©2013
Roberto I. González-Ibáñez

ALL RIGHTS RESERVED

A STUDY OF POSITIVE AND NEGATIVE AFFECTIVE STATES IN COLLABORATIVE INFORMATION SEEKING

By

ROBERTO I. GONZÁLEZ-IBÁÑEZ

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Communication, Information and Library Studies

written under the direction of

Dr. Chirag Shah

and approved by

____

____

____

____

New Brunswick, New Jersey

October, 2013

ABSTRACT OF THE DISSERTATION

A Study of Positive and Negative Affective States in Collaborative Information Seeking

by ROBERTO I. GONZÁLEZ-IBÁÑEZ

Dissertation Director:

Dr. Chirag Shah

Emotions and other affective processes have long been considered essential elements in people's lives. Whether during intimacy or in social contexts, human beings experience a wide spectrum of emotions every day, all the time. Despite emotion research conducted in various domains, little is known about the role of affects, emotions, feelings, and mood in the information search process, especially when this is carried out by teams. In this regard, this dissertation aimed to understand whether the affective dimension plays a role in collaborative information seeking (CIS) through four research objectives: (1) study how initial affective conditions influence information practices; (2) investigate what affective processes are typically expressed and experienced in information search; (3) examine how initial affective conditions and those derived from social interactions during the collaboration process influence team performance; and (4) study positivity ratio in collaborative search and their relation to team performance. To accomplish these research objectives, a controlled lab study with 135 participants distributed in fixed experimental conditions and a control group was conducted. In each experimental condition, participants were individually treated with affective stimuli in order to elicit positive and negative affective states.

Results from this study suggest that initial affective states may define and/or shape information processing strategies. Additionally, in collaborative settings, it was found that the interplay of similar or different affective processes could change the way searchers interact with each other, their frustration levels, affective load, and the quality of their work. This dissertation and the findings presented have theoretical implications in the study of collaborative and individual information seeking. Specifically, it gives the affective dimension a central role that could define the way people search, evaluate, and make sense of information. In terms of practical implications, if affective processes play such a key role in information seeking, this may redefine the design of information system by incorporating the ability to identify searchers' initial affective states and provide the necessary resources to support their information processing strategies. Finally, this dissertation also contributes with a research framework and a methodological approach to carry out experimental evaluations to investigate the role of the affective dimension in both collaborative and individual information seeking.

ACKNOWLEDGEMENT

Seven years ago, when I graduated from a Master program in Chile I thought about the possibility to pursue an academic career. It was not easy at first, but somehow things worked out and I undertook a trip to USA in order to learn, grow on a personal level, and contribute to the scientific development of my country, Chile. In the process there have been several people who have provided me with support, help, and guidance making this journey more bearable. In this regard I would like to thank the following people and organizations that in some way have supported me over the last five years. Without them, completing my studies would not have been possible. I apologize in advance if I accidentally do not mention someone.

- The Fulbright program in Chile for recognizing my work and giving me the opportunity to continue my graduate studies in USA.
- The National Commission for Scientific and Technological Research (CONICYT) for providing resources that undoubtedly allowed me to concentrate on completing my PhD studies.
- The Institute of International Education (IIE) for providing administrative support during my PhD and English studies.
- Rutgers University and especially the School of Communications and Information (SC&I), professors with whom I was fortunate to take classes and collaborate. Also to the administrative and technical staff, school directors, and Deans for their willingness to help.
- The Universidad de Santiago de Chile (USACH) for the training I received there and their support to continue my academic career.

- Committee members of my qualifying exam, Dr. Nicholas Belkin, Dr. Chirag Shah, Dr. Smaranda Muresan, and Dr. Mark Aakhus, who provided me valuable feedback to prepare my dissertation pre-proposal.
- Members of my dissertation committee, Dr. Chirag Shah, Dr. Nicholas Belkin, Dr. Mark Aakhus, and Dr. Ryen White, who contributed with their points of view, suggestions, and guidance in the development of my dissertation.
- The Institute of Museum and Library Services (IMLS) grant # RE-04-12-0105-12, and Yahoo! Campus Innovation Award 2011 that funded my dissertation work.
- My thesis advisor, Dr. Chirag Shah for providing me constant support and availability to carry out my research in a reasonable amount of time. Also thank his friendship and that of his family Lori and Sophie.
- My lab mates and the InfoSeeking group, with whom I shared during the last part of my dissertation process: Vanessa, Erik, Chathra, and Laure. Special thanks to Vanessa and Serife for helping me in proofreading my work.
- Dr. Chirag Shah, Muge Haseki, Dr. Smaranda Muresan, Dr. Nina Wacholder, Dr. Ryen White, Natalia Cordova-Rubio, Dr. Nicholas Belkin, and Dr. Jingjing Liu, with whom I collaborated on writing scientific articles that played a fundamental part in the development of my dissertation.
- My former advisor, Dr. Edmundo Leiva who guided me and motivated me to pursue an academic career.
- My friends in Chile for their unconditional friendship during these years.
- My friends and amazing people that I had the fortune to meet in USA: Patricio, Rosita, Natalia, Francisco, Brian, Sandra, Leonel, Karina, and Jessica Lingel.

- My parents, Roberto González and Edith Ibáñez, my siblings Paula, Marco and their families (Richard, Antonella, Giovanni, Paolo, Paola, Benjamin, Josefa), my grandmother Elisa, my parents in law Alejandro and Marta and all my family in Chile for their affection and support despite the distance. Thanks for their visits and keeping in touch, without their presence I might not have had the strength to finish.
- Finally, to my beloved wife Ema López, for being with me every day, for giving me strength, for understanding me, for loving me, for being a wonderful person, and for her courage to undertake this journey with me.

AGRADECIMIENTOS

Hace siete años, cuando completé el programa de magister en informática en la Universidad de Santiago de Chile pensé en la posibilidad de continuar una carrera académica. No fue fácil al principio, pero de alguna forma las cosas se fueron dando y me permitieron emprender este viaje a USA con el objetivo de aprender, crecer en el ámbito personal, y contribuir al desarrollo científico de mi país, Chile. En el proceso ha habido muchas personas que me han apoyado, ayudado y orientado haciendo que esta travesía sea más llevadera considerando las dificultades inherentes en un programa de doctorado. En este sentido me gustaría agradecer a las siguientes personas y organizaciones que de alguna forma me han apoyado durante los últimos cinco años, sin ellos/ellas completar el programa de doctorado no hubiese sido posible. Pido disculpa de antemano si por accidente olvido mencionar a alguien.

- Al programa Fulbright en Chile por reconocer mi trabajo y ofrecerme la oportunidad de realizar mis estudios de doctorado en USA.
- A la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) por proveer recursos que sin lugar a dudas permitieron concentrarme en completar en mi doctorado.
- Al Instituto Internacional de Educación (IIE) por brindar apoyo administrativo y facilitar muchos de los procesos relacionados con el doctorado y mis estudios de inglés.
- A la Universidad Rutgers y en especial a la Escuela de Comunicaciones e Información (SC&I), a los profesores con los quienes tuve la fortuna de tomar clases y colaborar y quienes me introdujeron a las ciencias de la información, bibliotecología, ciencias cognitivas, y las metodologías de investigación.

También al cuerpo administrativo y técnico, directores, y decanos por asistirme cuando lo necesité.

- A la Universidad de Santiago de Chile (USACH) por la formación que recibí y por abrir sus puertas para integrarme al cuerpo académico tras regresar a Chile.
- A los miembros del comité de mi examen de calificación, Dr. Nicholas Belkin, Dr. Chirag Shah, Dr. Smaranda Muresan, y Dr. Mark Aakhus, quienes también me brindaron retroalimentación muy valiosa para la presentación de mi propuesta de tesis.
- A los miembros del comité de mi tesis, Dr. Chirag Shah, Dr. Nicholas Belkin, Dr. Mark Aakhus, y Dr. Ryen White, quienes han contribuido diferentes puntos de vista y sugerencias que me orientaron en el desarrollo de mi tesis doctoral.
- Al Institute of Museum and Library Services (IMLS) grant # RE-04-12-0105-12, y Yahoo! Campus Innovation Award 2011 por financiar el desarrollo de mi tesis.
- A mi profesor guía, Dr. Chirag Shah por brindarme apoyo constante y completa disponibilidad para poder llevar a cabo mi investigación. Además por proveerme de recursos sin los cuales este trabajo no hubiese sido posible. Agradecer también su amistad y la de su familia Lori and Sophie.
- A mis compañeros de laboratorio y del grupo InfoSeeking con los que he compartido principalmente durante la escritura de mi tesis: Vanessa, Erik, Chathra, y Laure. Agradecimientos especiales a Vanessa y Serife por las correcciones gramaticales y léxicas de mi tesis.

- Al Dr. Chirag Shah, Muge Haseki, Dr. Smaranda Muresan, Dr. Nina Wacholder, Dr. Ryen White, Natalia Córdova-Rubio, Dr. Nicholas Belkin, and Dr. Jingjing Liu, con quienes he colaborado en la escritura de artículos científicos durante los pasados cuatro años:
- A mi primer profesor guía, Dr. Edmundo Leiva quien me orientó y motivo a tomar la decisión de seguir una carrera académica.
- A mis amigos en Chile por su cariño y preocupación durante estos años.
- A mis amigos e increíbles personas que tuve la fortuna de conocer en USA. Junto a mi esposa han sido parte de mi familia: Patricio, Rosita, Natalia, Francisco, Brian, Sandra, Leonel, Karina, y Jessica Lingel.
- A mis padres, Roberto González y Edith Ibáñez, mis hermanos Paula y Marco y sus respectivas familias (Richard, Antonella, Giovanni, Paolo, Paola, Benjamin, Josefa), mi abuelita Elisa, a mis suegros Alejandro y Marta y a toda mi familia en Chile por su constante preocupación, cariño, y apoyo a pesar de la distancia. Gracias por visitarnos o estar en contacto durante estos años, sin su presencia quizás no hubiese tenido las fuerzas para terminar.
- Finalmente, a mi amada esposa Ema López, por acompañarme cada día, por darme fuerzas, por cuidarme, por comprenderme, por quererme, por ser una persona maravillosa, y por su coraje para emprender este viaje conmigo.

DEDICATION

I dedicate this dissertation to my family and especially to my beloved wife. To my family in Chile, because despite the distance, they have been with me every day since I began this journey. To my wife, Ema López, for her unconditional support, patience, courage, and love during all these years. Thanks for your love and support.

DEDICACIÓN

Dedico esta tesis a mi familia y en especial a mi amada esposa. A mi familia, porque a pesar de la distancia han estado conmigo cada día desde que emprendí este viaje. A mi esposa, Ema López, por su apoyo incondicional, su paciencia, preocupación, valentía, y cariño durante todos estos años. Gracias por su amor y apoyo.

Table of Contents

List of Tables

List of Illustrations

List of Equations

Chapter 1. Introduction

Do we find (or do not find) relevant information because we feel a certain way? Or do we feel a certain way because we find (or do not find) relevant information? Both questions originally raised by González-Ibáñez et al. (2011) suggest a reciprocal relation between feelings and information search; however, little is known about what role, if any, feelings and related affective processes[1] such as emotions and moods (detailed definitions about affective processes and other terms are provided in section 2.1.1) play in the information search process. For many years affective processes were seen as irrelevant and negative for human reasoning (Damasio, 1994; Evans, 2001; 2002); nevertheless, during the past century scientists from different fields have shown that the affective dimension plays a fundamental role in different aspects of life such as survival, health, work, and social relations (Fredrickson & Levenson, 1998; Losada & Heaphy, 2004; Fredrickson & Losada, 2005; Gable & Haidt, 2005; Seligman et al., 2005). For example, some studies about adaptation and evolution from a psychological perspective have suggested that in dangerous situations, emotional responses may lead individuals, as well as other species, to escape, face, or freeze in order to preserve their lives (fight or flight response, Cannon, 1922). Likewise, research in neuroscience has shown that emotions play an important role in decision making (Bechara & Damasio, 2005), a process associated for centuries uniquely to rational thinking. In information science, on the other hand, while

[1] From here on affective processes and affective states will be used when referring in a general sense to the following specific terms: affect, mood, emotion, and feeling.
To help readers in the comprehension of specific terms used in this dissertation, 0 provides a list with the definition of key concepts that are often used in this and also in the following chapters.

the research emphasis has focused predominantly on cognitive and behavioral aspects, some authors such as Kuhlthau (1991) stated that feelings vary along the information search process (ISP) of individuals, yet it is unclear to what extent, if any, emotions shape the way people search, assess, evaluate, collect, and use information.

Additionally, there has been a predominant emphasis on individual users and their affective states as intrinsic factors of information search[2]. For example, Kuhlthau's ISP model states that at initial stages of information search, people experience uncertainty due to lack of knowledge. Then, while exploring different sources they may feel frustrated or confused. As they start collecting information, confidence may arise. Toward the end of the ISP and depending upon the outcomes, searchers may feel relieved, satisfied, or disappointed. From this perspective, affective states are strongly tied to different stages in the ISP; they vary as a result of information needs (Wilson, 1981), information encounters, and information access. Nonetheless, it is fundamental to recognize that emotions, moods, and feelings are also determined by external factors, thus becoming extrinsic to information search. Examples of such factors are: system design, past experiences, weather conditions, social interactions, and health. To better understand this perspective, consider the following examples[3]:

[2] Information search and search process are used interchangeably to refer to the process whereby individuals or groups search information. These terms are used to avoid confusion with specific models such as Kuhlthau's information search process (ISP).
[3] The cases of Jonh and April introduced in this section will be referred in the next chapters through their assigned codes, namely: $S_{John[-]}$, $S_{April[+]}$, $S_{John[-]\&April[+]}$.

John ($S_{John[-]}$) is an expert researcher at a major institution. During the weekend, he had a bad time at home due to an argument with his wife. As a result, unpleasant feelings derived from this situation quickly turned into negative mood that accompanies John for the rest of the week. Back to work on Monday, John is developing an important project that requires him to find information in order to complete the strategic plan. From this example the following question arises: how does being in a negative affective state change the way John searches, evaluates, and uses information?

Now, consider the case of April ($S_{April[+]}$), one of John's coworkers assigned to the same project. Unlike John, April had a wonderful weekend with a group of friends, which loaded her with positive attitude to start a new week. At work, April is helping John to complete the strategic plan of the project, which also requires her to find information. In this case, how does being in a positive affective state shape the way April searches, evaluates, and uses information?

In the above examples, affective states were determined by factors that are extrinsic to their information search processes (i.e. experiences during the weekend), yet the resulting affective states are present at the moment John and April have to deal with information. Whether affective states are positive or negative, is it possible that particular affective processes as preconditions to the ISP shape the way people interact with information? Few have addressed experimentally this question, however, experimental designs employed in such studies have not been effective in controlling and distinguishing affective processes as extrinsic or intrinsic factors with respect to information search. For instance, Lopatovska (2009a, 2009b) explored effects of mood in an online information search task performed by single users. In this study moods as well as other affective processes were not manipulated; instead they were discretely measured through self-reports. Additionally, expressed emotions were measures through facial expressions using specialized software. According to the author, neither positive nor negative moods have implications in the outcomes of information search.

While external factors that modify people's affective states typically occur outside the context search processes, in practical settings they may also change as a result of processes or events, which although external, are closely tied to information search. For instance, it has been stated that information seeking is not always a solitary activity, but it often involves interactions as well as collaboration with other people - which is the case of collaborative information seeking (CIS) (Twidale, Nichols, & Paice, 1997; Sonnenwald & Pierce, 2000; Reddy & Dourish, 2002; Morris, 2008). Interacting with others while searching information involves additional processes such as coordination and communication, the latter typically carried out through the exchange of messages that may or may not be related to the task (Gonzalez-Ibanez, Haseki, Shah, 2012a, 2012b, 2013). More importantly, such exchange of messages - whether it is performed through verbal or non-verbal communication - may lead to changes in individuals' affective states. To better illustrate this idea, let's return to the examples of John and April introduced above.

It was indicated that both John and April are searching information to complete a strategic plan for an important project. Being a common assignment at their workplace, John and April ($S_{John[-]\&April[+]}$) are required to collaborate with each other. Moreover, as stated earlier, John starts the week in a negative mood, whereas April starts it in a positive one. In both cases their affective states were determined by external events occurred during the weekend, yet they are still present at the moment of collaborating and searching information. One could see the search processes of April and John as independent of one another; however, as they interact they merge becoming one. In such scenario, affective states may change as a result of information-related events (e.g. being exposed to affectively-loaded content pages or having difficulty in finding relevant information) and also by external factors, such as room temperature, system performance, light conditions, and collaboration process. In the latter case, assuming that external factors but collaboration processes are carefully controlled, can John's affective states evolve to positive ones as a result of the interactions with April or vice versa? In turn, can such affective variations change the way John and April deal with information?

1.1 Problem statement

Whether information search is carried out individually or in collaboration with others, the above examples illustrate that affective processes are somehow present during information search. While John's and April's examples refer to a specific situation – namely, information search in work-related contexts - both cases could be extended to other situations where information is required, for example: planning a trip, working on a class project, or researching about a health issue, to name a few. No matter what the contexts or topics are, their affective states as a result of past events will be present when searching and interacting with information. In turn, in the process of acquiring information, people's emotions may or may not affect the way information is searched, interpreted, evaluated, and used.

Current information systems that support CIS, does not take into account affective processes and their potential involvement in the search processes carried out by groups. It has been argued, however, that in group contexts information assessment may be determined or biased by how people feel and share opinions or judgments about information with their collaborators (González-Ibáñez & Shah, 2010). At the present time, information systems are unable to mediate collaboration taking into account the affective variability of team members. In order to build systems with such a feature, it is first necessary to understand the participatory nature of affective processes in collaboration and information search.

When looking at relations between collaboration, information search, and affective dimension, multiple questions arise. For instance, is it possible that being in positive or negative affective states lead individuals to find (or not find) relevant information with different degrees of effort, efficiency, and efficacy? How do being happy or unhappy change the way people formulate queries and assess information? In collaborative contexts, can team dynamics and the affective processes of team members change how they deal with information? Or could prior affective states serve as predictors to anticipate search performance, success, or failure? This list of questions can be easily expanded, which make research in this domain quite complex. Consequently, it is necessary to narrow down the research by focusing on specific questions and problems.

Among the list of problems about information and affective dimension that can be investigated, this dissertation focuses on affective processes interactions and their implications in the domain of CIS. In this dissertation, the notion of affective states interaction refers to the process whereby group members who experience either positive or negative affective processes, engage in social practices and group dynamics carried out while collaborating to achieve common goals. This problem was introduced with the examples of John and April ($S_{John[-]\&April[+]}$), where both positive and negative affective states are determined by external (e.g. past events and social interactions) and internal (e.g. information practices) factors with respect to information search. While affective processes interactions have been investigated by some in business, teamwork, and also individual settings (Fredrickson & Losada, 2005); little is known about their role, if any, in CIS. Besides, unlike information search performed by solo users, CIS as a research topic offers ideal scenarios to examine affective processes as intrinsic and extrinsic factors in information search.

In particular, this dissertation focuses on four research objectives, namely, (1) Study how initial affective conditions influence information practices; (2) investigate what affective processes typically participate in information search; (3) examine how affective conditions derived from social interactions during the collaboration process influence team performance; and (4) study positivity ratio in collaborative search and their relation to team performance. In order to address these objectives, a controlled laboratory study with 135 participants distributed in six experimental conditions was conducted. In each experimental condition, participants were individually treated with affective stimuli in order to elicit positive and negative affective states. Data collected - which consists of users' affective responses (using expressive, physiological, and self-reported evaluations), social interactions, and information-related actions - was quantitatively and qualitatively analyzed.

1.2 Research questions

This dissertation investigates how affective processes participate in information search processes carried out by teams. In particular this study focuses on the following four research questions:

RQ1: Do initial affective states and their interactions shape the way team members collaborate when searching information, and if so, how?

RQ2: What affective processes are typically experienced and expressed (physically, physiologically, and verbally) by team members when collaborating in an information search task?

RQ3: To what extent, if any, do initial positive and negative affective states and those derived from the collaboration of individuals in an information search task influence team performance?

RQ4: To what extent, if any, does the relation 3-to-1 between positive and negative affective states (P/N) (Losada & Heaphy, 2004; Fredrickson & Losada, 2005) apply to CIS?

Note that these research questions are introduced in this chapter to list the specific research objectives of this investigation. Theoretical foundations and empirical studies that inspired their formulation (e.g. positive psychology and affect infusion) are discussed in the following two chapters.

1.3 Objectives

While the above research questions are the end goal of this dissertation in terms of research, addressing such questions requires working on intermediate steps that ultimately lead to additional contributions such as methods evaluation, a methodology, and tools, among others.

1.3.1 General objective

Investigate positive and negative affective states as initial conditions as well as their interactions in a CIS task.

1.3.2 Specific objectives

- Review past research that has addressed problems related to the affective dimension, in particular positive and negative affective processes, in workgroup and information search.
- Investigate, compare, and select theoretical frameworks for addressing the problem about interaction of positive and negative affective processes in CIS.
- Design and conduct a laboratory study to investigate affective states interactions in a CIS task.
- Design and develop instruments for data collection.
- Evaluate individuals' and teams' performance in the information search process.
- Study communication processes within teams.
- Evaluate individuals' affective responses during the information search process and explain how they relate to communication and team performance.

- Propose a model that explains implications of affective processes in CIS.

1.4 Summary

This first chapter introduced and described a research problem that focuses on affective processes and collaborative information seeking. In particular, it is suggested that there is a reciprocal relation between positive and negative affective states and information practices carried out by individuals and teams. Affective processes, whether they are intrinsic or extrinsic to information search, would play an active role in how information is sought, evaluated, and used. The dissertation pursues four major research goals, namely, (1) Study how initial affective conditions influence information practices; (2) investigate what affective processes typically participate in information search; (3) examine how affective conditions derived from social interactions during the collaboration process influence team performance; and (4) study positivity ratio in collaborative search and their relation to team performance. These objectives are addressed empirically through a controlled experimental study.

The remainder of this dissertation is structured in five chapters. First, Chapter 2 provides an in-depth review of definitions, theories, and previous works that have investigated the affective dimension in information science and collaborative domains. Following this, Chapter 3 introduces and discusses the research framework used to carry out this dissertation work. Then, Chapter 4 provides a detailed description of the methodology, system, instruments, and the experiment conducted to collect the data used in the analyses to respond to the research questions of this study. In Chapter 5 analyses and results are presented and discussed in regard to each research question. Finally, 0 offers an extensive

discussion about the results and their theoretical and practical implications.
Additionally, future research directions from this work are presented.

Chapter 2. Literature Review

This dissertation investigates affective processes in collaborative information seeking (CIS), in particular, it focuses on positive and negative affective states and their role, if any, in the information search process of teams. To better understand the rationale, motivations, and the research problem itself, a comprehensive review of related work is presented in this chapter. Relevant literature on this topic is scattered across different disciplines such as information science, psychology, computer supported cooperative work (CSCW), and communication. Each area has contributed in various ways such as providing definitions, evaluations, studies, methodologies, and theories. Among the various concepts and topics presented throughout this dissertation, there are two that guide the organization of this chapter, namely, the affective dimension and CIS. The next two sections review relevant literature about these two topics, including definitions, theories, evaluation methods, and relevant findings from empirical research. Finally, the third section of this chapter reviews connections between affective research and information seeking, with particular attention to CIS.

2.1 Affective Dimension

Scientific domains such as physics, chemistry, and biology usually describe phenomena by referring to dimensions of time and space. The dimension of time, for example, allows distinguishment of past, present and future events. The dimension of space, on the other hand, gives a sense of the location of the objects or phenomenon being described. Like time and space, there are other dimensions particularly relevant to dealing with phenomena involving human behavior. One such dimension is cognition, which has received particular attention for decades for attempting to explain rational thinking. Another equally important dimension but less investigated in some fields (e.g. information science) is the affective one, which represents what people feel, how they feel, and why they feel.

In the literature it is possible to find related terms that refer to the affective dimension, such as emotional dimension. But how is the affective dimension different from the emotional dimension? While the two may refer to the same notion, the underlying constructs have different meanings. Affective, on one hand, relates to the word affect, and both have their roots in the Latin word *affectus*, which relates to mood, emotion, and feeling. According to the Merriam-Webster (2013) dictionary, the term affective means "relating to, arising from, or influencing feelings or emotions"[4].

On the other hand, the words emotional or simply emotion are usually used as generic terms and are sometimes used interchangeably with related constructs such as mood, affect, humor, sentiment, and feeling. In psychology as well as in other disciplines, the term emotion is a controversial one. After decades and even centuries of research, to date there is no consensus among scientists about what emotion means (Palmero, Guerrero, Gómez, & Carpi, 2006).

To avoid confusions when referring to these terms in the following sections and chapters, this dissertation employs the terms affective dimension, affective processes, and affective states to refer in a broad sense to specific concepts such as emotion, feeling, mood, and affect. Such terms are approached in this work as

[4] http://www.merriam-webster.com/dictionary/affect (Merriam-Webster, 2013)

different constructs that may (or may not) be connected to the information practices of teams. Before jumping into a detailed review of the literature on the affective dimension, the first step in this section will be to establish leading definitions for key affective terms that will be used throughout this work.

2.1.1 Definitions of affective terms

Affective processes have been studied for centuries from different perspectives and disciplines such as philosophy, psychology, and neurophysiology, among others. These different currents of thought led to the production of hundreds of definitions for specific terms such as emotion or mood. At the same time, these definitions led to controversies that have not reached consensus yet within the scientific community. This situation has been widely recognized in different fields such as psychology (Palmero et al., 2006), human computer interaction (Peter & Herbon, 2006), and information science (Lopatovska, 2009b; Loptovska, 2011).

