrants with the starting and ending angles. Consequently, in the first test with both positive numbers, we expected an angle of the radius between 0° and 90°. In the second test, with positive Valence and negative Control, between 90° and 180°. And so on. As expected by TDD philosophy, all tests failed. At this point, we needed to build the algorithm for these tests to pass.

3.2.1.2 Calculating the Angle of the Slope

We considered our hypothetical radius as a line in a Cartesian plan. Consequently, it is possible to calculate the slope of a line for any given two points A and B on that line by the formula:

$$slope = \frac{A_y - B_y}{A_x - B_x}$$

[4] For our test environment, we used Karma (Dutta, Verma, & Prithviraj, 2015), a Test Runner for JavaScript, with the Jasmine framework notation (http://jasmine.github.io)

Since the aim of the dynamic mechanism is to determine the direction of a radius starting from the center of the circumflex, a point of the line is given by the origin (i.e., 0:0). The other is determined by the values of the two-dimensional sliders of the interface, where the Valence-value represents the x-coordinate, and the Control-value the y-coordinate. The slope thus results by the ratio between the Control and the Valence. Once obtained the slope, it is possible to compute its angle with respect to the x-axis using the arctangent (or inverse tangent) function:

$$angle = \arctan(slope)$$

Testing this function as the algorithm in our TDD environment highlighted the following problems. First, the angle of the slope is determined with respect to the x-axis, whereas we expected the angle referred to the y-axis clockwise. Second, the computed angle of all possible slopes is comprised between 90° and -90°, whereas we expected the angle covering the entire round angle. We therefore adapted the algorithm for the angle to be (1) symmetrically reflected from the x to the y-axis, and (2) rotated to the expected quadrant, given the positive/negative combination of dimensional values. The following JavaScript (Crockford, 2008) function illustrates the adapted algorithm that passed the test (the function uses radius that are converted in degree at the end of the block):

Listing 3-1. JavaScript function to compute an angle on a 360-degree range based on two-dimensional values.

```
function calculateSlopeAngle (x, y) {
    var slope = (Math.PI / 2) - Math.atan(y / x);
    if(x > 0 && y > 0) {
     slope = slope - 0.5 * Math.PI;
    } else if(x < 0 && y < 0) {
     slope = slope + 0.5 * Math.PI;
    } else if(x > 0 && y < 0) {
     slope = slope - 0.5 * Math.PI;
    } else if (x < 0 && y > 0) {
     slope = slope + 0.5 * Math.PI;
    }
    slope = slope * 180 / Math.PI + 90;
    return slope;
}
```

3.2.2 Generation of a Subset of Adjacent Emotions

The adopted algorithm is quite simple, which is at the same time a limit and an opportunity. A limit, since such a simple computational model cannot certainly account for the complexity of the relationship between appraisal processes and the corresponding feelings. An opportunity, since its simplicity requires minimal computational power and time. Therefore, the function could be called any time – even multiple times – without blocking concurrent processes. This is consistent with the idea of generating a subset of emotions according to variations in the Valence and Control sliders of the interface. Given that the algorithm provides a hypothetical angle representing the position of the evaluated situation on the circumflex, we can determine the circular distance between each emotion and this point. Let us say that a user evaluates an event minus 20 in Valence and minus 80 in Control. The computed slope equals 4, and the computed angle equals 194°. The closest emotion on the circumflex would be *sadness* (189°), followed by *guilt* (207°), then *compassion* (171°), and so on.

On a computational level, we could obtain a subset of adjacent emotions – that is, emotions considered similar based on appraised dimensions – by reordering all the emotions according to their proximity to the computed angle of the slope, and subsequently decide to limit the list to *n* items. This simple process requires, though, an adaptation due to the circularity of the round angle, which implies that a computed slope of 5° is closer to *anger* (351°) than to *amusement* (27°). We set up another test to validate a function, whose role consisted in ordering the emotions, provided the particularity of the items at the beginning and at the end of the circumflex. We consider the details of this function irrelevant for the scope of this contribution, but it is worth noting that the circular characteristic of a circumflex complicates the principle of proximity between adjacent emotions at the extremes of the 360-degree range.

3.2.3 Unit Test of the Dynamic Algorithm

We disposed at this point of a tested algorithm to obtain a subset of adjacent emotions, which would reduce the number of concurrent buttons present on the interface. The computational logic underneath this process, though, presented a structural shortcoming due to the calculation of the slope. In fact, the slope represents a ratio between the y-axis and x-axis values, which implies two interrelated consequences. First, a ratio is independent from the scale adopted: one observation measuring 20 on dimension x and 30 on dimension y, and another measuring 40 on dimension x and 60 on dimension y compute exactly the same slope of 1.5, with an angle of around 56°. Second, due to the proportional dependency, using two continuous integer scales within the same range implies that some ratios are more frequent than others, especially if the two scales have a limited scope in their granularity (e.g. two scales ranging from -10 to 10).

We therefore executed a unit testing simulating all the possible combinations using an integer range from -100 to 100 in order to:

1. Control that all the 360° degrees of the circumflex would be computed;
2. Check how many times each emotion would appear in a dynamically generated subset of three items.

The test confirmed that a range from -100 to 100 computes all the 360° of the circumflex. On the other hand, it also confirmed that not all emotions have the same probability to be part of the subset. Emotions situated in correspondence to the bisection of the four quadrants have more chances to appear in the subset, since they present a ratio close to one or minus one. On the contrary, emotions situated on the edges of each quadrant (i.e. close to the axis) are computed less frequently, since they need a wider gap between the x and the y values. **Figure 3.5** visually illustrates the unequal number of times each emotion appears in a subset of 3 items in first, second and third position cumulated. In our opinion, this shortcoming of the algorithm should nevertheless not affect the displaying function of the DEW, since the computed angle is determined by a contingent evaluation of the two dimensional criteria. In other words, no indicator suggests that users would encounter more frequently some specific emotions, unless they often evaluate the two dimensions with a similar proportion.

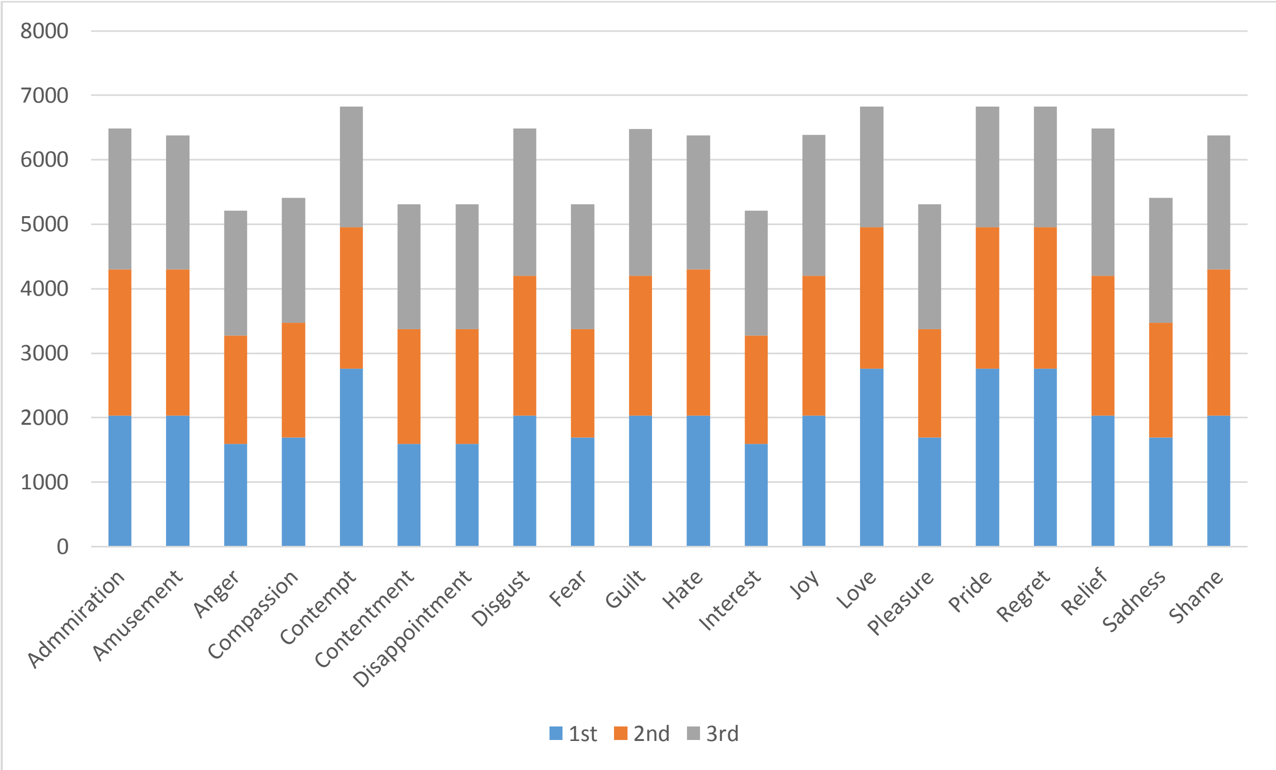


Figure 3.5. Histogram of the number of times each emotion appears as first, second or third item in a computed subset of emotions. Simulation executed with all combinations on a minus 100 to 100 range.

3.3 Description of the DEW's Interface

The Graphic User Interface of the DEW and the expected interaction workflow were shaped upon the aforementioned algorithm. We will first describe the layout of the tool. Then, we will illustrate the expected interactions allowing users to project their emotions into the interface, and perceiving their partner's emotions in the meantime.

3.3.1 The Layout

We decided to maintain the vertical-column layout of the EATMINT's first version of the tool, since it best fits with the necessity to bestow the larger part of the screen to the collaborative task. **Figure 3.6** illustrates the low-fidelity wireframe of the DEW. We separated the displaying (above) and monitoring (below) functions of the EAT to clearly define which part of the screen requires handling, and which just observation.

We maintained the displaying part of the interface as planned in the wireframe. For the monitoring part of the screen, we adopted a very simple approach that consisted in retrieving the data exactly in the inserted format, that is, without computational treatment. We created thus a zone for the subjective feeling data, and a zone (split in two) for the appraisal dimensions.

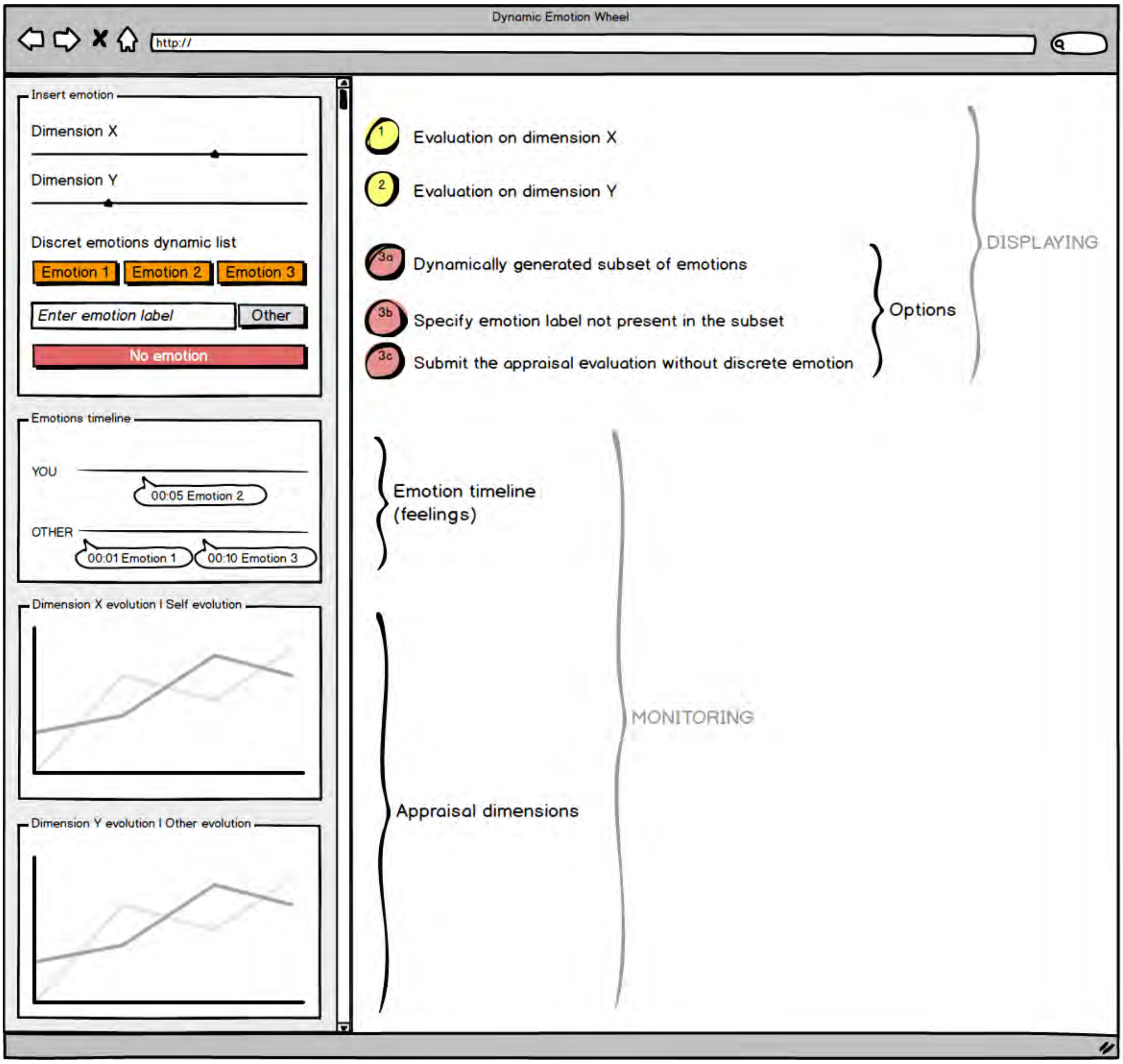


Figure 3.6. Low-fidelity wireframe of the entire layout of the Graphic User Interface.

3.3.1.1 Emotion Timeline for Subjective Feelings

For the subjective feelings, we created a widget named *Emotion Timeline*. The widget consists in two chronological timelines, the first belonging to the user, the second to his partner, where the discrete emotions aligned in a chronological order. To provide consistency with the combined chronology, the use of white-space margins allowed the synchronization of the two lines. Each line was intended to contain all the observations provided during collaboration, and therefore the widget must scroll on the horizontal axis. Each observation present: (1) the time of its manifestation, referring either to the time of the day or the elapsed time from the beginning of the task; and (2) the discrete-emotion label, either typed or clicked as a button.

3.3.1.2 Line charts for Appraisal Dimensions

The visualization of data – especially quantitative – is a subject about which cognitive sciences dispose of a thorough literature (e.g. Carpenter & Shah, 1998; Hegarty, 2011; Pinker, 1990). The basic principle concerning visualizing data is that "representations that are informationally equivalent (contain the same information) are not necessarily computationally equivalent" (Hegarty, 2011, p. 447). Although there is no golden rule about the best graphical form to use independently from the type of information presented and the planned use, some *design principles* may guarantee a better outcome. For instance, "visual displays can allow the offloading of cognitive processes onto perceptual processes" (Hegarty, 2011, p. 451).

For this reason, we decided to show appraisal dimensions values as line charts, which enhance the perception of trends over time. Considering that we disposed of quantitative data referring to two dimensions and provided by two different sources (i.e., participants), different type of visualizations were possible. The four lines could appear grouped on the same chart, coupled on two separated charts, or even separated on four charts. We speculated that two charts would strike a balance between the "offloading of cognitive processing" required to distinguish between the 4 lines, and the 4 separated charts that would use a lot of screen-space. The use of two charts also allows the combination of data in different ways.

1. Graphics can show data according to the underlying dimension (e.g. one graphic for Valence, one for Control), in which case users' values are compared on both charts.
2. Graphics can exhibit observations according to the participants, in which case a chart presents the two dimensions of the user, and the other the two dimensions of the partner.

In both cases, the principle aim of the chart is the perception of the variations in the evaluation of the situation. For this reason, we decided to keep the charts as simple as possible, providing contextual information (e.g. labels on the x or y-axis, legends, etc.) only if strictly necessary to understand what the chart intends to communicate.

3.3.2 Expected Interaction with the Tool

Once established the "static" layout of the tool, we focused on the expected interaction workflow, which refers to the way users handle the interacting elements of the interface. The tool allowed interaction is quite limited, since it consists in three steps, illustrated in **Figure 3.7**:

1. The user evaluates the situation/event moving the cursors of each slider on the left or on the right of a neutral point (i.e. a value of 0) until he judges the value proportional with respect to the minimal and maximal limits of the scale.
2. The user observes the computed subset of discrete emotions and assesses if any of them is suitable to communicate his feeling, in which case he clicks it. Otherwise, he can provide another emotion through a form-like element, or click on a "no emotion" button.
3. The user receives a confirmation of the performed action, which in the meantime has updated the monitoring information in the lower side of the screen.

The challenge of the interaction workflow does not concern, thus, the complexity of the demanded task, but rather the necessity of performing the action with the least possible effort. In other words, the tool must provide *micro-interactions* (Saffer, 2013) that facilitate the user's task: inserting new data, retrieving data meaningful for emotional awareness, and distinguish changes on the interfaces that signal new information available.

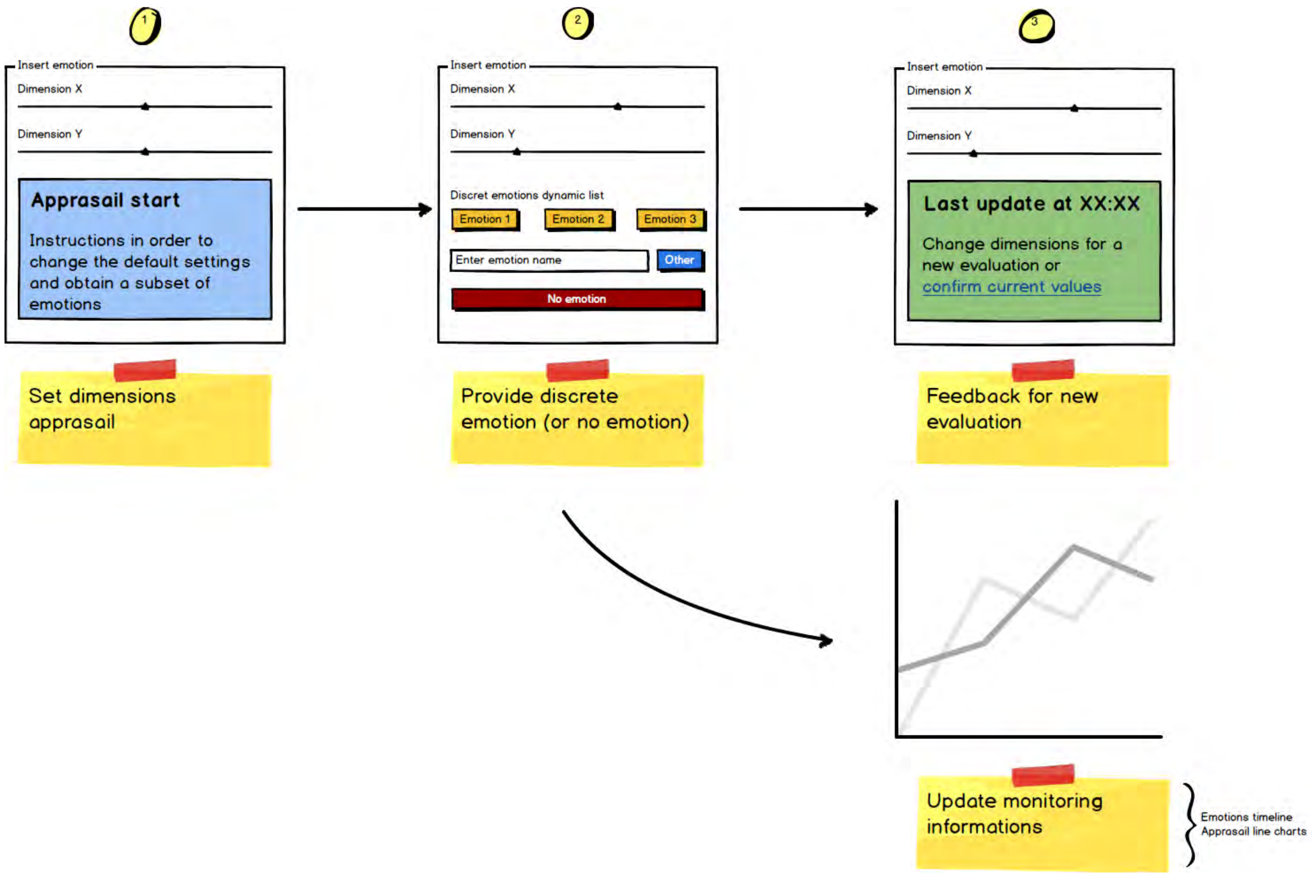


Figure 3.7. Expected interaction workflow.

3.3.2.1 How to Facilitate the Displaying Action

The most difficult challenge in asking people to assess their emotional state with respect to underlying dimensions is to find a meaningful way to translate the dimension in *everyday speech.* For instance, asking users "please, assess the valence of the event on a continuous scale between -100 and 100" would hardly facilitate the user introspection. In this regard, the GEW provides guidelines[5] to adapt the wheel to a different set of emotions. Among the options, the more economical consists in asking participants to judge the terms according to the Valence and Control dimensions. The following instruction is provided: "For each emotion word, please imagine that you are feeling the emotion or feeling referred to by that word and then rate how well each feature reflects how you would feel." Then two sliders are proposed:

1. For the Valence dimension, the slider displays on the left side *bad, unpleasant, negative* and on the right side *good, pleasant, positive*.
2. For the Control dimension, the slider displays on the left side *weak, powerless, low in control* and on the right side *strong, powerful, high in control*.

In the case of the DEW, the emotion word is given after the dimensional evaluation, so we adapted this instruction in a chronologically inverse perspective. We simply asked participants if they agreed on the following questions:

[5] Available at http://www.affective-sciences.org/gew

1. “Is the situation pleasant?”
2. “Is the situation under your control?”

In both cases, the slider presented, on the lower limit, the label “not at all”, and, on the upper limit, the label “yes, absolutely”. We decided to highlight the neutral point with an indicator, so that users were aware of the positive/negative distinction. Consistently with the SECs chronological order, we placed the Valence (i.e. *relevance* and *implications*) as the first slider, and the Control (i.e. *coping potential*) as second. We also preceded the question with the label 1 or 2 to enforce the expected order.

Below the sliders, another labelled instruction – with number 3 – asked users “Which emotion do you feel?”. Every time the users moved one of the sliders, the algorithm computed a subset of emotion-labelled buttons and ordered them with respect to the distance from the computed slope angle. The three buttons, appended under the instruction, changed every time the algorithm determined a different subset of buttons or a different order in the displayed subset. In the meantime, users could also provide an emotion that did not appear in the buttons. To speed up this action, the text input where users type the emotion also serve as an autocomplete dropdown list with all the emotions (see the interface on the right in **Figure 3.9**). To save vertical space, the other emotion form and the “no emotion” button were disposed on the same line. **Figure 3.8** illustrates the high-fidelity wireframe of the displaying interface (a) before and (b) after the choice of the subjective feeling.

To ease the use of this zone, the expected primary manipulations used the same orange color. The two questions for dimensional assessment were also preceded by the identification of the underlying dimension. The label was not intended to facilitate the comprehension of what was asked, but rather as a “priming” word for the line charts. In fact, the dimensional values needed a short term in the legend. If participants decided to provide an emotion through the “other emotion” text input, they also had to click on the “send” button to its right. Once they expressed an emotion – through either the button or the text input – the part of the screen representing the subjective feeling was replaced by a green alert informing participants that the emotion was saved (**Figure 3.8**b). The values on the dimensions did not return to the neutral point, but rather kept the same position for participants to have an indication about their last assessment. In this way, they could determine if, compared to the previous manifestation, their level of Valence or Control increased, decreased, or remained the same. In the last case, the feedback alert provided a link to “confirm the current values”, for the emotion button to re-appear even without moving the sliders. (This interaction could be useful when the same evaluation of the situation elicits two different emotions at the same time, so that users can express them in a row.) On the other hand, any variation on the slider will update the subset (as in **Figure 3.8**a).

Figure 3.8. Detail of the displaying zone of the DEW's interface. (a) On the left the displaying before the choice of the subjective feeling, and (b) on the right once the choice made.

3.3.2.2 How to Facilitate the Monitoring Action

Contrary to the displaying function, no direct manipulation involved neither the emotion timeline (except for scrolling back and forward), nor the line charts. We considered the objective of these zones fulfilled if: (1) the information provided was clear and easy to access; and (2) the user notice new information available, especially when it comes from his partner. For the first objective, we tried to use a combination of labels, fonts and colors that would facilitate the immediate perception of the *who*, the *what*, and the *when* of awareness, maintaining in the meantime the interface not too crowded.

For the second objective, we used small animations to highlight the appearance of new information into the interface. Human peripheral vision is particularly receptive to movements, and the EAT is not supposed to be the center of attention most of the time. Consequently, small animations can catch the attention in a non-obtrusive way. In this regard, we kept animations simple and limited to a short duration, otherwise they can distract users without conveying meaningful information, but just a graphical effect (Tversky, Morrison, & Betrancourt, 2002). When a new entry appeared:

1. The emotion timeline automatically scrolled to the right edge, for the last emotion word to be revealed.
2. Each line chart was implemented with a new segment, which quickly rotated for a few milliseconds before assuming the right position on the line. It is worth noting that: (a) the line charts contracted so that the entire evolution over time was visible without the need to scroll horizontally; and (b) since, for every observation, new data only concerned two lines out of four (i.e. Valence and Control of just one user), for the two other lines the last values were renovated – that is, a “flat” segment was added to keep the four lines of equal length.

Figure 3.9 illustrates two versions of the interface, on the right with charts regrouped by participants, and on the left regrouped by dimensions. The displaying zones show two different manipulations of the sliders to illustrate the dynamic mechanism; the interface on the left also shows the dropdown list with the remaining feelings.

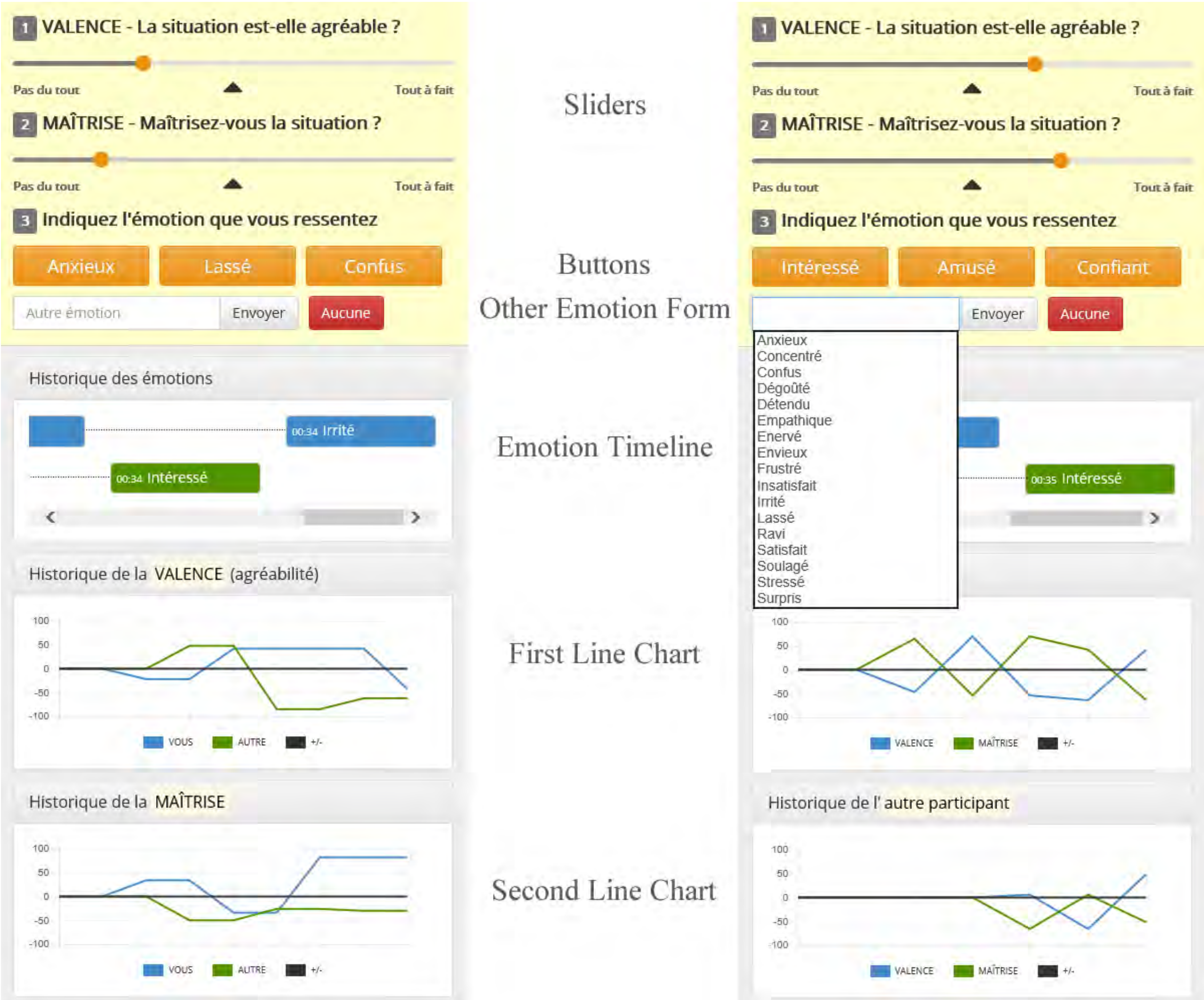


Figure 3.9. Two versions of the entire interface. On the left, the line charts are divided by dimensions; whereas on the right by participants. The right interface also shows the dropdown list with all the feelings. Please note that the two interfaces in this figure do not show neither the same, nor specular manipulations.

3.4 Architecture of the Application

In this section, we will describe the technologies used to develop the DEW. This part of the text will make extensive use of terms and abbreviation depicting technical features, for which we do not provide a thorough explanation, but only some references. Moreover, in the perspective of the iterative development process, we will limit the description to the general software architecture.

3.4.1 Underlying Technologies

Since one of the requirements concerned the adaptability to different experimental settings, we decided to develop the DEW according to two principles. First, experimenters should be able to adapt the interface to their needs. That implies, for instance, choosing a custom circumflex with different underlying dimensions and different discrete emotions; modifying the different labels on the interface; deciding which widget to show and in which configuration; etc. Second, the tool should maximize portability and independence from specific hardware or operating systems. We therefore developed the DEW as a web based application, referring to technologies such as HTML5 (MacDonald, 2013; W3C, 2014), CSS (W3C, 2015) and JavaScript

(Crockford, 2008). More specifically, we took advantage from HTML5 new features such as range input to provide the dimensional sliders for the interface, and web socket compatibility to synchronize data between the two members of the dyad. Web sockets have the advantage, compared to the HTTP protocol, to provide "full-duplex communication channels over a single TCP connection" (Wikipedia, 2015). In other words, it avoids the asynchronous practice to launch a request every few seconds to check for updates, since updates are automatically sent to the clients once they are detected by the socket. On the other hand, the shortcoming of web sockets concerns the fact that they are not available "out of the box" in every environment. However, relatively easy way to set up a web socket application are available, especially for the limited objective of experimental purposes. To minimize development time in the perspective of building a prototype for testing, we also used CSS and JavaScript open-source frameworks and libraries such as:

- Bootstrap (Otto & Thornton, 2015) and Font Awesome (Gandy, 2015) for CSS layout and some interactive features.
- JQuery (The jQuery Foundation, 2015) for interactive features.
- ChartJS (Downie, 2015) for line charts drawing and updating.
- AngularJS (Google, 2015) for the general architecture of the application (see following section).
- Deployd (Deployd, 2015), Firebase (Firebase Inc., 2015), or SailJS (McNeil, 2015) for socket connectivity and data saving. The three were tested as alternatives in the portability perspective.

3.4.2 The Model-View-Controller Pattern

To maximize portability and customization, we also decided to adopt a Model-View-Controller (MVC) design pattern (Leff & Rayfield, 2001). According to MVC practice, an application contains three interrelated layers:

1. The *Model* represents the underlying business logic of the application, for example with respect to data storage and retrieving.
2. The V*iew* represents the output of the application, which is what the user will see on the screen.
3. The C*ontroller* represents the *trait d'union* between the model and the view, updating data on the model based on the view manipulation, but also updating the view according to changes in the data.

Even though the MVC pattern was originally limited to the backend, and therefore computed on the server in the case of web applications, recent developments allow it at frontend-level as well. AngularJS (Google, 2015) is an example of JavaScript framework that computes the MVC architecture directly in the web browser.

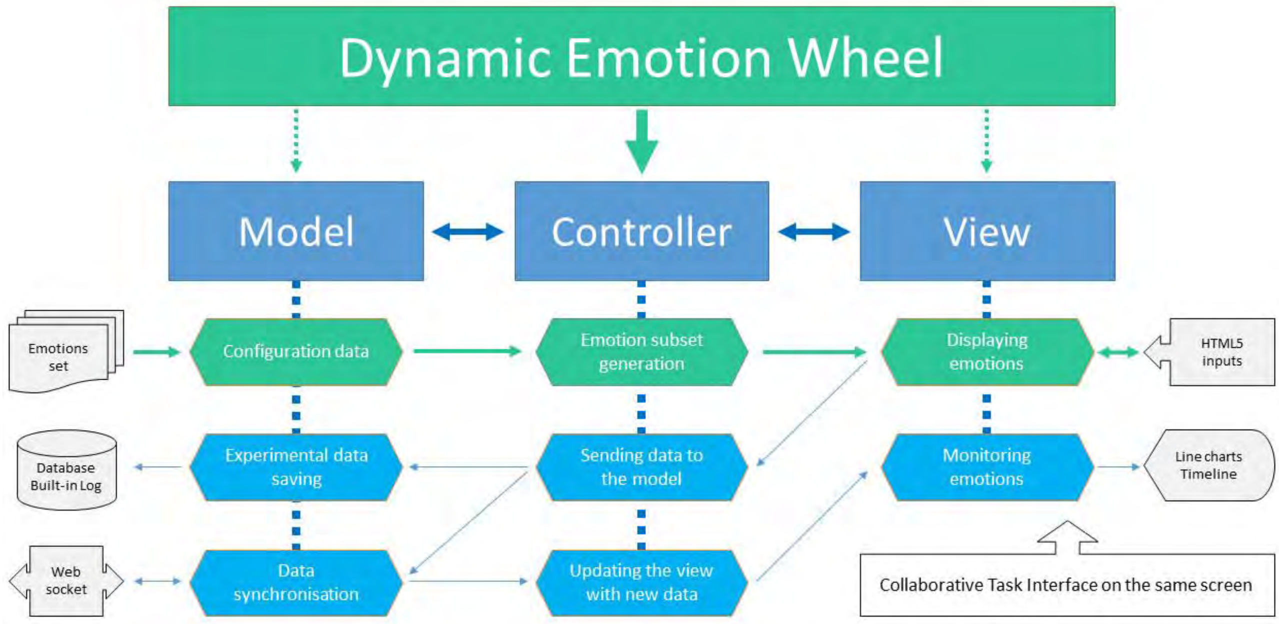


Figure 3.10. MVC architecture of the Dynamic Emotion Wheel.

Figure 3.10 illustrates the MVC architecture of the DEW application. The green boxes represent the distinctive features of the DEW, whereas the light-blue boxes refer to the necessary processes for the application to provide the expected behaviors. The core functions of the DEW architecture therefore relates to:

1. The possibility for the experimenter to configure different sets of discrete emotions organized in a circumflex (Model);
2. The possibility for the user to assess his emotional state on dimensional and discrete emotion approaches (View);
3. The possibility to correlate the dimensional and the discrete emotion approaches according to the user direct manipulation on the interface (Controller).

The rest of the architecture may vary without modifying the core principle of the DEW. Alternative graphical representations may replace the emotional timeline or the line charts, for instance because more than two users connect to the socket, and therefore the monitoring zone of the interface must represent the “overall” emotional state of a group.

3.5 What the DEW Measures

The different manipulations available through the DEW interface allow different measurements, both quantitative and qualitative. All measures are computed in real time and can be saved in different support such as database (using an API, for instance), log files on the server, or even a built-in log that registers the manipulation inside the session of the page. **Table 3.2** provides the list of measures with a brief description of what each measure represents. As highlighted by the requirements, the DEW provide dimensional measures with the two sliders (generally called X and Y, since they may be any appraisal dimensions) and the related subjective feeling. We considered the rest of the measures useful in experimental settings, even if some of them may be restrained to the purpose of the usability testing.

Table 3.2. Data measured by the DEW each time a participant submit an emotional episode.

Measure	Description
X	Value of the first slider that represents the horizontal dimension
Y	Value of the second slider that represents the vertical dimension
Last_dimension	Either the X or the Y value according to the last slider modified before sending data.
Slope_angle	Computed angle of the slope resulting from the ratio between the Y and the X values. It represents the bi-dimensional position of the evaluation of the situation on the underlying circumflex.
Slope_quadrant	Dividing the slope_angle by 90° and rounding the result to the upper integer provides in which of the four quadrants the computed radius crosses the circumflex.
Click	The action used by user to express the emotion. Possible values are "button", "other", or "none".
Emotion	The subjective feeling label of either the clicked button or the input text. The value is set to "none" if the "no emotion" button is clicked.
Emotion_angle	If the Emotion provided is part of the circumflex, its angle is retrieved and registered. The value is set to -1 if the emotion does not appear on the circumflex.
Emotion_quadrant	If the Emotion provided is part of the circumflex, the quadrant it belongs to is retrieved and registered. The value is set to -1 if the emotion does not appear on the circumflex.
Angle_diff	Absolute difference between the Slope_angle and the Emotion_angle. Due to the circularity problem of the circumflex, if the value is greater than 180°, the "reverse path" is calculated by subtracting the difference from 360. The value is set to -1 if the emotion does not appear on the circumflex.
Time	The absolute Unix time in milliseconds.
Elapsed_time	The elapsed time in milliseconds since the DEW is loaded.
User	A unique identifier for the user expressing the emotion.
Group	An identifier shared between two (and potentially more) users collaborating to the same task.

3.6 Requirements Check

We illustrated the comprehensive process of design and development of the Dynamic Emotion Wheel. Particular attention was drawn on the algorithm that combines dimensional and discrete emotion approaches, by computing a subset of adjacent emotions according to the evaluation of the situation/event on two-dimensional values such as Valence and Control. Compared to the first version of the EAT, the DEW introduces significant new features, which also requires, though, more manipulations and increases the amount of information on the interface. In particular, the DEW requires at least three manipulations to manifest an emotional episode, compared to a single manipulation of the first version. On the other hand, the increment in manipulation and information available derives from the requirement analysis. In this regard, we will briefly examine each requirement and illustrate to what extent the DEW fulfills it.

3.6.1 Theoretical Framework Requirement

The DEW is an adaptation of the Geneva Emotion Wheel, a tool deeply rooted in the Component Process Model theoretical framework and widely used as an emotion self-report tool in many contributions with different experimental settings. As an adaption of the GEW, the DEW manages to integrate the required quantitative measure, maintaining the usefulness of qualitative feelings. Notwithstanding, the DEW fails to adhere to the Control-Value Theory of Achievement Emotions suggested by the EATMINT project members. This is a deliberate choice deriving from the following considerations. First, since this is the first version of the DEW, we reckoned the advantage of *piecemeal engineering,* and – in contrast with the very title of this contribution – we did not intend to reinvent the wheel. The GEW served as a reference both for conceptual and testing purposes, and we think that the differentiation from a solid tool must be gradual and not abrupt. Second, the core principle of the DEW is independent from the circumflex and the underlying dimensions adopted, which leaves the possibility to implement a Control-Value perspective in the future. Third, we speculate that the difference between Control-Value and Control-Valence are not too deep, compared for instance to the Arousal dimension commonly used in dimensional spaces.

3.6.2 Cognitive Load Requirement

The number of concurrent emotion-buttons on the interface dropped from 20 to 3, even though all emotions are still available through the dropdown menu. If the dynamic generation of subset allows users to find the suitable feeling in the button, then the DEW succeeds in reducing the seek-and-evaluate task necessary in the first version of the EAT. On the other hand, the DEW introduces additional manipulations and information available on the interface, which increases the risk of cognitive overload. Without an empirical test, we are not able to determine if this requirement is fulfilled.

3.6.3 Inter- and Intra-Personal Comparison Requirement

Both the emotion timeline and the two line charts have been designed with the purpose of enhancing the comparability of emotional states, both at the intra-individual and inter-individual level. In this regard, though, we are not able to determine if these features facilitate the comparison without an empirical test. More specifically, we are unaware of which composition of line charts (the Dimension or the Participant division) is the most suitable.

3.6.4 Adaptability and Portability Requirement

As illustrated by the architecture of the application, both the technologies adopted and the MVC pattern should allow the adaptation of the DEW to different experimental settings and technical environments. Before developing a stable version of the tool, though, the result of the usability test shall be taken into account.

3.7 Summary

The development of a testable version of the DEW integrated a user-centered approach with development methodologies such as rapid and test driven development. The result is a highly iterative process that combines theoretical, technical and human-interaction principles in order to provide (a) users with a user-friendly tool through which displaying and monitoring emotional awareness, and (b) experimenters with a trustworthy tool for measuring emotions. To our knowledge, the Dynamic Emotion Wheel provides an innovative perspective in the field of affective computing, since it adapts the same emotion self-reporting principles of the Geneva Emotion Wheel to the purpose of emotional awareness. This adaptation mainly consists in two aspects.

1. The generation of a subset of discrete emotions based upon the users' evaluation of two appraisal dimensions, which combines quantitative and qualitative observations.
2. The same data – referring to the subjective feeling and appraisal components of the Component Process Model framework – is used for the monitoring function of the DEW, since the interface of the tool proposes:
    a. An emotion timeline for participants to dispose of the emotional meaning through natural language expression;
    b. Two line charts, for participants to dispose of the emotion through the cognitive evaluation of the situation according to appraisal criteria Valence and Control.

These distinctive features of the DEW are supposed to meet most of the requirements defined by the EATMINT project members. In order to establish if the DEW fulfills the expectations, though, a usability test is required.

4 Testing the Dynamic Emotion Wheel

In order to assess the utility and usability of the DEW, we planned to test it in a situation as close as possible to the experimental situation of the first EATMINT project empirical test. We wanted to combine a usability testing from the participant's point of view, and a utility testing from the experimenter's perspective. In other words, we needed a situation in which:

1. Participants engaged themselves in a real-time collaborative task, and manifested their emotional states throughout the activity;
2. Experimenters could have a sense of the accountability and interest of data gathered through the DEW.

To achieve this double result, we needed to strike a balance between the standardization of usability testing (i.e. exposing all participants to the same tasks) and the unpredictability of collaboration, where many variables interact at the same time (motivation of each participant, "alchemy" of the dyads, etc.). Consequently, we needed a task, which was *collaborative, but not too much* in order to expose participants to the same kind of stimuli. Since the test followed the protocol of a "real" experiment, we will present it accordingly. First, we will expose the objectives that we intend to assess in order to evaluate if the DEW fulfills the requirements. Second, we will describe the method we used to assess these objectives. Third, we will provide the results of the test. Fourth, these results will be discussed as a final assessment of the DEW, considering the limits of the test.

4.1 Objectives of the Test

As aforementioned, the DEW aims to accomplish two interrelated functions on the technical and theoretical level. The technical level concerns the reliability of the DEW as an Emotion Awareness Tool, whereas the theoretical level implies the evaluation of its use in a collaborative task, in order to assess if the information provided through the EAT is taken into account during the ongoing task.

4.1.1 Technical Objectives

From a technical standpoint, the DEW must provide participants with an intuitive and accessible interface through which they can share their emotions during a collaborative task. More specifically, the test must evaluate the DEW according to:

1. *Displaying*: the dynamic generation of a subset of emotions must integrate the cognitive evaluation of the situation (dimensional measure) with the subjective feeling (categorical measure) in a trustworthy representation of the participant's emotional state, which is useful for both participants and experimenters;
2. *Monitoring*: the graphic elements of the user interface must allow, in a clear and accessible manner, participants to perceive the information representing the emotional state (their own as well as their partner's ones), and notice the evolution during the task;
3. *Usability*: the tool must integrate both displaying and monitoring functions in a coherent way.

4.1.2 Theoretical Objectives

From a theoretical standpoint, the DEW must provide meaningful information that users would take into account during the collaborative task. More specifically, the test must evaluate the DEW according to:

1. *Emotion as information and interaction*: the tool must allow participants to reflect on their own emotions and the emotions of their partner. The information available through the interface must be instrumental to the collaborative task – that is, perceived as interesting and useful.
2. *Cognitive load*: the use of the tool must limit the mental effort and difficulty to use it, in order to avoid attention to be diverged from the collaborative task.

4.1.3 Comprehensive Assessment of the Objectives

Technical and theoretical functions intertwine, but present in the meantime some divergences that need a comprehensive assessment of the tool. For instance, the DEW could turn out to be a reliable instrument to display and monitor emotions, but its use too demanding in a context where the collaborative task is priority. Another aspect where technical and theoretical functions may diverge concerns the way by which gathered information is made available to participants. The DEW does not compute the data it saves, but rather shows exactly what it collects. Appraisal dimensions may be interesting as dimensional measures or as a way to compute a subset of emotions, but their representation in line charts could not be useful to participants. A goal-oriented approach is therefore necessary for a comprehensive assessment of the tool, combining objective measures gathered through the use of the tool in a collaborative task and subjective evaluation about its usability, interest and instrumentality.

4.2 Method

4.2.1 Participants and Design

16 university students (M=26,9, SD=6.1) coming from different disciplines both at graduate and undergraduate level participated to the test. In order to maintain consistency with the EATMINT experiment, both sexes were equally represented in the population: 8 women (M=24.4, SD=3.3) and 8 men (M=29.5, SD=7.4). Participants were not expert neither in the field of emotion psychology nor human-computer interaction. Two women nevertheless followed a specific course about emotion psychology in their bachelor program, and were thus familiar with concept such as Valence or Control. One man and one woman followed a specific course about human-computer interaction in their master program, and were thus familiar with concept such as usability and usability testing.

Two slightly different versions of the DEW interface were tested. Inside each group determined by the gender, participants were randomly assigned to a version where the two line charts of appraisal dimensions were grouped by Valence vs. Control (i.e. *Dimensions* version) or by Self vs. Other (i.e., *Participants* version). It is worth noting that the resulting repartition in groups does not constitute an intended experimental plan, but rather refers to:

1. A controlled variable with respect to emotional expression (Brody, 2000) in the case of gender: participants were assigned to a same-sex simulated collaborative partner during the task, even if they were induced to believe that the partner was real (more about this point in section 4.2.3.1).

2. An A/B test (Tullis & Albert, 2013) in the case of the interface type, since participants disposed of two slightly different versions of the tool.

In other words, these independent variables were chose in order to test different use contexts of the interface rather than their intrinsic effects. Furthermore, as suggested by Buder (2011), awareness tools effectiveness should not be tested comparing *with* and *without* the tool, but rather different tools or versions of the tool. In this perspective, the experimental plan (see **Table 4.1**) could also represent a pilot for future studies.

Table 4.1. Repartition of participants according to gender and interface tested.

		Gender		
		Male	Female	Total
Interface	Dimensions	4	4	8
	Participants	4	4	8
	Total	8	8	16

4.2.2 Experimental Task

In the experimental task, participants would collaborate to solve problems. Problem solving presents a series of advantages that fitted with the experimental needs to collaborate and elicit emotions. Problem solving is a cognitive process where emotions are known to play an important role (Belavkin, 2001). People usually get to a solution of a problem following a path that is not linear, but rather composed of trial-and-error pathways that are often disorganized and sudden (Metcalfe & Wiebe, 1987). This unpredictability is consistent with the idea of an ever-changing environment, and thus represented a potential source for emotional variety. Participant could in fact feel frustrated or ashamed if they cannot find the solution, happy or proud once they find it, and so on. The collaboration also added social-comparing emotions such as jealousy, shame, satisfaction or admiration if the other participant could solve a problem that the person herself could not achieve. Thus, the task could also create an interaction between the emotional states of the participants. Finally, solving a problem requires an effort that usually uses up the limited capabilities of the working memory. If we consider that participants are supposed to express their emotional feeling at the same time, we obtain a situation of dual-task (Pashler, 1994) that will assess at what extent the EAT interferes with the primary task.