Unlike general terms such as camera, lamp, or table, whose definitions can be found in most dictionaries, definitions of affective terms typically involve underlying paradigms, theories, and models that differ from one another. For example, the term emotion has been defined in many ways due to the basis of opposing perspectives, namely, discrete and dimensional theories.

While the lack of consensual definitions of affective terms is still a subject of debate, this has not impeded scholars from pursuing applied research involving the affective dimension. Some, however, have claimed that definitional issues need to be resolved before addressing research and developing new systems (Peter & Herbon, 2006).

This dissertation takes into account different facets of affective processes, in particular: emotion, mood, affect, and feeling. While debate and controversies around this topic persist, Palmero's et al. (2006) definitions supplemented with the perspectives of other authors are adopted in the development of this work. The following sections present formal definitions and examples for these four terms:

2.1.1.1 Affect

According to Palmero et al (2006), when comparing emotion, mood, feeling, and affect, the latter is the most general of all three. It is stated that affects are linked directly to physiological processes and are normally described under a dimensional approach (Russell, 1983; Scherer, 2005), which involves valence[5] (i.e. positive or negative) and intensity, which is typically linked to arousal[6] (Russell, 1983). These two dimensions are linked to neurophysiological and biological components. Affects are usually addressed toward external or internal aspects such as situations, people, or objects. These relations result in approximating pleasant conditions and estrangement from unpleasant ones. In Palmero et al.'s words "We could say that affect represents the essence of behavior, this understood from the most elemental formulation of approaching what likes, gratifies, or pleases, and estranging from what produces opposite consequences" [7] (2005, p. 17).

---

[5] Valence refers to the tone of an affective process, that is to say, positive or negative.
[6] Arousal refers to physiological activation or response (e.g. excited, relaxed, calm, etc.)
[7] This quote has been translated from the original source written in Spanish.

Duration of affects is difficult to determine; however, it has been indicated that they last longer than any other kind of affective process. To illustrate this, let's expand example $S_{John[-]}$ introduced in the previous chapter. John is a loving husband and father that enjoys sharing time with his family. For John, family is linked to positive affects, this in spite of some bad moments that he could experience with them throughout his life. On the other hand, when facing tragic events - for instance a sudden loss of a close relative – John's affects switch to negative ones and this new affective state may last weeks, months, or even years.

2.1.1.2 Mood

In a similar fashion, Palmero et al. (2006) described mood as a particular affective process. Like affect, mood is also characterized by valence and intensity. Unlike affects, mood lasts for a shorter duration, such as hours or days. Palmero et al. add "[mood] denotes the existence of a set of beliefs about the probability of a subject to experience pleasure or pain in the future; that is, of experiencing positive or negative affect"[8] (p. 17).

In $S_{John[-]}$ it has been stated that John had a bad weekend due to family issues. As a result of this situation, he experiences a negative mood that accompanies him during the following days at work. This negative mood, however, could be overcome depending upon John's coping skills or if the situation at home is solved. On the other hand, if similar events occur often, then John may begin to relate family time with negative affects.

[8] This quote has been translated from the original source written in Spanish.

2.1.1.3 Emotion

Emotions, on the other hand, have been described by Palmero et al. (2006) as a multidimensional response of individuals to external or internal stimuli. Emotions last for an even shorter duration than affects and moods. In fact, emotions are spontaneous and intense. In this dissertation, emotions are approached from a discrete perspective.[9] In other words, emotions can be categorized or labeled with specific names such as: happiness, disgust, anger, surprise, fear, and sadness (Ekman, 1977). Adding Bloch's (2002) perspective, emotions can be also classified as primary (also referred to as basic) and secondary. While the former (e.g. happiness, sadness, and surprise) are inherent and universal (Ekman, 1977; Izard, 1977), the latter (e.g. gratitude. envy, and ambition) are socially or culturally developed and expressed in regard to basic ones.

Another important aspect of emotions is their expressive component. Authors such as Ekman (1977) have indicated that emotions are physically expressed. In particular, Ekman identified in a cross-cultural study a small set of universal basic emotions (Izard, 1977; Plutchik, 1980) and facial expressions linked to them. Later, Bloch (1988) added that specific body postures and respiratory patterns are also linked to basic emotion. Other forms of expression of emotions include verbal communication (Fussell, 2003) and voice intonation (Pereira, 2000; Schröder et al., 2001; Gobl & Chasaide, 2003; Owren & Bachorowski, 2007).

---

[9] For more discussion about discrete and dimensional approaches refer to section 2.1.2.2.

Returning to $S_{John[-]}$, consider the situation that John experienced over the weekend. Both John and his wife were planning for several years to attend a dream event that takes place once every four years. John's wife agreed to make the reservations and buy the tickets. Unfortunately the tickets were sold out since John's wife forgot to purchase the tickets in advance as John had originally suggested. At the moment John received the bad news from his wife, his face changed abruptly, "his eyebrows [were] lowered and drawn together, [his] eyelids [were] tensed, [... his] eye appear[ed] to stare in a hard fashion [, and his] lips [were] tightly pressed." This description corresponds to the facial expression for anger explained by Ekman and Friesen (2003, p. 82). At the same time, John began breathing heavily through his nose, which according to Bloch, Lemeignan, and Aguilera (1991) is the respiratory pattern that expresses anger. Anger, however, showed up only for a brief moment (maybe for a few seconds), then, regulated by the context, John's face quickly changed as follows: "The inner corners of [his] eyebrows [...] raised and [...] drawn together. The inner corner of his upper eyelid [...] drawn up, and [his] lower eyelid [...] raised." In the lower area of his face, "[his] lips appear[ed] to tremble." This is the facial expression of sadness as described by Ekman and Friesen (2003, p. 117).

2.1.1.4 Feeling

Finally, according to Palmero et al. (2006), feelings are the subjective component of emotions. Compared to emotions, affects, and moods, feelings are at a higher order. Feelings appear when individuals become aware that they are experiencing a particular emotion. From this perspective, emotions precede feelings. As stated by Palmero et al. "[feelings are] probably the shortest of the various forms of affective processes. However, feeling can lead to a sustained experience over time, significantly more durable than the emotional process itself"[10] (p. 18).

Feelings can be expressed in different ways. If the person is able to consciously recognize his internal affective states, he or she may label them with specific names. For instance: "I'm happy!", "I'm sad", or "I feel disappointed." Sometimes, when it is unclear what one's internal affective states are, but it is still possible to describe their intensity and distinguish whether they are pleasant or unpleasant, people may use broad expressions linked to specific dimensions such as "I feel so bad", "I'm fine", and "It feels so strange." In other cases, individuals may hide their internal affective states by expressing something different (Gross & Levenson, 1997); this could happen due to social norms or cultural factors. For instance, someone who clearly recognizes his or her negative affective state could tell or show to others that he or she feels completely fine. Unfortunately, sometimes people are unable to accurately recognize their affective states. This may be a result of confusion, lack of self-awareness, brain injuries, mixed emotions, mood or mental disorders, or a particular condition such as Asperger syndrome and alexithymia (Taylor, 1984).

As noted above, John ($S_{John[-]}$) experienced two particular emotions when receiving the bad news from his wife about the event tickets: Anger (which was experienced for a very short moment and then sadness. Probably John did not even realize his first spontaneous emotional response, but later when his wife said, "I'm sorry, how do you feel about this?", John thinks for a moment and

[10] This quote has been translated from the original source written in Spanish.

respond: “I’m very disappointed and sad.” Alternatively, John could had hide his feelings to avoid making his wife feel bad. In this case, he consciously tries to put a smile in his face and say to his wife “I’m fine, don’t worry, we will make it some other time.”

Although the examples presented in this section have been developed around John’s situation ($S_{John[-]}$), similar situations could be illustrated for April’s case ($S_{April[+]}$) in which she shows positive affective experiences such as happiness, joy, amusement, and inspiration. In order to summarize the concepts and definitions introduced above, Figure 2.1 provides a hierarchical structure with key elements for affect, mood, emotion, and feeling. Moreover, a sketch of the duration and variations of different affective process with respect to their tone (valence) and intensity (arousal) is depicted in Appendix C.1. An expanded discussion about these two concepts is presented in section 2.1.2.2.

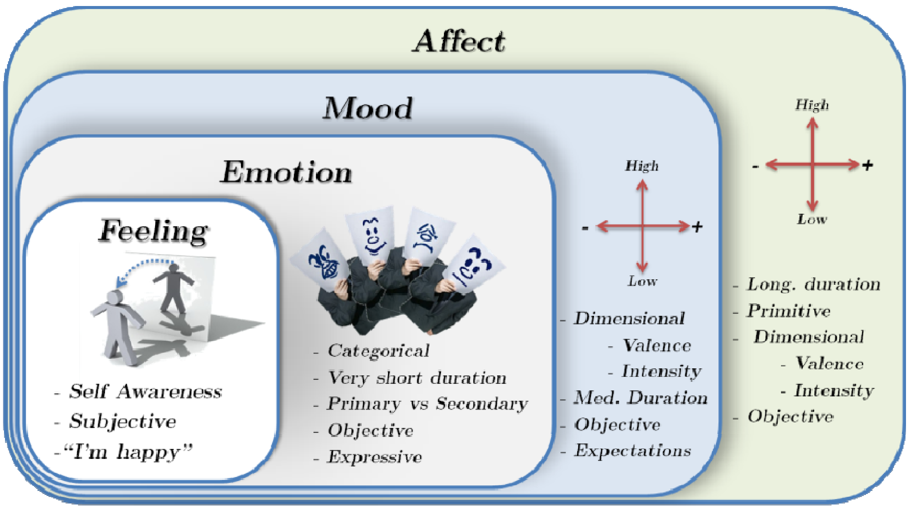


Figure 2.1. Hierarchical structure of affective processes.

2.1.2 Theories and approaches of affective processes

Affective processes have been widely studied for decades by psychologists, and even centuries by other disciplines (Reeve, 1994). During this time several theories and models attempting to explain the origin of affective processes, their meaning, and also psychological disorders have been developed (Reeve, 1994). In turn, methods to understand and overcome problems associated with affective states (e.g. psychological therapies) have been developed.

Within its origins, emotions were addressed by philosophers (Damasio, 1994; Evans 2001; Evans, 2002); however, today neuroscience and psychology are two of the disciplines with major contributions to the understanding of affective processes. As noted above, throughout history, contributions to this topic from different currents of thought have caused debates and controversies about what is meant by key terms such as emotion and feeling. Although the previous section provided definitions for key terms in the development of this dissertation (i.e affect, mood, emotion, and feeling), it is necessary to review different perspectives on the study of affective processes in order to explain why specific theoretical approaches and research methods have been chosen to address this study.

This section briefly describes theoretical aspects of affective processes, in particular, the biological-behavioral and structural approaches. In addition, a brief discussion about tone or affective valence, which is a fundamental topic in this research, is presented. Finally, basic foundations on positive psychology, which functions as one of the key theoretical framework of this work, are introduced.

2.1.2.1 Biological and behavioral approaches

For a long time, the study of emotions was considered to be irrelevant and taboo (Fredrickson, 2009). This can be attributed to the ambiguity and negative connotation of emotions inherited from the origins of philosophy (Damasio, 1994; Evans, 2001; 2002). One of the first set of approaches to study emotions focused on biological and behavioral aspects aiming to explain the source of emotions and how these are perceived. Following, five approaches are explained, namely, neuro-physiological, evolutionary, expressive, behavioral, and cognitive.

- ***Neurophysiological approach to emotions***

The James-Lange theory (Cannon, 1927) provided one of the first theoretical contributions in the affective domain and defied previous approaches that argued that emotions produce changes at the biological level. According to James-Lange's perspective, emotions (whose definition does not necessarily conform to that provided in the previous section) originate at specialized biological structures. Based on the authors' view, emotions are not directly triggered by the perception of stimuli. In contrast, stimuli triggers biological changes (e.g. blood pressure, muscle activation, etc.) and the perception of such changes results in emotions. According to James-Lange theory, different emotional stimuli cause different reactions in the body. It is also noted that non-affective stimuli do not produce bodily responses that can be related to affective processes (Cannon, 1927, Reeve, 1994).

Cannon-Bard (Cannon, 1922) presented a rival hypothesis which states that both emotion (in particular its physiological and subjective components) and changes at the biological level occur simultaneously and independently. However, biological changes as a result of affective stimuli are linked to expression of emotion. For example, in a dangerous situation, the blood stream is directed to the legs enabling individuals to face or get away from the threat (Cannon, 1922).

Biological approaches have indicated that specific areas of the brain participate in the generation of emotional responses. For example, the thalamus is linked to aggression or defense responses (Cannon, 1922; Damasio, 1994).

Evolutionary approach to emotions: The study of emotions has also been addressed through evolutionary approaches. From this perspective it is argued that emotions play a key role in the evolving function of species (Darwin, 1872; Plutchik, 1970; Ekman, 2006). According to this view, emotions have contributed to the preservation of species by allowing them to confront threats (with anger), get away from hazards (with fear and disgust), socialize (with acceptance and trust), and breed (with eroticism, tenderness, and love), thereby transferring learning and genes to future generations.

Expressive approach to emotions: Expressive approaches suggest that emotions have a communication function, whose purpose is to display internal affective states to others. Some argued that the expression of emotion is inherent to animals (not only the human species) (Darwin, 1872; Ekman, 2006). In addition, it has been suggested that emotion expression may be regulated as an adaptive resource by learning and socialization. While some argued that the expression of a certain group of emotions is inherent, Darwin (1872) indicated that emotion expressivity was originally learned. Due to its survival function, emotions remained for generations until becoming inherent.

Behavioral approach to emotions: The behavioral approach to emotions links them to motivation. According to this view, there exist a reciprocal relationship between motivated behavior and emotions, where the former is associated with goals, rewards, and structured plans to achieve them. By way of example, gratifying situations and the behavioral process to reach them can be learned; as a result those who learn this association tend to repeat the behavior. Likewise, behaviors leading to situations that displease will be avoided (Watson, 1924). This approach can be related to the dimensional approach, in particular to tone or affective valence, which is discussed later in this chapter.

Cognitive approach to emotions: Finally, emotions have been also studied from a cognitive perspective. This approach holds that cognition precedes emotion, and that emotional experience is necessary in order to understand what produces welfare and what does not (appraisal theory) (Scherer, Schnorr, & Johnstone, 2001). As noted in Reeves (1994), "each emotion implies a type of evaluation"[11] (p. 337).

2.1.2.2 Structural approaches

An alternative way to study affective processes has to do with the way these are categorized or structured. According to Peter et al. (2006), these approaches are of particular interest in technology-related fields such as human-computer interaction. Below, two groups of structural approaches are reviewed, namely, dimensional and discrete approaches.

---

[11] This quote has been translated from the original source written in Spanish.

Discrete approaches: Discrete approaches suggest that affective processes can be distinguished from one another through categories. Authors such as Ekman (1977), Izard (1977), and Plutchik (1980) argue that there exist a group of basic and universal emotions. Basic emotions are characterized as being independent from factors such as culture, gender, or age (Ekman, 1977; Izard 1977). Other discrete approaches note that there is also a group of secondary or mixed emotions, which are learned, developed, or acquired either culturally or socially (Ekman, 1977; Bloch, 2002). It is important to note that not only emotions, but also other affective processes such as mood, can be characterized under this approach.

For the particular case of emotions, the number of labels used to categorize them differs among authors. According to Peter and Herbon (2006), it is possible to find between 2 to 18 categories in the literature; however, most authors agree on six of them. This selected group of basic (or primary) emotions includes happiness, sadness, anger, surprise, fear, and disgust (Ekman, 1977).

Dimensional approaches: Dimensional approaches, as opposed to discrete ones, argue that affective processes can be described in terms of one or more dimensions (Russell, 1983; Scherer, 2005). According to this approach, instead of characterizing affective states by assigning specific labels that refer to them, each dimension represents a continuum that denotes aspects such as intensity, valence, and control (Bradley & Lang, 1994).

Two dimensions widely used under this approach are arousal and valence. Using these dimensions, affective processes can be characterized based on levels of physiological activation (arousal) and affective tone (valence), respectively.

Although arousal refers to aspects such as physiological changes triggered by the sympathetic nervous system (e.g. electrodermal activity), evaluations are typically conducted through questionnaires, in which participants report subjectively of how activated (aroused) they feel. Valence, on the other hand, distinguishes whether an affective process is pleasant or unpleasant, which is also referred to as positive and negative, respectively. Likewise arousal, evaluations of valence are typically obtained from self-reports.

Mixed approaches: Discrete and dimensional approaches have advantages and disadvantages. In some cases, having distinctive expressive patterns for each emotion under discrete approaches has made possible the development of automated systems capable of recognizing them. Significant development in this field can be found in the affective computing literature (Picard 1997; 2003) where techniques based on pattern recognition, digital image processing, and machine learning have been used to generate models that can automatically recognize facial expressions related to specific emotions.

A disadvantage of discrete approaches is that they may be limited when characterizing the intensity and valence of affective processes. For instance, determining the level of anger or sadness of an individual may be a complex task, though specialized coding systems offer some resources for that (Ekman & Friesen, 1978; Izard, 1983). Likewise, it is hard to establish whether these emotions are pleasant or not, unless a priori classifications are established (e.g. anger is to negative as happiness is to positive). More discussion about affective tone is presented below.

With the dimensional approach, on the other hand, it is possible to characterize both factors (i.e. intensity and valence), but it is limited when contextual interpretations are required. For example, what do high or low arousal mean in a given situation? In a similar fashion to discrete approaches, it has been possible to develop technology capable of measuring dimensional constructs such as arousal with specific operationalizations. Some examples include blood pressure, brain activity, and electrodermal activity.

Some authors have developed mixed models where aspects from discrete approaches are mapped onto dimensional ones. An example of this is the adaptation of the circumplex model of affect (Russell, 1983) created by Larsen and Diener (1992) (Figure 2.2). Another hybrid model with a focus on the development of digital systems was introduced by Peter and Herbon (2006);

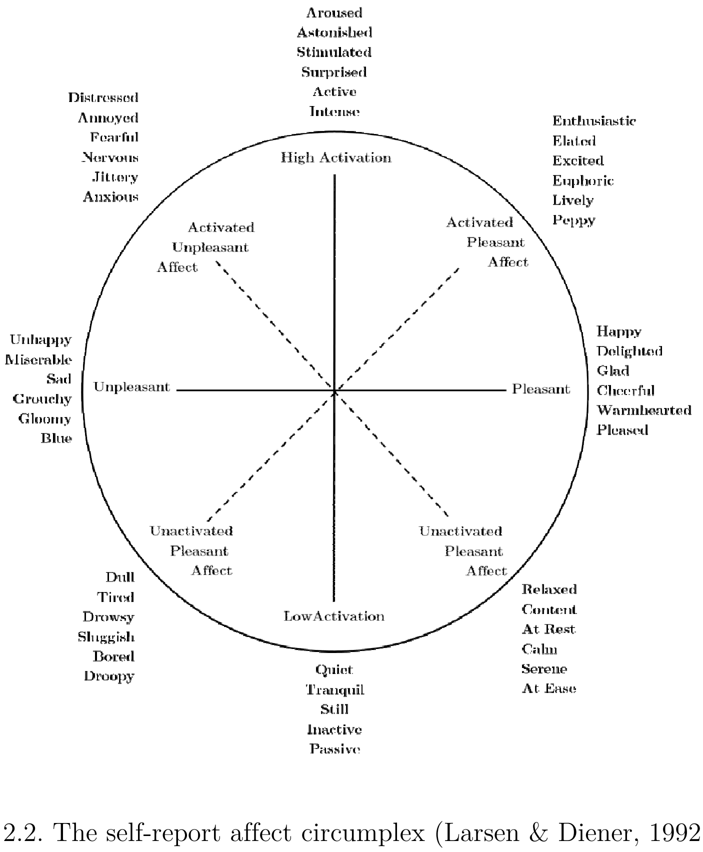


Figure 2.2. The self-report affect circumplex (Larsen & Diener, 1992, p.31)

however, as pointed out by the authors, the affective labels of this model distributed in a bi-dimensional plane are used only as a reference, given the mixed nature of emotions (Barret & Russell, JA, 1998).

- ***Positive versus negative: Contextual connotations***

As explained above, the dimensional approach indicates that affective processes can be characterized in terms of their valence or affective tone, which allows the classification of affective processes as positive or negative. This dichotomy, however, does not necessarily relate to contextual connotations. For example, determining if a particular category of emotion is contextually positive or negative, good or bad, pleasant or unpleasant, may depend on a number of factors (Tafrate, Kassinove, & Dundin, 2002). In other words, an emotion like anger could be classified either as positive or negative depending upon the context where this affective state is experienced (Friedman et al., 2004). To better illustrate this aspect, consider the following example:

In $S_{John[-]}$, it was explained that John experienced anger and sadness as a result of an issue with his wife. In this case both anger and sadness could be tentatively classified as negative. This classification derives from the antecedents provided about John's case. It is clear in this example that the affective states that John experiences are not pleasant for him, instead he would rather have gotten the tickets to attend the event and experience happiness on account of this situation.

Now consider a different scenario for John's anger, this time in the context of work. John is working on an important project. He recently had a meeting about the budget with the manager of his department. As a senior specialist in his field, John manages in detail the technical and financial specifications of the project. In an argument with the manager, who happens to disagree with John's approach, John experiences a controlled level of anger, enough to be able to defend and support his position. As John experience this emotion, he feels confident and empowered to face his boss without being disrespectful to him. From the perspective of John's boss, he is able to perceive the John's affective changes based on his facial expressions, tone of voice, body posture, and even the words he uses. In this particular scenario, which corresponds to a case of conflict resolution, anger ends up having a positive/pleasant connotation for John.

The classification of affective processes as positive or negative may be related to how they are experienced; in other words, whether they are pleasant or unpleasant. This classification may also depend on the context and ultimately on the consequences of experiencing and expressing a given affective state. In the following sections a discussion about how affective tone (valence) and also its intensity (arousal) are measured is presented.

2.1.3 Experimental methods to study affective processes

The methods used to study affective processes may be determined by approaches or theories (like the ones reviewed in the previous section) under which research is carried out. In reviewing the literature, it is possible to find a variety of approaches that are based on naturalistic studies and laboratory experiments.

On one hand, naturalistic studies consider affective processes in their natural context, that is to say, without intervention by the researcher. Although naturalistic approaches have significant advantages compared to experimental ones, since there are no interventions in the natural course of individuals' affective processes, it is limited due to the lack of control and reliable measurement. Recently, the availability of portable devices with limited levels of intrusion (e.g. *Q Sensor*[12]) has facilitated data collection for extended periods of time without people having to attend lab sessions. However, there are some limitations in terms of the type of measures collected and how these are interpreted.

Experimental methods are based on controlled or semi-controlled evaluations of affective processes with the use of one or more instruments. The main advantage of this kind of methods is the control and access it affords to various types of affective measurements. Paradoxically, the advantages derived from control and instrumentation may also constitute disadvantages. For example, participants' affective processes could be influenced unintentionally as a result of experimental conditions such as physical space, instrumentation, or the presence of the researcher. Considering the emphasis of this dissertation, this section focuses on a review of experimental methods for the evaluation of affective processes.

2.1.3.1 Emotion elicitation

Although emotion elicitation is an optional method in the study of affective processes, it is a practical one to investigate affective changes and their contextual implications. Emotion elicitation is based primarily on the application of stimuli (e.g. visual, auditory, cognitive, or social stimuli) that are expected to produce an internal emotional change that is usually predefined. The literature on elicitation of emotions is wide spread. Martin (1990) and Coan and Allen

---

[12] http://www.qsensortech.com/

(2007) provide a comprehensive review of different methods of emotion elicitation, explaining their theoretical basis, instruments, experimental procedures, applications, evaluation, and expected results. Below, some methods for emotion elicitation used in research are described.

Pictures: Under this approach participants are exposed to sets of pictures whose content is associated with different human experiences, objects, and situations, among others. In turn, the pictures have been previously cataloged by various affective categories as a result of experimental evaluations. Such categorization dictates the expected affective response after participants are exposed to them. Emotion elicitation through this method has been used for more than 15 years (Bradley & Lang, 2007), with the International Affective Picture System (IAPS) developed by Lang, Bradley, and Cuthbert (1997, 2008) as one of the most recognized due to its standardization.

Sounds: In a similar fashion to emotion elicitation through pictures, the induction of affective states through sounds is based on a series of acoustic stimuli previously classified in different affective categories. Once participants are exposed to acoustic stimuli, it is expected to elicit the affective state determined by their category. Bradley and Lang (1999) developed the International Affective Digital Sounds (IADS), which like IAPS is standarized and freely available for research purposes.

Music: Emotion elicitation through music, as described in Martin (1990) and Eich et al. (2007) is used in combination with participants' own efforts to achieve a particular affective state. As pointed out by Martin (1990), "[i]t is explained that the music alone may not automatically induce the desired mood state, and subjects are asked to use any other means they find effective to get into the appropriate mood" (p. 671). Procedures described in the literature suggest two implementations of this method, namely, (1) participants freely select the music (Sutherland, Newman & Rachman, 1982) and (2) all participants receive the same musical stimuli (Clark & Teasdale, 1985).

Films: Another method widely used is based on films. In some cases, films are edited excerpts that do not last more than five minutes. These short films contain material carefully selected with the specific goal to elicit a particular affective state. Unlike emotion elicitation through music, films are expected to be able to induce the expected affective state without having to ask participants to work on their own to achieve them (Martin, 1990; Rottenberg, Ray, and Gross, 2007).

Expressive patterns: Among the variety of techniques to elicit emotions, some procedures are based on the articulation of emotion via expressive patterns. These techniques are based on specific instructions that tell participants what to do with parts of their faces, posture, and even how to breathe, all without revealing the expected emotional response. For example, Ekman (2007) indicates that in an experiment the participants received instructions such as: "[1] Pull your eye brows down and together. [2] Raise your upper eyelids. [3] Now tighten your lower eyelids. [4] Narrow, tighten, and press your lips together, pushing your lower lip up a little" (p. 48) without telling them that such actions were linked to anger.