Problem solving, though, also presented some drawbacks. For instance, enigmas represent more a competitive than a cooperative challenge, since a person possesses, at any moment, the “keys” to solve the problem on its own. We therefore decided to enhance collaboration in two ways. First, inspired by knowledge awareness, we thought that participants could share their reasoning for solving the problems. Second, a system of points calculated by dyad was intended to motivate participants to articulate their reasoning in order to maximize the likelihood that both members of the dyad picked the right answer. The joint problem-solving task presented thus a suitable setting for standardize the test: sharing the reasoning to solve a problem maintains the task mainly individual, but integrates it in a collaborative perspective.

4.2.2.1 Joint Problem-Solving Task

The experimental task consisted in the resolution of 4 enigmas taken from a problem solving game box. Participants accessed a web page that presented two areas (**Figure 4.1**). The DEW occupied the left side of the screen, for participants to dispose of emotional awareness during the task. The right side of the screen proposed the elements for the problem-solving task:

- A progress bar, used as a timer, indicated the time with respect to three phases: a first phase to read the enigma (40 seconds); a second phase to solve it (3 minutes and 40 seconds); and a last phase to read the solution (1 minute).
- The text of the problem was provided in a light-blue box under the progress bar.
- Interactive elements to solve the problem, which were visible only during the second phase:
    - One text-area where participants could write the reasoning they thought suitable to solve the problem;
    - One text-area where participants could read their partner's reasoning;
    - One text-input where participants could write the solution to the problem.
- The solution of the partner and the expected solution were only visible during the third phase.

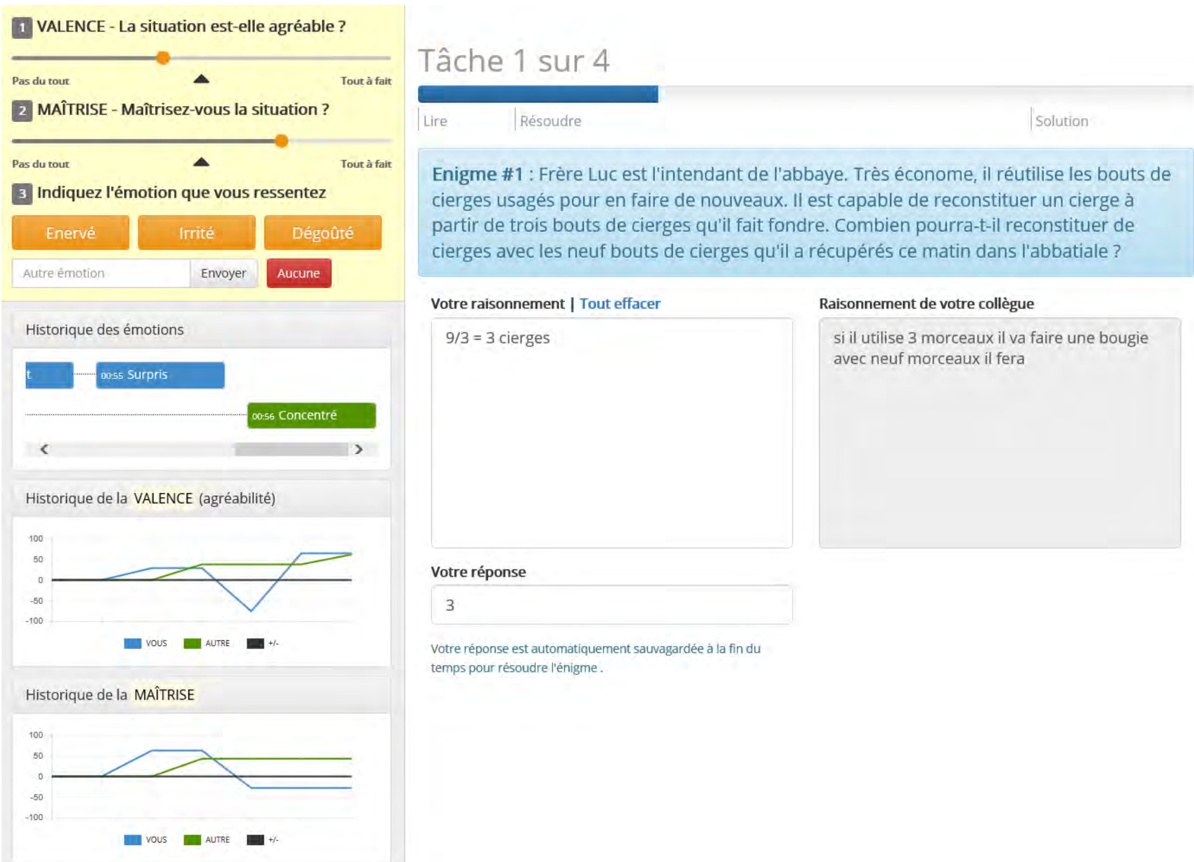


Figure 4.1. Screenshot of the experimental task. On the left side of the screen, participants dispose of emotional awareness through the DEW. On the right, the interactive elements to solve the problem.

4.2.2.2 Motivating the Collaboration

In order to motivate participants to collaborate and share their reasoning, a system of rewards was introduced. Participants would receive:

- 1 point for each right response in the dyad
- 0 point for each response not given in the dyad
- -1 point for each wrong response in the dyad

Consequently, participants should be encouraged not only to find the right solution, but also to express a convincing reasoning for their partner to reach the same conclusion, in which case the dyad would earn two points.

4.2.2.3 Limiting the Collaboration to Reasoning

In the meantime, participants received precise instructions to avoid direct communication with their partner. They were allowed only to express the logical passages of their reasoning through the corresponding text-area (a primitive form of knowledge awareness), but they could not "chat" with the partner. More specifically, they were instructed to avoid any kind of direct communication such as questions (e.g. "Is that right?") or comment about the other person's reasoning (e.g. "What you are doing is false"). This instruction was mandatory in order to introduce a simulated-participant (see 4.2.3.1), which could not reply to direct stimulation.

4.2.2.4 Manifestation of the Emotional State

Participants were also instructed to express their emotional state throughout the task. They were invited to provide at least one emotion for each of the three phases of the enigmas – that is, one during the reading, one during the solving, and one while the solutions were displayed. They could certainly provide more emotions if they wanted to, and they received precise indication that everything could be the source of emotion. In this regard, they did not have to justify the emotion they felt. For the purpose of the test, the DEW was adapted with respect to two aspects. First, the dropdown list showed all the emotions except those currently available through the buttons. In this way, each emotion appeared just in one place: either in the subset of buttons or in the "other-emotion" form. Second, a nonintrusive alert (positioned under the enigma) reminded participants to express their emotions after one minute from the last emotion provided.

4.2.3 Material

4.2.3.1 Simulated Partner

In order to standardize the test, we created a situation in which all participants were assigned to a simulated-partner of the same sex. Since the aim of the test is to evaluate the use of the DEW in an ongoing collaborative task, participants should be induced to believe that they were truly collaborating. To maximize this illusion, 4 fellow colleagues (2 men and 2 women) participated to a pre-experiment that consisted in the very same experimental task imagined for the test. The two dyads were actually collaborating in a synchronous setting, and their manipulation on the interface was registered. More specifically, we saved: (a) all the emotions manifested through the interface with respect to the two appraisal dimensions and the corresponding subjective feeling; (b) all the reasoning provided through the corresponding text-area element of the task interface; and (c) all the replies to the enigmas.

The elapsed time from the beginning of the task was attached to all manipulations. In this way, the gathered manipulations could represent the actions of a simulated participant synchronized with the participant to the test. We randomly chose the registered manipulations of one member of each dyad, and injected them, at the very same time they occurred, in the interface. Female participants accessed the manipulations of the female colleague, whereas male participants the ones of the men-dyad member's. Since the instructions for the experimental task stressed about the prohibition to communicate directly with the partner – an instruction also given in the pre-experiment – the intended result was a sort of delayed synchronicity.

4.2.3.2 Adapted Wheel of Emotions

To maintain consistency with the EATMINT first experiment, we decided to keep the same set of 20 discrete emotions. We therefore needed a way to position them in a circumflex for the dynamic generation mechanism to be effective. For the purpose of the usability test, we adopted a "rule of the thumb" process and combined different sources to assess the position of the 20 emotions on a valence-per-control circumflex. We looked for correspondences in the three versions of the GEW (Scherer, Shuman, et al., 2013), results of the GRID study (Fontaine & Scherer, 2013, p. 115 Figure 7.1), and in the alternative dimensional structures of the semantic space for emotions proposed by Scherer (2005, p. 720, Figure 1). Since we did not find any corresponding emotion for the "Reconaissant" (i.e. grateful) emotion, and during the disposition of the emotions, we felt short of a negative-valence/positive-control emotion, we introduced the "Dégoûté" (i.e. disgusted) label instead. **Figure 4.2** illustrates the final dispositions of the 20 emotions according to Valence and Control, and the corresponding quadrants of the circumflex.

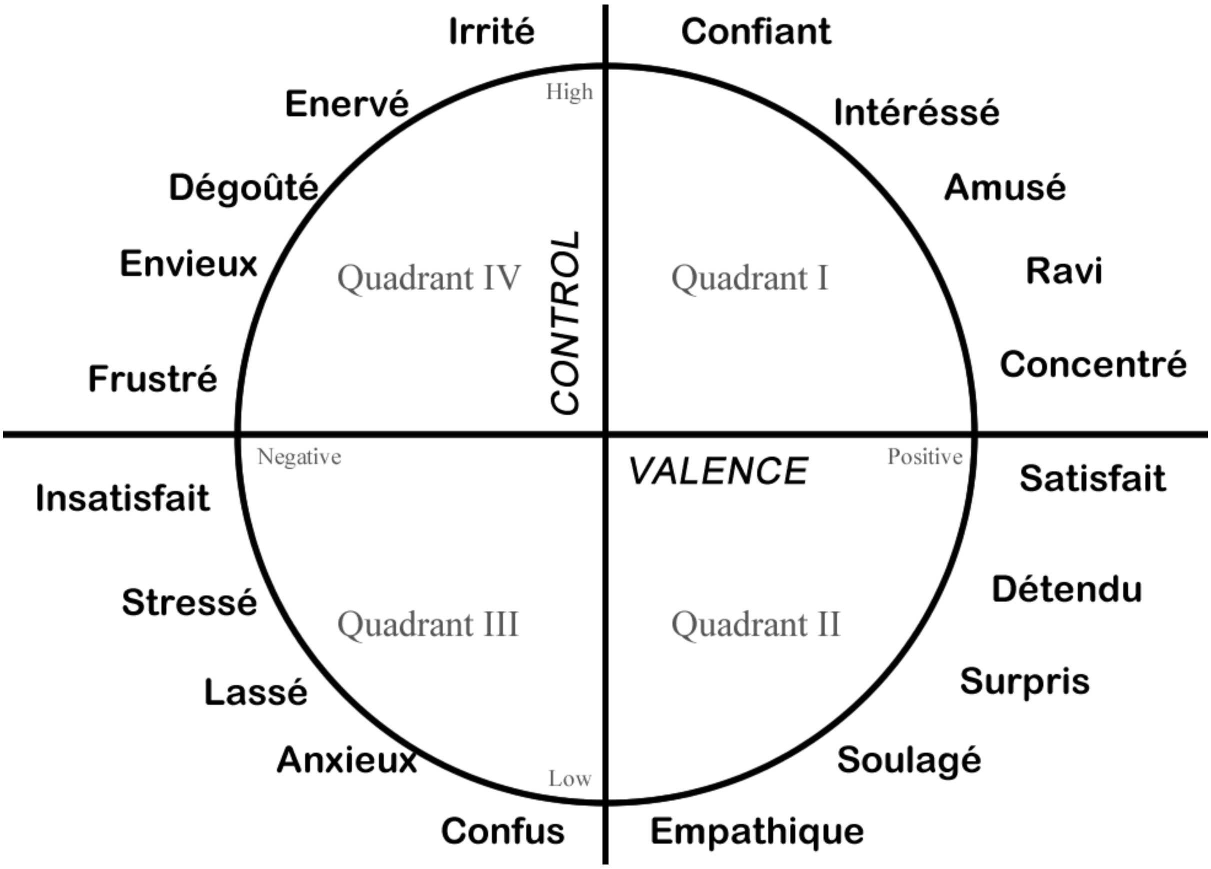


Figure 4.2. Adaptation to a valence-per-control circumflex of the 20 EATMINT discrete emotions. *Reconnaissant* (grateful) was replaced by *Dégoûté* (disgusted). English translation in Table 1.1, p. 7.

We are certainly aware that this “rule of the thumb” proceeding is not accurate enough for an empirical experiment where the influence of specific emotions is to be determined. For a more accurate disposition of emotions according to a two-dimension circumflex, we suggest the use of the CoreGRID or MiniGRID system (Scherer, Fontaine, et al., 2013), or the instruction provided by the GEW official material (*op. cit.*).

4.2.3.3 Eye Tracking and Computer Equipment

Eye tracking (e.g. Jacob & Karn, 2003; Poole & Ball, 2005) is a method of measuring the user experience that is increasingly being used in human-computer interaction. It refers to a “technique whereby an individual’s eye movements are measured so that the researcher knows both where a person is looking at any given time and the sequence in which their eyes are shifting from one location to another” (Poole & Ball, 2005, p. 211). With recent development in eye tracking equipment, this technique has become noninvasive, meaning that participants to a usability test involving eye tracking would hardly notice it. This technique provides meaningful information that are not available through simple self-report, since participants tend to under- or over-estimate the attention they really paid to particular information on the screen, or neglect to mention it altogether (Guan, Lee, Cuddihy, & Ramey, 2006; William Albert, 2010). Eye tracking provides different kind of metrics, among which the more frequently assessed are:

1. Fixation, which refers to “[a] relatively stable eye-in-head position within some threshold of dispersion (typically ~2°) over some minimum duration (typically 100–200 ms), and with a velocity below some threshold (typically 15–100 degrees per second)” (Jacob & Karn, 2003, p. 581). Longer fixation times may indicate either interest for the target, or difficulty to process the information (hence the necessity to look at it longer).
2. Saccades are eye movements between fixations, and they serve to move the attention across different zones of the screen. Saccades are usually indicators of research, or, in case of “backward saccades”, of complexity and misunderstanding.
3. Areas of interest are zones of the screen that the experimenters consider of particular importance for their analysis, and are therefore useful to capture movements specifically on these “targets”.

For this test, we used the Tobii T120 Eye Tracker, a device where the eye-movements sensor is directly embedded in a computer screen with a resolution of 1280x1024 pixels. The Tobii Studio software version 3.3.1 (Tobii AB, 2015) served to set up the test and collect eye-movements data. The eye-tracking procedure consisted in two steps. First, participants read a short text with the instructions for the eye-tracking phase, reminding them to keep their head in a steady position. Once they were ready, they clicked on a button and a web page with the application to perform the test was automatically launched in Internet Explorer 11 (Tobii Studio requires IE as default browser). We defined four areas of interest for our test, each of them on the DEW: (a) the entire displaying zone; (b) the emotion timeline; (c) the first appraisal line chart; and (d) the second appraisal line chart. The Tobii T120 was connected directly to the notebook of the experimenter (provided of external keyboard and mouse), where the web application was launched on a local webserver to minimize loading time.

4.2.3.4 Post-test Questionnaire

In order to assess the subjective perception of the various functions of the tool, we asked participant to fill a post-test paper survey (see 7.1, p. 96). The questionnaire presented a series of questions grouped in four main concepts of interest.

The perception of the various zones of the interface. A series of items, inspired by the System Usability Scale (Brooke, 1996), asked participants to evaluate their agreement on a 1 to 7 scale (1 = "total disagreement" and 7 = "total agreement"). Items were grouped in 4 blocks. The first block presented 13 questions about the displaying zone of the DEW, including the two-dimensional sliders, the generated buttons and the "other-emotion" form. The three following blocks concerned the 3 monitoring zones (i.e. the emotional timeline and the two line chart graphics) and repeated three times the same six questions.

The perception of the cognitive load, difficulty and *interest* related to (a) the problem solving task; (b) the displaying/monitoring activities during the collaborative task; and (c) the experiment as a whole. Consequently, the block proposed 9 items on a 1 to 7 scale (1 = "very low" and 7 = "very high").

The integration of emotion awareness during the collaborative task. Two blocks of 7 items – issued from a post-test questionnaire used to assess the correlation between self-reported emotions and perception of the interaction (Molinari, Chanel, et al., 2013, p. 6) – proposed the same statements about the frequency of some emotion-related behaviors on a 1-to-7 scale (1 = "Very infrequently", and 7 = "Very frequently"). In the first block, participants should assess their own behaviors, such as how much time they have spent in order to understand their partner's emotions. In the second block, on the other hand, participants were asked to imagine their partner's behavior, and the items were modified accordingly. For instance, the same item asked participants to evaluate, in their opinion, how much time the partner spent to understand the emotions of the participant himself.

The perceived usability of the DEW. The System Usability Scale (SUS, Brooke, 1996) was adapted for items to propose "tool" instead of "system". The SUS is a reliable tool to measure user experience (Tullis & Stetson, 2004). The scale presents 10 items on a 1 to 5 scale (1 = "total disagreement" and 5 = "total agreement")[6]. The items coding depends on their position in the list: even items are coded on a 0-to-4 points from left to right, whereas odd items are coded right to left. The sum of the points is multiplied by a factor of 2.5 to represent a percentage. A score below 50 is considered non acceptable, between 50 and 70 as marginal, and above 70 as acceptable (Bangor, Kortum, & Miller, 2009). The questionnaire also proposed two open questions that invited participants to express the strong and weak points of the DEW.

4.2.4 Procedure

All the experiments took place in the same room where the eye tracking system was available and followed the same steps. *Introduction* (5 minutes): the experimenter introduced the research global perspective (i.e.

[6] Due to a printing error, the SUS scale in the post-questionnaire proposed a 1-to-7 scale instead of 1-to-5. In the analysis of the results, the extremes of the scale have therefore been regrouped: 1 and 2 recoded to 1; 6 and 7 recoded to 5; and the resting figures decreased by one unit.

studying the influence of emotions on computer-supported collaboration) and then informed participants about the specific interest of the tool (i.e. usability test of an Emotion Awareness Tool). Participants were then invited to read and sign the consent form, where the guidelines of the experiment were reported.

Presentation of the task (5 minutes): participants then moved in front of the Tobii screen, where the experimenter projected a power-point presentation, which explained the experimental task. At this stage, participants were informed that the test included the joint problem-solving task, and were induced to believe that a same-sex participant was undergoing the same procedure in another university building. To enforce the illusion, the experimenter occasionally consulted his cellphone pretending to check the synchronization. During the presentation, participants received instructions about (a) the problem-solving task, and (b) the expected use of the EAT. Particular attention was drawn about the prohibition of direct chatting with the partner, since that would spoil the illusion of a real collaboration.

Practice with the DEW (5 minutes): participants could at this point directly manipulate a version of the tool that presented the same interface of the subsequent test. In the practice-version of the tool, the text of enigmas was missing and random emotions where injected in the DEW to give participants an overview of the general functioning. The experimenter showed the main features of the tool, especially with respect to the displaying zone of the interface. Subsequently, participants could manipulate the two sliders and see how emotion-labelled buttons updated accordingly. They were also informed about the possibility to find an emotion in the dropdown list, or to write an emotion that did not appear either in the buttons or in the dropdown menu. The instructions of the experimenter where nevertheless not too specific, since the aim of the test was to evaluate the "intuitive" use of the DEW, without a thoroughly explanation.

Test (25 minutes): once participants were ready to begin, the Tobii Studio software was launched and the calibration of the eye tracking took place. Participants received a five-digit code that they had to insert in a login-like form at the beginning of the test. This code was linked to the type of interface to be tested (Dimensions vs. Participants) and to the gender of the participant (Man vs. Woman in order to inject the corresponding manipulations gathered through the pre-experiment. Once the code provided, a "synchronization in progress" screen remained visible for about 20 seconds as a way to enforce the illusion of a real partner undergoing the same steps. The experimental task could eventually begin and last exactly 20 minutes (5 minutes for each of the four enigmas).

Post-test questionnaire and debriefing (15 minutes): at the end of the test, participants filled the paper survey. Since some of the questions referred to zones of the interfaces, participants also disposed of a paper with the screenshot of the interface used during the text. Once the questionnaire completed, the experimenter asked participants a general assessment of the experiment and some open questions about the DEW. Finally, the experimenter informed participants about the simulated partner, and provided explanations about the need to fake a collaboration, which related to the standardization of stimuli appearing on the interface in order to compare eye-tracking data.

4.3 Hypotheses

We define hypotheses with respect to (1) the displaying function of the tool; (2) the monitoring function of the tool; and (3) the overall function of the tool.

4.3.1 Test of the Displaying Function

The displaying function represents the core of the DEW itself, since the computational integration of dimensional and categorical measures is the innovating features of the tool. The test should therefore assess if the dynamic generation of a subset of emotions, according to the Valence and Control perceived over the situation, provides users with a useful and accessible way to express their emotion. In order to evaluate it, we postulate that:

A. Since the combined evaluation of the valence and the control perceived over the situation allows for the differentiation of the emotional response, we expect:
    1. The cognitive evaluation of the overall task to be independent from the predefined stimuli/event with respect to appraisal theories – that is, we do not expect to find a *collective appraisal pattern*.
    2. The two-dimensional values to be independent.
    3. The computed slope's angle to correlate with the subjective feeling's angle, and the *margin of error* between the two angles to be homogenous across the quadrants.
    4. The observed Valence and Control means associated to subjective feelings to have the corresponding positive and/or negative values with respect to the position of the emotion on the circumflex.
    5. The evaluation order of the dimensions to be respected (i.e. first Valence, then Control)

B. Participant will benefit from the dynamic generation of a subset of discrete emotions, since we expected them to find the word that best represent their subjective feeling more often in the three generated buttons than in the "other emotion" dropdown list.

C. Since the displaying zone of the interface does not provide new information, but just recombine the same according to the user manipulation, we expect users' fixation time to represent how long it takes to express an emotion. We expect this time to be in the order of a few seconds. Furthermore, we will also intercept the number of clicks inside the displaying area of interest. Since, potentially, users need 3 clicks to express an emotion in the quickest combination (one click for each slider and one for the button), we expect the average number of clicks to be close to 3.

D. We expect users to judge the displaying zone of the DEW as a useful, intuitive and accessible interface to share their emotional state in a satisfying and reliable way.

4.3.2 Test of the Monitoring Function

The monitoring function of the DEW is tightly related to the displaying function: the kind of data visualized depends upon the kind of data collected. Consequently, as a first evaluation of the tool, the test should foremost assess if the data gathered through the displaying function is also interesting for the monitoring function. In other words, the test should evaluate if the combination of appraisal dimensions and subjective feeling represents a suitable form to present information. The test of monitoring function should therefore assess if:

E. Considering that the DEW introduced subjective feeling as the most meaningful form of emotional communication, we expect participants to look at the emotion timeline more often compared to the appraisal line charts.

F. Users perceive the evolution of the appraisals dimensions depending on to the way data is presented. More specifically, we expect differences if the two line charts group the cognitive evaluation by dimensions (i.e. one graphic for the Valence, one graphic for the Control) or by members of the dyad (i.e. one graphic for the participant himself, one graphic for his partner).
G. The subjective feeling and the appraisal dimensions zones' utility differ depending on the objectives of the monitoring function. We expect feelings to be more useful in understanding and compare the emotional state of participants, and appraisal dimensions to be more useful for following the evolution in time.