Another technique is Alba Emoting (Bloch, 2002), which is based on facial expressions, body posture, and breathing patterns, which once reproduced by participants lead them to experience a group of basic emotions. The technique has been tested experimentally and today is used in areas such as therapy and theater.

Game-feedback or false-feedback: The elicitation of emotions through game-feedback (Martin, 1990), also referred to in the literature as false-feedback (Zhao, 2006), consists of providing feedback to participants about their performance during or after performing a particular task. The feedback provided, however, is independent of the actual performance achieved. In other words, if the objective is to elicit negative affective states, the feedback provided to participants indicates failure even if they solve the task successfully. On the other hand, if the goal is to elicit positive affective states, the feedback indicates success even if participants fail to solve the task (Martin, 1990). This technique must be carefully implemented because participants could identify that the feedback received is not accurate. Note that this technique is used in the experimental design of this dissertation. More details about its implementation are provided in Section 4.5.

Other emotion elicitation techniques: The list of methods to elicit emotions is rather large. This section briefly described some of the techniques that can be used to induce affective states in experimental settings. Other methods described in the literature suggest the use of text, social feedback, public speaking, autobiographical recall, and threats, among others (Martin, 1990; Coan & Allen, 2007).

2.1.3.2 Affective states evaluation

Just as there are several techniques to elicit emotions, there are various procedures to evaluate them. A comprehensive review on affective states evaluation can be found in Coan and Allen (2007). This section provides an overview of three approaches to evaluate emotions that are used in the experimental design of this dissertation, namely, self-reports, observational methods, and neurophysiological methods.

Self-reports: Affective states evaluation through self-reports consists of asking participants to indicate how they feel at a given moment or period of time. For example, in the application of emotion elicitation techniques like the ones described in the previous section, participants can indicate through questionnaires, interviews, diaries, or think-aloud protocols, how they feel before, while, and after receiving stimuli. Evaluation based on self-report relies on the assumption that participants are able to recognize, describe, and quantify their affective states (Martin, 1990; Coan & Allen, 2007). It is noteworthy that evaluations through self-report conform to the definition of feelings presented in Section 2.1.1.4.

Among the list of self-assessment methods, standardized questionnaires are widely used. Such questionnaires consists of questions such as, “How happy do you feel now?” or “What is the most happy you felt during the past 24 hours,” to which participants respond using specific scales (e.g. five-point Likert scale) to indicate aspects such as intensity or agreement, among others.

An example of a standardized questionnaire is the Positive Affect Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988; Gray & Watson, 2007), which focuses on the evaluation of affects through a set of 20 terms, half of which are linked to negative affects and the other half to positive ones. In this questionnaire, participants are asked to indicate how they have felt at a certain time (e.g. now, past few days, year, etc.). Participants' answers must be expressed on a five-point scale, where scores one and five correspond to "very slightly or not at all" and "extremely", respectively.

Other questionnaires or tests similar to PANAS are the Differential Emotions Scale (DES) (Izard, 1977) and a modified version of it (mDES) (Fredrickson, Tugade, Waugh, & Larkin, 2003), the latter also referred to as the Positivity Self Test (Fredrickson, 2009)[13]. In particular, the latter consists of a set of 20 questions, 10 of them linked to positive affective states and 10 to negative ones. Like PANAS, answers to this test are expressed in a five-point rating scale using a particular time frame. This questionnaire (see Appendix B.1) is incorporated in the experimental evaluation of this dissertation since it has been designed to measure the relationship between positivity and negativity as described in section 3.1.2.2.

Other self-report techniques are based on the use of pictorial instruments. For example, the Affect Grid (Russell, Weiss, & Mendelsohn, 1989) is an instrument to measure emotions under a dimensional approach. Through this instrument, participants can map their affective state through a graphical representation that consists of a grid with multiple subdivisions. The left and right sides of the

---

[13] An online version of this test can be found at http://www.positivityratio.com

grid represent unpleasant and pleasant feelings respectively. Furthermore, the upper and lower ends represent low and high arousal, respectively.

Another pictorial instrument is the self-assessment manikin (SAM) (Bradley & Lang, 1994) (see Appendix B.2), which is a non-verbal scale that allows individuals to self-report affective states or reactions across three dimensions, namely, pleasure (happy-unhappy), arousal (excited-calm), and dominance (controlled-in control). There are different variations of the scale; however the original one requires participants to self-report their affective states in each dimension using a 9-points scale. This instrument has been used in

the evaluation of the IAPS described in section 2.1.3.1. One important advantage of non-verbal instruments is the short time it takes for participants to provide their responses. For example, when used with IAPS, participants are given 15 seconds to respond the questionnaire (Bradley & Lang, 1999). As described in section 4.5, participants' affective states must be measured multiple times during the experiments carried out in this study; therefore SAM is selected as one of the key instruments to perform discrete measures during the experimental process.

Observer-based methods: Methods based on observations are mostly based on categorical approaches, where affective states are evaluated through their expressions (Cohn, Ambadar, Ekman, 2007). In this case, affective states conform to the definition of emotion presented in section 2.1.1.3.

Observer-based methods consider that emotions as having unique expressive patterns that allow them to be distinguished one another. From this idea, standardized coding systems such as the maximally discriminative facial movement coding system (MAX) (Izard, 1983) and the Facial Action Coding System (FACS) (Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002) have been developed. In particular, FACS allows qualified evaluators to code different facial regions, which are referred to as action units (AUs). This coding system enables not only the evaluation of presence or absence of AUs, but also their intensity.

As noted above, it is not only the face that allows affective evaluations. It has been recognized that other components such as body posture and gestures can be related to specific affective states (Bloch, 1988). Another expressive characteristic that can be used in the evaluation of affective states can be obtained from the voice (Pereira, 2000; Schröder et al., 2001; Gobl & Chasaide, 2003; Owren & Bachorowski, 2007) and verbal expression (Fussell, 2003).

By having standardized coding systems, it is possible to implement automated systems capable of classifying facial expression patterns through the use of specialized software (Cohn & Kanade, 2007). The area of affective computing (Picard, 1997; 2003) has driven and motivated the development of technologies capable of recognizing affective patterns such as specific facial expressions. As a result, different software-based solutions to code facial expressions have been developed. These tools are usually based on digital image processing and machine learning techniques to produce models often trained on a gold standard corpus of faces. These models are later used to process single images or videos with faces to detect components such as AUs, which are then linked to a particular emotion.

It is acknowledged, however, that such systems recognize facial expressions associated with emotions but not the emotions themselves. Examples of these systems are FaceDetect (Kueblbeck and Ernst, 2006; Face and Object Detection webpage, 2011), BMERS (González-Ibáñez, 2006), eMotion (Sebe et al., 2007), MindReader (El Kaliouby & Robinson, 2004), and Affdex (Affectiva, 2013). Note that these systems have some limitations that may affect their accuracy to detect facial expressions. For example, light conditions, face obstructions (e.g. hands, beard. and glasses), head rotation, and ethnicity are some of the factors that could alter the results provided by these systems (Pantic, 2000). In addition to self-report techniques, this dissertation also uses software to automatically recognize facial expressions. More details about the tools used in this study are presented in detail in section 4.7.

Neurophysiological methods: Neurophysiological methods are based on the use of instruments capable of measuring aspects at the brain and physiological levels. At the brain level, research has shown the existence of cerebral patterns linked to affective states. Access to brain activity is achieved through specialized instruments and procedures such as electroencephalogram (EEG), computed axial tomography (CAT), and functional magnetic resonance imaging (fMRI). Using these procedures provides access to high resolution data; however, elevated costs may limit their use in some types of studies. Recently, low-cost and portable devices to measure brain activity through EEG have been developed (e.g. EPOC neuroheadset[14] and MindWave[15]), nevertheless data resolution

[14] http://www.emotiv.com/
[15] http://www.neurosky.com/

collected with these kinds of instruments is not as good as that obtained with medical ones.

Other procedures are based on the access to physiological signals linked to brain structures. For example, electromyography (EMG) is used to measure electrical potentials at the muscle level (Nakasone Prendinger, & Ishizuka, 2005). EMG levels can be measured at the intramuscular level or on the surface of the skin. The former method is considered invasive since it requires the use of needles that are inserted into participants to reach their muscles. An interesting aspect of EMG, according to Nakasone et al. (2005), is that potential signals correlates with affective processes with negative valence.

Another technique used to measure affective changes in terms of physiological signals is electrodermal activity (EDA), which corresponds to different types of measures such as galvanic skin response (GSR) and skin conductance (SC). EDA is a response of the human body when the individual “becomes mentally, emotionally, or physically aroused” (Strauss et al., 2005, p. 701). Such response, as part of the sympathetic nervous system, which has been linked to affective processes, is expressed by the sweat glands in the skin (Boucsein, 1992). Instruments to capture and measure EDA have been around for decades, though usages have been limited by their costs, sensitivity to motion artifacts, and invasiveness levels. At the present time, portable devices, with little invasiveness levels, and high reliability are available (González-Ibáñez, 2011). The Q Sensor, for example, is a small device capable of measuring and recording participants’ EDA from their wrist and also from other body locations. The Q Sensor comprises two electrodes that are attached to the skin of participants. EDA measures between the electrodes (expressed in micro siemens) are possible due to the sweat produced by the glands, which happens to be a good conductor. According to Strauss et al. (2005) EDA measures can be used as “an indicator of [participants’] level of excitement or relaxation” (p. 701).

It is noteworthy that the neurophysiological methods are mainly linked to dimensional approaches. For instance, EDA-based analyses can be connected primarily to arousal levels. However, specific categories of emotions such as anger or surprise, as well as affective tone cannot be obtained directly from EDA. An exception would be EMG, briefly described above, and some measures based on brain activity that have shown patterns that could be linked to the expression of specific emotions and also to valence.

Taking into account the portability and low levels of invasiveness of the *Q Sensor*, the experimental procedures carried out in this study include the use of this instrument in order to dynamically measure arousal levels, which as stated in the literature, can be linked to levels of engagement and decision making (Figner & Murphy, 2011).

2.2 Collaborative Information Seeking (CIS)

The second major focus of this dissertation corresponds to CIS, a relatively new research topic in information science. The theoretical and conceptual foundations around this topic are in a premature stage of development, so there are still shortcomings in regard to basic subjects such as definitions. This section presents a literature review on CIS with a focus on theoretical frameworks and evaluation.

2.2.1 Theoretical foundations of CIS

2.2.1.1 Definition

Searching for information has long been described as an individual activity (Reddy & Jansen, 2008; Shah, 2010a). It is recognized, however, that the information search process can be also performed collaboratively (Twidale, Nichols, & Paice, 1997; Reddy & Dourish, 2002; Morris, 2007; 2008), that is to say, two or more individuals working together while searching, evaluating, collecting, and using information to accomplish a common goal. A situation like this was introduced in this dissertation earlier through the example $S_{John[-]\ \&April[+]}$.

As a relatively new research topic in information science, CIS lacks a consensual definition. For instance, Foster (2006) argued that defining CIS depends upon assumptions made by researchers from a particular discipline. According to Shah (2010a), CIS is usually used as interchangeable term with others such as collaborative information behavior[16] (CIB) (Reddy & Jansen, 2008), collaborative information retrieval (CIR) (Fidel et al., 2000), collaborative exploratory search (CES) (Pickens & Golovchinsky, 2007), social search (SS) (Evans & Chi, 2010), information sharing (Talja & Hansen, 2006), and collaborative search (Morris, 2008; Morris & Teevan, 2010). While there exist overlap among these topics in the literature, they all refer to different processes. Figure 2.3 shows one possible view of CIS with respect to related topics and areas.

[16] According to Pettigrew, Fidel, and Bruce (2001), information behavior is defined as "the study of how people need, seek, give, and use information in different contexts, including the workplace and everyday living" (p.44).

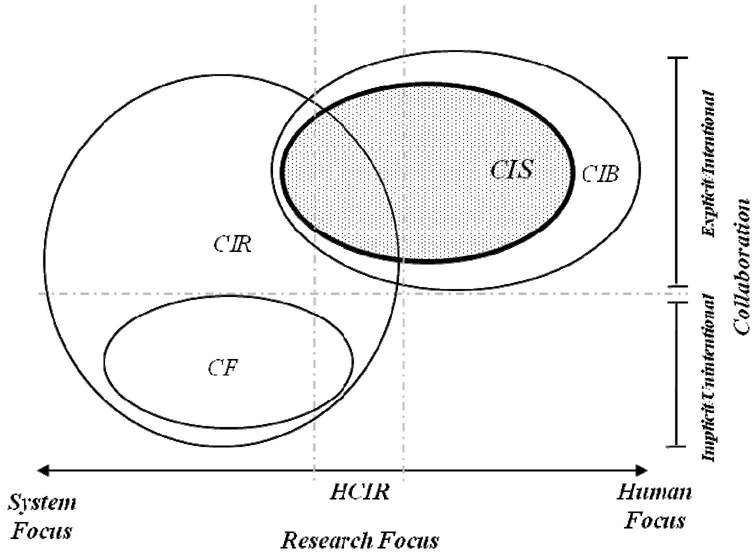


Figure 2.3. Collaborative information seeking (CIS) with respect to related research topics and areas: Collaborative information behavior (CIB), collaborative information retrieval (CIR), collaborative filtering (CF), and human computer interaction and information retrieval (HCIR). Horizontal axis represents the research continuum as described by Kelly (2009). Vertical axis categorizes research areas in terms of two views of collaboration.

Authors define CIS in different ways. Foster (2006), for example, defines CIS as "[t]he study of the systems and practices that enable individuals to collaborate during the seeking, searching, and retrieval of information" (p. 330). Shah (2012), on the other hand, defines CIS as "a process of information seeking that is defined explicitly among the participants, interactive, and mutually beneficial" [17] (p. 25). Both definitions are framed differently. While Foster's definition refers to CIS as a research topic, Shah's view describes CIS as a process. Considering

---

[17] Other authors, such as Karunakaran et al. (2010) defines CIS as "the purposive seeking of information by two or more individuals because of an information need in order to satisfy a shared goal". (p.3)

the complementary views of these definitions, in this dissertation both Foster's and Shah's definitions of CIS are adopted.

The above definitions, while providing a general perspective of what CIS is, do not elaborate on what is meant by key terms such as information seeking and collaboration. First of all, information seeking as defined in Shah (2012), corresponds to "the process or activity of attempting to obtain information in both human and technological contexts" (p.179).

With regard to collaboration, related research topics such as CIR are also defined around this concept, however, the views about collaboration differ among topics. For example, Shah's (2010a) definition implicitly delineates what is meant by collaboration by referring to explicitness, interaction, and mutual benefit. In CIR, however, collaboration does not necessarily imply interactions, explicit work, intention, joint goals, and mutual benefit. The underlying assumption in CIR is that the process of working with others may occur with or without individuals willing to do it (i.e. intentional and unintentional respectively) and with or without them being aware of the actions that make them part of the collaborative process (i.e. explicit and implicit respectively). So, how is that collaboration in CIS and CIR have different meanings?

Similar to emotions and related affective processes, it is also possible to find different definitions for the term collaboration. According to Shah (2010a), collaboration as a key term of CIS is usually used interchangeably with others such as cooperation. According to the Merriam-Webster (2013) dictionary, collaboration has the following meanings[18]:

1. The state of having shared interests or efforts.
2. The work and activity of a number of persons who individually contribute toward the efficiency of the whole.

The Oxford (2013) English dictionary, on the other hand, defines collaboration as follow[19]:

3. The action of working with someone to produce or create something.
4. Traitorous cooperation with an enemy.

Both dictionaries relate collaboration to cooperation. As noted in the fourth definition, cooperation is used to describe collaboration. On the other hand, the Merriam-Webster (2013) dictionary indicates that cooperation is a synonym of collaboration.

Cooperation, in particular, is defined in the Merriam-Webster (2013) dictionary as[20]:

5. Common effort.
6. Association of persons for common benefit.

Likewise, the Oxford (2013) English dictionary defines cooperation as:

7. The process of working together to the same end.
8. Assistance, especially by ready compliance with request.

---

[18] http://www.merriam-webster.com/dictionary/collaborate (Merriam-Webster, 2013)
[19] http://oxforddictionaries.com/us/definition/american_english/collaboration
[20] http://www.merriam-webster.com/dictionary/cooperation (Merriam-Webster, 2013)

From this list of definitions, it is difficult to differentiate collaboration from cooperation. Shah (2008, 2010a) takes a closer look into this definition problem by reviewing different aspects of collaboration and distinguishing it from related processes, namely, communication, contribution, coordination, and cooperation. Based on previous work by Denning and Yaholkovsky (2008) and Taylor-Powell, Rossing, and Geran (1998), Shah developed a set-based model that structures a hierarchical relationship among these processes.

Shah's model places communication in the center of a nested structure by positing it as an essential process to all others. Communication enables interaction and the exchange of information among group members. As a result of communication, contributions (e.g. ideas, products, opinions) can be made. Following these two processes, team members can coordinate their actions, which involves sub-processes such as conflict resolution and distribution of responsibilities. Cooperation, on the other hand, appears as nesting the previous processes. In order to cooperate, group members must communicate with each other in order to share their individual contributions and coordinate their actions toward a specific goal. While individuals can cooperate, the outcomes are not beneficial to all, at least not directly or explicitly. Finally, in Shah's model, collaboration is viewed as a major process that depends on cooperation, coordination, contribution, and communication. Compared to cooperation, collaboration implies mutual benefits and a synergic effect as a result of group members "[going] beyond their own individual expertise and vision by constructively exploring their differences and searching for common solutions" (Shah, 2010a, p. 6) and also "[the creation of] a solution or a product that is more than the sum of each participant's contribution" (p. 6).

In this dissertation, collaboration is defined as a social process in which two or more individuals *intentionally* and *explicitly* work together, in order to accomplish common goals that are mutually beneficial. The collaboration process as defined here is supplemented with related terms according to Shah's (2008) hierarchical sub-set model, that is to say, communication, contribution, coordination, and cooperation.

There are additional aspects to consider about CIS in regard to collaboration. Like emotions, collaboration can be characterized in terms of different dimensions such as time, space, interactivity, control, awareness, trust, cognition, and affective processes. For example, the space dimension indicates *where* collaboration takes place. This dimension can represent a continuum expressing different degrees of space with two extremes, namely, same location (co-located) and different locations (remotely located).

The time dimension, on the other hand, indicates *when* collaboration takes place. This relates to the concept of synchrony that indicates that processes can take place either at the same time (synchronous work) or at different times (asynchronous work). Similar to space, this dimension can also represents a continuum with levels such as parallel, concurrent, and sequential work.

Taking into account the conceptual definitions discussed above, the following section reviews some models of CIS and CIB.

2.2.1.2 Theories, models, and empirical research in CIS

While CIS has been developed mostly during the past decade, there are indications of interest in this topic in the late 90s (Churchill, Sullivan, Golovchinsky, & Snowdon, 1998). The history of CIS, however, goes back to early studies in collaboration and computer-supported cooperative work (CSCW) (Shah, 2010a). Although CIS has not been explicitly identified and described in the literature as a research topic in itself, CIS can be found as an implicit subprocess of different group activities that are performed with or without the support of computer systems.

It is acknowledged in this work that information seeking, whether is performed individually or in a group, is typically placed in context. There are many activities and situations that are developed around information, for example, planning a trip, patent research, school assignments, research projects, and health issues, to name a few (Hansen, & Järvelin, 2004; Shah, 2010c; 2012). In all these examples, there are implicit and explicit information needs that often play a key role in decision making and in the achievement of respective goals; however, the process to find the required information is not the goal itself.

This literature review does not cover general models of collaboration and CSCW, instead, it focuses on those developed specifically for CIS and CIB, the latter as a parent research topic of CIS as depicted in Figure 2.3. CIS and CIB studies have contributed along with empirical findings toward the development of models, theories, and systems. While general models and theories seem to be ideal to explain CIS in a broad sense and develop system-based solutions that support users in every possible scenario, research around these topics has shown that different domains (e.g. education, health, and legal) have unique characteristics that are usually not captured by general approaches. Ideally, a general model should include guidelines about what CIS is and at the same time constitute a core to the production of models for particular domains. The rest of this section explores some general models, the applicability of individual models to describe CIS, and models derived from specific domains.

- ***General CIS models***

A system-oriented model of CIS introduced by Shah (2008) resulted as an extension of a model of individual information seeking introduced by the same author in the same work. The model consists of four layers (Figure 2.4), which are explained as follow: The first layer (Layer-1) corresponds to information (referred to as sources in Shah, 2012) and represents a collection of information objects such as pages, videos, people, and databases where information can be obtained. The second layer (Layer-2) represents the tools (e.g. search engines) that are used to access the information; in CIS this layer may also include

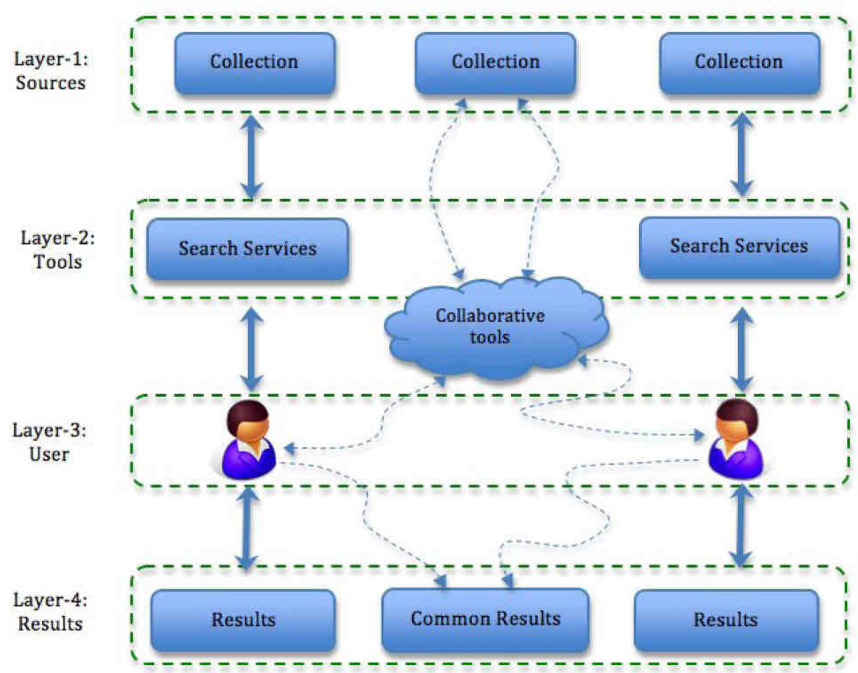


Figure 2.4. A model for collaborative information seeking (Shah, 2008, p.4).

specialized tools to support collaboration, information organization, and information sharing. The third layer (Layer-3) corresponds to the users that engage in collaborative practices. Finally, the fourth layer (Layer-4) symbolizes the relevant information that users in the previous layer were able to find either individually or collaboratively.

Shah's model (2008) does a good job in separating its components into modular layers and indicating how they are interconnected. In particular, according to this model the focus of the connection between Layer-1 and Layer-2 is on information retrieval (IR), whereas the connections between Layer-2 and Layer-3, and Layer-3 and Layer-4 focus on human computer interaction and personal information management, respectively. Additionally, Shah's model is characterized for being domain-independent in spite of its system orientation. Being domain-independent makes the model general, however it does not capture fundamental aspects contained in specific layers such as Layer-3. This layer, in particular, consists of a number of aspects linked to human factors such as behaviors, cognition, affective processes, and how these relate to information-related practices.

In an early work by Jäverlin & Ingwesen (2004), the authors propose a general model that integrates information seeking and retrieval (IS&R) (Figure 2.5). A particular aspect of this model is the incorporation of teams as cognitive actors in information search and their subsequent interactions with information and systems. According to Ingwersen (2005), "information actor (or team of actors) operate in, and is influenced by, a dual contextual frame: that of the IT and information spaces surrounding the actor(s)" (p. 216). While the authors

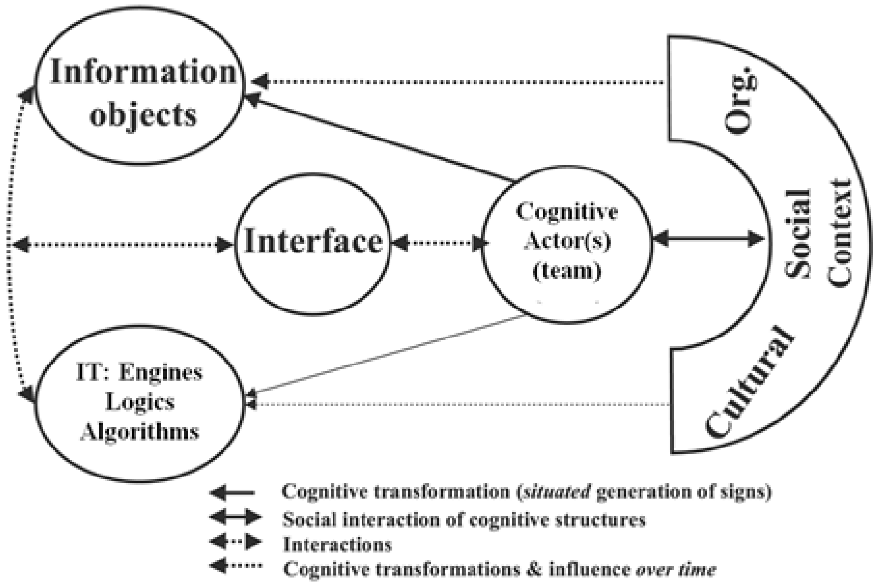


Figure 2.5. General model of cognitive information seeking and retrieval (Jäverlin, K., & Ingwesen, 2004). Adapted from Tenopir (2003, p.15) and Ingwersen (2005, p.216).

acknowledge the importance of cognitive and affective factors in the integration of IS&R as stated in the following quote, these are not developed in the general model.