4.3.3 Test of the Overall Function

The overall function of the DEW consists in providing emotional awareness to participants in a collaborative task, which means that the information provided should be (1) instrumental to the task, and (2) accessory to the task. We therefore postulate that:

H. Users considered emotional awareness information provided by the interface, which means that their attention was focused on the EAT during the task, but for a limited time compared to the overall task. We expect the time users spent on the DEW – observing and manipulating – between 10 and 15% of the total time of the task.
I. The use of the DEW does not increase either the mental effort or the difficulty of task, but rather increases the interest.
J. Users globally assess the usability of the tool as acceptable, which – according to Bangor and collaborators (2009) – requires a score above 70 on the SUS.

4.4 Results

All 16 tests were successfully completed and provided exploitable data with respect to the experimental task. Two participants, though, encountered technical problems with the web page used to test the DEW. In the case of the participant identified with the code 16146, the dropdown menu of the "other-emotion" form stopped to work around 10 minutes into the task. Consequently, the participant could not choose from the list anymore, but would have to type even the emotions included in the EATMINT set. This drawback has most certainly influenced the fact that this user was the only one to use only the generated buttons. Since he noticed the problem, it is very likely that he was about to use the feature and could not, even though that occurred in the middle of the task, suggesting that for at least 10 minutes he found a suitable emotion within the buttons. The participant with code 21713, on the other hand, could not get rid of the contextual menu of Internet Explorer (the one appearing on a web page clicking on the right button of the mouse). Fortunately, the menu covered a peripheral zone of the screen and only impaired the participants to read the expected solution to an enigma.

With respect to the eye tracking, 2 participants provided low quality recordings, calculated using the number of eye tracking samples that are correctly identified. One participant scored 26% and the other 36%. Fortunately, they belonged to two different groups, so we excluded the lowest sampling score for each of the 4 groups, keeping 12 participants with a sampling score between 68% and 89%. **Figure 4.3** shows a general overview of eye tracking measures using a heat map of the entire interface used during the experimental task (even though the image shows a Dimensions-like interface, fixations concern both the interface types). “A heat map uses different colors to show the number of fixations participants made in certain areas of the image or for how long they fixated within that area. Red usually indicates the highest number of fixations or the longest time, and green the least, with varying levels in between” (Tobii AB, 2015, p. 65). In this particular case, red areas indicate the zones of the interface where the attention of the 12 participants focalized for longer periods. The range of the maps goes from 0 seconds (transparent) to 50 seconds (red). A first assessment of the eye tracking suggests an unequal distribution of attention across the DEW: the displaying zone and, to a lesser extent, the emotion timeline received more attention than the line charts.

Figure 4.3. Eye tracking collective heat map representing how long participants (N=12) fixated within that area. The range of the maps goes from 0 seconds (transparent) to 50 seconds (red). Even though the image shows a Dimensions-like interface, fixations concern both the interface types.

The combination of objective (manipulation and eye tracking) and subjective (post-test questionnaire) measures resulted in a substantial amount of data. Data gathered through the DEW and the post-test survey was analyzed using STATISTICA (version 12), whereas the eye tracking data with Tobii Studio (version 3.3.1) for the general extraction, and again STATISTICA for specific analysis. We will present the results according to the triple function of the DEW also used to state the hypotheses: (1) the displaying function of the tool; (2) the monitoring function of the tool; and (3) the overall function of the tool.

4.4.1 Assessment of the Displaying Function

Participants expressed 301 emotions in total: 166 clicking on one of the generated buttons, 131 through the "other emotion" form, and only 4 by clicking the "no emotion" button (which happened by accident, as the participants reported in the debriefing). 292 emotions appeared in the EATMINT 20-emotion list, and only 9 were typed in by participants: *triste* (sad), *curieux* (curious), *intrigué* (intrigued), *honteux* (ashamed), *collaboratif* (collaborative), *décontenancé* (disconcerted), *déçu* (very close to disappointed, which was in the list), and two times *etonné* (astonished). Due to the small number, we decided to exclude from the analysis the 4 "no emotion" cases, and the 9 emotions typed by participants, leaving a sample of 288 emotions in total. All emotions in the circumflex where selected between 4 (*relieved*) and 50 (*confused*) times. With respect to gender differences in emotional expression, a *t* test showed that men and women did not differ in the number of emotions expressed: 17.3 on average for women, 18.6 for men ($t(15) = 1.24$, $p > .05$). For this reasons, we will not assess all analyses according to gender, but only occasionally present significant differences.

4.4.1.1 Hypotheses A.1: Subjective Appraisal of the Situation Independent from the Events

In order to assess the use of the two-dimensional sliders, we first checked if – consistently with appraisal theories of emotions – participants evaluate similar events in different ways. Since all men on the one side, and all women on the other were exposed to the same stimuli on screen, if a direct link exists between the event and the emergent emotions, we would expect a "collective emotional patterning" – in which case the interest of measuring the appraisal would be useless, since it could be predicted by the event. Results showed that, for instance, men and women assessed both dimensions in different ways: women evaluated the task in general to be less pleasant and they felt less in control. Women's mean in Valence equals 7.6, against men's 25.1 ($t(286) = -3.18$, $p < .05$); and women's mean in Control equals -21.6, against men's 6.2 ($t(286) = -4.32$, $p < .05$). Furthermore, we used the elapsed seconds from the beginning of the task to find a possible correlation with both dimensions (i.e., if participants made similar assessment at specific moments of the task). Results show no correlations for women neither for Valence ($r = 0.12$, $p > .05$) nor Control ($r = -0.15$, $p > .05$); whereas for men no correlation for Valence ($r = -0.08$, $p > .05$) and a weak correlation for Control ($r = -0.30$, $p < .05$). Results confirm the interest of having an appraisal evaluation of events, since participants evaluate them in different ways.

4.4.1.2 Hypotheses A.2: Independent Assessment of the Two-Dimensional Sliders

Once excluded the hypotheses of a collective appraisal pattern, we shall expect the two dimensions to be independent, since the combination of Valence and Control allows for the discrimination of emotions on the entire circumflex. Contrary to our expectations, we found a general positive correlation between Valence and Control ($r = 0.54$, $p < .05$). In this regard, if Valence and Control positively correlate, we shall expect the computed slope's angles to belong more often to the first and third quadrants – that is when both dimensions are either positive or negative. We therefore counted the number of times that the computed slope's angle belong to each of the four quadrants, and crosschecked this count with the number of times that the feeling's angle belong to each quadrant. Results, showed in the histogram in **Figure 4.4**, suggest that (1) the correlation between the two dimensions results from an underrepresentation of the fourth quadrant, and (2) the second and fourth quadrants present a discrepancy between appraisal and subjective feeling.

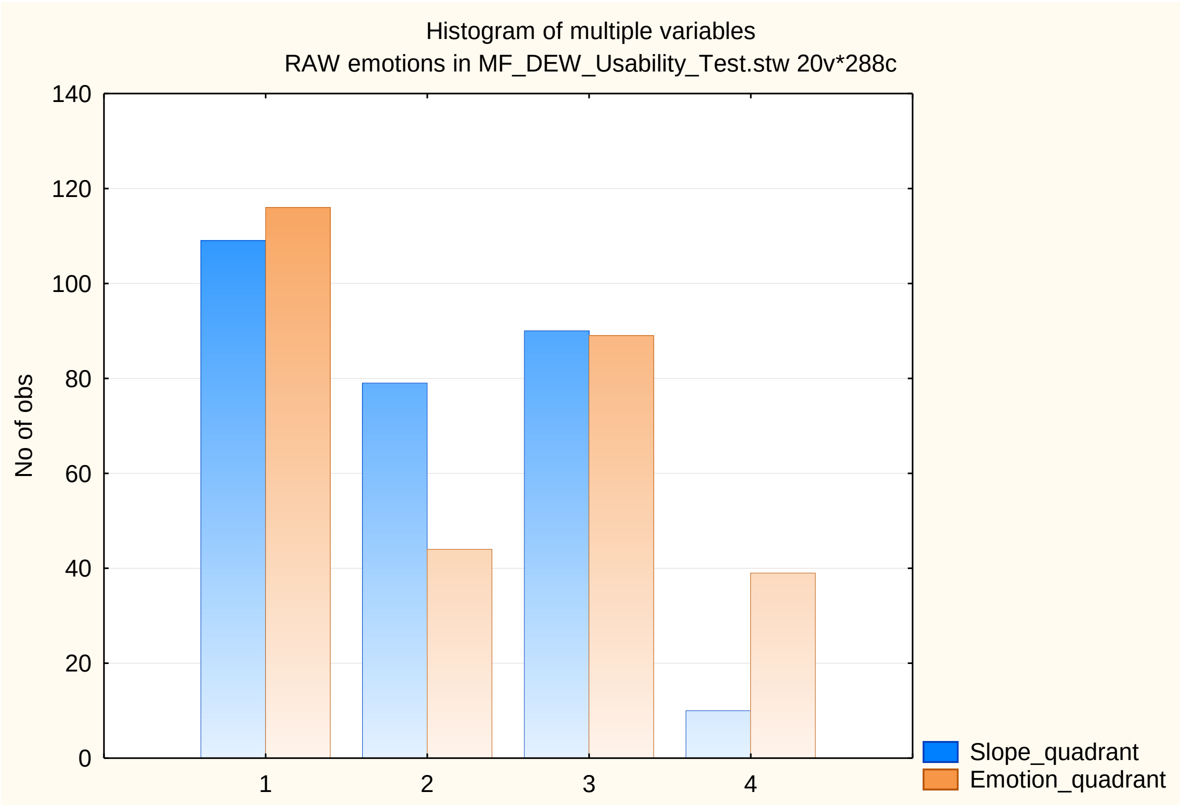


Figure 4.4. Histogram of the number of observations for each quadrant of the circumflex, based upon the computed slope' angle and the feeling's angle. N=288 for slope's and feeling's quadrants (see Figure 4.2 for the disposition of emotions in the quadrants**)**.

4.4.1.3 Hypotheses A.3: Difference between Appraisal Dimensions and Subjective Feeling

Results of hypotheses A2 suggest heterogeneity among the four quadrants in terms of the relationship between appraisal and subjective feeling. We therefore tested if the slope's angle computed by the combination of the sliders' values correlates with the angle given by the position of the subjective feeling on the circumflex. Hypothetically, the correlation should be close to 1, meaning that the cognitive evaluation of the situation perfectly matches the subjective feeling. Results, though, show a correlation of only $r = 0.54$ ($p < .05$). We therefore tested if the difference between the two angles was homogenous across the quadrants, in order to assess if all the possible combinations of Valence and Control contributed in similar ways to weaken the correlation. The distribution of the difference, though, was skewed on the left, since the difference between the two angles was defined in absolute numbers. We therefore transformed the data using a square-root transformation, obtaining a more normal distribution – necessary considering the unequal sample sizes (Howell, 2010) – that we renamed *margin of error*. A one-way analysis of variance revealed that there is a significant difference among the means of the four quadrants ($F(3,284) = 15.72$, $p < .05$). Even without the use of contrasts, the resulting means (illustrated in **Figure 4.5**) suggest that the fourth quadrant present the biggest margin of error. Hypotheses A3 is therefore partially rejected: if, on the one hand, the computed slope's and the feeling's angles correlate, this correlation is weakened by an uneven distribution of the margin of error, which is significantly more elevate for emotions belonging to the fourth quadrant.

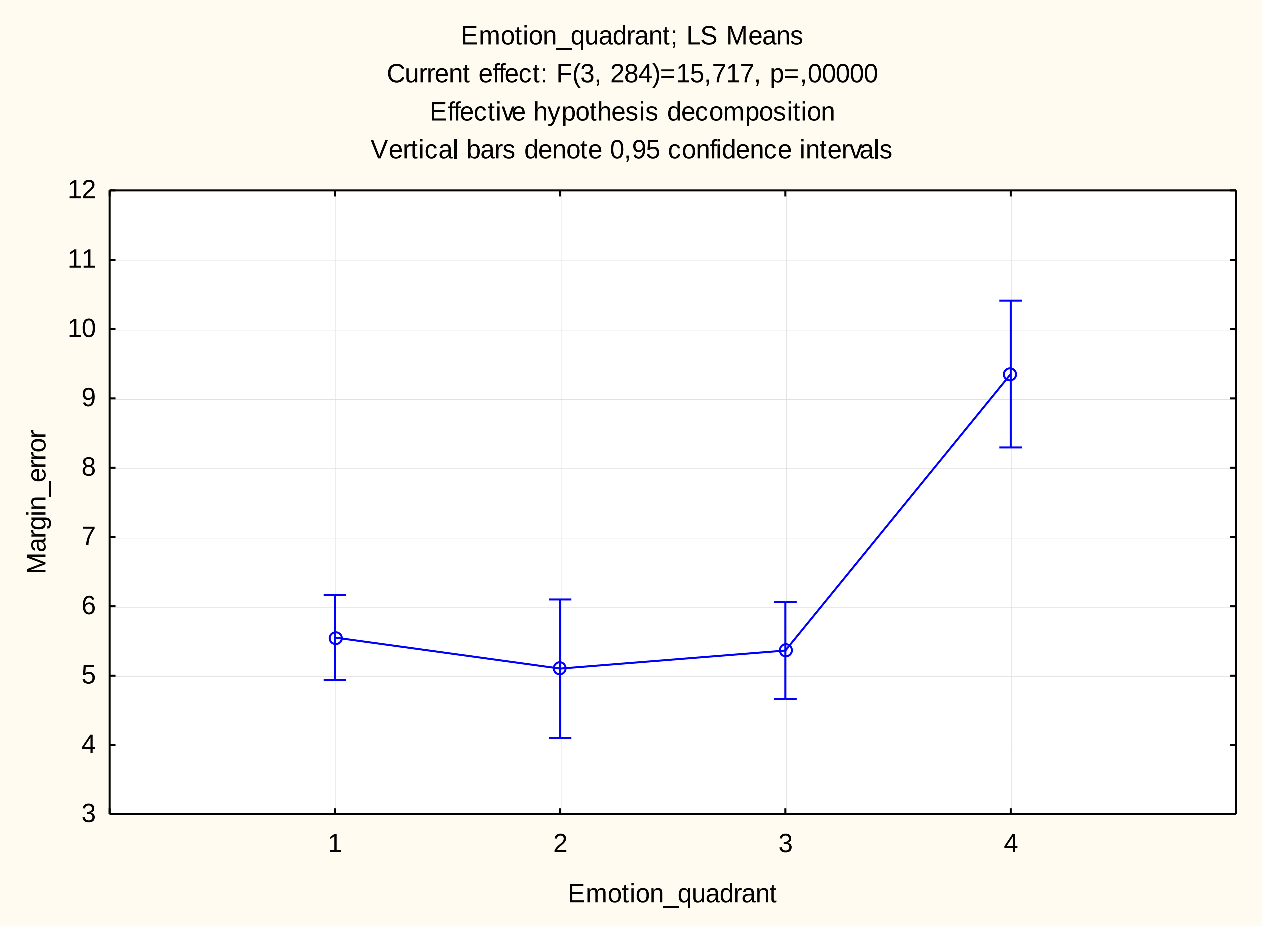


Figure 4.5. ANOVA of the margin of error between the computed slope's angles and the subjective feeling's angles with respect to the quadrants of the circumflex (N=288).

4.4.1.4 Hypotheses A.4: Valence and Control Means Match the Feeling's Position on the Wheel

The margin of error roughly signifies that participants choose a subjective feeling with an evaluation of the situation that does not correspond in terms of expected values for the Valence and Control dimensions. With the DEW measures, it is possible to establish for each feeling on the circumflex: (1) how many times the associated slope's angle belongs to each of the circumflex quadrant; (2) the means of the Valence and Control dimensions; (3) the mean of computed slope's angle; and (4) the mean difference between the slope and the feeling's angles. **Table 4.2** shows all of these measures for the 20 emotions in the circumflex, highlighting unexpected results with a red background of the cell. Individual analysis corroborate the problem with emotions belonging to the fourth quadrants (i.e. *annoyed*, *envious*, *disgusted*, *frustrated*, and *irritated*) which present at least one problem with the expected values of the dimensions or the computed slope. Other emotions that present some issues are *attentive*, *anxious* and *confused* (their mean slope does not belong to the same quadrant of the feeling), and *satisfied*, for which the mean of the Control dimension is positive, whereas a negative value is expected.

Table 4.2. Observed dimensional values (N=288) compared to the expected subjective feeling's position in the circumflex for all the 20 emotions on the wheel. Cell with a red background highlight infractions to the attended results (e.g. a positive mean in one dimension where a negative was expected).

Q#	Emotion	N	Q1	Q2	Q3	Q4	Val.	Con.	Slope	Error
Q1	Amused	40	28	8	4	0	46	20	87°	47°
Q1	Attentive	32	19	7	5	1	32	12	108°	48°
Q1	Happy	5	5	0	0	0	65	58	51°	12°
Q1	Interested	25	19	5	1	0	51	40	65°	43°
Q1	Confident	14	12	0	0	1	22	45	62°	29°
Q2	Empathic	10	0	7	3	0	23	-38	165°	21°
Q2	Relaxed	7	3	4	0	0	59	-3	94°	28°
Q2	Relieved	4	1	3	0	0	72	-40	116°	37°
Q2	Satisfied	7	3	4	0	0	54	36	97°	32°
Q2	Surprised	16	2	9	5	0	15	-25	157°	45°
Q3	Anxious	7	2	0	5	0	-14	-50	144°	58°
Q3	Bored	11	0	2	8	1	-32	-51	222°	29°
Q3	Confused	50	6	15	28	1	-7	-38	169°	37°
Q3	Disappointed	13	1	3	8	1	-19	-38	201°	57°
Q3	Stressed	8	1	1	6	0	-32	-26	200°	39°
Q4	Annoyed	4	0	1	3	0	-64	-52	215°	118°
Q4	Disgusted	5	2	1	1	1	-6	-9	151°	105°
Q4	Envious	6	2	3	1	0	48	12	127°	117°
Q4	Frustrated	15	1	5	6	3	-7	-41	194°	84°
Q4	Irritated	9	2	1	5	1	-11	-12	176°	96°
-	*Total*	*288*	*109*	*79*	*90*	*10*	-	-	-	-

With the observed means of the Valence and Control dimensions, we created a scatterplot to illustrate where each emotion would position in a hypothetical circumflex based upon participants' assessments (**Figure 4.6**). Compared to the expected disposition, 6 emotions (in red italic on the plot) result in a different quadrant. The 5 emotions of the fourth quadrant appear elsewhere: 4 emotions in the third quadrant, which means that these emotions have been evaluated low instead of high in Control; and 1 emotion (*envious*) in the first quadrant, which means that it has been evaluated positive instead of negative in Valence. *Satisfied*, on the other hand, appears in the first quadrant instead of the second, because it has been evaluated as high instead of low in Control. Hypotheses A4 is therefore rejected: the observed means of the Valence and Control dimensions do not correspond to the expected values given the position of the subjective feeling on the circumflex for 6 out of 20 emotions.

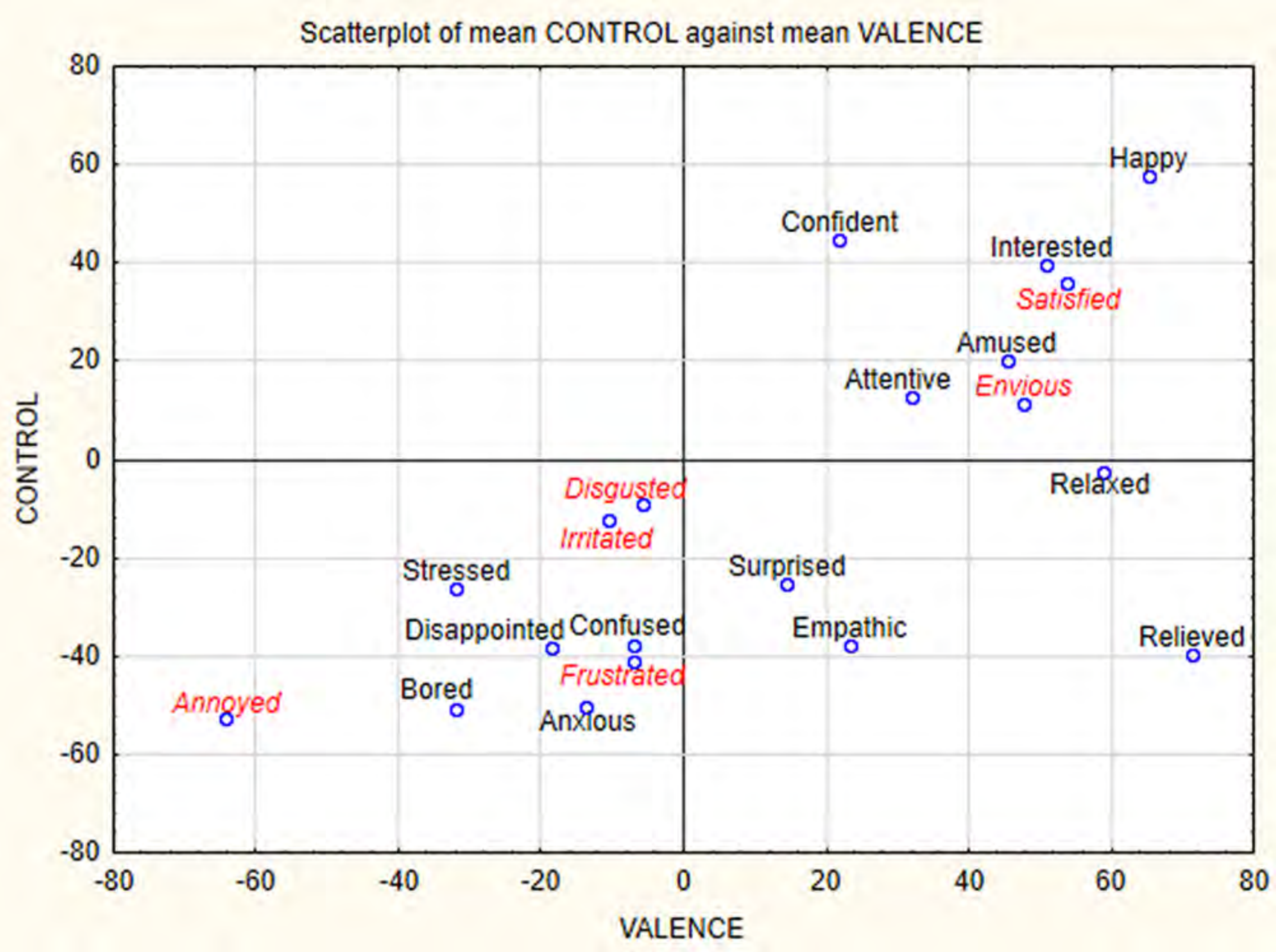


Figure 4.6. Scatterplot of the observed means of Valence and Control, composing a hypothetical circumflex where 6 emotions are misplaced compared to the expected position on the adopted circumflex (N=288).

4.4.1.5 Hypotheses A.5: Evaluation Order of the Dimensions

With respect to the SECs fixed order in evaluating an event, we labelled the Valence dimension with number one, and Control dimension with number two on the interface, expecting participants to follow this order in the manifestation of emotional episodes. We therefore calculated for each participant the ratio of the times participants modified the Valence dimension or the Control dimension as the last one. The result of a *t* test ($t(15) = -2.21$, $p < .05$) confirmed that participants assessed the Control dimension more frequently as the last one, with a ratio of 0.60 against the 0.38 of the Valence dimension.

4.4.1.6 Hypotheses B: Subset of Generated Buttons vs. Dropdown List

Even though the analysis of the relationship between the cognitive evaluation and the subjective feeling highlighted some serious issues, the dynamic generation of a subset of discrete emotion could still allow participant to save time in the manifestation of their emotional episodes. We therefore tested for each participant the ratio between the times they manifested the subjective feeling through a button, and the times they had to resort to the dropdown list. Results to a *t* test showed no differences ($t(15) = 1.24$, $p > .05$) between the mean of the button ratio (0.56) and that of the dropdown list ratio (0.44). Hypotheses B is therefore not confirmed, andin this regard, we crosschecked each emotion with respect to the number of times it was expressed through the buttons or through the dropdown list. Results show that all five emotions of the fourth quadrant are chosen more often within the dropdown list. So did the emotions *disappointed* and *attentive*, whereas all the remaining received more clicks through the buttons (see **Figure 4.7**).

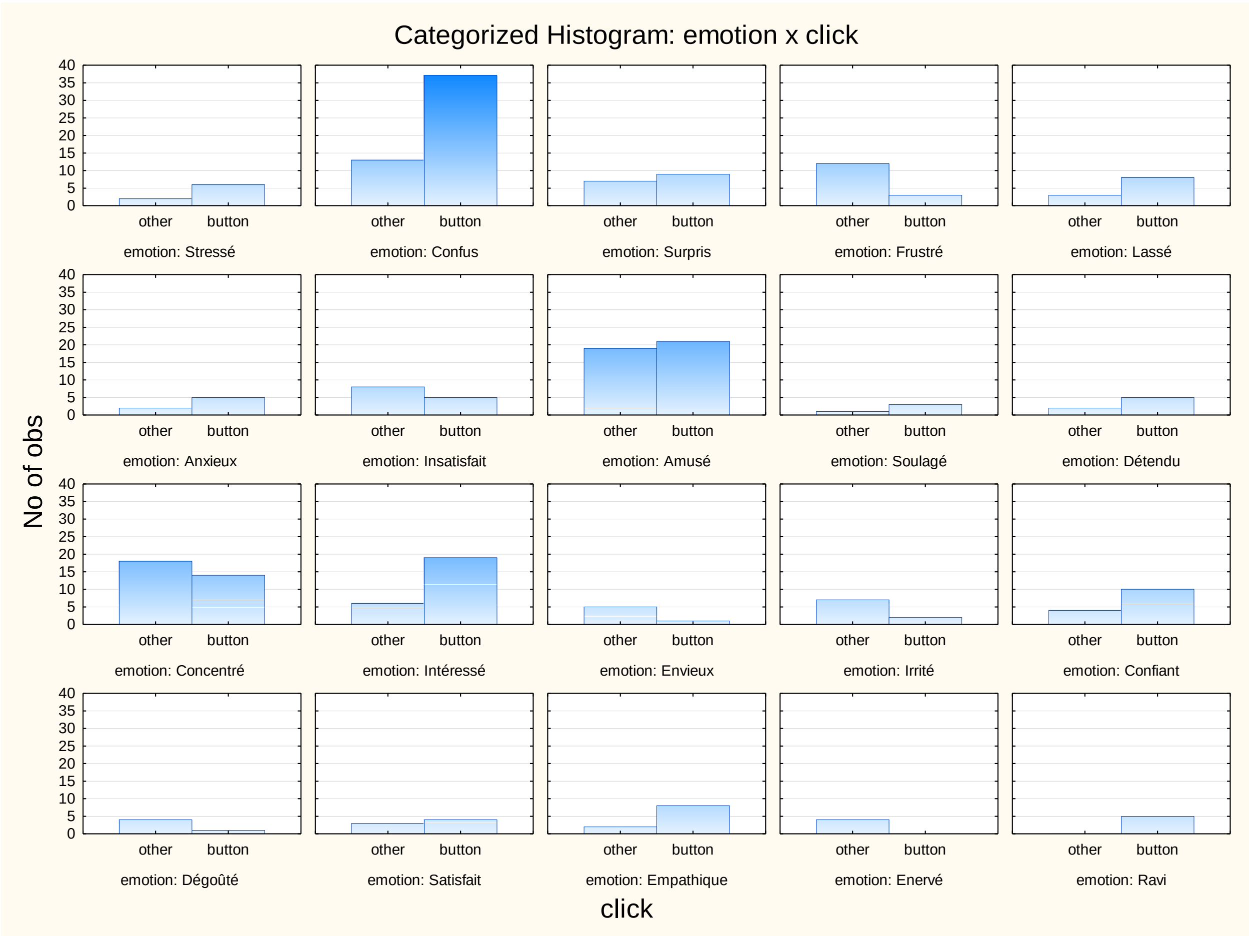


Figure 4.7. Categorized histogram of the cross tabulation between emotions and the type of click to express it: button or "other" for the dropdown list (N=288)

4.4.1.7 Hypotheses C: Time and Manipulations Required to Express an Emotion

Tobii Studio allows to measure the total duration of visits inside an area of interest (AOI) and also the total number of clicks executed inside the zone. A visit is defined as "the time interval between the first fixation on the active AOI and the end of the last fixation within the same active AOI where there have been no fixations outside the AOI" (Tobii AB, 2015, p. 107). This measure represents, thus, how much time (in seconds) participants looked inside a determined area. In the case of the displaying zone, participants should not have any other reason to look at this zone if they are not manifesting an emotional episode. We therefore divided each participant's visit time by the number of emotion provided. The result represents a hypothetical average time necessary to express an emotion. This measure must be considered indicative for two reasons. First, the dropdown menu spills over into the emotion timeline zone, so the time participants spent to choose within the list is not taken into account in this measure. Second, participants could have looked at the displaying zone even without the objective to manifest an emotion.

Notwithstanding, we consider this measure to be still revealing even if in a loose perspective. Participants (N=12) required in average 9.8 (SD=2.55) seconds in order to express an emotion. We also intercepted the number of clicks inside the displaying area for each participant, and divided it by the total number of emotion they expressed. Results suggest that participants required in average 4.00 (SD=0.7) clicks to express an emotion. It is worth noting that this measure also suffers from the trespassing of the dropdown list. The average time and clicks number, on the other hand, strongly correlate ($r=0.80$, $p < .05$), which corroborates the consistency of the measures: participants taking more time to express emotions also clicked more often. Hypotheses C is therefore confirmed: the expression of an emotion takes less than 10 seconds, and requires around 3 clicks.

4.4.1.8 Hypotheses D: Perception of the Displaying Zone

The post-test questionnaire presented a block of 13 questions on 1-to-7 scale similar to the SUS items, but adapted to the specific objective of the displaying zone of the DEW. Participants reported to encounter few problems in handling the zone (M=2.0, SD=1.03), and that the use of the zone is rather intuitive (M=5.3, SD=1.7). They also judged that the dynamically generated buttons allowed them to save time during the expression of emotions (M=5.25, SD=1.57), rejecting the idea of a better compromise if all the emotions were listed in the dropdown menu (M=2.38, SD=1.71). Finally, participants reported that, in general, they have understood how the displaying zone works (M=6.06, SD=1.34), and that the zone facilitated the expression of their emotions (M=5.69, SD=1.45). Contrary to objective measures suggesting some issues with the Control and Value dimensions, participants did not report particular difficulty in evaluating each dimension (M=3,69, SD=1.35 for Valence and M=3.25, SD=1.48 for Control). On the other hand, it emerged from the post-test interview that most participants perceived the use of the sliders as difficult and demanding in terms of effort and attention required. For instance, a participant reported in the open question about the weaknesses of the tool that “giving the valence and the control is a waste of time, it would be better to give only the name of the emotion through the buttons”. Another participant complained that “where [he] moved the sliders were a bit arbitrary – it is difficult to quantify an emotion (especially on a scale of 200 points)”, even though the same participant reported, in the strengths of the tool, that “the two sliders are easy to manipulate”.

4.4.2 Assessment of the Monitoring Function

The monitoring zone of the DEW requires minimal manipulation, since its main purpose is providing participant with meaningful and accessible information. To assess the utility and usability of the monitoring zones, thus, we will mainly use eye tracking and subjective data.

4.4.2.1 Hypotheses E: the Interest of Subjective Feeling vs. Appraisal Dimensions

The DEW integrated subjective feeling through a single emotion-related adjective, for the resort to language remains the most viable way to communicate emotions. Consequently, we tested if participants focused their attention more often on the emotion timeline compared to the two appraisal line charts (independently from the interface type). We determined the mean of total fixations on the emotion timeline and the two combined line charts for the 12 participants with eye tracking data. The emotion timeline received an average of 161.75 fixations during the test, whereas the two line charts an average of 77.58. A *t* test confirmed hypotheses E, according to which the timeline received significantly more fixations ($t(11) = 4.45$, $p < .05$, **Figure 4.8**). The heat map in **Figure 4.3** visually confirms the results also with respect to the duration, since the emotion timeline presents more red-orange zones, especially in correspondence with the last emotions expressed both by the participant himself and by the partner. Results are also corroborated by subjective evaluation, since we asked participants (N=16) to rate to what extent they judged each one of the monitoring zone as useless (i.e., a low score signifies a high perceived utility). The emotion timeline obtained a mean of 3.00, against similar score of 4.85 and 4.75 for the two line charts ($t(15)=-1.86$ and -1.75, $p < .05$).

Figure 4.8. The emotion timeline received more overall fixations (M=161.75) compared to the two line charts combined (M=77.58), corroborating the usefulness of conveying the subjective feeling (N=12).

4.4.2.2 Hypotheses F: Difference in the Line Charts Grouping

We tested two different interfaces of the DEW that diverged with respect to grouping of appraisal dimensions on line charts. We expected a different perception of the line charts depending on how Valence and Control were subdivided: by Participants or by Dimensions. We used the total number of fixations as an indicator of the interest each chart elicited in the 12 participants with eye-tracking data. First, though, we checked for differences in the number of fixations between the two types of interfaces. Results to two *t* tests conducted suggest that participants equally fixated the line charts regardless of the type of interface. The first (top) line chart was fixated on average 43.33 times in the Participants version, and 49.67 in the Dimensions version ($t(10) = -0.58$, $p > .05$). The second (bottom) line chart was fixated on average 34.83 times, in the Participants version and 27.33 times in the Dimensions version ($t(10) = 0.59$, $p > .05$). We then crosschecked the differences between users within the Dimensions subdivision and users with the Participants subdivision. The Dimensions subdivision presented a significant difference between the two line charts. In fact, the top graphic, representing the two Valence dimensions of each participants, received an average of 49.67 fixations, against 27.33 for the bottom line chart representing the Control dimensions ($t(5) = -4.98$, $p < .05$). On the contrary, this difference was not significant in the Participants subdivision. The top graphic, representing both Valence and Control for the user himself, received on average 43.33 fixations, whereas the bottom graphic, representing both dimensions of the partner, received 34.83 fixations ($t(5) = -0.66$, $p > .05$). Results suggest thus that users with the Dimensions subdivision paid more attention to the first graphic (Valence dimension) compared to the second (Control dimension), whereas users with the Participants subdivision equally shared their attention between the two line charts.

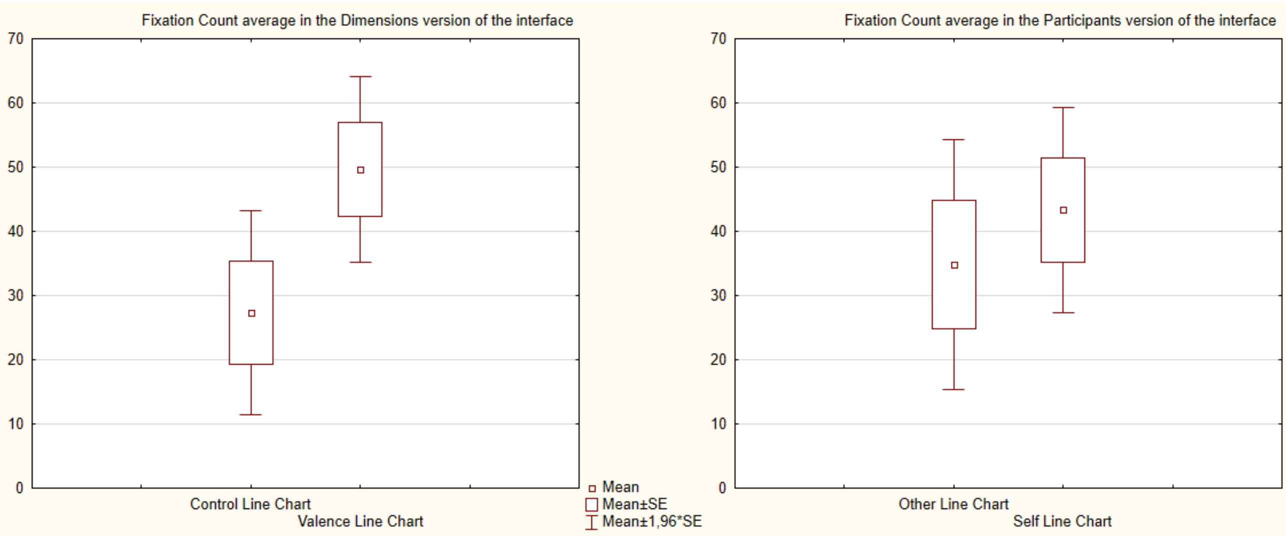


Figure 4.9. Average fixation count on the two line charts depending on the type of interface tested (Dimensions on the left, Participants on the right). The Valence line chart received significantly more fixations than the Control line chart in the Dimensions interface, whereas in the Participants version the number of fixations was equally distributed (N=12).

Furthermore, in the post-test questionnaire, we also asked participants (N=16) the perceived utility of the different zones of the monitoring area with respect to emotional awareness. Each zone was assessed on 6 items, each on a 1-to-7 scale, and the items concerned the perceived utility in understanding, comparing, and following the evolution of participant's and their partner's emotional states. We created a general index of utility of the zone with the mean of each item, and compared these indexes with respect to the participant's type of interface. The emotion timeline, which was the same in both interfaces, did not present a significant difference in the perceived utility: the users with the Participants interface assessed its utility at 4.92, and the users with the Dimensions interface at 4.44 ($t(14) = 0.66$, $p > .05$). On the contrary, the two appraisal line charts presented significant differences. With respect to the two top line charts in the respective interfaces, the Self Line Chart utility index equals 3.69, against 1.69 for the Valence Line Chart ($t(14) = 3.2$, $p < .05$). Similar results for the bottom line charts, since the Other Line Chart index equals 3.79 and the Control Line Chart 1.65 ($t(14) = 3.15$, $p < .05$). Combining objective and subjective measures, the Dimensions subdivision of appraisal dimensions was perceived less useful than the Participants subdivision and the Valence Line Charts received significantly more fixations compared to the Control Line Charts. On the other hand, the Participants interface was perceived more useful compared to the Dimensions interface and the two line charts received a similar number of fixations.

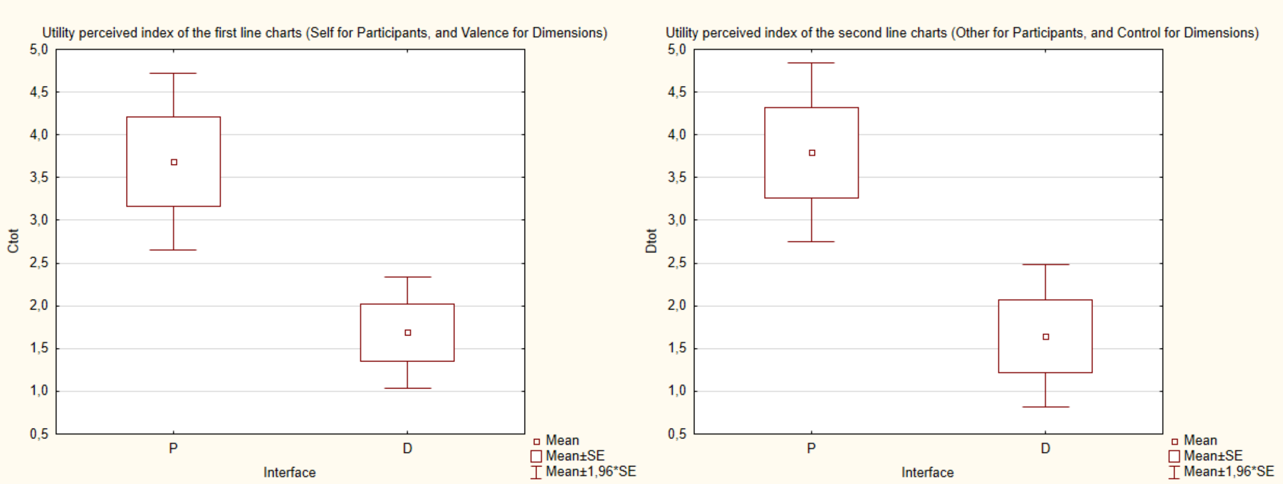


Figure 4.10. Perceived utility index of the first (left) and second (right) appraisal Line Charts depending on the interface. Results suggests that the two line charts have been significantly assessed more useful in the Participants interface type, and the judgment to be homogenous for the two charts in both interfaces (N=16).

4.4.2.3 Hypotheses G: Perception of the Monitoring Functions Depending on the Objectives

In the post-test questionnaire, we asked participants (N=16) to assess, on a 1-to-7 scale, each zone of the monitoring area with respect to its utility in (a) understanding the participant's own emotional state; (b) understanding the partner's emotional state; (c) comparing the two emotional state; and (d) following the evolution of the emotional state in time. We expected the emotion timeline to be perceived more useful for objectives (a), (b) and (c); whereas the appraisal dimensions for objective (d). Since, as we have seen, the evaluation of the line charts depend from the interface type, we tested the score of the emotion timeline with the highest mean obtained by either the first or the second line chart, regardless of the interface type. As expected, emotion timeline was assessed more useful ($t(15) = 2.55$, $p < .05$) in understanding participants' own emotional state compared to the best line chart zone (M=3.63 against M=2.25). It was also perceived more appropriate ($t(15) = 2.44$, $p < .05$) for understanding the emotional state of the partner (M=5.06 against M=2.94). Finally, consistently with our expectations, the emotional timeline resulted more pertinent ($t(15) = 3.63$, $p < .05$) even for comparing the emotional state between users (M=5.44 against M=3.00). On the other hand, contrary to our expectations, the appraisal dimensions were not perceived as more useful ($t(15) = 1.91$, $p > .05$) for perceiving the evolution of emotional episodes in time (M=3.19 against M=4.44 of the emotion timeline). Users therefore perceived the emotion timeline more useful for understanding and comparing emotional episodes; and equivalent to the line charts for following the evolution of the emotional state in time.

4.4.3 Assessment of the Overall Function

4.4.3.1 Hypotheses H: the Accessory Role of Emotional Awareness

The total duration of the task in second was 1'200. Participants spent on average 171.13 seconds on the displaying zone, 35.16 on the emotion timeline, 10.55 on the first line chart, and 7.86 on the second, for a total of 224.7 seconds during which users looked at the DEW. Compared to overall duration of the task, users spent 18.73% of their time either displaying or monitoring emotional awareness, which is slightly above the 15% threshold expected. It is nevertheless worth noting that the displaying time (171.13) is much longer than the monitoring time (53.57 for the three monitoring zones combined). The heat map in **Figure**

4.3 visually corroborated these results: the displaying area shows more "heated" zones (especially the buttons) compared to the monitoring areas (especially the line charts).

4.4.3.2 Hypotheses I: Mental Effort, Difficulty and Interest of Emotional Awareness

In the post-test questionnaire, participants rated the mental effort, the difficulty, and the interest of the task with respect to (a) the resolution of enigmas, (b) the emotional awareness during the resolution of enigmas; and (c) the experience as a whole. We therefore tested if the reported means of the three dimensions were significantly different within the a-type, the b-type, and the c-type questions. The *t* tests conducted ($t(15)$, $p < .05$) suggested a rather homogenous perception of the three type of question for the three dimensions. Significant differences were detected only between the mental effort for emotional awareness (M=4.31) and mental effort of the overall task (M=4.94); the difficulty of problem solving (M=5.44) and the difficulty of the overall task (M=4.5); the interest of emotional awareness (M=4.13) and the interest of the overall task (5.56). Results do not suggest, thus, that emotional awareness increased the mental effort or the difficulty of the task, but neither indicate a particular interest of having it in a problem-solving task.

4.4.3.3 Hypotheses J: Perceived Usability of the Emotion Awareness Tool

As a general evaluation of the DEW, we added the System Usability Scale (SUS, Brooke, 1996) to the post-test questionnaire. Participants' mean rating of the tool was 77.03 (SD=17.35) with a minimum score of 45 and maximal of 100. The SUS result was independent from gender and type of interface. According to Bangor and collaborators (2009), a total score above 70 is considered acceptable, which corroborates hypotheses J: users assessed the overall usability of the DEW as acceptable.

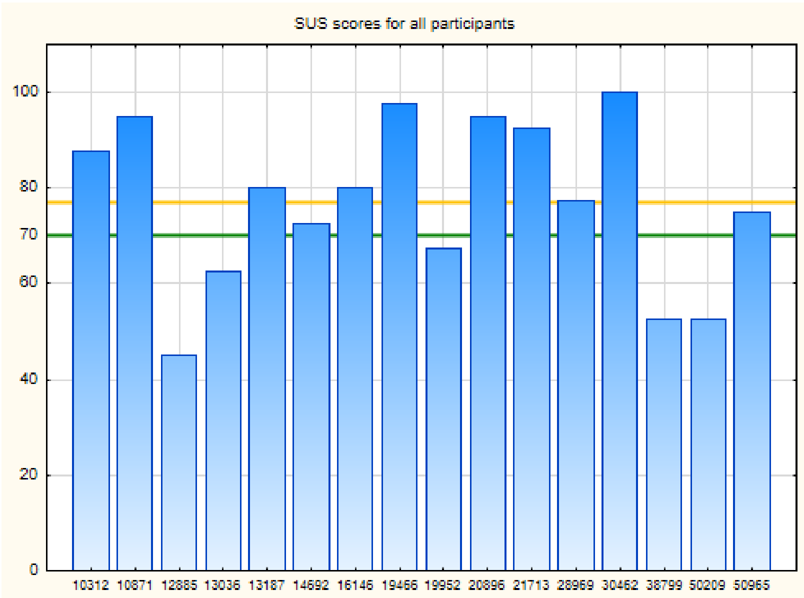


Figure 4.11. Score to the System Usability Scale (Brooke, 1996) for all the participants to the test (N=16). The threshold of 70, representing acceptable usability (Bangor et al., 2009), is marked by a green line, and the rounded mean obtained by the DEW – 77 points – is marked in orange.

An analysis of the single items of the scale revealed that participants were particularly doubtful about the first question (i.e., *I think that I would like to use this tool frequently*), which obtained on average 1.8 points out of 4, and the ninth question (i.e., *I felt very confident using the tool*), which obtained on average 2.4 points (all the other questions obtained more than 3 points on average). In this regard, most of the participants expressed doubts, in the post-test interview, about the possibility to use the DEW during their collaborations and integrate it with the tools they normally use. For instance, in the weaknesses of the tool, a participant reported that “[she] did not see how the tool could be integrated on websites like Moodle” (a free and open-source software learning management system often used by universities).

With respect to the 5 participants that rated the SUS below the 70-point threshold, 2 of them clearly defined that they would not use the tool frequently (0 points on the first SUS question) and that the tool was unnecessary complex (0 points on the second SUS item). 3 of them also reported in the open question about the weaknesses that the tool interfered with the primary collaborative task. As one of them stated it, “having to constantly divert the attention from the enigma to use the tool [caused] a loss of concentration on the problem”.

On the other hand, in the open question about the strengths of the tool, 10 participants out of 16 cited the interest of disposing of emotional awareness through the tool. In particular, a participant stated that the strength of the tool was to “become aware of our emotions, asking ourselves what we feel, [so that] we can take some distance from negative emotions”. Five participants were appealed by the possibility to “compare the emotions with the other participant”, and another one added among the strong points “the tracking of emotions (personal and of the other) in real time”.

4.5 Discussion

The usability test produced mixed results, since many hypotheses were totally or partially rejected, and just a few corroborated. The overall outlook may nevertheless be more encouraging compared to the ratio between confirmed and rejected hypotheses mainly because participants have positively assessed the overall usability of the tool, which was the primary objective of this contribution. This is no reason, though, to neglect the conceptual shortcomings emerged, in particular with respect to the core principle of the Dynamic Emotion Wheel: the generation of a subset of discrete emotions according to the appraisal of two-dimensional values highlighted in fact some limitations. In the discussion, we will therefore begin with answering to two fundamental questions. First, *does the dynamic algorithm of the DEW facilitate the expression of an emotional episode*? Second, *does the combination of appraisal dimensions and subjective feelings convey meaningful emotional awareness to participants*? In our responses, we will integrate the results with the strengths and the weaknesses of the tool reported by participants in the post-test survey open questions, as well as feedbacks and suggestions emerged during the final debriefing. Third, we will list some usability issues discovered during the test, and provide some minor suggestions for changes. Finally, we will assess the limitations of the current usability test. A more comprehensive assessment about the limitations and future development of the tool will be part of the general conclusion to the contribution.