> (IS&R) is a process of cognition for the information-seeking actor(s) or team in context. Algorithmic and IIR [(interactive information retrieval)], as well as information seeking (IS), involve cognitive and emotional representations from a variety of participating actors. Such representations are seen as manifestations of human cognition, reflection, emotion, or ideas forming part of IS&R components and kinds of interaction in context (Ingwersen, 2005, p. 215)

In the context of CIB, Karunakaran, Spense, and Reddy (2010) proposed a model of CIB for organizational contexts. The model, as depicted in Figure 2.6, consists of three phases and activities across them. The first phase corresponds to the identification of the problem by a group, who "create a shared representation collaboratively to arrive at a shared understanding of the situation" (p. 2). This shared representation can be made through "conversations, verbal communication and representations via artifacts and more" (p. 3). The transition between this phase and the second one is possible through triggers, which according to Reedy and Jansen (2008), in the case of CIB, corresponds to "Complexity of information need, Fragmented information resources, Lack of domain expertise and Lack of immediately accessible information" (Karunakaran et a., 2010, p. 3).

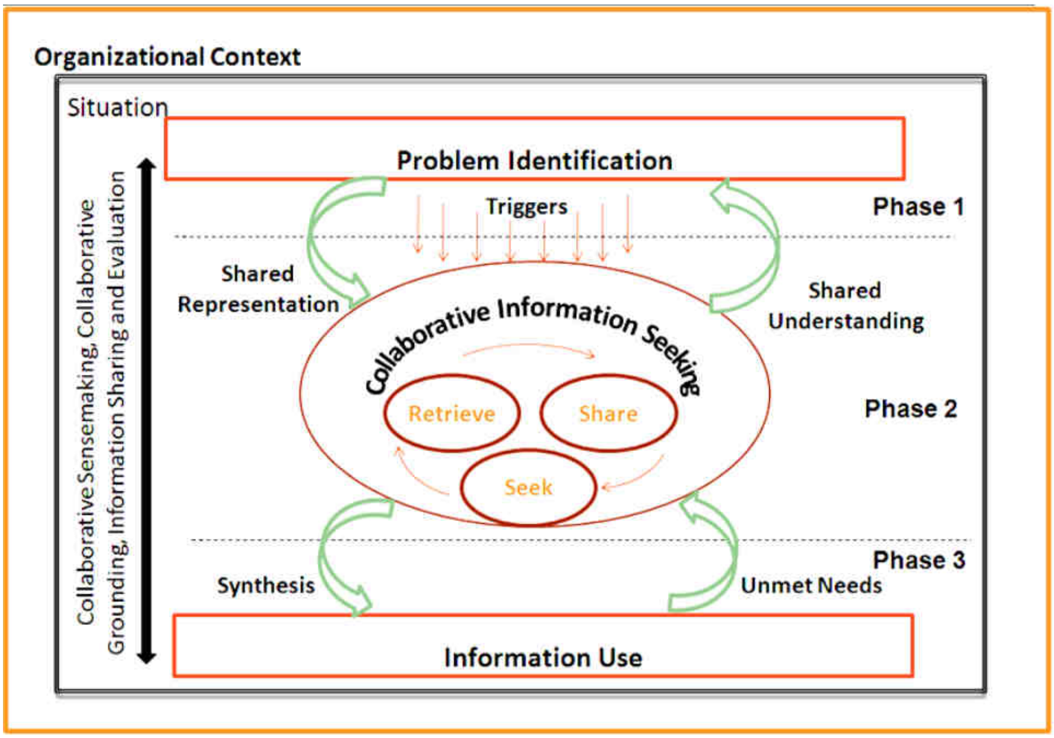


Figure 2.6. A model of collaborative information behavior (Karunakaran, Spense, & Reddy, 2010, p.3).

Once the problem is defined and a shared representation of it has been specified by the group, the process triggers the second phase, which starts with CIS. This process in particular involves different actions that are performed by group members, namely, retrieval, sharing, and seeking. As the authors indicate, "problem identification reciprocally shapes and gets shaped by the subsequent phase through the continuous process of moving from shared representation of the problem to shared understanding of the situation" (Karunakaran et a., 2010, p. 3).

In the third phase, group members proceed to use the synthesized information collected in the previous phase. During this process the goal is to incorporate "the information found into the group's existing knowledge base in order to achieve common understanding" (Karunakaran et a., 2010, p. 3). If some information needs are not met, then activities in previous phases can be performed again. In addition to main phases, the model also includes supporting activities, namely, information sharing and evaluation, collaborative grounding[21], and collaborative sensemaking.

A key question to address toward the development of a general model of CIS is whether or not there are aspects related to human factors and information practices that are common across all CIS scenarios. Of course responding to this question by studying CIS in all possible situations is unpractical and likely impossible. However, there may be common elements that in spite of not representing all CIS situations, at least could express a close approximation to

[21] "[T]he active construction by actors of a shared understanding that assimilates and reflects available information" (Hertzum, p.958)

most of them. The previous models, though general, have limitations and differ in terms of resources and perspectives on how CIS and CIB are framed.

- ***Individual information seeking models and CIS***

With a longer research history, studies in individual information seeking have produced several models attempting to explain in broad terms how people search, evaluate, and collect information, along with details regarding cognitive, behavioral, and affective aspects. A seminal piece in this domain is Kuhlthau's (1991) information search process (ISP), a model developed as a result of extensive empirical work that resulted in a clear organization of the information search process of individuals into six stages, namely, *initiation*, *selection*, *exploration*, *formulation*, *collection*, and *presentation*. Each stage of the ISP is characterized across four major dimensions: feelings, thoughts, actions, and appropriate tasks. For example, according to Kuhlthau, in the initiation stage searchers feel uncertainty, their thoughts are general/vague, their actions focus on seeking background information, and the appropriate task at this level is to recognize the information need.

A fundamental aspect of Kuhlthau's (1991) ISP, like other individual information seeking models such as Berrypicking (Bates, 1989), is its focus on the searcher. Taking into account this characteristic, some researchers have attempted to evaluate the applicability of ISP in collaborative domains. For example, Hyldegård (2006a; 2009) explored the behaviors of two groups of information science students working on a project over seven weeks. Rather than proposing a new model, the author compared collaboration practices of searchers to those of individual by using Kuhlthau's (1991) ISP model as a reference.

While similarities were found between group and individual practices, the author stated that differences in behaviors could be attributed to social factors. For instance, at the cognitive level "[g]roup members were generally regarded as reliable information sources for problem solving, hence held a cognitive authority, which implied that suggestions and recommendations from other group members were trusted" (Hyldegård, 2009, p. 153). Such factor found in collaborative contexts is not expressed in the ISP.

Based on data collected in a laboratory study, similar results were found by Shah and González-Ibáñez (2010). The authors attempted to map information behaviors of teams working on an exploratory search task to Kuhlthau's ISP stages. Although the authors identified the six stages of the ISP model, they found overlaps for three of them, namely, *exploration*, *formulation*, and *collection*. According to the researchers, these overlaps were explained in part due to specific group members' behaviors such as coordination and topic discussion. In terms of affective processes as well as relevance judgments, it was found that group members verbally express their feelings with respect to the situation as well as information when a communication channel is enabled. Based on the same study, it was later suggested by González-Ibáñez and Shah (2010) that relevance judgments are socially constructed through both objective and affective discourse; meaning that team members share their opinions (e.g. "This page contains useful information"), reactions (e.g. "I loved this page"), and objective comments (e.g. "This information came from the president of the company") with respect to the information they find. Furthermore, it was argued that social interactions carried out when selecting relevant material may dynamically shape feelings, engagement, and the confidence of team members.

The studies of Hyldegård (2006a; 2009) and Shah and González-Ibáñez (2010) demonstrate the potential limitations of mapping collaborative information behaviors into models intended to describe the individual search process. In particular, a general model like Kuhlthau's ISP is intrinsically limited by its structure since it does not consider social components. In order to properly map collaboration onto Kuhlthau's ISP, changes to its structure would be required, such as adding intermediary stages and a social dimension.

There have also been attempts to investigate collaboration through methodologies intended mostly for individuals. For example, Paul and Morris (2009) focused on sensemaking (Dervin, 1983; 1998), and while this approach can be "applied to entities other than individuals (e.g., collectives)" (Dervin, 1992, p. 277), the authors identified several challenges in the process of supporting collaborative search based on sensemaking. Some of the challenges were related to awareness and temporality, which were then addressed and evaluated experimentally with a system called Co-sense. Without generalizing, this shows that individual approaches to studying information seeking may be strongly limited in describing collaborative behaviors.

- ***Domain-derived models***

Yue and He (2009) investigate the behaviors of groups while working on a TREC[22] tasks in the Legal Track. The tasks referred to as *e-discovery* are characterized for their complexity when performed by an individual, and

[22] Text RETrieval Conference

therefore require collaboration to be properly addressed. As described by the authors,

> [i]n real e-discovery processes, there is a lead attorney who is in charge of overseeing a large document-review effort and for vouching for the completeness and accuracy of the produced collection. The attorney often hires an e-discovery firm or team to gather all the relevant documents from the full document collection implicated by the matter (Yue & He, 2009, p. 1).

As a result of their previous participation in the Legal Track 2008, the authors found three key aspects of CIB in e-discovery tasks, namely, communication, division of labor, and awareness. As part of their participation in TREC 2009, the authors conducted an experimental evaluation. Based on the results obtained from analyses at the task level, the authors proposed a CIB model for an e-discovery scenario.

Like Kuhlthau's (1991) ISP, the CIB model depicted in Figure 2.7 consists of a set of stages including initiation, exploration, division of labor, and final results. Since this model is task-dependent, the authors include division of labor as a key stage in the process.

Yue and He's (2009) model also describes cognitive aspects of CIS for each stage. Additionally, the authors included elements from Shah's (2008) set-based model of collaboration described in section 2.2.1.1, which are depicted as part of collaboration levels and related to each stage in the model. Additional aspects in

Stages	Initial & Exploring ------> Division of labor ------> ...... ---------->Final result		
Cognitive Load	--> adaption to the system ----------------------> getting familiar with other collaborators		
Collaboration Levels	Collaborative -----------------> Individual -----------------> Collaborative Coordination / Communication Cooperation / Contribution		
Collaboration Type	Synchronous --------------------> Asynchronous ---------------------> Synchronous		
Subtask complexity	Easy ---------------------------------> difficult ------------------------------> easy		
Types of supports needed	Support for frequent communication and keeping awareness	Implicit collaborative support (recommendation) support for relevancy judgment (topic authority)	shared space for common results

Figure 2.7. Collaborative information behavior model for an e-discovery scenario (Yue & He, 2009, p.8).

this representation are the temporal dimension (referred to as collaboration type), subtask complexity, and types of support.

There are some restrictions to consider about this model. First, is the fact that the model is domain-dependent and task-dependent. Second, the experimental evaluation was limited to specific resources to support team members' search processes, information sharing, and communication. For example, the authors used a search tool called "email explorer." Collaboration, on the other hand, was supported with Skype text-chat, screen sharing, and a wiki. Moreover team members were co-located in the same room but using different computers, thus verbal communication was allowed. The use of such resources without proper evaluation of their appropriateness to support collaborative information behaviors may make this model also technology-dependent.

Other domain-dependent models of CIB have been developed. For instance, Hansen and Järvelin (2004) introduced a CIR framework derived from an information seeking and retrieval model. The model is divided into three levels, to be precise, work task level (i.e. initiation, preparation, and planning), information seeking task level, and retrieval task level. The model provides detailed specification of the activities carried out by actors in each of the stages; however, such level of detail limits the possibilities to expand or adapt the model to other domains.

Reddy and Jansen (2008) proposed a CIB model focusing on general aspects; however, the model was derived from empirical research conducted in the health domain. The model is described in terms of two axes, behavior and context. Behavior, on one hand, distinguishes information search[23] from seeking[24]. Context, on the other hand, refers to whether information behaviors[25] are individual or collaborative. These two dimensions are depicted in Figure 2.8.

Unlike the models described above, Reddy and Jansen's (2008) model was developed around dimensions. Each dimension in this model is a continuum. While there are not clear distinctions of stages, orders, or specific activities performed by actors, the model illustrates the possibilities to move along these

---

[23] According to Reddy and Jansen (2008), seeking is "tactical maneuvering" (p.266)
[24] According to Reddy and Jansen (2008), searching is "strategic maneuvering" (p.266)
[25] According to Reddy and Jansen (2008), information behavior is "philosophy of seeking and use" (p.266)

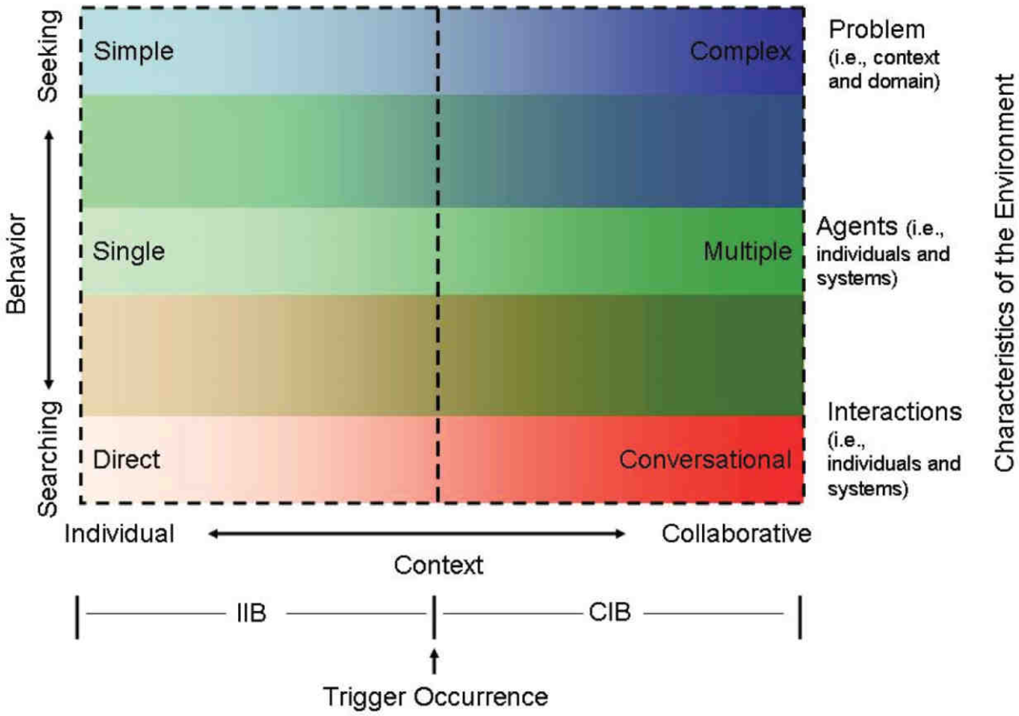


Figure 2.8. Individual versus collaborative information behavior (Reddy & Jansen, 2008, p.266).

dimensions based on particular characteristics of the environment such as the kind of problem, the number of agents, and the types of interactions. Nevertheless, like the previous general models reviewed above, this one fails to incorporate internal aspects of the participating agents such as cognition, affective processes, and actions.

The model is effective in showing that depending upon the specifics of each scenario, actors may engage in individual information behaviors (IIB) or in collaborative ones (CIB). For example, if the task is simple[26] this can be

---

[26] The notion of simple and complex is acknowledged by Reddy and Jansen (2008) to be subjective and possibly subject to debate.

performed individually. However, as the task become more complex, multiple agents may be involved.

This section has reviewed general models of CIS describing their advantages and limitations. There was also an overview of some attempts to explain collaborative information behaviors through models originally intended for individuals. Finally, a few models that were created from research findings in particular domains were discussed. A discussion on CIS evaluation is presented below.

2.2.2 CIS evaluation

One of the major challenges around CIS is evaluation. According to Foster (2006), CIS is defined as "[t]he study of the systems and practices that enable individuals to collaborate during the seeking, searching, and retrieval of information" (p. 330). In order to perform such studies, it is necessary to perform qualitative and/or quantitative evaluations. According to Shah (2012), CIS evaluation is a challenging process due to the complexity of its structure (i.e. multiple users, specialized systems to support collaboration, and the interactions carried out during the collaboration process). In a more general sense, it is acknowledged in the CSCW literature that one of the major issues of evaluation is due to the different approaches applied, which come from investigators with diverse research backgrounds (Neale, Carroll, & Rosson, 2004). In CIS, for example, Shah (2012) describes different methodological approaches of evaluation; in particular, he mentions the evaluation of usability through user studies, system-based training, qualitative evaluation, and task or application based evaluation. This section provides an overview of evaluation approaches in CIS relevant for the development of this work, which are categorized into three major groups: product, process, and system.

2.2.2.1 Product-based evaluations

The evaluation of products sees CIS as a black box, that is to say, a transformative process that receives a set of inputs and produces certain outputs. The outputs, in this case, correspond to the end products that will be evaluated. In a CIS task, product types may depend on aspects such as task and domain. Typically product evaluation is based on performance measures, for instance, the time that group members spend in accomplishing a given task (Baeza-Yates & Pino, 1997), the number of relevant documents they collect, and/or information synthesis. Performance measures are based on aggregated values that summarize the overall results of a process or subprocess. These measures can act as indicators of collaboration costs (Klein & Adelman, 2005), the quality of collaboration (Baeza-Yates & Pino, 1997), and the products derived from it. The term indicator used above denotes the intrinsic limitations of product-based evaluations. For instance, describing the quality of collaboration based on specific measures of the outputs may be questionable since it does not consider internal aspects of the process.

Product evaluation in CIS can be performed in both natural and experimental settings. While naturalistic studies focuses mainly on the process, they may also evaluate its products. For example, in studies carried out in educational settings, students' grades could be used as a product measure; however, this may be limited by the particular conditions of the situation. As reported by Hyldegård (2006a), grades were not used as part of her evaluation because students "were allowed to choose between two types of grade systems [...] for their assignments" (p. 284).

In the general field of CSCW, Baeza-Yates and Pino (1997) developed a framework to evaluate collaborative work. As part of this framework, the authors describe performance analysis, which focuses on the quality of the work. In order to carry out this evaluation, the authors divided the overall process into several stages or subprocesses and conducted evaluations on each of them. According to an experimental evaluation, the authors found that "quality improvement rate decreased with every stage" (Baeza-Yates & Pino, 1997, p. 57). In addition, the authors described that adding more members to groups did not contribute to improving the quality of the work. The authors also indicated possible performance differences based on aspects such as parallelism and location. For instance, Baeza-Yates and Pino (1997) argued that co-located work should take less time than distributed work since group members working in the same space would have less problems communicating with each other.

As pointed out by Baeza-Yates and Pino (1997), different collaboration conditions may produce different outcomes. An effective way to investigate such differences is through experimental evaluations, which enable researchers to control different aspects of collaboration such as levels of parallelism, communication, awareness, and space. These kinds of experiments, however, have been criticized due to the inherent limitations of experimental setups (Neale et al., 2004). For the case of CIS, it is possible to find studies carried out in experimental settings, semi-controlled environments, and also naturalistic studies.

Shah and Gonzalez-Ibáñez (2011; 2012), for example, carried out a user study to investigate the effects of different collaboration conditions in terms of space and communication resources in an exploratory search task. In this study, the authors focused on the collaborative products of pairs of users through an evaluation framework that consists of IR measures adapted for CIS, information diversity, and user measures such as questionnaires to determine cognitive and affective load. It is noteworthy that all these measures were computed as end products.

In the case of IR measures, Shah and Gonzalez-Ibáñez (2011) used adapted versions of precision, recall, and F-measure (an approach similar to this was previously taken by Smyth et al., 2005), which were computed with respect to an universe of pages formed by distinct documents visited by all participants in the study. The authors also used complementary measures such as likelihood of discovery, which indicates how difficult or easy it would be for a user to find a given document. The organization of information coverage based on these measures is depicted in Figure 2.9.

A measure of usefulness was also implemented in this study. This measure consists of the time that users spend in webpages (dwell time) with a threshold of 30 seconds as an indicator that the page is useful (Fox, Karnawat, Mydland, Dumais, & White, 2005; White & Huang, 2010).

As a result of this study, Shah and Gonzalez-Ibáñez (2011) found significant differences across five experimental conditions: single users, pairs of users working co-located at the same computer, pairs of users working co-located at different computers, remotely-located users using text-chat communication, and

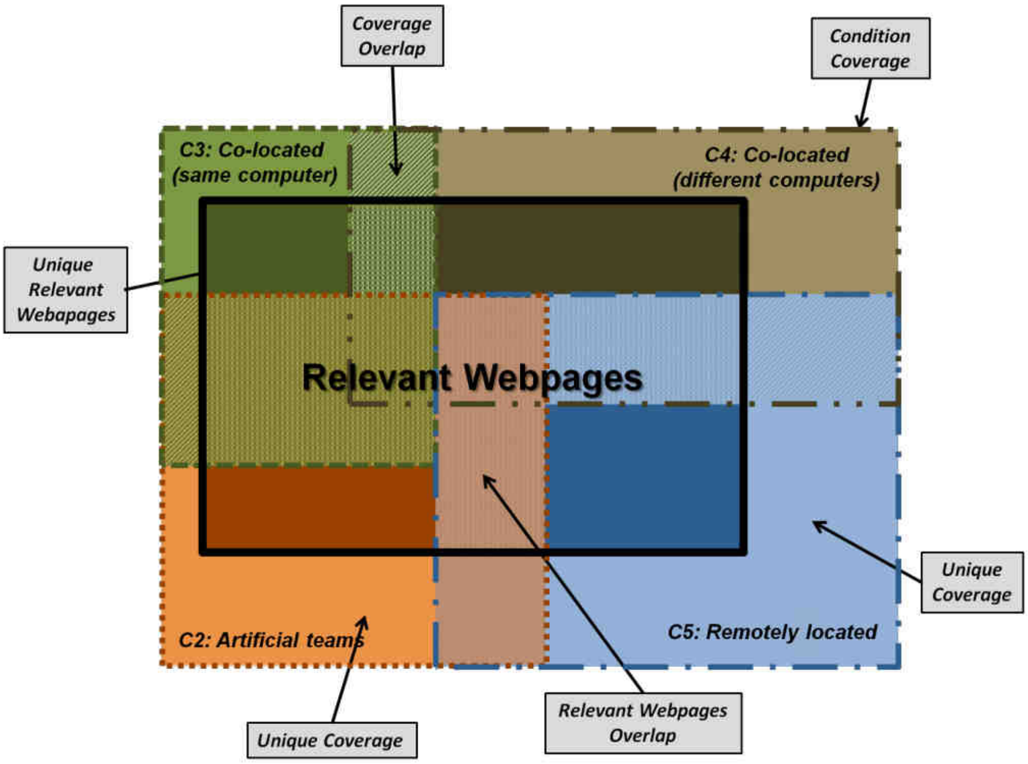


Figure 2.9. Depiction of coverage by various collaborative conditions (Shah & Gonzalez-Ibáñez, 2011, p.6)

simulated pairs. Single users and simulated pairs were used as baselines to perform comparisons and evaluate the synergic effect of collaboration. Among the significant findings, the authors described that working co-located at the same computer was similar to individual work with the added value of higher precision levels. The authors also found that working remotely located enabled users to reach more independency, thus reaching higher levels of coverage of diverse and useful information. The researchers also showed that working collaboratively, in particular users remotely-located or co-located but using different computers, reported better results than simulated pairs with the added cost of higher levels of cognitive load, albeit the latter measure for the case of simulated pairs was based on that of single users.

In a follow-up study, González-Ibáñez, Haseki, and Shah (2012a; 2012b; 2013) investigated the effects of time and space. In this new study, the authors discarded the individual condition and that of simulated pairs. In turn, they added two additional collaborative conditions: remotely-located using voice communication and asynchronous collaboration. The evaluation framework in this case considered the following five factors: communication, productivity, information synthesis, cognitive load, and affective load. Product-based evaluations of communication were performed in terms of volume, effort, and balance. For productivity, the authors used the evaluation framework of Shah and Gonzalez-Ibáñez (2011). Information synthesis, on the other hand, was performed using a grading system and two readability measures (i.e. Flesch Kincaid Grade Level and Flesch Reading Ease; Kincaid, Fishburne, Rogers & Chisson, 1975). Like Shah and Gonzalez-Ibáñez's (2011) study, evaluation of cognitive and affective load was conducted using the NASA-TLX (Hart & Staveland, 1988) and the PANAS (Watson et al., 1988), respectively. Significant results were found at the levels of communication, cognitive load, and affective load. For instance, coordination efforts of pairs working co-located at the computer were significantly higher than participants working with text-based communication. Another finding indicates that participants working asynchronously or with text-chat support were more task-oriented than other conditions, in which a significant amount of communication efforts were dedicated to task-social and non-task related conversations. A summary of the overall evaluation framework used in Shah and Gonzalez-Ibáñez's (2011) and Gonzalez-Ibáñez et al.'s (2012a, 2012b) studies is depicted in Figure 2.10.