4.5.1 Did the Dynamic Mechanism Fulfill the Expectations?

The generation of a subset of discrete emotions according to specific appraisals on underlying dimensions is the distinctive feature of the Dynamic Emotion Wheel – its *raison d’être*. Without the algorithm that reduces the number of emotions on screen, the DEW would be limited to a simple web form with two sliders and a bunch of useless buttons. The answer to the fundamental question about the usefulness of the dynamic mechanism must therefore be assessed very carefully, since a negative answer will obliterate the concept altogether.

4.5.1.1 Why Did It Fail – If It Failed

The results of the usability test and qualitative information obtained through open questions and feedbacks indicate two main aspects that negatively affected the expected behavior of the dynamic algorithm. The first is a structural issue, which broadly relates to the generalized problem of emotion discrimination, in particular with respect to two-dimensional spaces. The second issue relates to the least possible effort, which comprises both conceptual and usability aspects.

With respect to the former, the displaying function of the DEW highlighted a clear problem with the subjective feelings positioned in the fourth quadrant of the circumflex. According to the underlying circumflex, these feelings are characterized by the combination of negative Valence and positive (high) Control. As illustrated in **Figure 4.6**, the *annoyed*, *disgusted*, *envious*, *frustrated* and *irritated* feelings were assessed, on average, with either the Valence or Control values differing from the expectations. This problem may be attributed to the “rule of the thumb” composition of our underlying circumflex, but in this case at least the *disgusted* and *irritated* feelings should be spared. They appear, in fact, in the same quadrant on the third version of the DEW (if we consider anger and irritation sufficiently close), whose disposition is determined by the GRID comprehensive experiment. We rather think that the problem with this quadrant results from a conceptual difficulty to assess a situation upon which participants feel to have control, even though they judge the event as unpleasant. We remind that only 10 observations out of 288 proposed a negative Valence and positive Control appraisal (of which only 5 cases presented an associate feeling of the same quadrant), despite the fact that participants chose 39 times the feelings belonging to the fourth quadrant (see **Table 4.2**). The feelings of the fourth quadrant also presented the highest margin of error between the computed slope and the subjective feeling’s angles. We therefore speculate that the Control dimension becomes a sort of duplicate of the Valence dimension when the situation is appraised as unpleasant. Four of the fourth quadrant feelings, in fact, would rather appear – according to the observed means – in the third quadrant, whose feelings are precisely characterized by negative values for both dimensions. In our opinion, thus, the interdependence of the two sliders primarily derives from this issue concerning the fourth quadrant of the circumflex. We think, in fact, that without the poor performance of the fourth quadrant in terms of correspondence between the computed slope and subjective feeling’s angles, the ratio of emotions expressed through the buttons or through the “other-emotion” form would have surpassed the fifty-fifty threshold in favor of the buttons.

The second issue detected with the overall displaying function is that participants tended to minimize the effort of expressing emotions, especially because they were deeply involved in the problem-solving task. In this regard, the majority of participants reported that the dimensional appraisal required substantial mental effort, considering in the meantime the stress and difficulty in resolving the enigmas. Consistently with the literature and the objective measures of this test, they perceived the subjective feeling as more representative of their emotional state and more practical to assess, compared to the bi-dimensional appraisal of Valence and Control. In other words, they considered the subjective feeling more accessible and economical, since it conveyed sufficient information without the need to add the cognitive evaluation of the situation. In this regard, at least three participants clearly stated that, if they "seriously assessed" the two sliders, they did find the corresponding subjective feeling in the button list. Nevertheless, since they knew the subjective feeling was also available through the dropdown list, the "serious assessment" represented an additional effort that could be easily avoided. Two participants also found that the sliders were not enough "fine-tuned", since they did not observe any change in the subset even after changes in the values. This is most certainly related to the ratio-dependency of the algorithm, which determines that some emotions are more frequently computed (see 3.2.3). Contrary to our expectations that this shortcoming would pass unnoticed, participants rather perceived that sometimes their accuracy in evaluating the dimensional values was not "rewarded" by new choices, and would therefore be more inclined to switch to the dropdown shortcut. All things considering, thus, the integration of quantitative measures was perceived more as a nuisance than an opportunity. We nevertheless think that this assumption shall be put in a double perspective that could at least mitigate its extent. First, the usability test perspective aimed to figure out the intuitive use of the tool, and did not stress about the importance of an accurate and pondered assessment of the two-dimensional sliders. In other words, we think that this shortcoming could be avoided if experimenters give more specific instructions about the use of the sliders. For instance, the simple use of numbers to illustrate the expected order of manipulations produced the expected behavior, even though the experimenter did not stress about this more than he did for the use of the sliders. "Force" people to do things is nevertheless a clumsy solution if we plan to use the tool in ecological situations. For this reason, we think that a second reason that may have affected the use of the dimensional sliders is that participants did not benefit from the corresponding graphical representation. As the eye tracking data pointed out, participants seldom looked at the line charts, whose perceived utility was significantly inferior compared to the emotion timeline. We will assess the monitoring function more in detail below in this section, but it is worth noting at this point that the two functions are mutually dependent.

4.5.1.2 Why Did It Succeed – If It Succeeded

Without any doubt, the most encouraging indications about the usefulness of the dynamic mechanism come from participant's subjective evaluation. The post-test survey and interview confirmed that the interaction process of the displaying zone appealed to users. They reported that the dynamically generated buttons allowed them to save time in the expression of their emotion, rejecting in the meantime both the idea of having all the buttons on the interface and the possibility to maintain exclusively the dropdown list. We therefore think that at least the core principle of the Dynamic Emotion Wheel is valid in terms of usability. In this regard, the ratio between the use of buttons compared to the resort to the "other-emotion form" could also be revaluated in statistical terms. In fact, we rejected the hypotheses concerning the usefulness of the dynamic algorithm since the use of the buttons and the dropdown list was equally distributed. We can nevertheless consider that only three emotions out of twenty were displayed as buttons. In other words, if the choice of the emotion were random, participants would have 15% of chances to find the corresponding subjective feeling in the list of buttons compared to 85% in the dropdown list. Of course we shall consider that (a) emotion were not randomly assigned to buttons or dropdown list, and therefore the gain effect of the mechanism must be assessed well above the 15% threshold; and (b) the buttons have a clear "least effort" appeal compared to the dropdown list, consequently participants could be more complaisant with a subjective feeling proposed as a button if it is close enough to the one they had in mind. Still, a 50% chance to find the right feeling in the subset may be considered sufficient according to the *ex negativo* perspective of the DEW: the algorithm allows to exclude unrelated feelings half of the time. Roughly speaking, thus, users can spare time displaying their emotions every other time, so that the expression of their emotional state takes on average around ten seconds.

4.5.1.3 The Final Verdict

All things considering, we are not in the condition to make a final yes-or-no statement about the effectiveness of the core principle of the DEW. On the one hand, results to the test highlighted conceptual and structural problems that we cannot simply dismiss because of the flexibility and adaptability of the tool (see 3.4). In particular, we could divert the problem with the underlying dimensions suggesting that it is possible to adopt a different circumflex with different underlying dimensions. Nevertheless, this assertion would simply represent an act of deferred responsibility: the algorithm must be based upon solid evidence that exists, and not upon speculations on future achievements. On the other hand, the very same principle shall be considered from a different perspective: if every approach to emotion measuring present some issues, it would be presumptuous to solve all problems with a few lines of code. An overall assessment of the algorithm must therefore take into account, in a goal-oriented perspective, what is the final aim of the DEW. Since we pointed out that the DEW adopts an *emotion as interaction* approach to affective computing (see 1.2.5), we believe that a certain degree of indulgence towards the accuracy of the algorithm might be granted. The role of the DEW consists in sharing emotions, regardless of their "true" nature and intentionality. In this regard, we can assume that the dynamic generation of a subset of discrete emotions according to the appraisal of two-dimensional criteria positively contributed to sharing emotional episodes during an ongoing collaborative task.

4.5.2 Did Quantitative Data Add Useful Information to the Monitoring Function?

Another important feature integrated in the DEW is the use of both qualitative and quantitative data to convey emotional awareness. The question concerning the added value of appraisal dimensions instead of the exclusive resort to subjective feeling is nevertheless less crucial compared to the usefulness of the dynamic algorithm. In the case of the monitoring function of the DEW, we believe that flexibility and adaptability can play a major role, and changes at this level do not undermine the core principle of the tool. As aforementioned, the monitoring function tightly depend upon the displaying function, but may also be to some extent independent. No imposition states that an awareness tool must show everything it collects, even though it would be more appropriate, since the act of displaying requires effort from the users. In this regard, eye tracking data and subjective reports suggest that the balance between the effort to assess the appraisal dimensions and the utility of the line charts is disproportionate. The subjective feeling words of the emotion timeline have been evaluated more economical and meaningful elements for understanding and comparing the emotional state of and between participants. We do believe, nonetheless, that quantitative data might still be useful under some circumstances, especially with respect to the evolution of the emotional state in time. The visualization of quantitative data through graphics can efficiently contribute to put information into perception, facilitating a bottom-up approach to information processing.

In the post-test interview, though, the attitude towards the line charts was quite controversial. Some participants kindly admitted that they did not even look at them, because they were not interested or they did not understand what information they conveyed – a phenomenon corroborated by eye tracking data. Two participants even confessed that the line charts contributed to make the interface overcrowded and intimidating. Other participants, on the other hand, manifested more interest, but confessed that they did not dispose of enough time to observe them "seriously". The interest for line charts is corroborated by the significant differences detected with respect to the type of interface tested. Users clearly preferred a subdivision of the line charts according to which one chart represented both appraisal dimensions of the participant himself, and the other chart both appraisal dimensions of his partner. (Even though a participant assigned to the Participants interface spontaneously proposed to organize the line charts in the Dimensions' subdivision, without knowing that different types of interface were tested.) Another user with the Participants subdivision, on the other hand, suggested to switch places between the Self (top) and the Other (bottom) line charts because "[she] will be more interested in my partner's line chart, because [she] does recall [her] evaluations, but [she] knows nothing about [her] partner". In fact, the position of the line charts on the screen affects their "casual" visibility. Even if in the participant subdivision the count of fixations was not statistically different, the mean of the top (Self) line chart was nevertheless greater than the bottom (Other) line chart. Considering the assumption of the participant that found the Other line chart more useful, we should have expected the contrary, if it were not for the bias of the on-screen position. Another user also suggested that line charts could be available upon specific request (e.g. by clicking on a button) instead of being visible all the time. This is in fact consistent with the idea of collaboration as a seamless cycle between displaying and monitoring functions: participants may exploit moments in which they are less involved in the collaborative task to look into the line charts.

All things considering, the monitoring zones of the DEW can be improved and the usability test provides useful information for the purpose. The emotion timeline was not spared some critiques either, since at least two participants admitted they did not understand neither its function nor its utility. It is also worth considering, in a more general scope, that the monitoring function might provide emotional awareness to a group comprised of more than two participants, in which cases both the emotion timeline and the line charts should be completely revised. This possibility will be further assessed in the conclusion of this contribution among the future developments for the tool.

4.5.3 Usability Issues

Even though the general perception of the DEW's usability was considered acceptable, with an average SUS score of around 77 points, the test also highlighted some issues. We will briefly list the most significant ones with respect to the ergonomic criteria proposed by Scapin and Bastien (1997) and suggest possible solutions. Other issues discussed in the assessment of the dynamic algorithm and the monitoring functions can be referred to with ergonomic criteria, but we do not report them here for two reasons: (a) they usually spread over different criteria (for instance the problem with the sliders pertains at least to the workload, explicit control, adaptability and compatibility criteria); and (b) they do not have "simple" ergonomic causes, therefore we are not in the condition to suggest precise corrections.

4.5.3.1 Error Management Issue in the Displaying Function

Users clicked four times on the "no-emotion" button, even though they reported it as an error in the post-test interview. This issue highlights two kind of problems in error management: (a) error protection – the "other" and the "none" buttons are very close one another; (b) error correction – the DEW does not provide the possibility to undo an action, which can represent a serious problem. To solve this issue we therefore propose a more clear separation of the "other" and "none" buttons. If the overall space on the screen allows it, they may be put on separate lines altogether. The error correction is more tricky to resolve. A possible solution implies a general way to undo the last action, like a button or keyboard shortcut; whereas another may imply that the last emotion on the timeline also present a "delete button", in which case, though, the clear separation between handling and observing zones of the DEW will be infringed. The idea of a general way to undo the last action seems the more economical solution.

4.5.3.2 Guidance and Consistency issue in the Monitoring Zone

A participant correctly pointed out an incoherence in the use of colors for the Participants interface type, which relates to the criteria of guidance, more specifically grouping and distinguishing items by format, and consistency. In fact, the emotion timeline uses the blue color for the emotion of the participant himself and the green color for the partner. These colors are used also for the line of the charts: the blue for the Valence and the green for the Control dimensions, creating thus a mismatch between participants and dimensions colors (see **Figure 3.9**). Consequently, it would be better to choose different colors for the timeline and the two lines of the charts.

4.5.3.3 Guidance Issue in the Emotion Timeline

The emotion timeline indicates that the blue emotions represent the participant's own emotions and the green emotions those of his partner with two labels at the beginning of the respective rows. Nevertheless, these indicators quickly disappear once the timeline start scrolling to the right, which affects the guidance, more specifically the grouping and distinguishing items criterion. Two possible solutions may be adopted: (a) inserting a legend at the bottom of the timeline as the ones used for the line charts; or (b) fixing the indicator labels so that they do not scroll and remain always visible. Considering the limited horizontal space at disposal in the timeline, the first solution seems more appropriate.

4.5.4 Limits of the Usability Test

The aim of the usability test was to evaluate the use of the DEW in a computer-mediated collaborative task and, for this reason, we simulated a synchronized problem-solving task in experimental settings. If this solution provided standardized conditions, useful for comparison and statistical analysis, it highlighted three major drawbacks. First, the problem-solving task implied a limited time to resolve the task. This interfered with the possibility for participants to take their time and explore the interface, in particular the monitoring zones of the DEW. The short time at disposal is also inconsistent with "real" collaborative tasks, where participants usually dispose of larger amount of time, which facilitate the cycles of active involvement and reflexive/metacognitive activities. Second, the majority of the participants explicitly doubted about the presence of another user on the other side of the screen. This phenomenon has most certainly affected the interest and instrumentality of emotional awareness in the task. Third, the problem-solving task did not represent a situation for which effective collaboration was required. Consequently, it is difficult to assess the instrumental role of emotional awareness in the computer-mediated collaboration. Considering the primary interest in usability, though, we nevertheless consider the created scenario useful for most of the objectives of the test.

5 Conclusion

We will conclude this *mémoire* by illustrating some limits of the present contribution and by providing future developments of the tool. In both cases, we will assume a broader approach, assessing the overall development process and the general context in which the tool will eventually be used. Finally, in a general conclusion, we will resume the main aspects issued by the development and test of the Dynamic Emotion Wheel.

5.1 Limits of the Present Contribution

We identify three major limitations in the present contributions. Two specific limits concerns the primary usability perspective adopted in testing the DEW, and the underlying theoretical framework of the tool with respect to emotion self-reporting. The third limitation, on the other hand, concerns the extent of the entire contribution.

5.1.1 The Limits of Usability Perspective

Usability tests are an invaluable source of insight for developers, since they provide meaningful information about the concrete use of the application by potential end-users. On the other hand, focusing the test of the DEW primarily on usability excluded some alternatives that could have contributed to better assess the core principle of the tool. For instance, we planned to test two different versions of the displaying zone of the DEW (as we did for the monitoring zone). In one version, the subset of discrete emotions would be generated with the DEW algorithm, whereas in the second version with a random mechanism. In these conditions, the added value of the dynamic mechanism would have been more evident. On the other hand, though, the perceived usability of the DEW for the group with the random mechanism would have been biased by an exclusive experimental factor, since there is no interest in testing the random mechanism.

5.1.2 An Excess of Cognition in Emotion Self-Reporting

In the development section of this contribution, we stated that the DEW represented a piecemeal evolution of the Geneva Emotion Wheel, because the modification of a valid tool must be gradual and not abrupt. This principle should not justify, alone, the theoretical consequences of the choice. The Geneva Emotion Wheel is a tool deeply rooted in the Component Process Model framework, which has been criticized for an excessive importance given to cognition in emotions (see Roseman & Smith, 2001 for an overview). In this regard, we believe that an Emotion Awareness Tool, for its *awareness* purpose, inevitably implies a substantial role of cognition, especially since emotions are voluntary shared and not automatically detected by a visual decoding device or physiological sensor. The interest of awareness in computer-mediated collaboration is to provide information that are instrumental to the joint task. For this reason, the users are not supposed to “throw” their emotional state into the interface, regardless of its causes and consequences. For emotional awareness to be effective, users should perform metacognitive operations on both their emotional episodes and those of the other participants. Consequently, even admitting that the cognitive evaluation of the situation play no role in emotion elicitation (and we do not share this position), it certainly play a significant role in understanding and comparing emotions both at intra- and inter-individual levels. The DEW, at least in its intentions, makes the link between appraisal and subjective feeling so explicit, that the limit of the tool would rather be – and in part is – the opposite: the independence of generated buttons from the dimensional sliders. Consequently, the tight relationship between the DEW and the GEW is not simply sustained by developmental affinity, but is deeply connected with the underlying theoretical framework, and is therefore exposed to the same critiques.

5.1.3 Extent in Scope and Method

This contribution integrates concepts of different disciplines and methodologies. More specifically, the development of an Emotion Awareness Tool used in experimental settings refers at the same time to two types of *mémoire* usually proposed in the MALTT: the development and the experimental types. This contribution, especially in its early stage, suffered from the uncertainty about which of the two aspects should be privileged. In other words, would the contribution be an experiment that makes extensive use of a tool developed for the purpose; or would it rather concern the development of an application, which would be used in experimental settings? As it shall be clear at this stage of the reading, the contribution was ultimately focused on the development process, integrating an experiment in the perspective of the usability test. Although the interest of this procedure has been widely explained and sustained, the overall result may still suffer from the ambivalence. This contribution is neither a thoroughly illustration of a well-defined piece of software, ready to be integrated – or already integrated – in a "production" environment, nor an experiment that aims to establish the mediating effect of some variables over specific behaviors. On the one hand, the software architecture of the DEW is simply outlined in its general features, since the tool is not yet in a stable version. On the other hand, the experimental settings defined for the usability test comprise many elements that are usually applied to "serious" experiments (e.g. a pre-experience, controlled variables, pseudo-independent variables, and so on) but do not investigate the effect of experimental factors on behaviors, not even the effect on emotional awareness of the tool itself. The result is thus a "hybrid" contribution, which – also considering its length – was probably too ambitious in both scope and method.

5.2 Future Developments

In spite of the aforementioned limitations and the partial results of the usability test, we believe that some encouraging elements have nevertheless emerged throughout the comprehensive development process of the DEW. As a conclusion to the entire work, we will put these aspects into the perspective of future developments in the broad meaning – that is, not restrained to "feature development", even if we start with them.

5.2.1 Development of a Version for Experimental Settings

The next step in the development of the DEW requires a stable version of the tool that can be integrated in an experimental design implying participants to share their emotional episodes during a remote collaborative task. For this purpose, we plan to develop an "experimenter-friendly" version of the EAT, which would allow experimenters to easily integrate the tool with the rest of their material. This procedure requires different steps. First, the DEW must be easy to set up without the need of overcomplicated hardware and/or software requirements. Second, the tool must be easy to configure to the particular needs of the experimenter (e.g. subset of emotions, widgets visible on the interface, etc.). Third, the interface of the DEW must allow the integration of the primary collaborative task on the same screen, regardless of the nature of this task. Fourth, experiments must have access to results, and should be able to export and analyze them in the way they prefer. We believe that the most practical and economical way to fulfill these requirements is the development of a web-based admin area, through which experimenters configure the suitable settings, and dispose of a link (not necessarily on the web, it could be a private network or even the localhost) that proposes the complete interface. In the same admin area, experimenters shall also dispose of the results of their test, at least with respect to data collected through the DEW.

5.2.2 Square the Circle: Modifications to the Core Algorithm of the DEW

The core algorithm of the DEW presented two main issues from a structural point of view: the circularity problem of the circumflex, and the ratio-dependent computation of the slope representing the appraisal dimensions. To avoid these problems, we imagined to use scatterplots of emotions (as the one proposed in **Figure 4.6**) instead of a circumflex as the underlying model. The GRID experiment provides in this regard different scatterplots of emotions arranged according to two dimensions (Fontaine & Scherer, 2013), which can be used as basis for an even simpler algorithm that the one used by the DEW. Instead of calculating the angle of a radius starting from the center of the Cartesian plan, we calculate a *confidence-interval surface* around the point established by the x-axis and y-axis appraisals. All the emotions comprised in the squared surface will be part of the subset. End-users will not notice the difference, since the underlying model would not be visible, and therefore the tidy circular order of the GEW does not provide any usability advantage.

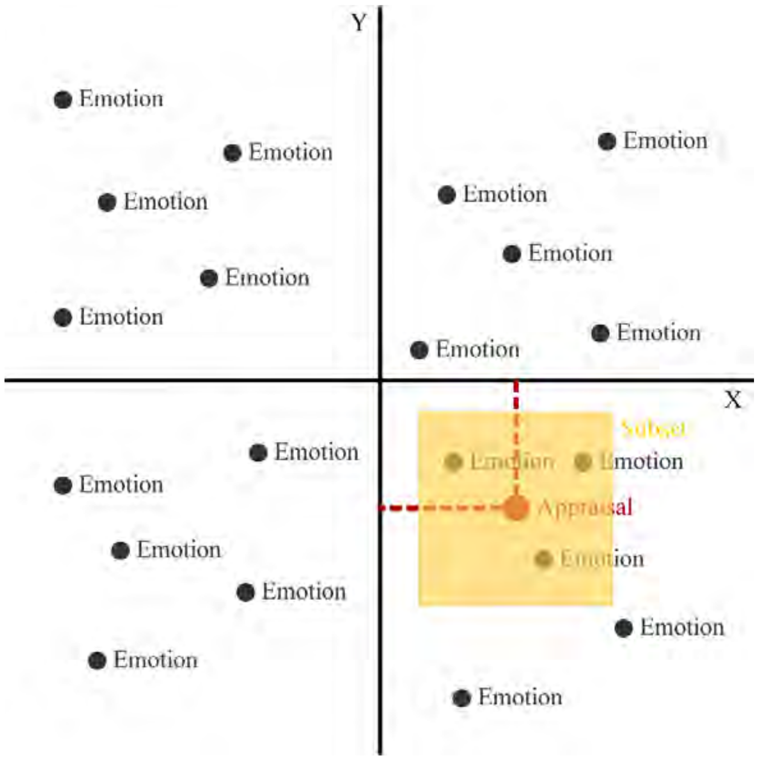


Figure 5.1. Use of squared surfaces instead of circular segment for determining the subset of dynamically generated emotions according to appraisal dimensions.

5.2.3 Use of the DEW as a Self-Report Tool for Measuring Emotions

Considering that the interface of the DEW is modular, we can even imagine the tool deprived from the monitoring functions of awareness, in which case the DEW can eventually be used as a “simple” emotion self-report tool. Even if we do not encourage this use of the tool, since the DEW primary objective does not concern the accuracy and authenticity of the expressed emotions, we reckon that its interface could be useful in situation where participants are supposed to provide many different emotions over a short period. If the scope of the experiment privileges real-time quantity over delayed-time accuracy, considering the perceived usability of the tool, the Dynamic Emotion Wheel could in fact represent an option.