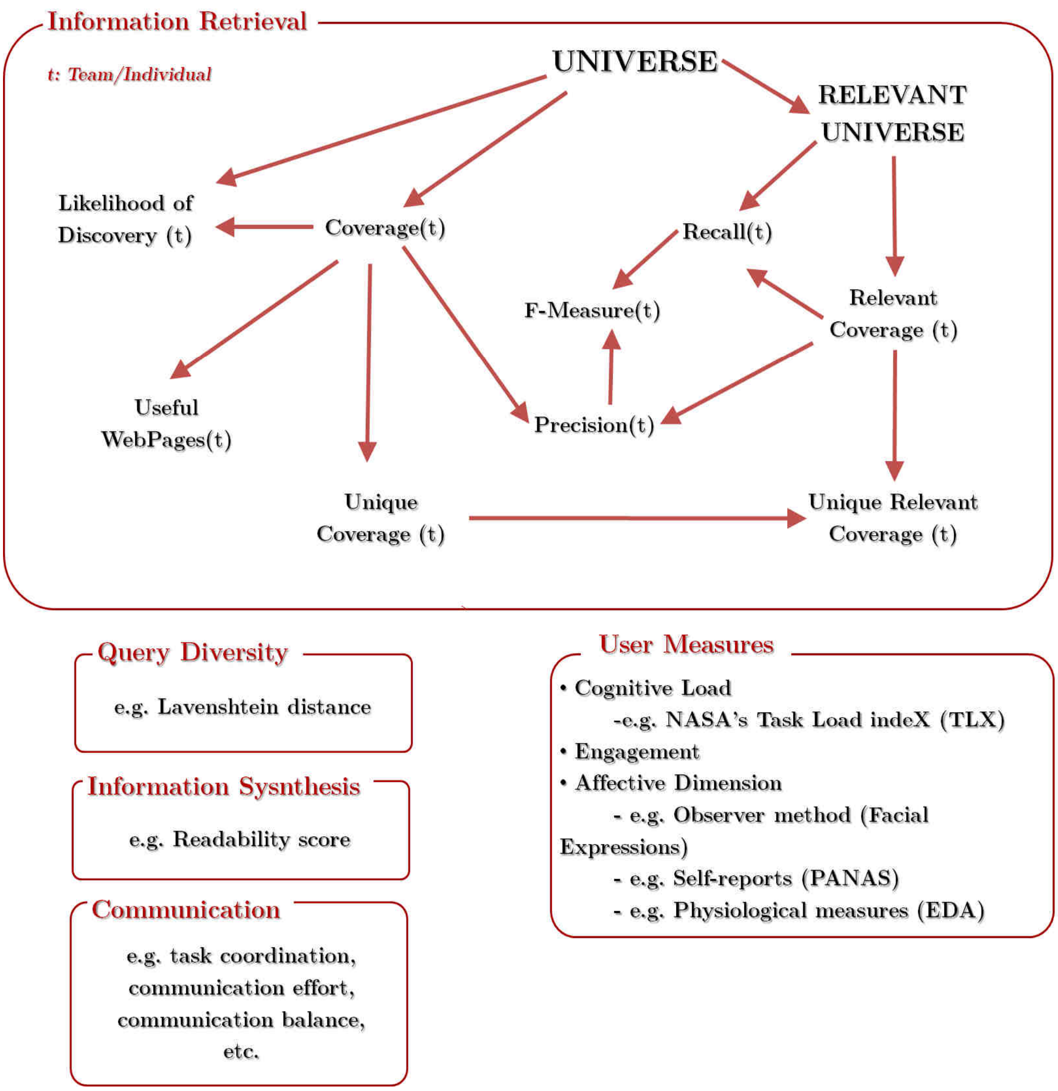


Figure 2.10. Summary of evaluation framework. Adapted from Gonzalez-Ibáñez, Haseki, and Shah (2011).

Specialized measures have also been used to measure effectiveness and efficiency. For example, González-Ibáñez, Shah, & White (2012) described a method to simulate collaboration among users in order to provide timely and pertinent recommendations from systems. In order to evaluate the simulated collaborative process, the authors defined the following two measures:

$$Effectiveness(u,t) = \frac{UsefulCoverage(u,t)}{OverallCoverage(u,t)} \qquad \text{(Eq. 2.1)}$$

and

$$Efficiency(u,t) = \frac{Effectiveness(u,t)}{NumQueries(u,t)} \qquad \text{(Eq. 2.2)}$$

where u represents an individual user or a team, and t the time up to where the measure is computed. More details about these measures are provided in section 3.2.2.1. The authors showed these measures to be helpful in evaluating information search in terms of products and process, the latter by comparing collaborative products at different stages of the ISP. More details about process-based evaluations are discussed in the following section.

2.2.2.2 Process-based evaluations

Process-based evaluations take a closer look at CIS by studying aspects such as procedures, practices, strategies, human factors, actions, and interactions. In other words, based on the black box metaphor introduced in the previous section, a process-based evaluation is performed by opening the black box and exploring its different layers. Following this approach, the granularity of analyses will depend on how deep the researcher wants to go in exploring the process. For instance, upper layers require coarse-grained analyses, whereas deeper layers involve fine-grained analyses as well as higher sampling rates and data resolution during the data collection stages of the research process (see Appendix C.4).

Process-based evaluations at different layers can be performed in terms of product-based measures. For example, Baeza-Yates and Pino (1997) divided the collaboration process into multiple stages and analyzed each of them in relation to their partial products. Similarly, González-Ibáñez, Shah, & White (2012) computed effectiveness and efficiency in each minute with the aim to evaluate the search processes of simulated pairs. Gonzalez-Ibáñez et al. (2012a; 2012b), on the other hand, used product-based measures to describe communication processes in each minute in order to describe how people interact around an exploratory search task. In particular, this latter study showed that in spite of different collaboration conditions leading to similar results, their respective processes differ significantly at the level of interactions, conflict resolution, and strategy definition, to name a few.

Other approaches to studying the collaboration process are based on quantitative self-reports and also qualitative approaches such as ethnographic studies, observational research, interviews, diaries, and focus groups. For example, Haseki, Shah, and González-Ibáñez (2012) conducted a diary study in order to investigate the collaborative practices of groups of students in a masters-level human information behavior class. Participants in this study completed weekly reports of their individual and collaborative processes around class activities and their final projects. Diary studies are particularly useful in helping people to self-report their activities, thoughts, feelings, and progress in a detailed and unstructured (or semi-structured) manner in mid and long-term research projects.

The authors also complemented data collected through diaries with interviews conducted with some of the participants in the study. One of the main findings of this study was that communication preferences of groups switched from asynchronous to synchronous forms, in particular face to face, as they got close to deadlines. It was also reported by participants that the more they interacted, the more satisfied they felt.

Hyldegård (2006a) also approached the study of CIS with a focus on the collaborative search processes carried out by a group of information science students. The author used both quantitative and qualitative approaches. The quantitative approach was based on the use of a questionnaire to collect nominal, ordinal, and scale forms of data about a wide range of factors such as demographics, collaborative experience, cognition, and feelings. Using information obtained through this questionnaire, the author conducted the first of three interviews. The first interview was conducted two weeks after initiated the project. The second and third interviews took place in the middle and at the end of the project, respectively.

Moreover, during the process that the project took place, each student was required to complete a diary for a period of five weeks (this technique is further described in Hyldegård (2006b)). With this methodological approach, the researcher collected information about the process under the assumption that participants in the study were willing to take an active part in the study by providing reliable information during the period the study took place.

Observational research in collaboration has been applied in both experimental and natural settings. This kind of approach consists of covert or overt observations of the collaboration process and with or without participation of the researcher. Observations can be conducted while the process is performed, or later through logs and video recordings of the collaborative situation. The latter method is more suitable in experimental settings due to ethical constraints. In general, observational approaches applied in experimental settings help researchers to describe the process as events go on; take field notes, and if followed by interviews, formulate contextual questions; and finally, support the interpretation and discussion of findings derived from other data sources.

Evaluation frameworks of collaborative work also suggest the evaluation of additional dimensions such as control, awareness, coordination, and conflict resolution (Klein & Adelman, 2005). The focus on these dimensions can be approached with product-based and also process-based analyses. Usually mixed-methods (both quantitative and qualitative) may be required to properly explore and describe the collaboration processes around these factors with the hope of identifying patterns and key themes.

2.2.2.3 System-based evaluations

The final type of evaluation described in this section corresponds to evaluations performed around a given system. There are different studies in the literature that describe the implementation of a CIS system followed by a study of users' experiences around these software-based solutions (Pickens et al., 2008; Shah, 2010b; Golovchinsky et al., 2012).

CIS systems used in evaluations often play multiple roles within the evaluation process. For example, they are intended to support the collaboration process, and provide mechanisms for information organization and information sharing; they also provide communication channels and awareness resources, information collection tools, and no less important, instruments to collect data during information search activities.

Although system-based evaluations can be performed in terms of products, typically they focus on processes with a final goal of measuring the performance of the system and determining how this can be improved. In addition, these kinds of evaluations may result in the development of new models and hypotheses, as well as provide technology to lay people. The following section provides a brief review of some systems and tools that support CIS and have been used in the past for research purposes.

2.2.3 Computer-supported CIS: Systems and tools

An exploration of the literature in CIS indicates that most systems and tools were created for scientific purposes and that very few of these software-based solutions are released for public usage. This section provides an overview of three systems and tools found in the literature that support CIS, and focuses on their common as well as unique features.

2.2.3.1 Coagmento

Coagmento (Shah, 2010b; González-Ibáñez & Shah, 2011) is a highly modular web platform that provides fully functional logging capabilities. Coagmento offers support for CIS in terms of awareness, communication, information sharing, and information synthesis, among other features. This system has been used in multiple studies (Shah, & González-Ibáñez, 2010; Shah & González-Ibáñez, 2011; González-Ibáñez, Shah, & Córdova-Rubio, 2011; González-Ibáñez, Haseki, & Shah, 2013), and though it was originally developed for research purposes, after several iterations the system evolved and it was finally released for public usage[27]. A screenshot of the main components of the public version of Coagmento is presented in Figure 2.11.

Coagmento consists of five major components, specifically, (1) a web-based platform that provides access to projects and information collected, (2) an extension for the Mozilla Firefox web browser which contains tools to save

---

[27] *http://www.coagmento.org*

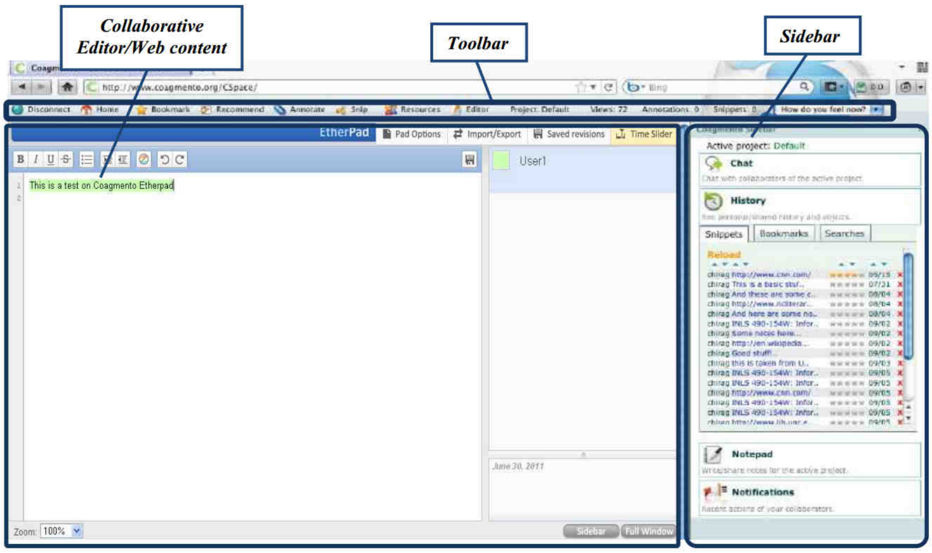


Figure 2.11. A screenshot of Coagmento (González-Ibáñez & Shah, 2011).

pages, make relevance judgments, collect snippets, make annotations, and provide direct access to other resourcesa (3) a sidebar, which is also part of the Firefox extension; this sidebar provides access to a text-based communication channel, search and browsing logs, notes, notifications, and shared documents, (4) a real-time collaborative editor that supports collaborative information synthesis, and finally, (5) a mobile application for Android devices that provides users ubiquitous access to their projects.

2.2.3.2 SearchTogether

Originally designed as a desktop prototype as a result of preliminary surveys on search practices (Morris, 2007), SearchTogether (Morris & Horvitz, 2007) later became publicly available as a plug-in for the Microsoft Internet Explorer (IE) web browser (Figure 2.12). SearchTogether supports collaborative search either synchronously or asynchronously. Similar to Coagmento, this tool offers resources to support awareness, communication, and "[persistent] representation of search" (Morris & Horvitz, 2007, p. 4). A feature that characterizes

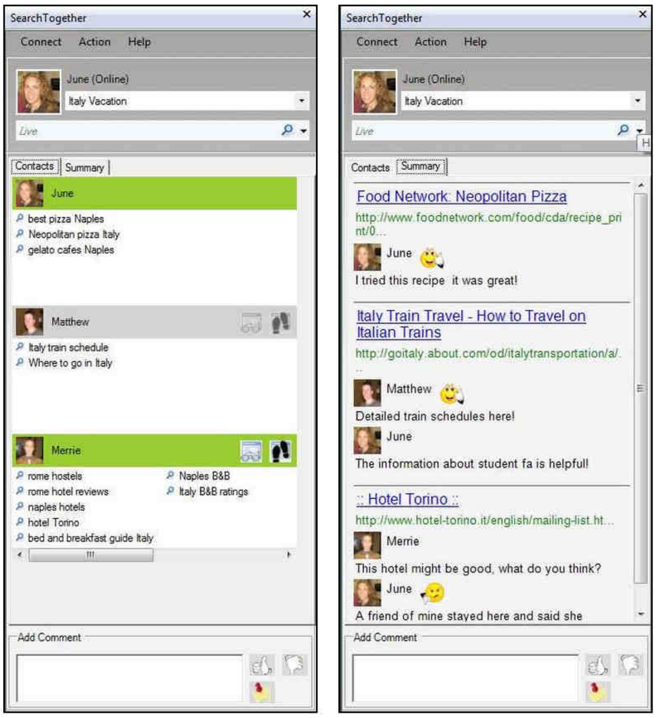


Figure 2.12. Screenshot of the plug-in version of SearchTogether (2008).

SearchTogether is the "Split Search" function, which consists of a procedure that automates division of labor. The authors, however, found that this feature was not often employed by users in the study.

2.2.3.3 Querium

Golovchinsky, Dunnigan, and Diriye (2012) introduced Querium, a web-based solution to support session-based collaborative search. Unlike Coagmento and SearchTogether, Querium's appearance resembles that of a web search engine, nonetheless, it is not an IR system. As described by the author, "Querium organizes search activities into task, and augments the search view by adding a resizable embedded document pane [...], rich query history displays, additional controls on the search results, a variety of filters, and a notepad/chat window" (p. 1799).

Besides these features, what distinguishes Querium from other systems is the incorporation of what the authors call "algorithmic mediation for collaborative exploratory search" (AMC) (Pickens, Golovchinsky, Shah, Qvarfordt, & Back, 2008) implemented in the form of a software-based agent that takes active part in the collaboration process of users by providing mechanisms for: automatically requesting information on behalf of users, re-ranking, clustering, and query fusion. AMC was previously implemented in Cerchiamo (Pickens et al., 2008), which unlike Querium, Coagmento, and SearchTogether, was specially designed to support role-based collaboration.

2.3 Affective Dimension and information seeking

The final section of this chapter focuses on reviewing the literature that connects the affective dimension with information seeking. Previously in this chapter, the literature reviews on the affective dimension and CIS were presented independently. It is acknowledged, however, that overlaps between these two research topics exist, but research problems that connect affective processes and CIS are barely explored in information science and other disciplines.

Some information scientists have identified and partially addressed the affective dimension in individual domains. It is often the case that researchers use concepts such as the affective dimension, emotions, feelings, moods or affects to indicate that affective processes are somehow linked to the information practices carried out by searchers. In other cases, umbrella terms (Savolainen, 2007) are used to refer to affective processes combined with other aspects such as cognition. For example, in regard to the concept information behavior in a review of conceptual frameworks used in the information science literature, Pettigrew et al. (2001) explain that "[…] information behavior endorses research that examines the cognitive and emotional motivations for information behavior that carry across contexts or are independent of context" (p. 46).

This situation suggests that it is possible that information scientists have used information behavior as an umbrella term in their studies to implicitly refer to both cognitive and affective aspects without making a clear distinction between them. Regrettably, very few have actually conducted research with a strong focus on the affective dimension and its role in information seeking.

The following sections present relevant theories and models, as well as empirical studies that have incorporated or at least considered the affective dimension as a key component in both individual and collaborative information seeking.

2.3.1 Affective dimension and individual information seeking

Researchers in information science have recognized the participation of affective processes in different areas related to information behavior. However, it is acknowledged that these studies posit a noticeable focus on cognition. For example, Fidel, Pejtersen, Cleal, and Bruce (2004) indicate that "[…] the prevailing approach among researchers in information behavior [… is …] the psychological, focusing primarily on cognitive factors and to a much lesser degree on others, such as affective and perceptual factors" (p. 940).

Although some authors in information science discuss affective-related aspects (e.g. motivations and satisfaction) as part of their theories, models, and empirical research, the majority of these works have not been designed giving the affective dimension a central role. It is often the case that the affective dimension is just mentioned in a much larger context within the discussion. As Fidel et al. (2004) point out, "[w]hile a large number of studies investigated psychological variables, the field is not prepared as yet to express variables from other dimensions on a specific, measurable level" (p. 940).

Information science is relatively new compared to other scientific disciplines. As part of its development as a science, information science-based empirical studies of phenomena have allowed the identification and description of new elements that were not previously considered to be relevant (e.g. the affective dimension). Regarding this fact, Fidel and colleagues (2004) adds:

> [a]nother tactic has been to uncover these dimensions through field studies. Solomon (1999), for instance, observed and analyzed information behavior in three different contexts. He created an information mosaic for each context that represented patterns of action. When he compared these mosaics he found that common to all were the factors: action preference; way of thinking (cognitive); knowledge about task, problem, etc.; and response to the actions of others (affective) (p. 941).

Two decades ago, very few scholars highlighted the potential role of affective processes to explain the way people behave within information-related situations. A seminal piece in this matter is Kuhlthau's (1991) ISP model, which borrowed elements from Kelly's (1967) personal construct theory to investigate information search from the perspective of individuals. As a result, one of the sub-products of her work was the identification of a list of specific feelings that individuals experienced at different stages of the ISP (Table 2.1). The feelings listed by Kuhlthau include uncertainty, optimism, confusion, frustration, doubt, clarity,

Table 2.1. Information search process (ISP) (Kuhlthau, 1991, p.367).

Stages in ISP	Feelings common to each stage	Thoughts common to each stage	Actions common to each stage	Appropriate task according to Kuhlthau model
1. Initiation	Uncertainty	General/Vague	Seeking Background Information	Recognize
2. Selection	Optimism			Identify
3. Exploration	Confusion/Frustration/Doubt		Seeking Relevant Information	Investigate
4. Formulation	Clarity	Narrowed/Clearer		Formulate
5. Collection	Sense of Direction/Confidence	Increased Interest	Seeking Relevant or Focused Information	Gather
6. Presentation	Relief/Satisfaction or Disappointment	Clearer or Focused		Complete

confidence, sense of direction, relief, satisfaction, and disappointment. The ISP model was developed as a result of empirical studies. In fact, Kuhlthau conducted different studies using both quantitative and qualitative methods (e.g. interviews, journals, flowcharts, search logs, and questionnaires) and the model was derived from content analyses of the data collected.

Regarding the affective dimension, the model states that as individuals progress in their search processes, they experience specific feelings[28] that are derived from each ISP stage. One particular aspect of this model is the assumption that searchers start by experiencing a negative feeling, that is to say, uncertainty.

[28] Note that Kuhtlthau (1991) uses the term feeling, which according to the methods she used to collect and analyze the data, conforms to the definition provided in section 2.1.1.4.

Then, as they move toward understanding, they may be able to reduce uncertainty and at the end of the ISP, find relief, satisfaction, or, alternatively, disappointment.

The ISP model makes a key contribution to the study of human information behavior, in particular information seeking, by placing the affective dimension as a relevant piece of the puzzle that explains how people search for information. After Kuhlthau's work, others started paying attention to the affective dimension in their information-related studies. Today, it is possible to find a considerable amount of work discussing affective aspects in the general field of information science. A good compilation of research about the affective dimension in information seeking as well as other information-related problems can be found in the book *Information and Emotion*, edited by Nahl and Bilal (2009). The works presented in this book explore the affective dimension within various domains yet research centered on affective aspects (i.e. studies specially designed to study the affective dimension in information-related problems) is still limited.

Another interesting model that emphasizes the role of the affective dimension as comprising information behaviors was developed by Nahl (2009), which is referred to as the "socio-biological information technology model of information behavior" (p. 3). In order to develop this model, the author considered and integrated aspects from other disciplines such as neuroscience, cognitive science, affective computing, human-computer interaction, and information science. Besides the integration of affective and cognitive factors, another key aspect of this model is the participation of the technological component. As the author states,

> [a] unified theoretical framework needs to integrate the three intersecting zones of the information environment as it exists today, namely, the intersection of technology, human biology, and social structures that define group practices and community values. An information behavior theory should therefore describe a symbiotic integration of technological affordances, social practices, and biological activity such as the user's sensorimotor, cognitive, and affective procedures during information reception and use (Nahl, 2009, p. 3).

A representation of Nahl's model is depicted in Figure 2.13. According to this diagram, there is a "recurrent flow of information behavior in the context of

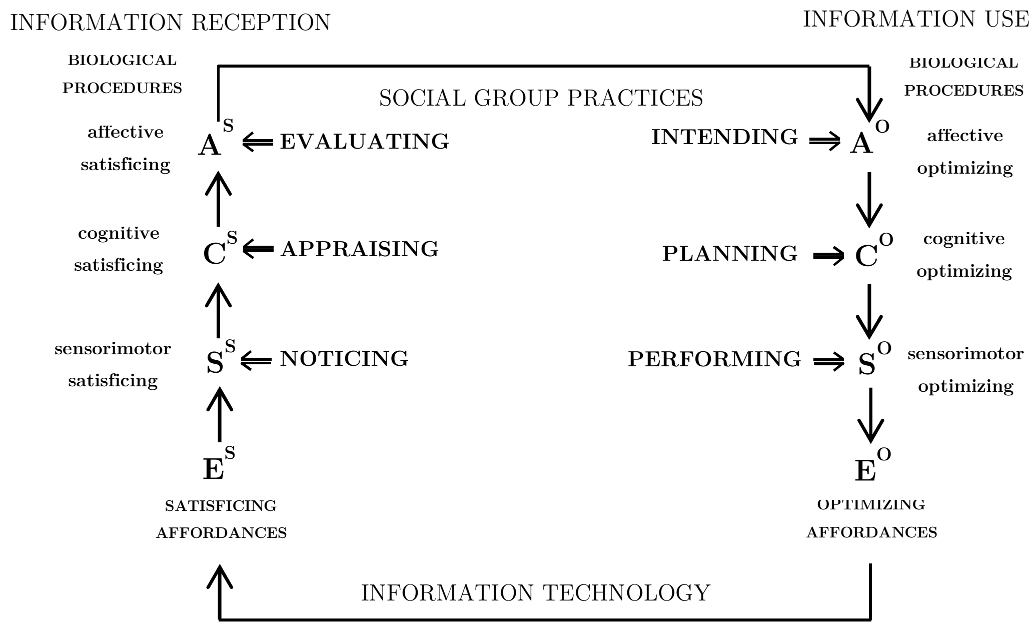


Figure 2.13. Socio-biological information technology (Nahl, 2009, p.7).

[information] reception and use" (p. 7). The flow, as expressed in the diagram denotes a sequential order in the way processes are activated. This sequence starts with information reception in the sensorimotor layer, followed by the cognitive, and finally the affective one. The process then continues backwards (i.e. first affective, followed by cognitive, and finally sensorimotor) as part of information use. Note that in this model "social group practices," as illustrated in Figure 2.13, do not relate to collaborative work, instead the concept is developed around the idea of affordances in social contexts.

As explained by Nahl (2009), "there are two basic biological functions of the human affective system in relation to information behavior. One function is adapted for reception and evaluation of information, while the other has evolved for the use of information in the purposeful planning of tasks" (p. 5).

Nahl's theoretical approach is well supported by empirical evidence as well as theoretical frameworks from other disciplines. It is important to note, however, that the model relies on underlying assumptions that have been the object of debate as discussed early in this chapter. For example, the sequential relationship between sensorimotor, cognitive, and affective layers conforms to cognitive approaches in the study of emotions as described in section 2.1.2.1 and differs from other biological approaches such as the neurophisiological one from Cannon-Bard (Cannon, 1922).

In an earlier work, Nahl (2005) introduced affective load theory (ALT) as a way to "[identify] underlying habits of thinking and feeling while engaging in information behavior" (p. 39). According to Nahl, both positive and negative affective states play a fundamental role in maintaining or interrupting cognitive behaviors. As stated by the author,

> [a]ffective behavior initiates, maintains, and terminates cognitive behavior (Isen, Daubman, & Gorgolione, 1987; Carver & Scheier, 2001). For instance, when searchers lose the motivation to continue a task, they begin thinking about something else. Or, if they unexpectedly find some new information they want, they switch activity midstream. The new affective behavior interrupts and takes over the ongoing activity and continues in a new direction with new cognitive activity" (Nahl, 2005, p. 40).

In terms of evaluation, Nahl's (2005; 2009) framework includes methodological elements to collect, code, and analyze data. For example, it provides a set of measures to operationalize constructs such as affective load (AL), which is expressed as "uncertainty (U) multiplied by felt time pressure (TP). Uncertainty is defined as the combined degrees of irritation, frustration, anxiety, and rage (Nahl, 2004)" (Nahl, 2005, p. 41) and represented by:

$$AL = U\ [irritation + frustration + anxiety + rage] \qquad \text{(Eq. 2.3)}$$

One potential limitation of Nahl's framework is the fact that it relies heavily on subjective components such as self-reports, think-aloud protocol, and human judges in charge of coding and analyzing the data. For instance, when people search for information, there are many affective reactions that people are unable to perceive, hence they are not reported at the moment of describing their experiences. The method, however, is particularly useful to investigate affective processes in information seeking at a higher order, that is to say, at the level of feelings according to the hierarchical structure of affective processes illustrated in Figure 2.1.