5.3 General Conclusion

This contribution presented the comprehensive development of an Emotion Awareness Tool, whose objective is the enhancement of emotional awareness during computer-mediated collaboration. The developed EAT is supposed to replace a first version of a tool used in the EATMINT project, and its scope is therefore mainly experimental. For this reason, we integrated – in an Interaction Design process – both experimenters and users’ perspectives by adapting an existing emotion self-report tool – the Geneva Emotion Wheel – to the purpose of emotional awareness. More specifically, we took advantage of the GEW’s combination of dimensional (i.e., the Valence and Control dimensions) and discrete emotions approaches to emotion self-reporting. We used this principle as inspiration for the test-driven development of an algorithm that allows the dynamic generation of a subset of discrete emotions according to the users’ evaluation of the situation on two-dimensional values. Since this idea is the core principle of the tool, we named the EAT Dynamic Emotion Wheel, and we shaped the overall interface of the tool upon the aforementioned algorithm. The result is an Emotion Awareness Tool deeply rooted in the Component Process Model definition of an emotion, which shares the same principles of the Geneva Emotion Wheel, applied to an emotion as interaction approach to affective computing. The main objective of the DEW consists, in fact, in voluntary sharing emotions so that participants to a computer-mediated collaborative task may integrate them as meaningful information, regardless of authenticity, in an instrumental perspective.

Considering the DEW, at least initially, an element of the experimental material used in empirical tests, we conducted a usability test in experimental conditions. 16 participants were induced to believe they would collaborate with a same-sex partner in a remote computer-mediated collaborative task consisting in the joint resolution of four enigmas. During the collaborative task, they were asked to share their emotional state through the Dynamic Emotion Wheel, and received in the meantime the emotional episodes recorded in a pre-experiment in which four fellow colleagues executed the very same task, but in a synchronized situation. In order to assess the utility and perceived usability of the tool, we gathered both objective and subjective data. More specifically, we conducted the usability test on a computer screen equipped with an eye tracker in order to assess if users paid attention to the different elements of the DEW. We also registered the emotional episodes they voluntary expressed through the interface, providing a practical demonstration of what kind of data experimenters can collect with the DEW. It is worth reminding that, for the purpose of the test, we arranged the 20 emotions used in EATMINT experiments on a circumflex characterized by the same underlying dimensions of Valence (i.e. “is the situation pleasant?”) and Control (i.e. “is the situation under your control?”) as the GEW.

The usability test provided mixed results. The algorithm at the core of the DEW highlighted some major drawbacks that deeply questioned the utility of the generation of a subset of discrete emotions according to appraisal dimensions. Specially, the use of the two-dimensional sliders was particularly puzzling. On the one hand, the capacity of the algorithm to provide the “authentic” subjective feeling corresponding to the appraised dimensions was weakened by feelings characterized by negative Valence and positive (high) Control. On the other hand, users perceived the sliders more as a nuisance than an opportunity to convey meaningful information. Consistently with the literature, they privileged the expression of emotions through the holistic subjective feeling, which was often judged as “self-sufficient”. The same principle also affected the monitoring function of the DEW, since the emotion timeline (conveying subjective feelings) was fixated more often – suggesting more interest – and perceived more useful than the two line charts (providing appraisal dimensions). The usability test, though, did not provide only negative evidence. On the contrary, the participants appreciated the tool especially in terms of usability (with a mean score of 77 on the System Usability Scale) and manifested an overall positive attitude towards the use of the DEW as Emotion Awareness Tool. We therefore postulate that, in the perspective of an emotion as interaction approach, the DEW could efficiently enhance emotional awareness during computer-mediated collaboration.

6 Bibliography

Andriessen, J., Baker, M., & Van der Puil, C. (2011). Socio-cognitive tension in collaborative working relations. In S. Ludvigsen, A. Lund, I. Rasmussen, & R. Saljo (Eds.), *Learning across sites: new tools, infrastructures and practices*. London, UK: Pergamon.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839. doi:10.1038/nrn1201

Baker, M., Hansen, T., Joiner, R., & Traum, D. (1999). The role of grounding in collaborative learning tasks. In *Collaborative learning: Cognitive and Computational Approaches*.

Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. S. (2002). Computer-Mediated Communication and Group Decision Making: A Meta-Analysis. *Organizational Behavior and Human Decision Processes*, *87*(1), 156–179. doi:10.1006/obhd.2001.2961

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114–123. doi:66.39.39.113

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, *12*(5), 1161–79. doi:10.1037/a0025827

Barsade, S. (2002). The ripple effects: emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(47), 644–75.

Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, *55*(1), 30–40. doi:10.1016/j.bandc.2003.04.001

Belavkin, R. V. (2001). The Role of Emotion in Problem Solving. In *Proceedings of the AISB'01 Symposium on emotion, cognition and affective computing* (pp. 49–57). Heslington, UK.

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, *65*(4), 275–291. doi:10.1016/j.ijhcs.2006.11.016

Bradley, M., & Lang, P. J. (1994). Measuring Emotion : the Self-Assessment Semantic Differential Manikin and the. *J Behav Ther Exp Psychiatry*, *25*(I), 49–59.

Brody, L. R. (2000). The socialization of emotion: Display rules, infant temperament, and differentiation. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 24–47). Cambridge, UK: Cambridge University Press.

Broekens, J., & Brinkman, W. P. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human Computer Studies*, *71*(6), 641–667. doi:10.1016/j.ijhcs.2013.02.003

Brooke, J. (1996). SUS: a quick and dirty usability scale. In B. A. Weerdmester & I. L. McClelland (Eds.), *Usability evaluation in industry*. London, UK: Taylor & Francis.

Brosch, T., Scherer, K. R., Grandjean, D., & Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Medical Weekly*, *143*(May), 1–10. doi:10.4414/smw.2013.13786

Buder, J. (2011). Group awareness tools for learning: Current and future directions. *Computers in Human Behavior*, *27*(3), 1114–1117. doi:10.1016/j.chb.2010.07.012

Caicedo, D., & van Beuzekom, M. (2006). *How do you feel? An assessment of existing tools for the measurement of emotions and their application in consumer product research. Delft University of Technology, Department of Industrial Design*.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*(2), 75–100. doi:10.1037/1076-898X.4.2.75

Carroll, J. (1997). Human–computer interaction: Psychology as a science of design. *International Journal of Human-Computer Studies*, *46*(4), 501–522.

Cereghetti, D. (2013). *Le feedback émotionnel en situation de collaboration médiatisée par ordinateur: étude sur les représentations des utilisateurs*. Geneva University. Retrieved from http://tecfa.unige.ch/tecfa/maltt/memoire/Cereghetti2013.pdf

Chanel, G., Molinari, G., Cereghetti, D., Pun, T., & Bétrancourt, M. (2013). Assessment of Computer-Supported Collaborative Processes using Interpersonal Physiological and Eye-Movement Coupling. In *ACII 2013, 5th Int. Conf. Affective Computing and Intelligent Interaction* (pp. 2–5). Geneva, Switzerland.

Cheshin, A., Rafaeli, A., & Bos, N. (2011). Anger and happiness in virtual teams: Emotional influences of text and behavior on others' affect in the absence of non-verbal cues. *Organizational Behavior and Human Decision Processes*, *116*(1), 2–16. doi:10.1016/j.obhdp.2011.06.002

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (Vol. 13, pp. 127–149). doi:10.1037/10096-006

Cooper, A., Reinman, R., & Cronin, D. (2007). *About Face 3: The Essential of Interaction Design*. Indianapolis, IN: Wiley Publishing Inc.

Crockford, D. (2008). *JavaScript: The Good Parts. Unearthing the Excellence in JavaScript*. Sebastopol, CA: O'Reilly.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. doi:10.1016/j.learninstruc.2011.10.001

Damasio, A. R. (2006). *Descartes' Error*. London, UK: Vintage.

Darras, B. (1998). L'image, une vue de l'esprit. Étude comparée de la pensée figurative et de la pensée visuelle. *Recherche En Communication*, (9), 77–100.

Davidson, R. J., Scherer, K. R., & Goldsmith, H. (Eds.). (2003). *Handbook of Affective Sciences*. New York, NY: Oxford University Press.

Deployd. (2015). Deployd. Retrieved May 9, 2015, from http://deployd.com/

Derks, D., Fischer, A. H., & Bos, A. E. R. (2008). The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior*, *24*(3), 766–785. doi:10.1016/j.chb.2007.04.004

Desmet, P. M. A. (2005). Measuring emotion: Development and application of an instrument to measure emotional responses to products. In M. A. Blythe, A. F. Monk, & P. C. Overbeeke, K., Wright (Eds.), *Funology: From usability to enjoyment* (pp. 111–123). Dordrecht, Netherlands: Kluwer Academic Publishers. doi:10.1007/1-4020-2967-5

Dillenbourg, P. (1999). What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative learning Cognitive and computational approaches*. doi:10.1.1.167.4896

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In *Technology-enhanced Learning* (pp. 3–19). doi:10.1007/978-1-4020-9827-7_1

Doherty, R. (1997). The emotional contagion scale: A measure of individual differences. *Journal of Nonverbal Behavior*, *21*(2), 131–154. doi:10.1023/A:1024956003661

Doise, W., Deschamps, J. C., & Mugny, G. (2010). *Psychologie sociale expérimentale*. Paris, France: Armand Colin.

Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication*, *40*(1-2), 33–60. doi:10.1016/S0167-6393(02)00070-5

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work - CSCW '92*, 107–114. doi:10.1145/143457.143468

Downie, N. (2015). Chart.js | Open source HTML5 Charts for your website. Retrieved March 17, 2015, from http://www.chartjs.org/

Dutta, P., Verma, A., & Prithviraj, J. (2015). Karma- The Test Runner , for Automated Testing of Web Based Applications. *International Journal of Innovative Research in Science & Engineering*, *3*, 31–37.

Eco, U. (2013). *Trattato di semiotica generale* (XX.). Milano: Bompiani.

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

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science (New York, N.Y.)*, *164*(3875), 86–88. doi:10.1126/science.164.3875.86

Eligio, U. X., Ainsworth, S. E., & Crook, C. K. (2012). Emotion understanding and performance during computer-supported collaboration. *Computers in Human Behavior*, *28*(6), 2046–2054.

Ellis, C. A., Gibbs, S. J., & Rein, G. L. (1991). Groupware: some issues and experiences. *Communications of the ACM*, *34*(l), 38–58.

Ellsworth, P. C. (1994). William James and emotion: is a century of fame worth a century of misunderstanding? *Psychological Review*, *101*(2), 222–229. doi:10.1037/0033-295X.101.2.222

Engelmann, T., Dehler, J., Bodemer, D., & Buder, J. (2009). Knowledge awareness in CSCL: A psychological perspective. *Computers in Human Behavior*, *25*(4), 949–960. doi:10.1016/j.chb.2009.04.004

Feidakis, M., Daradoumis, T., & Caballé, S. (2011). Endowing e-learning systems with emotion awareness. *Proceedings - 3rd IEEE International Conference on Intelligent Networking and Collaborative Systems, INCoS 2011*, 68–75. doi:10.1109/INCoS.2011.83

Feinman, S. (1982). Social Referencing in Infancy. *Merrill-Palmer Quarterly*, *28*(4), 445–470. doi:10.2307/23086154

Firebase Inc. (2015). Firebase - Build Realtime Apps. Retrieved May 9, 2015, from https://www.firebase.com/

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. doi:10.1037/0003-066X.34.10.906

Fontaine, J. R. J., & Scherer, K. R. (2013). The global meaning structure of the emotion domain: Investigating the complementary of multiple perspectives on meaning. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of Emotional Meaning: A sourcebook* (pp. 106–128). Oxford, UK: Oxford University Press.

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050–1057.

Fontaine, J. R. J., Scherer, K. R., & Soriano, C. (2013). The why, the what, and the how of the GRID instrument. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of Emotional Meaning: A sourcebook* (pp. 83–97). Oxford, UK: Oxford University Press.

Frijda, N. H., & Scherer, K. R. (2009). Emotion definitions (psychological perspectives). In D. Sander & K. R. Scherer (Eds.), *The Oxford Companion to Emotion and the Affective Sciences*. New York, NY: Oxford University Press.

Gandy, D. (2015). Font Awesome, the iconic font and CSS toolkit. Retrieved March 17, 2015, from http://fontawesome.io/

Garrett, J. J. (2011). *The Elements of User Experience* (Second Edi.). Berkley, CA: New Riders.

Geneva Emotion Research Group. (1988). Appendix F. Labels describing affective states in five major languages. In K. R. Scherer (Ed.), *Facets of emotion: Recent research* (Version re.). Hillsdale, NJ: Erlabaum. Retrieved from http://www.affective-sciences.org/researchmaterial

George, B., & Williams, L. (2004). A structured experiment of test-driven development. *Information and Software Technology*, *46*(5 SPEC. ISS.), 337–342. doi:10.1016/j.infsof.2003.09.011

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors*, *40*(1), 79–91.

Google. (2015). AngularJS - Superheroic JavaScript MVW Framework. Retrieved March 17, 2015, from https://angularjs.org/

Gottesdiener, E. (1995). RAD REALITIES: Beyond the Hype to How RAD Really Works. *Application Development Trends*, *2*(8), 28–38.

Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, *17*(2), 484–95.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences, 281–291. doi:10.1017.S0048577201393198

Gross, T., Stary, C., & Totter, A. (2005). User-centered awareness in computer-supported cooperative work-systems: Structured embedding of findings from social sciences. *International Journal of Human-Computer Interaction*, *18*(3), 323–360.

Grudin, J. (1994). Computer-Supported Cooperative Work: History and Focus. *IEEE Computer*. doi:10.1109/2.291294

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1253–1262). ACM. doi:10.1145/1124772.1124961

Gutwin, C., & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work*, *11*, 411–446. doi:10.1023/A:1021271517844

Gutwin, C., Greenberg, S., & Roseman, M. (1996). Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation. *People and Computers*, 281–298. doi:10.1145/257089.257286

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, *3*(3), 446–474. doi:10.1111/j.1756-8765.2011.01150.x

Howell, D. C. (2010). Simple Analysis of Variance. In D. C. Howell (Ed.), *Statistical Methods for Psychology* (7th Editio., pp. 317–362). Belmont, CA: Wadsworth.

Isomursu, M., Tähti, M., Väinämö, S., & Kuutti, K. (2007). Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *International Journal of Human Computer Studies*, *65*, 404–418. doi:10.1016/j.ijhcs.2006.11.007

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in Human–Computer interaction and usability research: Ready to deliver the promises. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 573–605). doi:10.1016/B978-044451020-4/50031-1

Jacquinot, G. (1993). Apprivoiser la distance et supprimer l'absence ? ou les défis de la formation à distance. *Revue Française de Pédagogie*. doi:10.3406/rfp.1993.1305

Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: do students regulate emotions evoked by social challenges? *The British Journal of Educational Psychology*, *79*(Pt 3), 463–81. doi:10.1348/000709909X402811

Jézégou, A. (2010). Créer de la présence à distance en e-learning. Cadre théorique, définition, et dimensions clés. *Distances et Savoirs*, *8*(2), 257–274. doi:10.3166/ds.8.257-274

Johnson-Laird, P. N. (1980). Mental Models in Cognitive Science. *Cognitive Science*, *4*(1), 71–115. doi:10.1207/s15516709cog0401_4

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In Guilford (Ed.), *The handbook of emotions* (pp. 226–235). New York, NY.

Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer‐mediated communication in distance education. *American Journal of Distance Education*, *9*(2), 7–26. doi:10.1080/08923649509526885

Karau, S., & Williams, K. D. (1993). Interpersonal relations and group social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*(4), 681–706. doi:10.1037/0022-3514.65.4.681

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York, NY: MacGraw-Hill.

Leff, a., & Rayfield, J. T. (2001). Web-application development using the Model/View/Controller design pattern. *Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference*, 118–127. doi:10.1109/EDOC.2001.950428

Leventhal, H., & Scherer, K. (1987). The Relationship of Emotion to Cognition: A Functional Approach to a Semantic Controversy. *Cognition & Emotion*. doi:10.1080/02699938708408361

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556. doi:10.1016/j.jesp.2009.12.019

Lund, K., Molinari, G., Séjourné, A., & Baker, M. (2007). How do argumentation diagrams compare when student pairs use them as a means for debate or as a tool for representing debate? *International Journal of Computer-Supported Collaborative Learning*, *2*(2-3), 273–295. doi:10.1007/s11412-007-9019-z

MacDonald, M. (2013). *HTML5 the missing manual* (2nd editio.). Sebastopol, CA: O'Reilly.

McNeil, M. (2015). Sails.js | Realtime MVC Framework for Node.js. Retrieved May 9, 2015, from http://sailsjs.org/

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, *15*(3), 238–246. doi:10.3758/BF03197722

Molinari, G. (2011). La visualisation partagée des connaissances individuelles pendant la construction collaborative d 'une carte conceptuelle : Quels effets sur l'apprentissage et la perception des connaissances ? *Environnements Informatiques Pour l'Apprentissage Humain*, 341–352.

Molinari, G., Bozelle, C., Cereghetti, D., Chanel, G., Bétrancourt, M., & Pun, T. (2013). Feedback émotionnel et collaboration médiatisée par ordinateur: Quand la perception des interactions est liée aux traits émotionnels. In C. Choquet, P. Dessus, M. Lefevre, J. Broisin, & P. Catteau (Eds.), *Environnements Informatiques pour l'apprentissage humain, Actes de la conférence EIAH 2013* (pp. 305–326). Touluse, France: IRIT Press.

Molinari, G., Chanel, G., Bétrancourt, M., Pun, T., & Bozelle, C. (2013). Emotion feedback during computer-mediated collaboration: Effects on self-reported emotions and perceived interaction. In *Computer-Supported Collaborative Learning Conference, CSCL* (Vol. 1, pp. 336–343). Madison, WI.

Molinari, G., Sangin, M., Dillenbourg, P., & Nüssli, M. A. (2009). Knowledge interdependence with the partner, accuracy of mutual knowledge model and computer-supported collaborative learning. *European Journal of Psychology of Education*, *24*(2), 129–144. doi:10.1007/BF03173006

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759. doi:10.1037/0033-2909.125.6.737

Nickerson, R. S. (2001). The Projective Way of Knowing: A Useful Heuristic That Sometimes Misleads. *Current Directions in Psychological Science*, *10*(5), 168–172. doi:10.1111/1467-8721.00141

Novak, J. D., & Cañas, a J. (2008). The Theory Underlying Concept Maps and How to Construct and Use Them. *IHMC CmapTools*, 1–36. doi:Technical Report IHMC CmapTools 2006-01 Rev 2008-01

Ogata, H., & Yano, Y. (2000). Combining Knowledge Awareness and Information Filtering in an Open-ended Collaborative Learning Environment. *International Journal of Artificial Intelligence in Education (IJAIED)*, *11*, 33–46.

Osgood, C. E. (1952). The Nature and measurement of meaning. *Psychological Bulletin*, *49*(3), 197–237. doi:10.1037/h0021468

Otto, M., & Thornton, J. (2015). Bootstrap. The world's most popular mobile first and responsive front-end framework. Retrieved March 17, 2015, from http://getbootstrap.com/

Pančur, M., & Ciglarič, M. (2011). Impact of test-driven development on productivity, code and tests: A controlled experiment. *Information and Software Technology*, *53*(6), 557–573. doi:10.1016/j.infsof.2011.02.002

Parkinson, B. (1996). Emotions are social. *British Journal of Psychology*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.2044-8295.1996.tb02615.x/full

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, *116*(2), 220–244. doi:10.1037/0033-2909.116.2.220

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. doi:10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. doi:10.1016/j.cedpsych.2010.10.002

Picard, R. W. (2000). *Affective computing*. Cambridge, MA: MIT press.

Picard, R. W. (2009). Affective computing. In D. Sander & K. R. Scherer (Eds.), *The Oxford Companion to Emotion and the Affective Sciences* (pp. 12–15). New York, NY: Oxford University Press.

Pinker, S. (1990). A theory of graph comprehension. *Artificial Intelligence and the Future of Testing*, 73–126. doi:10.1145/2046684.2046699

Poole, A., & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. *Encyclopedia of Human-Computer Interaction*, 211–219. doi:10.4018/978-1-59140-562-7

Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology*, *79*(3), 410–424. doi:10.1037//0022-3514.79.3.410

Rime, B. (2009). Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review. *Emotion Review*, *1*(1), 60–85. doi:10.1177/1754073908097189

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interction Design: Beyond Human-Computer Interaction* (3rd editio.). Chichester, UK: John Wiley & Sons.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *ComputerSupported Collaborative Learning* (pp. 69–97). doi:10.1145/130893.952914

Roseman, I. J., & Smith, C. A. (2001). Appraisal Theory. Overview, Assumptions, Varieties, Controversies. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal Processes in Emotion. Theory, Methods, Research* (pp. 3–19). New York, NY: Oxford University Press.

Rosenbaum, A. (2009). *Leçons d'introduction à la philosophie des sciences*. Paris, France: Les Presses de l'ENSTA.

Rouet, J.-F., & Tricot, A. (1995). Recherche d'informations dans les systèmes hypertextes : des représentations de la tâche à un modèle de l'activité cognitive. *Revue Sciences et Techniques Éducatives*, *2*(3), 307–331.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*. doi:10.1037/h0077714

Sacharin, V., Schlegel, K., & Scherer, K. R. (2012). *Geneva Emotion Wheel rating study (Report)*.

Saffer, D. (2013). *Microinteractions*. Sebastopol, CA: O'Reilly.

Sangin, M., Molinari, G., Nüssli, M.-A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior*, *27*(3), 1059–1067. doi:10.1016/j.chb.2010.05.032

Scapin, D. L., & Bastien, J. M. C. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, *16*(4-5), 220–231. doi:10.1080/014492997119806

Scherer, K. R. (1982). Emotion as a process: Function, origin and regulation. *Social Science Information*, *21*(4-5), 555–570.

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. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal Processes in Emotion. Theory, Methods, Research* (pp. 92–120). New York, NY: Oxford University Press.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.

Scherer, K. R. (2009). Component process model. In D. Sander & K. R. Scherer (Eds.), *The Oxford Companion to Emotion and the Affective Sciences* (pp. 93–94). New York, NY: Oxford University Press.

Scherer, K. R. (2010). The component process model: Architecture for a comprehensive computational model of emrgent emotion. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for Affective Computing. A Sourcebook* (pp. 46–70). Oxford, UK: Oxford University Press.

Scherer, K. R. (2013). Measuring the meaning of emotion words: A domain-specific componential approach. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of Emotional Meaning: A sourcebook* (pp. 7–30). Oxford, UK: Oxford University Press.

Scherer, K. R., & Ceschi, G. (1997). Lost Luggage : A Field Study of Emotion-Antecedent Appraisal. *Motivation and Emotion*, *21*(3), 211–235.

Scherer, K. R., Fontaine, J. R. J., & Soriano, C. (2013). CoreGRID and MiniGRID: Development and validation of two short versions of the GRID instrument. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of Emotional Meaning: A sourcebook2* (pp. 523–541). Oxford, UK: Oxford University Press.

Scherer, K. R., & Meuleman, B. (2013). Human Emotion Experiences Can Be Predicted on Theoretical Grounds: Evidence from Verbal Labeling. *PLoS ONE*, *8*(3). doi:10.1371/journal.pone.0058166

Scherer, K. R., Shuman, V., Fontaine, J. R. J., & Soriano, C. (2013). The GRID meets the Wheel: Assessing emotional feeling via self-report. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of Emotional Meaning: A sourcebook* (pp. 281–298). Oxford, UK: Oxford University Press.

Schmidt, K. (2002). The Problem with “Awareness.” *Computer Supported Cooperative Work*, *11*, 285–298.

Schorr, A. (2001). Appraisal. The Evolution of an Idea. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal Processes in Emotion. Theory, Methods, Research* (pp. 20–34). New York, NY: Oxford University Press.

Shaver, P., Schwartz, J., Kirson, D., & O’Connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3598857

Siemer, M., Mauss, I., & Gross, J. J. (2007). Same situation-different emotions: how appraisals shape our emotions. *Emotion*, *7*(3), 592–600.

Smith, C. a., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences*.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. doi:10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive Load Theory , Learning Difficulty , and Instructional Design. *Learning and Instruction*, *4*, 295–312. doi:10.1016/0959-4752(94)90003-5

The jQuery Foundation. (2015). jQuery. Retrieved March 17, 2015, from http://jquery.com/

Tobii AB. (2015). Tobii Studio. Danderyd. Sweden.