In recent years, with the development of new technologies, there has been a new turn in the way affective processes are studied in different domains. Today it is possible to find systems and instruments capable of recognizing facial expressions, measuring electrodermal activity, capturing and tracking eye movements, and capturing EEG signals with portable devices. Research on information search has not been an exception in the use of such technologies; for example, Arapakis, Jose, and Gray (2008) used a facial expression recognition system to investigate emotions in implicit relevance feedback. The authors were motivated by the intrinsic limitations of implicit approaches to assess documents' relevance (Kelly & Belkin, 2001), thus they conducted a user study to evaluate the participation of emotions as a potential source to enhance current methods of implicit relevance feedback.

In order to investigate emotions, Arapakis et al. (2008) used subjective evaluations based on self-reports and objective measures based on facial expressions. The latter were analyzed using the *eMotion* (Sebe et al., 2007) software, which can recognize a group of seven facial expressions linked to basic emotions (i.e. happiness, surprise, anger, disgust, fear, and sadness) and neutrality, with different levels of accuracy. In the experimental evaluation, users worked on search tasks with different levels of difficulty determined by the retrieval collections linked to each task.

The authors used different resources to reduce noise in the data such as the use of high thresholds, in order to consider the expression of an emotion to be reliable. In addition, as part of self-assessments, researchers evaluated to what extent participants masked their emotional expressions. In terms of facial expressions, the authors found that the most frequent one was surprise.

One potential limitation of studies like the one of Arapakis et al.'s (2008) is that basic emotions in human-computer interactions may be less expressive through users' faces than emotion expression in other scenarios such as outdoor and social activities. Facial expression recognition systems like *eMotion* are often trained on a gold-standard corpus that contains labeled facial expressions for each basic emotion. For the particular case of surprise, which was the most prominent facial expression found in this study, Arapakis et al. do not elaborate on the episodes where the facial expression linked to this emotion was detected.

Along the same lines, Lopatovska (2009a; 2009b) took a closer look to investigate relations between expressivity of emotions and individual information seeking. With a study design similar to that of Arapakis et al. (2008), which used two tasks with different topics and difficulty levels, and data collection procedures based on self-assessments, interviews, search logs, and actions logs, the author investigated three major aspects: (1) expressive patterns linked to basic emotions that occur around specific search behaviors such as mouse clicks, mouse scrolling, visiting search result pages (SERPs), and content pages; (2) relationships between users' emotions and search performance expressed by measures such as number of queries, query length, and time; and (3) relationships between emotional expression and subjective measures such as search experience, interest in the task, and topic familiarity.

Like Arapakis et al. (2008), Lopatovska (2009a; 2009b) used eMotion to perform automatic facial expression recognition. In addition, she applied similar procedures in terms of expressive threshold with the aim of reducing noise in the data. As a result, Loptovska found that the most prominent facial expression corresponded to surprise and then neutrality. With respect to the former, this was consistent with Arapakis et al.'s (2008) findings. Not surprisingly, in Loptavoska's works, this expression was also the most frequent when analyzed in terms of the search behaviors stated above. Interestingly, the author gives a negative tone to this expression by indicating that "[t]he high frequency of surprise might represent the reality of the search experience, characterized by the high levels of uncertainty, high levels of surprise about the (non)findings, retrieved results, and reviewed information" (Lopatovska, 2009b, p. 125). The author also pointed out that detections of frequent expressions of surprise might be also due to software (i.e. *eMotion*) inaccuracy.

When looking at self-reports, Lopatovska (2009a) states that positive mood could be related to specific search behaviors; however, she states that neither negative nor positive moods would be an indicator of performance in terms of the quality of the search results. From an expressive point of view, on the other hand, Lopatovska (2009b) found positive correlations between the likelihood to express happiness and the time spent searching. Additionally, the author indicates that an increased frequency of expressed sadness and a decreased frequency of expressed happiness would relate to an anticipated end of the search.

In addition to empirical findings, Loptavoska (2009b) also contributed a conceptual model that summarizes different aspects found in the literature about information studies and the affective dimension (Figure 2.14). This model situates affective processes as participating factors at different stages of information search: first, as preconditions that characterize users' states before they interact with an IR system; second, as a component subject to change during information search; and finally, as a resulting change depending upon the outcomes of this process.

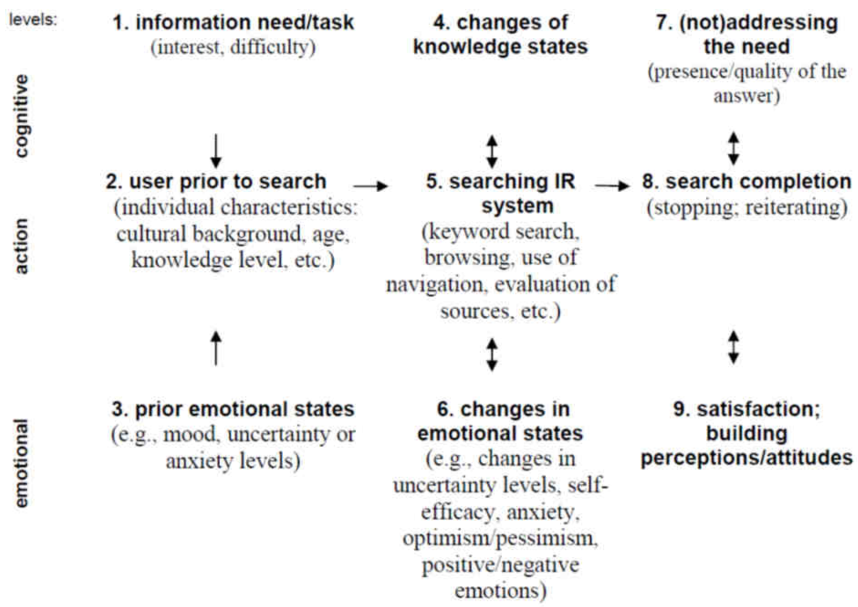


Figure 2.14. Model of emotional information search (Lopatovska, 2009b, p.40).

With a focus on positive psychology, González-Ibáñez, Shah, and Córdova-Rubio (2011) investigated expressed happiness of information searchers in solo and collaborative settings. The evaluation of happiness was conducted using the "smiling analysis method" (p. 5), which consists of a procedure to normalize data, a weighted average measure, and the use of multiple software for facial expression recognition. In particular, the authors used *FaceDetect* (Kueblbeck and Ernst, 2006; Face and Object Detection webpage, 2011), *BMERS* (González-Ibáñez, 2006), and a customized version of *eMotion* (Sebe et al., 2007) that provides results frame by frame[29]. Like Lopatovska (2009a; 2009b) and Arapakis et al. (2005), González-Ibáñez et al. (2011) applied a high threshold to discriminate facial expressions to reduce noise in the data. However, the threshold was slightly reduced from 90% to 80%. As the authors explain, this decision was made because high thresholds are effective to distinguish prominent facial expressions in users' faces (e.g. users laughing) - which are typically uncommon in certain human-computer interaction scenarios such as information search - but ineffective when dealing with more subtle facial expressions.

In addition to employing a different expressivity threshold, the objective of using three different tools for processing face data was to implement the idea of software-based judges (or coders) and then look for agreement among them. In this sense, a procedure similar to inter-rater reliability, used when different human judges perform a coding task, was conducted.

After performing this method to detect smiles in all the study participants, the authors conducted a minute-by-minute analysis. As a result, they found that participants smiled few times during the search session and that smile-related

[29] The original version of eMotion provides analyses based on groups of frames, which hinders synchronization with other data sources. Although Arapakis et al.'s (2008) and Lopatovska's (2009a, 2009b) studies describe some analyses based on the number of frames, it is not clear whether they used the standard version of this software or if they requested changes to obtain outputs frame by frame.

measures were not correlated to search performance, the latter based on the evaluation framework introduced by Shah and González-Ibáñez (2011).

In addition to software-based methods to recognize facial expressions, González-Ibáñez et al. (2011) included self-assessments based on the PANAS instrument before and after the search sessions. The researchers found that smiles may be related to two specific feelings: proud and hurry, both measured through self-reports once search sessions were completed. Another interesting finding derived from self-assessment is that two particular negative affective states (i.e. feeling upset and hostile) before starting the task were positively correlated to two particular performance measures, namely, relevant coverage and unique-relevant coverage. As the authors claim, affective processes as preconditions (like the two indicated above) could serve as indicators of how successful searchers will be in finding relevant information.

González-Ibáñez et al.'s (2011) study constitutes one of the first attempts to investigate findings from positive psychology in information seeking. As Agarwal (2012) indicates, happiness has been understudied as part of information seeking behaviors. In fact, besides the study of González-Ibáñez and colleagues (2011), happiness has only been partially addressed by others such as Gwizdka and Lopatovska (2009) and Lopatovska (2009a; 2009b).

As noted above, González-Ibáñez's et al. (2011) work investigated expressed happiness in individual and collaborative information seeking. Their results in CIS are discussed in the following section, which provides a review of relevant literature about the affective dimension in CIS.

2.3.2 Affective dimension and CIS

As shown in the previous section, studies of the affective dimension in information seeking have been rather limited. Not surprisingly, as presented below, studies of affective processes in CIS are even scarcer. An extensive exploration of scientific databases using key terms such as "collaborative information seeking," "collaborative information behavior," and "collaborative information retrieval," in combination with others such as "emotion," "affect," "feeling," and "mood," showed a lack of research about this research topic.

Even though relevant works about affective processes and individual information seeking have been produced over the years, their findings do not necessarily apply to information-related situations in which collaboration takes place. The participation of the social dimension in CIS makes research about affective processes more complex than in individual settings. Such heightened complexity is due to the fact that affective states in such scenarios may derive from and/or may influence not only information-related processes, but also group dynamics. As noted by Wilson and Wilson (2010),

> "Most commonly, in CIS, we model relationships between collaborating searchers in terms of roles. We expect there to be searchers who differ in terms of search expertise and knowledge [13]. Further, we can consider them to be taking different tactical approaches to divide up the tasks [4]. While these role-focused models account for behavioural changes, they have so far not modeled affective changes. How do people feel being watched, when they have different roles or abilities? How do people perform if they feel anxious or judged by collaborators?" (p. 1)

While some research in CIS acknowledges the participation of affective processes as part of collaborative practices carried out by team members when searching, collecting, evaluating/assessing, and using information, the majority of such works refer tangentially to the affective dimension; that is to say, the authors suggest possible links between affective states and information-related practices of team members but do not study nor develop these ideas. As discussed in the previous section, this situation is rather common in the general field of information science, where very few scholars have conducted research that assigns the affective dimension a central role.

A formal study partially addressing the affective dimension in collaborative information behavior was carried out by Hyldegård (2006a; 2006b; 2006c; 2009). In this study, the author investigated group members' information behaviors in an academic context. In particular, two groups of students in information science were studied while working on their respective term projects. During this period, data collection was based on the use of a questionnaire, interviews, and diaries.

In Hyldegård (2006a), the author discussed differences between individual information behaviors and collaborative ones in relation to information search. The author used Kuhlthau's (1991) ISP as a comparative reference. According to Hyldegård (2006a), students in each group exhibited cognitive experiences similar to those of individuals as depicted by Kuhlthau's (1991) ISP. However, when looking at the affective dimension, the author found differences especially toward the end of the search processes investigated. In particular, it was found that group members did not show relief or satisfaction at the final stage, instead, some of them were frustrated and disappointed. These negative feelings were associated with group member's incompatibilities in terms of their motivations and expectations. In regard to the causes of such differences, Hyldegård (2006a) stated,

> [a]lthough the individual group-members' focus formulation became clearer during the process, also resulting in positive feelings as the ISP model prescribes, it became evident in the last interview that *intragroup aspects* demonstrated as a divergence in motivations and ambitions between group members had contributed to the negative feelings such as frustration and disappointment. When "uncertainty" was experienced at the end of the project it seemed to be associated with the quality of the product (the assignment). Hence, also the *nature* of the work task turned out to have an impact on the feelings experienced by the group members. For instance, lack of time was often mentioned as an explanation for noting "stress" as an "other" feeling in the diary. Accordingly, a noted "relief" in the diary at the end of the project was often explained in the last interview by statements such as "end of stress" and "end of pressure" and not solely by cognitive factors in response to gained clarity and focus formulation, as the ISP model states (p. 293)

Findings from this study suggest the existence of additional factors contributing to affective variations in group members performing a search task. While aspects around information search still take active part in the affective experiences of people, it seems to be that causes beyond cognition (e.g. clarity provokes a feeling of relief) take over in influencing group members' affective states. For example, social interactions during the collaboration process, conflict resolution, decision making by group members, personal motivations as well as individual affective states, are some of the possible factors in group scenarios that could influence the affective experience of people while collaborating with others on an information search task. Hyldegård (2006c) expressed this phenomenon in her group member in context (GMIC) model (see Appendix C.2) by characterizing groups as “forming/storming” and “norming/performing.” The former “implying conflicts and difficulties in finding a group identity” (p. 346), and the latter referring to groups that “generally formulated a focus and experienced positive feelings of confidence and low levels of uncertainty and frustration” (p. 346).

Hyldegård‘s (2006c) findings are consistent to what Allen (1997) expressed in his ‘person-in-situation’ model. According to Allen (1997), there are four perspectives of information needs that should be considered in research, namely, cognitive, social, socio-cognitive, and organizational, each one linked to influential factors in both individual and group practices around information behaviors. The interaction between these factors was expressed by Hyldegård (2006c) in a modified version of Allen’s (1997) model as illustrated in Figure 2.15.

As pointed out by Allen (1997), this integrated view emphasizes the importance of considering the situations that originate information needs, but at the same time he noted the challenges for new research addressing such a larger context around information behaviors since it “will require more complex research designs, and more sophisticated data analyses, than those studies that simply focus on situational or on individual variables” (p. 121).

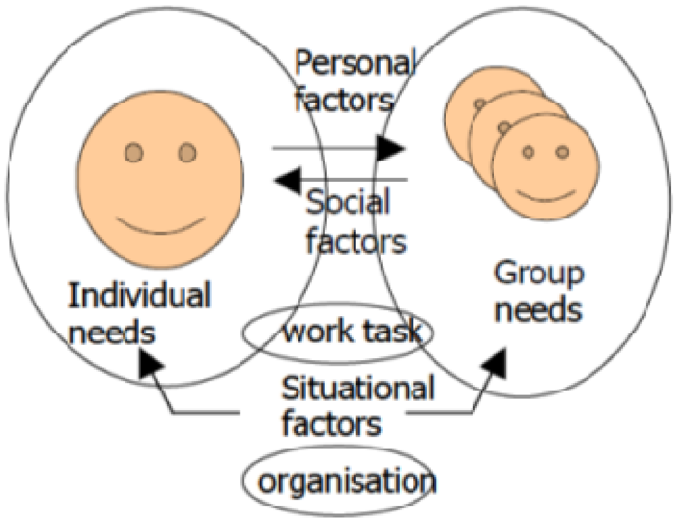


Figure 2.15. The 'person-in-situation' behavior model. (modified after Allen, 1997, p. 112) (Hyldegård, 2006c, p.19)

Hyldegård's (2006c) findings have also been investigated in the lab. In particular, Shah and González-Ibáñez (2010) conducted an exploratory user study with pairs of users performing exploratory search tasks. The authors attempted to describe group members' search behaviors using Kuhlthau's (1991) ISP, which, as indicated previously, was designed for individuals. Using communication logs and a coding scheme, the authors identified coordination messages, strategy definition, expressed feelings, and perceived relevance during the collaboration process. With this information, the authors found the presence of the six stages in the ISP (i.e. *initiation*, *selection*, *exploration*, *formulation*, *collection*, and *presentation*), however, they noted the existence of overlaps between particular stages such as *formulation* and *collection* due to participants that often switched between these two stages while collaborating.

With regard to the affective dimension, the authors found mixed feelings (i.e. positive and negative) during the initiation and selection stages, and also during the transitions between stages. In the remaining stages, however, positive feelings were found to be predominant. Another interesting aspect involving affective processes was identified in the selection of relevant information. According to Shah and González-Ibáñez (2010), "[i]t was [...] observed that the selection of relevant information was first done by an individual and then subjected to the group's judgment and reflection" (p. 7). This phenomenon was later referred to in González-Ibáñez and Shah (2010) as group's affective relevance (GAR), where the authors claim that relevance judgments are socially constructed through both objective and affective discourse, meaning that team members share their opinions (e.g. "This page contains useful information"), reactions (e.g. "I loved this page"), and objective comments (e.g. "This information came from the president of the company") with respect to the information they find and share.

In a preliminary evaluation of GAR, the authors "found that the closer the distance between the number of positive, negative, and neutral information judgments, the higher the performance of teams as measured by precision. This indicates an interesting correlation between expressed emotions and performance of a group" (González-Ibáñez & Shah, 2010, p. 318). Furthermore, it was argued that social interactions carried out when selecting relevant material may dynamically shape feelings, engagement, and the confidence of team members in their actions within the group.

In order to improve the emotional awareness (García, Favela, & Machorro, 1999) of group members in CIS, the authors proposed a model to implement this idea in CIS systems, which is depicted in Figure 2.16.

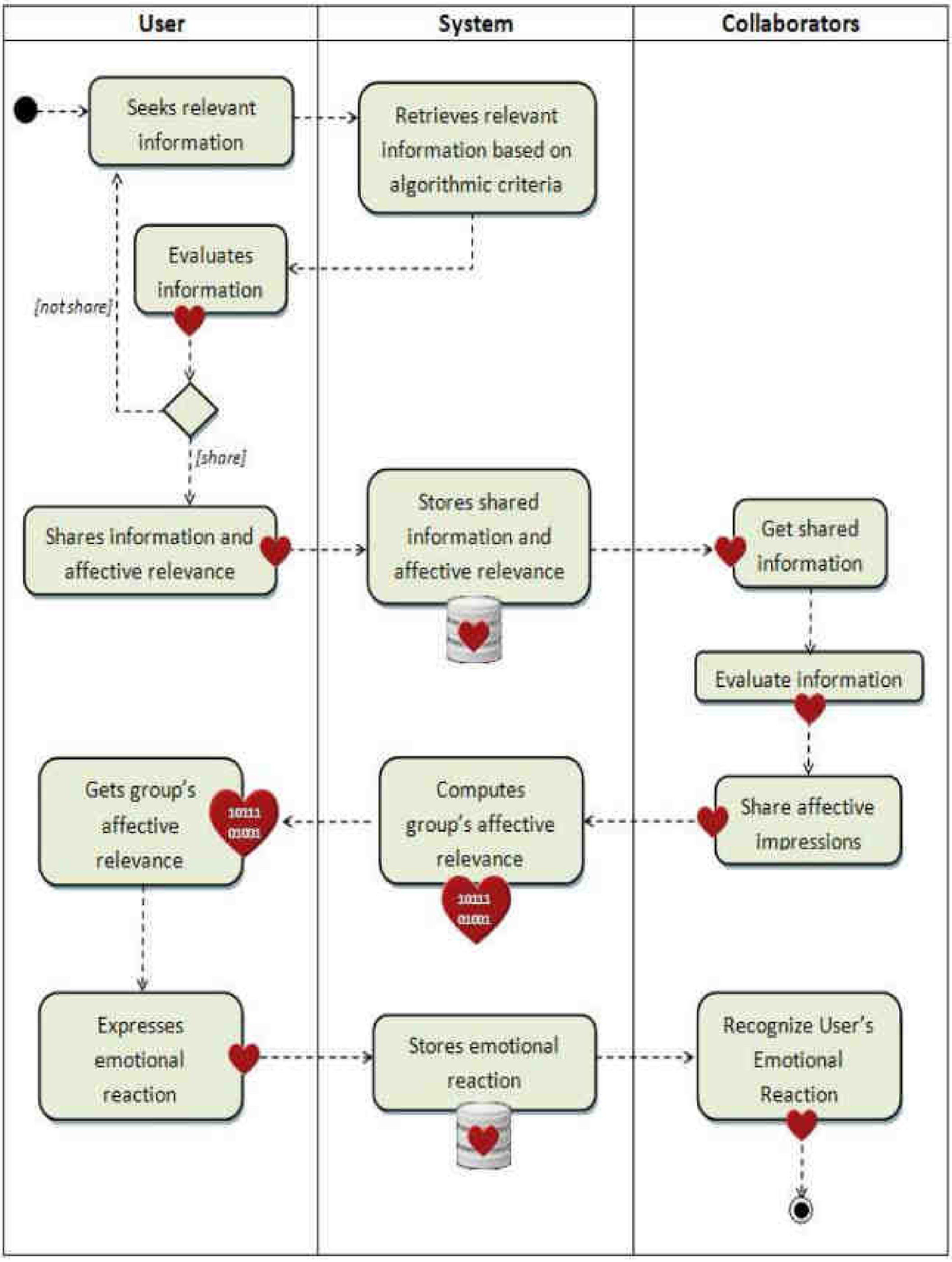


Figure 2.16. Model for implementing group's affective relevance (GAR) in CIS systems (González-Ibáñez & Shah, 2010, p.318).

Shah and González-Ibáñez (2010) concluded that the ISP is a good starting point to investigate CIS. Moreover, they also indicated that

> [i]n addition to Hyldegard's finding that ISP model lacks social element in a collaborative setting, [...] various ISP stages in CIS setting also need to be considered in the light of affective dimension for the collaborators, as well as group's affective relevance (p. 8).

It is acknowledged in the literature that in collaborative contexts, group members not only share with their peers what they think, but also how they feel. As part of social interactions, the communication of feelings plays an important role in how collaboration is carried out. Even in the absence of rich communication channels such as face-to-face interactions (Daft & Lengel, 1984), group members adapt and find ways to express themselves (Jaffe, Lee, Huang, Oshagan, 1995; Provine, Spencer, & Mandell, 2007; Derks, Bos, & Grumbkow, 2007; 2008). For example, in a study of CIB in a math discourse community, Zhou and Stahl (2007) found that

> [e]moticons and abbreviations are frequently seen being used in the chat as ways to convey emotions and tones to the text posting [...] Providing such contextual information is important not only because it builds the sense of presence but also plays significant roles in facilitating the collaboration process by bringing the group up to date of one's status in working on the problem (p. 8).

As Neal and McKenzie (2011) pointed out, the communication of affective states by individuals would play a key role in how others perceive and assess the information found in online communities. The authors used the term affective authority to refer to “the extent to which users think the information is subjectively appropriate, empathetic, emotionally supportive, and/or aesthetically pleasing” (p. 131). The authors found that affective authority and cognitive authority (Wilson, 1983) differed in the claims expressed in textual data. For example, sources associated with affective authority “provid[ed] more personal and therefore more emotionally supportive information” (Neal & McKenzie, 2011, p. 131), this by showing similar experiences, encouragement, and a positive writing style.

In a more specialized domain of collaborative search, Smeaton, Lee, Foley, and McGivney (2007) investigated pairs of users working co-located on a video search task using a tabletop system. One of the topics addressed by the authors was the study of how matching personalities affect the interactions within teams and their performance. The authors found that personality matching within a team did not guarantee efficient work and high performance in completing the video search task. Conversely, “[t]wo people who are both improvisers (perceiving) and both empathic (feeling) but whose personalities are both extraverted and idealistic may not be best matches for a long relationship, but might work well together in a searching task” (p. 382)

Finally, as indicated in the previous section, González-Ibáñez, Shah, and Córdova-Rubio (2011) investigated expressed happiness (i.e. smiles) in both individual and collaborative information seeking. Contrary to findings in individual settings, the authors found that group members smiled significantly more than individuals while working in the search task. A process-based analysis revealed that smiles of group members were positively correlated with communication episodes and at the same time, these correlations were mixed when compared to the smiles derived from the interaction with information. Additionally, regardless of whether participants in each team were placed in different rooms so that they could not see nor hear each other, text-based communication proved effective in allowing team members to express their affective experiences, add affective emphasis or tone to their messages, and provoke affective responses in their peers.

2.4 Summary

This chapter reviewed relevant literature about two major research topics, the affective dimension and collaborative information seeking (CIS). For each topic, definitions that will structure the discussions in the following chapters were presented. Furthermore, research approaches were discussed. In terms of the affective dimension, different techniques to elicit and evaluate affective processes were described. The elicitation of affective states included techniques based on the use of films, pictures, music, and game feedback, to name a few. Evaluation approaches, on the other hand, were based on three major groups, specifically, self-reports, observation methods, and neurophysiological methods.

With respect to CIS, different models were discussed. Evaluation approaches were then described and categorized into three major groups, product-based, process-based, and system-based evaluations. Finally, three systems that support CIS offering common and unique features were described.

The last part of this chapter focused on exploring the literature that connects affective dimension with individual and collaborative information seeking. It is concluded that while the affective dimension is acknowledged as taking an active part in information seeking (both individual and collaborative), research in this domain is very scarce and limited in terms of the methodological approaches undertaken. Additionally, no research has been designed and conducted to investigate the role of positive and negative affective processes in individual and collaborative information seeking. Therefore, considering the specialized background of other fields that study the affective dimension, it is suggested that psychological approaches should be implemented in experimental designs focused on the study positive and negative affective processes and their role in both individual and collaborative information seeking.

The next chapter introduces the research framework used in this dissertation, which includes supporting theories to address the research questions presented in section 1.2 and the evaluation framework used in the experimental evaluation.

Chapter 3. Research Framework

In the previous chapter a comprehensive review of the major topics in this research (i.e. affective dimension and CIS) and their connections were presented. It was concluded that no research has been specifically designed nor carried out to explore positive and negative affective states in CIS. Despite the lack of research about this topic, existing theories in different fields, as well as findings from closely related studies and preliminary studies provide elements to establish a research framework for designing and conducting empirical investigation on this matter.