Tran, V., Páez, D., & Sánchez, F. (2012). Emotions and Decision-Making Processes in Management Teams: A Collective Level Analysis. *Revista de Psicología Del Trabajo Y de La Organizaciones*, *28*(1), 15–24. Retrieved from http://scielo.isciii.es/pdf/rpto/v28n1/original2.pdf

Tullis, T. S., & Albert, B. (2013). *Measuring the User Experience. Collecting, Analyzing, and Presenting Usability Metrics* (Second edi.). Waltham, MA: Morgan Kaufmann.

Tullis, T. S., & Stetson, J. N. (2004). A Comparison of Questionnaires for Assessing Website Usability. *Usability Professional Association Conference*, 1–12.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation : can it facilitate ? *International Journal of Human-Computer Studies*, 247–262. doi:10.1006/ijhc.1017

Valstar, M. F., Mehu, M., Pantic, M., & Scherer, K. R. (2012). Meta-Analysis of the First Facial Expression Recognition Challenge. *IEEE Transactrions on Systems, Man, and Cybernetics*, *42*(4), 966–979. doi:10.1109/TSMCB.2012.2200675

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science*, *18*(3), 184–188. doi:10.1111/j.1467-8721.2009.01633.x

Van Kleef, G. A. (2010). The Emerging View of Emotion as Social Information. *Social and Personality Psychology Compass*, *4*(5), 331–343.

Van Kleef, G. A., Anastasopoulou, C., & Nijstad, B. a. (2010). Can expressions of anger enhance creativity? A test of the emotions as social information (EASI) model. *Journal of Experimental Social Psychology*, *46*(6), 1042–1048. doi:10.1016/j.jesp.2010.05.015

W3C. (2014). HTML5. W3C Reccommendation 28 October 2014. Retrieved March 17, 2015, from http://www.w3.org/TR/2014/REC-html5-20141028/

W3C. (2015). Cascading Style Sheets. Retrieved March 17, 2015, from http://www.w3.org/Style/CSS/

Wikipedia. (2015). WebSocket. Retrieved May 9, 2015, from http://en.wikipedia.org/w/index.php?title=WebSocket&oldid=658881077

William Albert, D. T. (2010). Reliability of Self-Reported Awareness Measures Based on Eye Tracking. *Journal of Usability Studies*, *5*(2), 50–64.

7 Annexes

7.1 Survey

Dynamic Emotion Wheel : test d'interface d'un outil d'awareness émotionnelle

LAISSER VIDE **Code participant** [ ] **Date de passation** [ ] **Interface** [ ]

Dans ce questionnaire vous allez répondre à des questions qui portent sur le test d'interface que vous venez d'accomplir. Les questions peuvent se référer exclusivement à l'outil d'awareness émotionnelle ou a l'ensemble de la tâche expérimentale. Parfois, les questions font références à des zones spécifiques de l'écran. À ce propos, vous disposez d'une feuille avec une capture d'écran de l'interface qui propose également une legende à laquelle vous pouvez faire référence pour répondre à ces questions. L'expérimentateur reste à votre disposition pour toute explication ultérieure.

Veuillez d'abord répondre à quelques questions sur vous.

Vous êtes

- O Une femme
- O Un homme

Quel est votre âge ?

[ ]

*Les questions suivantes se réfèrent **exclusivement à l'outil d'awareness émotionnelle** situé sur la gauche de l'écran lors du test. Les questions font référence à une zone spécifique de l'outil. Référez-vous à ce propos à la feuille avec la capture d'écran de l'interface.*

Dans l'échelle numérique proposée, 1 = Pas du tout d'accord et 7 = Tout à fait d'accord.

Questions sur la **ZONE A** de l'interface de l'outil d'awareness émotionnelle (expression des émotions)

	1	2	3	4	5	6	7
La manipulation de cette zone m'a posé beaucoup de problèmes	O	O	O	O	O	O	O
L'utilisation de cette zone est assez intuitive.	O	O	O	O	O	O	O
J'ai eu l'impression de répéter trop souvent les mêmes actions.	O	O	O	O	O	O	O
J'ai trouvé difficile d'évaluer la VALENCE dans les critères d'évaluation de la situation (premier "slider").	O	O	O	O	O	O	O
J'ai trouvé difficile d'évaluer la MAÎTRISE dans les critères d'évaluation de la situation (deuxième "slider").	O	O	O	O	O	O	O
J'ai parfois modifié les critères d'évaluation avec l'objectif de générer un bouton avec une émotion précise.	O	O	O	O	O	O	O
J'ai toujours respecté l'ordre de manipulation indiqué pour les critères d'évaluation.	O	O	O	O	O	O	O
J'ai parfois cliqué sur un des boutons parce que l'émotion était assez proche de celle que j'avais en tête.	O	O	O	O	O	O	O
Les boutons générés dynamiquement m'ont permis de gagner du temps dans l'expression des émotions.	O	O	O	O	O	O	O
J'aurais préféré que toutes les émotions s'affichent exclusivement sous forme de boutons.	O	O	O	O	O	O	O
J'aurais préféré que toutes les émotions s'affichent exclusivement dans le menu déroulant.	O	O	O	O	O	O	O
En général, j'ai bien compris le fonctionnement de cette zone.	O	O	O	O	O	O	O
En général, cette zone a facilité l'expression de mes émotions.	O	O	O	O	O	O	O

Questions sur la **ZONE B** de l'interface de l'outil d'awareness émotionnelle (timeline émotionnelle)

	1	2	3	4	5	6	7
J'ai trouvé cette zone inutile.	O	O	O	O	O	O	O
J'ai bien compris l'information affichée dans cette zone.	O	O	O	O	O	O	O
Cette zone a été utile pour comprendre mon ressenti émotionnel.	O	O	O	O	O	O	O
Cette zone a été utile pour comprendre le ressenti émotionnel de l'autre participant.	O	O	O	O	O	O	O
Cette zone a été utile pour comparer mon ressenti émotionnel avec celui de l'autre participant.	O	O	O	O	O	O	O
Cette zone a été utile pour suivre l'évolution du ressenti émotionnel dans le temps.	O	O	O	O	O	O	O

Questions sur la **ZONE C** de l'interface de l'outil d'awareness émotionnelle (premier graphique)

	1	2	3	4	5	6	7
J'ai trouvé cette zone inutile.	O	O	O	O	O	O	O
J'ai bien compris l'information affichée dans cette zone.	O	O	O	O	O	O	O
Cette zone a été utile pour comprendre mon ressenti émotionnel.	O	O	O	O	O	O	O
Cette zone a été utile pour comprendre le ressenti émotionnel de l'autre participant.	O	O	O	O	O	O	O
Cette zone a été utile pour comparer mon ressenti émotionnel avec celui de l'autre participant.	O	O	O	O	O	O	O
Cette zone a été utile pour suivre l'évolution du ressenti émotionnel dans le temps.	O	O	O	O	O	O	O

Questions sur la **ZONE D** de l'interface de l'outil d'awareness émotionnelle (deuxième graphique)

	1	2	3	4	5	6	7
J'ai trouvé cette zone inutile.	O	O	O	O	O	O	O
J'ai bien compris l'information affichée dans cette zone.	O	O	O	O	O	O	O
Cette zone a été utile pour comprendre mon ressenti émotionnel.	O	O	O	O	O	O	O
Cette zone a été utile pour comprendre le ressenti émotionnel de l'autre participant.	O	O	O	O	O	O	O
Cette zone a été utile pour comparer mon ressenti émotionnel avec celui de l'autre participant.	O	O	O	O	O	O	O
Cette zone a été utile pour suivre l'évolution du ressenti émotionnel dans le temps.	O	O	O	O	O	O	O

*Les questions suivantes se réfèrent à l'**ensemble de l'expérience** que vous venez de passer, y compris la tâche de résolution de problème que vous avez effectuée lors du test d'interface.*

Dans l'échelle numérique proposée, 1 = Extrêmement bas(se) et 7 = Extrêmement élevé(e)

Evaluez votre position sur chaque dimension, concernant l'ensemble de l'étude.

	1	2	3	4	5	6	7
L'effort mental nécessaire pour résoudre les énigmes.	O	O	O	O	O	O	O
L'effort mental nécessaire pour exprimer/tenir compte des émotions pendant la résolution des énigmes.	O	O	O	O	O	O	O
L'effort mental global nécessaire pour mener à bien l'expérience dans son ensemble.	O	O	O	O	O	O	O
La difficulté dans la résolution des énigmes	O	O	O	O	O	O	O
La difficulté d'exprimer/tenir compte des émotions pendant la résolution des énigmes	O	O	O	O	O	O	O
La difficulté globale pour mener à bien l'expérience dans son ensemble.	O	O	O	O	O	O	O
L'intérêt de résoudre des énigmes avec un coéquipier.	O	O	O	O	O	O	O
L'intérêt d'exprimer/tenir compte des émotions pendant la résolution des énigmes avec un coéquipier.	O	O	O	O	O	O	O
L'intérêt global de l'expérience dans son ensemble	O	O	O	O	O	O	O

*Nous cherchons maintenant à connaître votre avis sur l'**intégration des informations concernant votre propre ressenti émotionnel et celui de votre coéquipier(ère)** pendant la tâche de résolution des énigmes.*

Dans l'échelle numérique proposée, 1 = Très peu et 7 = Beaucoup

En ce qui vous concerne, avec quelle fréquence avez-vous émis ces différents comportements ?

	1	2	3	4	5	6	7
Comprendre les émotions de votre partenaire.	O	O	O	O	O	O	O
Essayer d'adapter votre comportement aux émotions de votre partenaire.	O	O	O	O	O	O	O
Comprendre vos propre émotions.	O	O	O	O	O	O	O
Prendre en compte vos émotions pour améliorer votre comportement.	O	O	O	O	O	O	O
Se demander comment votre partenaire va réagir en découvrant vos émotions.	O	O	O	O	O	O	O
Comparer vos émotions à celles de votre partenaire.	O	O	O	O	O	O	O
Se demander comment paraître apte à maîtriser vos émotions.	O	O	O	O	O	O	O

D'après vous, avec quelle fréquence votre coéquipier(ère) a-t-il(elle) émis ces différents comportements ?

	1	2	3	4	5	6	7
Comprendre vos émotions.	O	O	O	O	O	O	O
Essayer d'adapter son comportement à vos émotions.	O	O	O	O	O	O	O
Comprendre ses propres émotions.	O	O	O	O	O	O	O
Prendre en compte ses émotions pour améliorer son comportement.	O	O	O	O	O	O	O
Se demander comment vous allez réagir en découvrant ses émotions.	O	O	O	O	O	O	O
Comparer ses émotions aux vôtres.	O	O	O	O	O	O	O
Se demander comment paraître apte à maîtriser ses émotions.	O	O	O	O	O	O	O

*Enfin, nous amerions recueillir vos impression générales sur l'**outil d'awareness émotionelle** que vous avez testé. Pour exprimer votre avis, nous vous prions de tenir également compte du contexte d'utilisation. En d'autre termes, nous vous prions d'évaluer l'outil en fonction du fait qu'il s'intègre à une autre tâche collaborative médiatisée par ordinateur.*

Dans l'échelle numérique proposée, 1 = Pas du tout d'accord et 7 = Tout à fait d'accord.

Veuillez évaluer votre degré d'accord avec les propositions suivantes.

	1	2	3	4	5	6	7
J'aimerais utiliser cet outil fréquemment.	O	O	O	O	O	O	O
Je trouve cet outil inutilement complexe.	O	O	O	O	O	O	O
Je pense que cet outil est facile à utiliser.	O	O	O	O	O	O	O
J'aurais besoin d'un support technique pour pouvoir utiliser cet outil.	O	O	O	O	O	O	O
Les différentes fonctionnalités de cet outil sont bien intégrées.	O	O	O	O	O	O	O
Cet outil est truffé d'incohérences.	O	O	O	O	O	O	O
Le grand public peut apprendre à utiliser cet outil très rapidement.	O	O	O	O	O	O	O
Cet outil est lourd à utiliser.	O	O	O	O	O	O	O
J'ai confiance en cet outil.	O	O	O	O	O	O	O
J'ai dû apprendre beaucoup de choses avant de pouvoir utiliser cet outil.	O	O	O	O	O	O	O

Veuillez décrire quel sont, à votre avis, les **points forts** de l'outil d'awareness émotionnelle.

Veuillez décrire quel sont, à votre avis, les **points faibles** de l'outil d'awareness émotionnelle.

Merci de votre participation !

7.2 Outputs

Hypotheses A1

T-tests; Grouping: Gender (RAW emotions in MF_DEW_Usability_Test.stw)
Group 1: F
Group 2: M

Variable	Mean F	Mean M	t-value	df	p	Valid N F	Valid N M	Std.Dev. F	Std.Dev. M	F-ratio Variances	p Variances	Mean 1 - Mean 2	Confidence -95,000%	Confidence +95,000%
X	7,6187	25,06711	-3,17615	286	0,001656	139	149	49,12147	44,09138	1,241182	0,196651	-17,4484	-28,2614	-6,6355
Y	-21,1511	6,24161	-4,32268	286	0,000021	139	149	58,69140	48,66830	1,454308	0,025456	-27,3927	-39,8657	-14,9197

Aggregate Results
Correlations (Spreadsheet in Workbook2_(Recovered).stw)
Marked correlations are significant at p < ,05000

Variable	Gender	Elapsed_seconds
X	F	0,120395
Y	F	-0,153495
X	M	-0,076477
Y	M	-0,295877

Hypotheses A2

Correlations (RAW emotions in MF_DEW_Usability_Test.stw)
Marked correlations are significant at p < ,05000
N=288 (Casewise deletion of missing data)

Variable	Means	Std.Dev.	X	Y
X	16,64583	47,31818	1,000000	0,542442
Y	-6,97917	55,36952	0,542442	1,000000

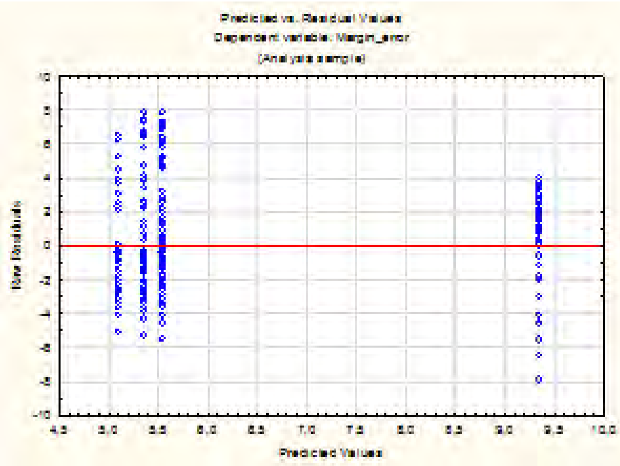
Hypotheses A3

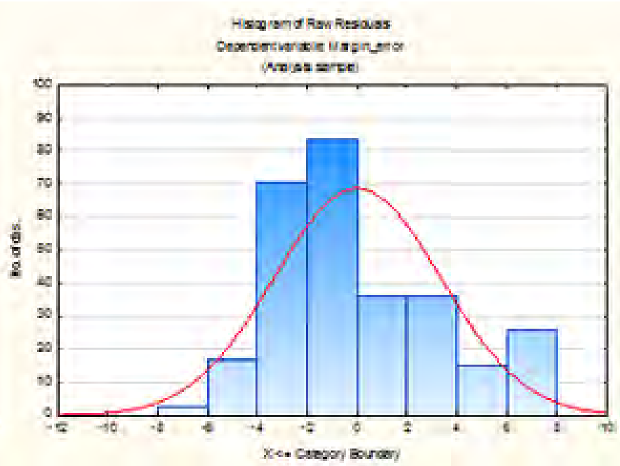
Correlations (RAW emotions in MF_DEW_Usability_Test.stw)
Marked correlations are significant at p < ,05000
N=288 (Casewise deletion of missing data)

Variable	Means	Std.Dev.	Emotion_angle	Slope_angle
Emotion_angle	146,6875	97,37789	1,000000	0,544860
Slope_angle	134,0660	80,07231	0,544860	1,000000

Univariate Tests of Significance, Effect Sizes, and Powers for Margin_error (RAW emotions in MF_DEW_Usability_Test.stw)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	SS	Degr. of Freedom	MS	F	p	Partial eta-squared	Non-centrality	Observed power (alpha=0,05)
Intercept	9432,157	1	9432,157	835,7080	0,000000	0,746362	835,7080	1,000000
Emotion_quadrant	532,162	3	177,387	15,7169	0,000000	0,142384	47,1507	0,999989
Error	3205,345	284	11,286					





Hypotheses A5

T-test for Dependent Samples (Questionnaire and cumultated emotions in MF_DEW_Usability_Test.stw)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
RatioLastX	0,378937	0,202329								
RatioLastY	0,597915	0,204380	16	-0,218979	0,396034	-2,21171	15	0,042924	-0,430010	-0,007947

Hypotheses B

T-test for Dependent Samples (Questionnaire and cumultated emotions in MF_DEW_Usability_Test.stw)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
RatioButton	0,556402	0,182028								
RatioOther	0,443598	0,182028	16	0,112805	0,364056	1,239423	15	0,234231	-0,081187	0,306797

Hypotheses C

Correlations (Eye Tracking in MF_DEW_Usability_Test.stw)
Marked correlations are significant at p < ,05000
N=12 (Casewise deletion of missing data)

Variable	Means	Std.Dev.	TimePerEmotion	ClickPerEmotion
TimePerEmotion	9,803423	2,547247	1,000000	0,798911
ClickPerEmotion	4,001024	0,697196	0,798911	1,000000

Hypotheses E

T-test for Dependent Samples (Eye Tracking in MF_DEW_Usability_Test.stw)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
Fixation Count_Test_Emotion Timeline_Mean	161,7500	69,81029								
Fixation Count Two Line Chart	77,5833	32,43583	12	84,16667	65,48814	4,452132	11	0,000975	42,55749	125,7758

Hypotheses E

T-test for Dependent Samples (Questionnaire and cumultated emotions in MF_DEW_Usability_Test.stw)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
B1	3,000000	1,788854								
C1	4,875000	2,125245	16	-1,87500	2,655184	-2,82466	15	0,012806	0,46015	3,28985
B1	3,000000	1,788854								
D1	4,750000	2,236068	16	-1,75000	2,620433	-2,67131	15	0,017432	0,35367	3,14633

Hypotheses F

T-tests; Grouping: InterfaceType (Eye Tracking in MF_DEW_Usability_Test.stw)
Group 1: participants
Group 2: dimensions

Variable	Mean participants	Mean dimensions	t-value	df	p	Valid N participants	Valid N dimensions	Std.Dev. participants	Std.Dev. dimensions	F-ratio Variances	p Variances
Fixation Count_Test_Top Line Chart_Mean	43,33333	49,66667	-0,577137	10	0,576608	6	6	19,95662	18,00741	1,228207	0,827041
Fixation Count_Test_Bottom Line Chart_Mean	34,83333	27,33333	0,585047	10	0,571480	6	6	24,35912	19,81582	1,511121	0,661568

Aggregate Results
T-test for Dependent Samples (Spreadsheet in MF_DEW_Usability_Test.stw)
Marked differences are significant at p < ,05000

Variable	InterfaceType	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
Other Line Chart	participants	34,83333	24,35912								
Self Line Chart	participants	43,33333	19,95662	6	-8,50000	31,34805	-0,664177	5	0,535974	-41,3978	24,39776
Control Line Chart	dimensions	27,33333	19,81582								
Valence Line Chart	dimensions	49,66667	18,00741	6	-22,3333	10,98484	-4,98007	5	0,004175	-33,8612	-10,8055

Variable	T-test for Dependent Samples (Questionnaire and cumultated emotions in MF_DEW_Usability_Test.stw) Marked differences are significant at p < ,05000 Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
B1	3,000000	1,788854								
C1	4,875000	2,125245	16	-1,87500	2,655184	-2,82466	15	0,012806	0,46015	3,28985
B1	3,000000	1,788854								
D1	4,750000	2,236068	16	-1,75000	2,620433	-2,67131	15	0,017432	0,35367	3,14633
B2	5,500000	1,673320								
C2	3,750000	2,489980	16	1,75000	2,569047	2,72475	15	0,015661	-3,11895	-0,38105
B2	5,500000	1,673320								
D2	3,687500	2,414367	16	1,81250	2,561738	2,83011	15	0,012666	-3,17755	-0,44745
B3	3,625000	1,995829								
C3	2,250000	1,527525	16	1,37500	2,156386	2,55056	15	0,022173	-2,52406	-0,22594
B3	3,625000	1,995829								
D3	1,500000	1,032796	16	2,12500	2,093641	4,05991	15	0,001026	-3,24062	-1,00938
B4	5,062500	2,143789								
C4	1,812500	1,641899	16	3,25000	2,768875	4,69505	15	0,000288	-4,72543	-1,77457
B4	5,062500	2,143789								
D4	2,937500	2,322893	16	2,12500	3,480900	2,44190	15	0,027475	-3,97984	-0,27016
B5	5,437500	1,824600								
C5	3,000000	2,449490	16	2,43750	2,682505	3,63466	15	0,002446	-3,86691	-1,00809
B5	5,437500	1,824600								
D5	2,875000	2,418677	16	2,56250	2,707243	3,78614	15	0,001793	-4,00509	-1,11991
B6	4,437500	2,159282								
C6	3,187500	2,455436	16	1,25000	2,620433	1,90808	15	0,075709	-2,64633	0,14633
B6	4,437500	2,159282								
D6	3,062500	2,434988	16	1,37500	2,629956	2,09129	15	0,053931	-2,77640	0,02640
C1	4,875000	2,125245								
D1	4,750000	2,236068	16	0,12500	0,619139	0,80757	15	0,431957	-0,45492	0,20492
C2	3,750000	2,489980								
D2	3,687500	2,414367	16	0,06250	0,442531	0,56493	15	0,580471	-0,29831	0,17331
C3	2,250000	1,527525								
D3	1,500000	1,032796	16	0,75000	1,390444	2,15758	15	0,047585	-1,49092	-0,00908
C4	1,812500	1,641899								
D4	2,937500	2,322893	16	-1,12500	1,995829	-2,25470	15	0,039529	0,06150	2,18850
C5	3,000000	2,449490								
D5	2,875000	2,418677	16	0,12500	0,341565	1,46385	15	0,163876	-0,30701	0,05701
C6	3,187500	2,455436								
D6	3,062500	2,434988	16	0,12500	0,806226	0,62017	15	0,544451	-0,55461	0,30461

Hypotheses I

Variable	T-test for Dependent Samples (Questionnaire and cumultated emotions in MF_DEW_Usability_Test.stw) Marked differences are significant at p < ,05000 Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
EXP1	5,000000	1,032796								
EXP2	4,312500	1,778342	16	0,68750	1,990603	1,38149	15	0,187361	-0,37322	1,748217
EXP1	5,000000	1,032796								
EXP3	4,937500	1,340087	16	0,06250	1,526161	0,16381	15	0,872068	-0,75073	0,875734
EXP2	4,312500	1,778342								
EXP3	4,937500	1,340087	16	-0,62500	1,024695	-2,43975	15	0,027591	-1,17102	-0,078979
EXP4	5,437500	1,547848								
EXP5	4,937500	1,691892	16	0,50000	1,788854	1,11803	15	0,281139	-0,45321	1,453213
EXP4	5,437500	1,547848								
EXP6	4,500000	1,591645	16	0,93750	1,388944	2,69989	15	0,016462	0,19738	1,677616
EXP5	4,937500	1,691892								
EXP6	4,500000	1,591645	16	0,43750	1,590335	1,10040	15	0,288506	-0,40993	1,284930
EXP7	5,000000	2,129163								
EXP8	4,125000	1,857418	16	0,87500	1,857418	1,88434	15	0,079049	-0,11475	1,864748
EXP7	5,000000	2,129163								
EXP9	5,562500	1,364734	16	-0,56250	1,504161	-1,49585	15	0,155434	-1,36401	0,239011
EXP8	4,125000	1,857418								
EXP9	5,562500	1,364734	16	-1,43750	1,504161	-3,82273	15	0,001664	-2,23901	-0,635989