This chapter introduces the research framework used in this dissertation, which consist of four major components. First, a theoretical framework that comprises a group of theories, models, and empirical results from different disciplines that establish the basis of how the research questions are approached. Second, an evaluation framework, which entails a methodological approach to design and conduct experiments as well as recommendations for data collection and analyses. Third, preliminary studies comprising empirical evaluations conducted during the preparation of this dissertation with the aim of evaluating additional factors in CIS such as communication, system usability, and affective stimuli. Fourth, a set of hypotheses derived from the theoretical framework, research questions, and preliminary studies, which express different relations between affective processes and CIS. Each of the components of the research framework is presented individually in the following sections.

3.1 Theoretical framework

Early in this dissertation, a group of four research questions were introduced. Motivated by the desire to investigate possible reciprocal relations between affective states and the information search process of teams, these questions were formulated to address different perspectives of the problem, in particular: causality,reciprocity, socio-affective factors, performance, and interaction between affective processes.

With the exception of specific models like the ISP (Kuhlthau, 1991), the socio-biological information technology model of information behavior (Nahl, 2009), and affective load theory (Nahl, 2005), information science lacks theoretical foundations about affective processes as part of information behaviors. Fortunately, other disciplines such as psychology, communication, and neuroscience provide a generous source of theoretical frameworks that can help in addressing the inquiries of this work and also in the interpretation of findings derived from empirical evaluations.

As illustrated in $S_{John[-]\&April[+]}$, one of the major challenges of studying the affective dimension in CIS is that affective states in collaborative scenarios can be derived from different sources. If external sources of affective influence such as room temperature, noise, or system performance, are properly controlled, then the sources of affective variability in CIS can be narrowed to: (1) The collaboration process itself, especially as a result of social interactions that are not necessarily related to the task being solved by the group (extrinsic factors); and (2) information-related processes (intrinsic factors) such as information search, information assessment, and sense-making, performed either individually or collaboratively.

In order to explain the affective variability in CIS generated from these two major sources, this section introduces a theoretical framework composed by theories, models, and empirical findings from three different disciplines: information science, psychology, and communication.

3.1.1 Information Science

In the previous chapter a review of different studies about affective processes in information seeking was presented. In particular, Kuhlthau's (1991) ISP, the socio-biological information technology model of information behavior (Nahl, 2009), and affective load theory (ALT) (Nahl, 2005) were discussed as theoretical products derived from empirical research in information science. However, as pointed out, these models were derived mainly from individual information behaviors, thus limitations may exist when explaining collaborative information behaviors.

Regardless of the potential limitations of using these models in CIS, it is recognized that collaborative practices involve individual components. In this sense, the ISP model, socio-biological information technology model of information behavior, and ALT could all help to explain to some degree the relationship between affective processes and information seeking.

On one hand, the ISP structures information search into six different stages, each one linked to cognitive and affective processes, as well as actions performed by searchers. With a focus on the affective dimension, this organized view depicts what affective processes are expected to appear during information search. For instance, in a given search situation it would be anticipated a progression from negative to positive affective states with variations in the middle (e.g. uncertainty$^{(-)}$ -> optimism$^{(+)}$ -> confusion$^{(-)}$ -> clarity$^{(+)}$). Moreover, affective states in each stage are linked to particular cognitive experiences such as an increase in interest and a gain in focus.

Although the ISP consists of multiple stages, the model offers a macro view without explaining how particular affective changes relate to cognitive judgment and vice versa. According to ALT (Nahl, 2005), affective process would modify the way people deal with a given information need. During information search, affective states determine when cognitive behavior starts and finishes. As illustrated in the socio-biological information technology model of information behavior (Figure 2.13), cognitive processes linked to the reception of information are ultimately managed by affective ones. On the other hand, when it comes to the use of received information, affective processes trigger the cognitive ones.

Why and how affective processes influence cognitive ones is not addressed in detail by Nahl (2009). To answer this question, the following section looks at specific theories in psychology that offer possible explanations about the relationship between traits[30], affective, cognitive, and social processes.

3.1.2 Psychology

[30] *Trait affect* “is defined as a tendency to respond to specific classes of stimuli in a predetermined, affect-based manner. Therefore, an affective trait is considered a relatively stable characteristic of personality” (Sonnentag & Sparr, 2007)

Psychologists have investigated the affective dimension for over a century. During this period, different theories and models explaining the role of emotions as well as other affective processes in human reasoning have been developed. Different authors have argued that affective states play a major role in decision making (Forgas, 1995; Yuen & Lee, 2003), a claim that has been also stated by neuroscientists such as Damasio (1994).

According to this claim, affective processes may influence how people make decisions when facing multiple options (e.g. risk-taking decisions when gambling) (Damasio, 1994). It has also been stated that prior affective states can condition rational thinking. While some authors have supported these claims from a valence-based approach (i.e. positive and negative), others have indicated that particular affective states, from a discrete approach, could have different effects on judgment (Lerner & Keltner, 2000).

In a study with affective states elicitation through movie clips, researchers showed that participants in an induced negative mood made less risky decisions than those in a neutral or positive mood when engaged in a task involving a set of realistic dilemma questions (Yuen & Lee, 2003). From a discrete approach, Lerner and Keltner (2000) investigated how specific affective states (i.e. sadness and anger) as well as a general negative valence relate to the perception of risk. In this study, the authors found that both sadness and anger were more effective in predicting the judgments to estimate risk perception than a valence-based approach.

There are multiple theories that have attempted to explain how affect influences rational thinking. The following section provides an overview of one of the referential theories of this dissertation, the affect infusion model.

3.1.2.1 Affect infusion model (AIM)

One model in social psychology that accounts for the influence of affective processes in cognition is the Affect Infusion Model (AIM) (Forgas, 1995; Forgas & George, 2001). The AIM relates moods (which, as described in section 2.1.1.2, are longer in duration than emotions), decision making, and information processing. According to Forgas and George (2001), affect infusion refers to "the process whereby affectively loaded information exerts an influence on, and becomes incorporated into, a person's cognitive and behavioral processes, entering into their constructive deliberations and eventually coloring the outcome in a mood-congruent direction" (p. 9). This model is based on the notion of affect-priming, which indicates that "affect may influence the availability of memories, constructs, and associations that people implicitly rely on to produce a reaction […] [a]ffect may thus prime thoughts and cognitive constructs as it selectively activates memory structures to which is connected" (Forgas & George, 2001, p. 8)

The influence of mood depends on information processing strategies, which may be selected due to additional factors such as familiarity with the situation, target complexity, specific motivations, cognitive capacity, and situational pragmatics (Forgas, 2009). As pointed out by Forgas and George (2001), "[s]ometimes, moods have no influence on the content of cognition and action and may even have an inconsistent, mood-incongruent[31] influence (Erber & Erber, 2001; Forgas, 1991; Parrot & Sabini, 1990; Sedikides, 1994)" (Forgas & George, 2001, p. 9). In an earlier work, Forgas (1995) described four processing strategies: direct access, motivated processing, heuristic processing, and substantive processing. Below, explanations and examples of these processing strategies as well as possible relations with information seeking are provided.

- ***Direct access strategy***

This strategy refers to situations in which individuals can act by using pre-existing information or pre-existing judgments (Ciarrochi & Forgas, 2000). Consider the example of a person that is asked about his home address and phone number. In such a scenario (assuming that there are no background troubles in providing such information), the access to information is direct, with low personal involvement, and elaboration, hence, not subject to affect infusion (Ciarrochi & Forgas, 2000; Forgas, 2009). In an information search scenario, this strategy could be related to situations in which people have to provide information that is already part of their knowledge or that is accessible through external repositories. For instance, returning to John's example ($S_{John[-]}$), there are different aspects of the strategic plan John is working on that he can manage with existing and objective information (e.g. number of employees, salaries), even if he has to retrieve such information from external resources such as

---

[31] Mood-congruency indicates that actions, decision making, judgments, and expressions are all consistent with one's mood. For example, novice Target customer representatives in a positive mood are more helpful with customers (Forgas, Dunn, & Granland, 2008). Mood-incongruence indicates that behavior and cognitive processing occurs in an opposite direction to mood. For instance, a person in a negative mood sees the outcomes of a risky decision making process as more favorable than a person in a positive mood.

information repositories of the company. In this scenario, John's information processing strategy is direct, and thus it is not infused by his negative affective state as a result of a bad weekend at home.

- ***Motivated processing strategy***

A motivated processing strategy occurs in situations with pre-existing motivations and goals. "In these circumstances, people are likely to engage in highly selective, guided, and targeted information search strategies to support a motivational objective" (Forgas, 2009, p. 104). An example of this processing strategy is described by Forgas (1991), in which individuals motivated to recover from sadness were not infused by their particular affective state when facing this task (Ciarrochi & Forgas, 2000), that is to say, they showed mood-incongruent reaction. In the example $S_{John[-]\&April[+]}$, John could be self-motivated to do his best with the project and not let his negative affective state interferes with his collaboration with April. As a result, John may make an effort to represent himself as a helpful, polite, and optimistic partner.

As pointed out by Forgas (2009), "[b]ecause many experimental studies employ artificial and uninvolving tasks, motivated processing is [...] rarely demonstrated in the literature" (p. 104).

- ***Heuristic processing strategy***

Heuristic processing strategy takes place when decision making is the result of a simplified operation to determine the course of action in the absence of motivation and pre-existing evaluations. As Forgas (1991) describes,

> [h]euristic processing is most likely when the target is simpler or highly typical, the personal relevance of the judgment is low, there are no specific motivational objectives, the judge has limited cognitive capacity, and the situation does not demand accuracy or detailed consideration (p. 47).

For example, as part of the strategic plan, John and April ($S_{John[-]\&April[+]}$) have to provide an evaluation of the facilities of the company. Both John and April are not aware of the different units, so they quickly look around in order to provide their respective evaluations with a minimal effort. On one hand, April provides an overall positive evaluation infused by her positive mood. Conversely, John's ($S_{John[-]}$) negative affective state may inadvertently participate in his judgments of the same places that April rated positively, resulting in an overall negative evaluation of the facilities.

- ***Substantive processing strategy***

This strategy is used in situations that require individuals to "select, learn, and interpret novel information about a target and to relate this information to preexisting knowledge structures" (Forgas, 2009, p. 105). As the authors indicate, this kind of strategy requires more effort and it is likely to be used in complex situations and where no motivations exist. The author adds that substantive processing takes place only when other less demanding approaches "prove inadequate to the judgmental task" (p. 106). Yet, judgments resulting from information processing are likely to be biased by affective states.

For instance, after comparing their individual judgments and their evaluation approach of the company's facilities, John and April realize that their strategy was not effective. They then proceed to perform a more exhaustive evaluation, which consists of collecting information about the facilities available in the company's databases, visiting different units, and asking to some employees for their impressions of their workplaces. Moreover, they may link information collected with their pre-existing knowledge in the matter, which is derived from personal experiences in the company. Regardless of this adjusted evaluation protocol, the individual evaluations of John and April may still be influenced by their respective affective states. For instance, while reviewing blueprints of the company, April (who is in a positive affective state) may be amazed by the spaciousness of the different units in the company as well as the availability of multiple resources such as a gym, a kitchen, and a parking space for almost every employee. On the other hand, John (who is in a negative affective state) may react negatively to the same blueprints by indicating that the gym is not relevant considering the demanding nature of the company's work, the kitchen is just an additional resource of the company to maximize the time that employees spend at work, the space allocated for each unit is not properly used, and the amount of parking spaces are not sufficient for all employees.

- ***Affect infusion in information seeking***

In the context of information seeking, affective states could govern the way information is processed and in turn determine its perceived relevance with respect to the task being performed. In simple precision-oriented tasks with easy access to information - either as part of prior knowledge or through exact keywords issued in an IR system (e.g. find the capital of Morocco) - affect infusion should not take place. Therefore information processing strategy in these kinds of scenarios could be attributed to direct access.

When the complexity of a precision-oriented task increases, for example the required information is not part of the searcher's pre-existing knowledge nor it can be retrieved using simple keywords in a search engine (e.g. "What is the name of the salmon subspecies that never goes to sea before they get stuck inland?"), then other strategies must be implemented in order to obtain the information. In a motivated situation, no matter what the searchers' prior affective states are, they will attempt to exhaust the possibilities when looking for relevant information. Searchers might also judge the relevancy of pages in a mood-incongruent manner. For example, even though a person is in a negative mood, he/she may make a positive assessment of a page if the source is reliable. Conversely, even if a person is in a positive mood, he/she still may judge pages containing related information to what is being sought negatively if the sources are not trustworthy.

In the absence of motivations and pre-existing knowledge when dealing with a complex precision-oriented task, information evaluation can be performed following either heuristic or substantial strategies. Under a heuristic processing strategy, searchers in a negative affective state may explore some sources and judge the information they find as irrelevant for the task being solved. After some attempts they may terminate their search processes early and feel disappointed about their performance. On the other hand, a searcher in a positive affective state may consult potentially unreliable sources, such as question-answering systems, and judge some of the answers he/she finds positively regardless of their reliability.

If the use of substantial strategies is applied, then information search is more meticulous. Searchers proceed with an exhaustive search process, which consists of visiting, evaluating, and contrasting the content of multiple pages as part of a constructive process. Searchers in a negative affective state can be more critical when evaluating the sources they visit, whereas those in a positive affective state will be more relaxed with their judgments. While negative affective states may have positive implications toward the identification of reliable information, there is a trade-off in terms of the cognitive effort and time required to reach a satisfaction point.

When the task is recall-oriented, affect infusion may have different implications. Exploratory search, for example, involves a high level of uncertainty, which means that information to accomplish a particular goal is neither available through prior knowledge nor immediately through IR systems. In this kind of scenario, a direct access strategy is not possible. Accomplishing goals in exploratory search tasks implies a constructive process in which information is searched, evaluated, selected, interpreted, consolidated, and used.

Other information processing strategies (i.e. motivated, heuristic, and substantial) may be used. For example, in a health-related situation where patients look for information about cures for their diseases (e.g. cancer), a motivated processing strategy could have positive or negative implications. If the searcher is in a negative affective state, but motivated to get better, he/she may attempt different searches and judge information as relevant in spite of its source, but with the hope that such information constitutes a possible cure or a promising treatment.

Without motivated goals, on the other hand, high-school students working on a biology assignment about cancer could use a heuristic processing strategy. Students in a positive affective state may be more optimistic about finding treatments and cures for this disease. Rather than exploring multiple sources, these students may use question-answering systems as a shortcut to find information that supports their optimistic view. On the other hand, students in a negative affective state may be more pessimistic, and after visiting some pages, terminate their search process with a negative conclusion about treatments and cures for cancer.

Finally, a substantial processing strategy could be used by other outsiders to the cancer community. For example, a journalism student preparing an essay about treatments and future approaches to cure cancer could proceed with a comprehensive review of available resources in this matter. Students in positive affective states, with positive background experiences of relatives that successfully overcame cancer, may be more optimistic about possible cancer cures based on the information they collect. On the other hand, students in negative affective states, with background experience of relatives that passed away as a result of cancer, may be influenced negatively when assessing information related to cancer treatments.

The above examples have been expressed under a valence-based approach, that is to say, only references to positive and negative affective states as broad categories have been made. However, as pointed out by Lerner and Keltner

(2000), particular affective processes under a discrete approach could have different effects as part of affect infusion.

Additionally, the previous examples have been developed around individual information behaviors, nonetheless, Forgas and George (2001) argued that affect infusion is especially relevant in social contexts such as group work. For instance, social interactions in collaborative settings such as communication and coordination could be influenced by prior affective states. This could be indicated by the way group members express themselves, the way they evaluate each other's findings and performance, and how engaged they are in the collaborative process, to name a few.

In the example of John and April ($S_{John[-]\&April[+]}$), removing motivated goals from both parts, the interactions from John toward April could be influenced by his negative affective state. As a result, April's affective state could be contaminated by John's attitude or it may be the case that April's positive attitude helps John to recover from his negative one.

If both John and April start working in a congruent affective state, either positive or negative, common reactions and perceptions would be shared within the team. In this matter, it has been found that "consistent homogenous mood states within groups may be meaningfully viewed as autonomous group-level constructs with subsequent implications for group members' behavior" (Forgas & George, 2001, p. 16). The authors refer to this collective phenomenon as "group affective tone." According to the authors past research has shown that "enduring positive group affective tone is also likely to function as a useful resource, allowing the group to cope better with adverse information and aversive situations (Trope et al., 2001)" (p. 17). In applied situations it was found that "negative group affective tone of the groups decreased the extent to which group members were helpful toward customers" (p. 17), whereas "positive affective tone was related to reduced absenteeism in these groups" (p. 17).

The next section will further discuss about the influence of positive and negative affective states from a positive psychology perspective.

3.1.2.2 Positive psychology

Classical psychology has been concerned for decades with the study of people's affective ailments. Issues such as depression, anxiety, or specific disorders, have been widely investigated. Among the products derived from these studies, there are models, theories, and practical applications such as therapies.

As pointed out by Fredrickson (2009), the focus of classical psychology is on negativity, in other words the understanding of what is wrong, its consequences, and methods for addressing the problems. Positive psychology, on the other hand, focuses on what is good, which encompasses factors such as affective processes, attitudes, and social relations that have a positive connotation (Fredrickson & Levenson, 1998; Fredrickson, Tugade, Waugh, & Larkin, 2003, Seligman, Steen, Park, & Peterson, 2005; Gable & Haidt, 2005; Waugh & Fredrickson, 2006, Fredrickson, 2009).

As a relatively new branch in psychology, positive psychology has brought the formulation of new research questions, hypotheses, models, and theories. For example, the broaden-and-build theory proposed by Fredrickson (2004) focuses on a particular group of positive emotions (e.g. gratitude, joy, interest, and hope, among others), their characterizations, and roles. According to the author, "positive emotions appear to *broaden* peoples' momentary thought–action repertoires and *build* their enduring personal resources" (p. 1369), for example, "joy sparks the urge to play, interest sparks the urge to explore, contentment sparks the urge to savour and integrate, and love sparks a recurring cycle of each of these urges within safe, close relationships" (p. 1367).

Another implication of positivity is the ability to help individuals recover from the effects caused by negative emotions (Fredrickson, Mancuso, Branigan, & Tugade, 2000). This is particularly relevant in information-related situations in which searchers may experience negative affective states such as anxiety or frustration at different stages of their search processes. As expressed in RQ1[32] and RQ2[33], one problem to investigate when collaboration is added to the equation is how negative affective processes during information search are regulated by interactions with positive ones derived from information encounters as well as social interactions among team members.

While positive psychology focuses on positivity, it also considers negativity as a necessary component. Empirical studies in this domain have been able to evaluate and demonstrate the existence of specific proportions between positive and negative affective states. In particular, the ratio 3-to-1 (three positive over one negative) has been linked to success and flourishing in various contexts such

[32] **RQ1:** How do initial affective states and their interactions shape the way team members collaborate when searching information?

[33] **RQ2:** What affective processes are typically experienced and expressed (physically, physiologically, and verbally) by team members when collaborating in an information search task?

as health, marriage, learning, leadership, business, and team work (Fredrickson & Losada, 2005). For example, Losada and Heaphy (2004) and Fredrickson and Losada (2005) investigated the effects of positive emotions in collaborative contexts. The authors argued that the ratio between positive and negative emotions defines a specific zone (Losada's zone = [2.9013, 11.6346]) that characterizes high performing teams. Therefore, one problem to investigate is to what extent Losada's zone and in particular the 3-to-1 ratio apply to CIS scenarios. This problem is of particular interest in the development of this dissertation as expressed in RQ3[34] and RQ4[35].

Research findings of the AIM in social contexts have also shown the potential implications of positive affective states. As discussed in Forgas and George (2001), a study in a retail organization of customer service showed positive correlations between positive affective states and group performance in terms of the behaviors of group members. As pointed out by the authors,

> [l]aboratory evidence suggests that shared positive mood also tends to promote greater group confidence and positivity when making group decisions, more cooperative and helpful responses when dealing with others, and more cooperative and integrative bargaining strategies when dealing with another group (Forgas, 1990, 1998a, 1999a) (p. 17).

---

[34] RQ3: To what extent, if any, do initial positive and negative affective states and those derived from the collaboration of individuals in an information search task influence team performance?

[35] RQ4: To what extent, if any, does the relation 3-to-1 between positive and negative affective states (P/N) (Losada & Heaphy, 2004; Fredrickson & Losada, 2005) apply to CIS?

The theories described in the previous sections provide a reasonable theoretical framework to hypothesize about the role of affective states in the information search process of individuals and teams. However, there are complementary processes in CIS that the above theories do not account for. These processes include communication, coordination, and common ground. The following section focuses on communication theories to bridge this gap.

3.1.3 Communication and Computer Mediated Communication (CMC)

Communication, as described early in this document, is a cornerstone in CIS; however, this component has been barely studied in this domain. Communication enables team members to exchange different kinds of information, such as thoughts, judgments, and affective states. Regarding the latter, it has been argued that these are not only a personal phenomenon, but also a social and communicative one (Bartsch & Hübner, 2005). How communication is carried out may depend on specific conditions such as the communication channel and common ground shared by participants.

According to media richness theory (MRT) (Daft & Lengel, 1984), the particular characteristics of the task may determine the communication channel to be used. According to this theory, the more ambiguous the situation, the richer the communication channel required. In regard to the richness of communication channels, this can be determined in part by the types of information that can be communicated through them.

For example, in face-to-face (f2f) communication, people can convey content and meaning through verbal and non-verbal communication. The content of messages, facial expressions, gestures, tone of voices, and physical contact are some of the information resources that can be used to communicate through this channel. When f2f communication is not possible due to spatial and/or temporal constraints, other channels may be used. However, as stated in MRT, particular communication channels may not be appropriate for some tasks.

In addition to f2f interactions, group members could use technological support to reach similar levels of richness. For instance, video conference enables interlocutors to exchange information similar to that in f2f. Yet, restricted view, spatial differences, and lack of physical contact act as filters of some of the communication cues inherent to f2f interactions. In terms of affective expressivity, participants using video conference can still communicate verbally using positive and negative words. They can also see each other's facial expressions. However, while access to expressive gestures is enabled, it may be restricted to a specific visualization area. Additionally, access to acoustic cues such as tone of voice is also available.

Voice communication eliminates access to visual cues, allowing interlocutors to verbally communicate adding affective tone to their messages through the use of positive and negative words as well as prosody[36]. Text-based communication is even more restricted. Here, the expressivity is constrained to verbal communication where judgment of affective tone relies on the use of specific words, punctuation, and paralinguistic cues such as emoticons (Derks et al., 2007; 2008; Provine et al., 2007).

---

[36] "The rhythmic and intonational aspect of language" http://www.merriam-webster.com/dictionary/prosody (Merriam-Webster, 2013).

With the exception of f2f, the above communication alternatives can be used in different temporal (i.e. synchronous and asynchronous) and spatial (i.e. co-located and remotely located) conditions. However, adding these variables introduces additional constraints to communication such as lack of immediate access to information and restricted awareness. In terms of affective processes in asynchronous communication, temporal-contexts may influence affective perception. For instance, if a person (T_x) sends a message with a negative tone at time t_1, the perception of the message by the receiver (R_x) at time t_2, may be different than what R_x wanted to express at time t_1. Asynchronous communication as illustrated in this example does not support timely emotional awareness. Other possible issues associated to asynchrony relate to content regulation (e.g. selective use of words and message editing) and contextual interpretation. Table 3.1 provides a summary of different communication channels and associated information types.

Table 3.1. Relation between communication channels and affective information types. (A) Available, (L) Limited, (N) Not available.

Information types \ Channel	f2f	Video	Audio	Text chat	Text email
Facial expressions	A	A	N	N	N
Gestures	A	L	N	N	N
Physical contact	A	N	N	N	N
Prosody	A	A	A	N	N
Verbal	A	A	A	A	A
Paralinguistic cues[37]	N	N	N	A	A

[37] Some examples of paralinguistic cues include emoticons, punctuation, and exclamation marks.

While the richness of communication channels is important with respect to the supported information types, there are situations in which awareness plays a fundamental role. In CIS, for example, group members may need to access each other's information collections in order to be conscious of their current progress in the task. They may also need to know the current page that they are visiting, which portions of the page they are currently looking at, and what their impressions about that page are. Social presence theory (SPT) (Short, Williams, & Christie, 1976) accounts for the different levels of awareness required to accomplish a task. Different levels of social presence may be related to how interactions with others are handled and how discussions are perceived and carried out (Sonnenwald & Pierce, 2000).

In the literature some authors indicates that the highest level of social presence is offered by f2f communication. Lower levels of social presence can be found in video, audio, and text-based communication (Shaw, 1981). Interestingly awareness can be supported in different ways in spite of the richness of the underlying communication channel. For example, in Coagmento (Shah, 2010b; González-Ibáñez & Shah, 2011) communication is performed via text chat, but additional awareness resources are provided. In this system, group members have access to a common repository with notifications of recent events, notes, queries that have been issued by each group member, the pages they have visited, the ratings they have assigned to each page, bookmarks, and also specific snippets of the pages that they find to be relevant for the task being performed.

Richer communication channels or higher levels of social presence do not guarantee success. In the case of social presence, for example, researchers have observed that a mismatch between social presence and the particular characteristics of the task could lead to negative effects in areas such as communication performance (Mennecke, Valacich, & Wheeler, 2000; Chou & Min, 2009).

Another important aspect to consider is how relationships between collaboration partners influence their confidence, affective expressivity, and group performance. It has been argued in experimental evaluations that partners without common ground are less likely to express themselves naturally, thus affecting communication and performance (González-Ibáñez et al., 2012a; 2012b; 2013).

Introduced by Clark & Brennan (1991), grounding in communication is a construct that refers to the set of beliefs, knowledge, and assumptions that are shared by a group of individuals, which are fundamental in order to communicate and coordinate. Hertzum (2008) indicated that grounding as part of collaborative practices performed around information is essential to enable searchers to collect, share, and use information. This mutual knowledge is updated during communication through a process that the authors call *grounding*, which varies according to the particular medium used to communicate. Gupta, Duff, and Tranel (2011) also provided empirical evidence that supports the hypothesis that socio-emotional signals, which are communicated verbally and not verbally, play a fundamental role in *grounding*. Another important aspect of common ground is the ability of individuals to coordinate what is said and what is meant in terms of the content that is being shared within a given moment (Grice, 1975).

In order to address questions about the appropriateness of communication channels in CIS, an experimental study was conducted to compare multiple communication channels in a CIS task. The study, which constitutes one of the bases of this dissertation, is discussed later in this chapter (see section I.2). The following section focuses on the evaluation framework used in this dissertation.

3.2 Evaluation framework

As described in Shah (2012), evaluation is one of the major challenges in CIS. In the previous chapter, different evaluation approaches were discussed. In particular, product-based, process-based, and system-based evaluations were presented as three major branches in which to perform evaluations around CIS.

This dissertation presents an evaluation framework based on methodological approaches and measures introduced by previous studies. The methodological approach consists of recommendations for experimental designs, data collection procedures, and analysis strategies.

3.2.1 Data collection

González-Ibáñez's (2010) approach is adopted for data collection. As depicted in Figure 3.1, this mixed-method strategy consists of out-process and in-process procedures oriented to collect information about behaviors, affective processes, interactions, and performance within a CIS task.

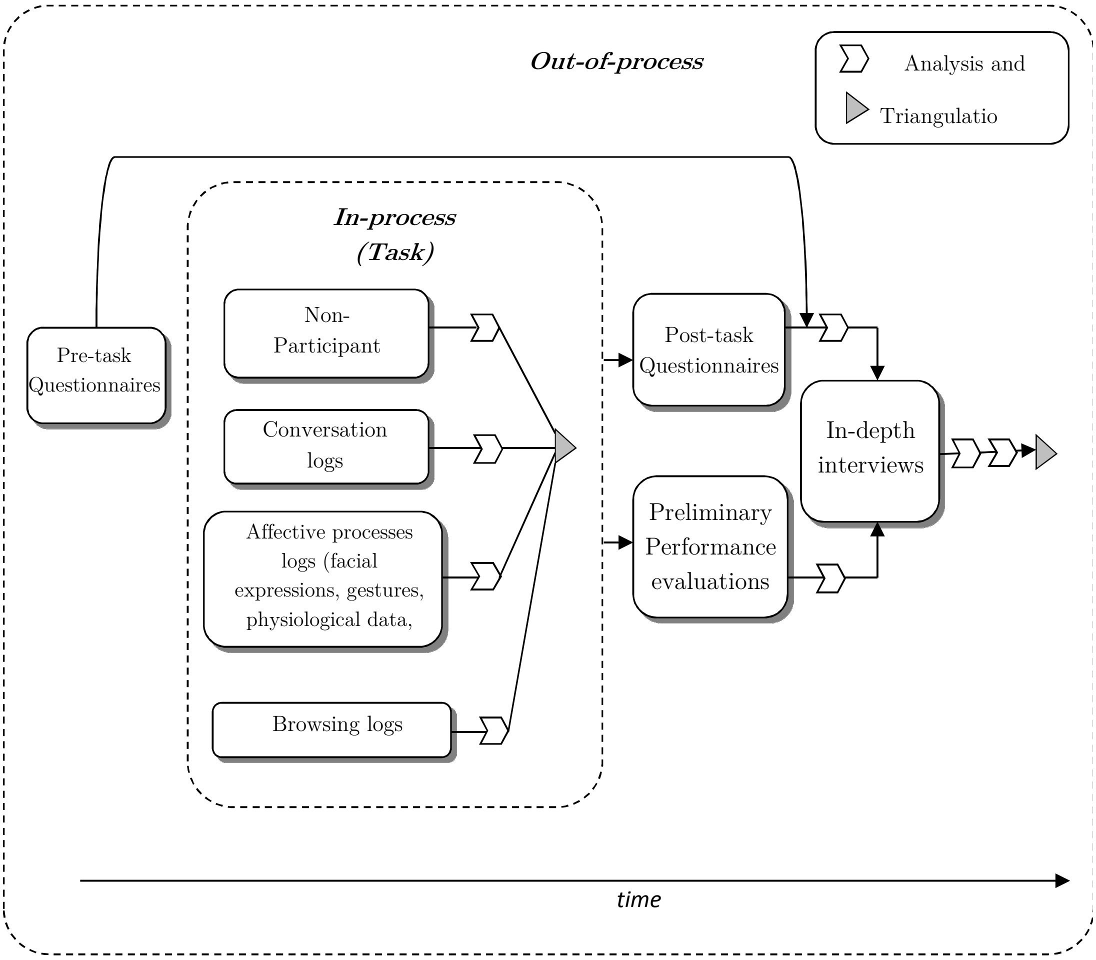


Figure 3.1. A mixed-methods approach to collect data about affective processes in CIS. Adapted from González-Ibáñez (2010, p. 3).

Following the diagram from left to right, the study starts with questionnaires (quantitative) in order to collect information about pre-conditions of participants at the moment of starting the experimental sessions. Next, in-process data collection procedures are performed. This includes non-participant observations (qualitative), communication logs (quantitative/qualitative), recording of affective processes related-data (e.g. facial expression, electrodermal activity, and self-reports) (quantitative), behavioral features (e.g. keystrokes, mouse activity, desktop activity, and eye movements) (quantitative/qualitative), and browsing logs (e.g. visited pages, ratings, bookmarks, SERPs, queries, and timestamps, to name a few). After completing in-process data collection, participants are prompted to complete post-task questionnaires. In addition, preliminary performance measures that can be used to frame questions for the following data collection stage may be available at this time. The final step comprises semi-structured in-depth interviews, which can be conducted in tandem with other techniques such as think-aloud protocol (qualitative).

The purpose of combining multiple data collection methods is to gain access to the information search process of both individuals and teams from different points of view. Non-participant observation, for example, helps in the interpretation of quantitative evaluations from the eyes of the researcher. On the other hand, in-depth interviews provide information from the perspective of participants, who can explain how they felt at a given moment and why they felt that way. Having multiple perspectives of information search enables the researcher to validate and consolidate the results derived from each data source.

The granularity of the data collected is linked to the black box metaphor used in section 2.2.2 to describe evaluation approaches in CIS. The deeper the evaluations of information search, the finer the granularity of the data collected. For example, data obtained at higher levels of evaluation such as overall search performance is considered to be coarse grained, whereas data from lower levels of evaluation such as EDA are collected with higher sampling rates and with varying resolution levels, thus considered to be fine grained. This continuum of granularity is depicted in Appendix C.4.

The data collection procedures illustrated in Figure 3.1 provide different data types that must be analyzed with proper methods. However, before conducting analyses, evaluation measures are needed. From a quantitative standpoint, evaluation measures provide a numeric representation of different constructs linked to the data collected. The following section describes the evaluation measures used in this research.

3.2.2 Evaluation measures

As discussed in section 2.2.2, different studies have developed or adapted existing measures to evaluate different aspects of information search in CIS such as performance, affective processes, and social interactions. However, most of the attention has focused on performance. Since this dissertation investigates not only the products of information search but also process, it is necessary to implement measures and instruments to operationalize constructs around behavioral factors, the affective dimension, and cognition.

The research included within this dissertation comprises a group of studies that evaluated different aspects of CIS such as performance, communication processes, information synthesis, and effects of time and space. Aside from findings derived from these studies (which are briefly discussed in the following section), the measures and instruments used will be discussed as they proved adequate in describing different aspects of the information search process of both individuals and teams, and can be used to inform similar studies.

These measures are computed using the data collected in experimental sessions according to the approach depicted in Figure 3.1. The complete set of evaluation measures incorporated in this dissertation is summarized in Figure 3.2, while details of IR measures as well as some communication and user measures are provided below. Descriptions of the remaining measures and instruments are

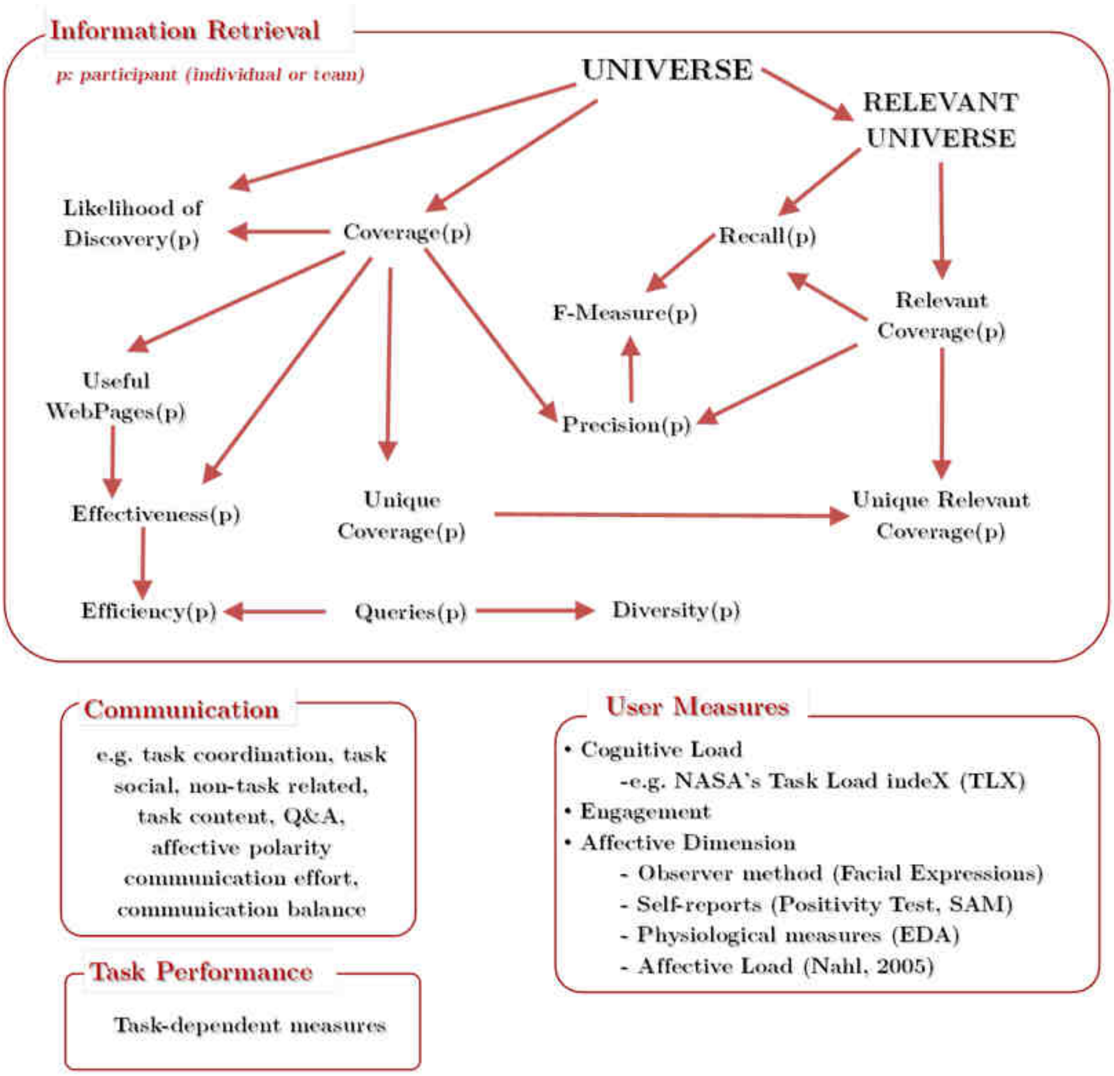


Figure 3.2. Evaluation framework used in this research. Adapted from Gonzalez-Ibáñez, Haseki, and Shah (2011).

presented in Chapter 5 due to their close relationship with the experimental design of this study.

The evaluation framework presented in this section includes sets of measures that can be applied to both individuals and teams (IR measures), only to teams (communication measures), and only to individuals (user measures).

3.2.2.1 IR performance measures

In terms of IR, this dissertation adopts a set of measures introduced in Shah and González-Ibáñez (2011). All these measures assume that the participants perform their searches on the World Wide Web (WWW).

The measures described in Shah and González-Ibáñez (2011) are computed with respect to a fixed universe of webpages, which consists of the "union of all the webpages visited by all [the] participants" (p. 916) in the study. This is expressed in the following expression:

$$U = \bigcup_{p \in P} Coverage(p) \qquad \text{(Eq. 3.1)}$$

Where *Coverage(p)* represents the webpages wp visited by a given participant p, which is expressed by:

$$Coverage(p) = \{wp_i : wp_i \text{ was visited by } p \wedge wp_i \in U\} \qquad \text{(Eq. 3.2)}$$

Within the universe of pages U, there is a subset of relevant pages, which the authors refer to as U_r and express as follow:

$$U_r = \bigcup_{p \in P} RelevantCoverage(p) \qquad \text{(Eq. 3.3)}$$

Here *RelevantCoverage(p)* corresponds to the set of pages within *Coverage(p)* that participants in the study find to be relevant. This perceived relevance can be denoted with the below expression:

$$RelevantCoverage(p) = Coverage(p) \cap U_r \qquad \text{(Eq. 3.4)}$$

Shah and González-Ibáñez (2011) also explored regions of U that were reached by one participant or one team. The authors called this area unique coverage and expressed it as:

$$UniqueCoverage(p) = Coverage(p) \setminus \bigcup_{p' \in (P \setminus \{p\})} coverage(p') \qquad \text{(Eq. 3.5)}$$

Likewise, the unique coverage within U_r corresponds to the set of relevant pages visited by one participants or one team.

$$UniqueRelevantCoverage(p) = UniqueCoverage(p) \cap U_r \qquad \text{(Eq. 3.6)}$$

While unique pages constitute one form of information diversity, Shah and González-Ibáñez (2011) also included two additional measures to estimate diversity of exploration. The first one corresponds to *likelihood of discovery*, or the probability of a participant/team to find webpages that "are difficult to reach and probably beyond the first results page of search engines" (p. 917). The authors describe the procedure to compute this measure as follows:

> In order to operationalize this idea [(*likelihood of discovery*)], we used a formulation similar to that of inverse document frequency (IDF). Using the frequency of each webpage in our log data, we computed its likelihood to be visited; in addition, we multiplied each webpage's likelihood by -1 in order to denote the IDF. As a result, each webpage was assigned with a normalized value between -1 and 0. In this sense, those webpages with a value close to 0 are rare (and even unique) to be reached by teams/participants, while those close to -1 are more likely to be visited (p. 917)

The second measure of diversity described by Shah and González-Ibáñez (2011) corresponds to *query diversity*. This measure expresses diversity from the perspective of participants based on the way they approached their searches. The authors implemented this measure using Levenshtein edit distance to compute query diversity between pairs of users by comparing the queries they issued during the process. The larger the values derived from this procedure, the more different the queries. While the authors applied this procedure within teams, it can also be applied to explore query diversity of each individual.

As part of the set of evaluation measures, the authors also included traditional IR measures, namely, precision, recall, and F-Measure, which were adapted in terms of the above measures. The adapted formulations for these measures are presented below:

$$Precision(p) = \frac{RelevantCoverage(p)}{Coverage(p)} \qquad \text{(Eq. 3.7)}$$

$$Recall(p) = \frac{RelevantCoverage(p)}{U_r} \qquad \text{(Eq. 3.8)}$$

$$F(p) = \frac{2 * Precision(p) * Recall(p)}{Precision(p) + Recall(p)} \qquad \text{(Eq. 3.9)}$$

In addition to measures described in Shah and González-Ibáñez (2011), the evaluation framework of this dissertation also incorporates a pair of measures introduced by González-Ibáñez, Shah, and White (2012).

The first of these measures corresponds to *effectiveness*, which is denoted by:

$$Effectiveness(p) = \frac{UsefulCoverage(p)}{Coverage(p)} \qquad \text{(Eq. 3.10)}$$

This measure is a variation of the *precision* measure presented above. However, instead of using relevant coverage, this measure is based on useful coverage, which consists of the set of pages within *Coverage(p)* in which users spent 30 or more seconds viewing their content. This implicit measure to estimate usefulness has been used in previous studies (Fox et al., 2005; White & Huang, 2010; Shah & González-Ibáñez, 2011). Moreover, González-Ibáñez, Shah, and White (2012) reported that "70% of the pages found useful according to the dwell-time threshold [of 30 seconds] were also relevant according to participants' explicit judgments" (p. 3).

The second measure corresponds to *efficiency*, an implicit measure of search effort expressed by the ratio between effectiveness and the number of queries that searchers have to formulate in order to reach such effectiveness. Here the number of queries corresponds to the cardinality of the set of queries that a given participant formulates and issues during a search session. Efficiency as described above is symbolized by:

$$\text{Efficiency(p)} = \frac{\text{Effectiveness(p)}}{\text{n(Queries(p))}} \qquad \text{(Eq. 3.11)}$$

3.2.2.2 Communication

In terms of communication, González-Ibáñez et al. (2012a; 2012b, 2013) describe a framework to evaluate communication in CIS. This evaluation framework comprises a coding scheme and set of measures. The coding scheme adapted from Strijbos, Martens, Jochems, and Broers (2004) consists of a group of four major categories of messages: task coordination (TC), task content (TN), task social (TS), and non-task related (NT).

The first of these categories TC, corresponds to "[s]tatements involving decision making about how the task should be performed" (González-Ibáñez et al., 2012b. p. 5). The second category TN , represents messages where participants discuss issues around the task such as topic and evaluation of sources to address the task. The third category TS, corresponds to "statements that concern group functioning, effort, or attitude as well as opinions in regard to information obtained or information sources" (p. 5). Finally, NT is comprised of messages with "a social orientation that are not related to the assignment or regarding technical issues of system being used" (p. 5).

Within each category, messages can be classified in particular subcategories such as questions, answers, control, awareness, information seeking, and polarity-based classification in order to determine the affective tone of the messages.

While the above categories are broad enough to characterize a wide spectrum of communication messages, the coding scheme also includes a fifth category (non-codable) to denote messages that cannot be classified following the criteria specified in the other four categories. The procedure to code messages as described by the authors should be performed by two or more coders in order to check inter-rater reliability.

In addition to classifying messages according to the coding scheme above, the authors described a set of quantitative measures that can be used to measure the balance and effort of the interactions during the collaboration process. The first of these measures is communication volume (Vol), which corresponds to the overall number of messages issued by an individual participant p during the collaboration process. This measure is denoted by:

$$Vol(p) = n(messages(p))$$ (Eq. 3.12)

The second measure attempts to evaluate the effort of participants while communicating. Assuming text-based communication, González-Ibáñez et al. (2012b) expressed the communication effort within each minute as the sum of the individual efforts to produce each message, which was represented as the proportion between the number of words in each message and average words per minutes (*wpm*) that users are expected to type. This measure is expressed as follow:

$$CommEffort(p) = \sum \left(\frac{\#words(msg) * 60}{wpm} \right)$$ (Eq. 3.13)

While the authors reported communication effort for the overall communication taking place in the collaboration process, this measure, as well as the measure of communication volume, can be also applied to individual categories of messages. For example, it would be possible to compute the communication volume and effort of TC, TN, TS, and NT.

The authors also described a measure to investigate how communication was distributed among team members. González-Ibáñez et al. (2012b) referred to this measure as communication balance (*B*), which can be computed with respect to the volume of messages or with respect to the effort required to produce the messages. Both approaches to describe communication balance are expressed below:

$$B_{Volume}(p) = |Vol(p_1 \in p) - Vol(p_2 \in p)|$$ (Eq. 3.14)

$$B_{Effort}(p) = |CommEffort(p_1 \in p) - CommEffort(p_2 \in p)|$$ (Eq. 3.15)

Like communication volume and effort, these measures can be independently computed for each message category.

3.2.2.3 User measures

As depicted in Figure 3.2, this dissertation involves multiple users' evaluations based on the use of different instruments such as facial expression recognition, EDA, and positivity test. Some of these instruments provide well-established procedures to compute scores, which will be discussed in Chapter 5.

This section only introduces Nahl's (2009) affective measures, which are less established measures. The first of these measures corresponds to affective load, which was introduced in the previous chapter. *Affective Load* (AL) is operationalized as follows:

$$AL = Uncertainty * TimePressure \quad \text{(Eq. 3.16)}$$

In this formulation, *Uncertainty* corresponds to levels of irritation as a result of anxiety, frustration, or rage, experienced during the search task. *Time Pressure*, on the other hand, indicates the time expected by the searcher to complete the search task before starting the search session minus his/her perception of task length after completing the task. Time pressure is denoted by:

$$TimePressure = ExpectedLenght - FeltLenght \quad \text{(Eq. 3.17)}$$

The following section provides additional details on the studies from which some of the measures introduced above were obtained.

3.3 Preliminary studies

The design decisions of this dissertation study were achieved after a research processes involving three user studies and specific investigations conducted on the data obtained from them. These studies are briefly described in the following sections and emphasis is put on their relevance for the study design introduced in the following chapter.

Different studies were carried out in the process of framing the research topic addressed on this dissertation. These studies focused on different aspects of CIS such as search processes (Shah & González-Ibáñez, 2010), synergy (Shah & González-Ibáñez, 2011), communication processes and space (González-Ibáñez, Shah, & Haseki, 2013), and time (González-Ibáñez, Shah, and Haseki, 2012a; 2012b). Each of these studies explored the affective component at a broad level shedding light on potential roles of affective processes in CIS. For example, a study conducted by Shah and González-Ibáñez (2010) about the applicability of Kuhlthau's ISP on CIS derived on the notion of group's affective relevance (González-Ibáñez & Shah, 2010), which states that in CIS relevance judgments are socially constructed through both objective and affective discourse. In other studies it was shown that affective load of searchers varies depending on the temporal and spatial conditions in which CIS takes place (González-Ibáñez, Shah, & Haseki, 2012a; 2012b, 2013). A more detailed description of each study in the context of this dissertation is provided in Appendix I.

3.3.1 Pilot study

Prior conducting the main study, a pilot study was carried out with a preliminary experimental design. This pilot study allowed performing an evaluation of design decisions, experimental system, research protocol, instruments such as questionnaires and devices, laboratory setup, task, and affective stimuli.

The preliminary experimental design was developed considering findings from previous studies as well as lessons derived from them. For example, the decision of using text chat support in collaborative conditions was made based on the findings derived from the time/space study presented in the previous section. As noted above, pairs in $C4_{Audio}$ outperformed those in other conditions in a number of measures; however, conditions with audio-based communication introduced a source of noise for automatic facial expression recognition.

In the list of analyses, it was found that pairs in $C3_{Text}$ showed similar results to those found in $C4_{Audio}$. The main difference between these two conditions was that $C3_{Text}$ showed higher communication balance than $C4_{Audio}$. Moreover, it was found that in spite of the limitations of text-based channels to carry affective information, communication was still able to generate affective reactions in the participants.

This comparison between multiple communication channels provided sufficient evidence to support the use of text chat as a communication channel in the design of the study presented in this dissertation.

Unlike the studies reviewed in the above sections, the pilot study used a precision-oriented task. Studies in CIS described so far have been conducted around recall-oriented tasks such as exploratory search. In this type of tasks, searchers usually need long periods of time and even multiple sessions to gather a reasonable amount of information. Although investigating affective processes under this type of tasks is feasible, the evaluation through self-reports is restricted to specific stages of the experimental sessions, which are usually several minutes apart from each other.

Moreover, for the particular case of CIS, it was found in the above preliminary studies that division of labor is a common strategy followed by group members when facing exploratory search tasks. That is to say, each member addresses a group of subtopics within the task, thus search processes are almost independently performed. Interactions in these scenarios are mostly related to task coordination and in a much lesser degree about information. While this kind of group behavior does not necessarily represent what occurs in long term projects, it may be possible that time constraints in laboratory studies shape the way group members address the task in order to optimize the use of time.

Contrary to recall-oriented tasks, the pilot study showed that a precision-oriented tasks, in particular multiple-step fact finding, can be a good alternative to promote interactions about information during the collaborative process. It was also found, that this kind of tasks can be completed in shorter periods of time, thus allowing multiple observations of affective processes through self reports.

In terms of performance, the evaluation of collaboration products can be performed more objectively. That is to say, other than looking at perceived relevance and information coverage, it is possible to determine whether or not the searchers are able to find the facts.

Despite the advantages of this precision-oriented task for the purpose of this study, some challenges must be considered. For example, if using the open Web and well-known search engines, the task must be carefully framed so that the answers are not retrieved with a simple query. For example, finding the answer to a question like “what is the capital of Russia?” (i.e. Moscow) can be easily obtained in Google and Bing by simply copying and pasting the question.

In the pilot study, special questions that require multiple steps to find the answer were used. To evaluate the questions, data from 12 participants (three pairs and six individuals) was collected in order to analyze perceived complexity, familiarity, difficulty, confidence levels, response time, and interactions. Based on the application of different affective stimuli to elicit positive and negative affective processes, differences were found around these factors.

For example, despite the use of a different type of task, analyses from the data collected in the pilot study partially supported the notion of synergy by showing that response precision (number of correct answers divided by the total number of questions addressed) of collaborative pairs was higher than that achieved by individual users. In terms of communication, it was found that participants who received negative stimuli communicated less than those who received positive stimuli.

A detailed report of the pilot study is presented in Appendix J. Part of the information described in this report is referenced in the following chapter in order to justify some of the decisions made at the methodological level.

3.3.2 Ten lessons

This section described different studies carried out in the preparation of this dissertation. Each study contributed not only with scientific findings, but also with particular lessons derived from observations and the experience of the researcher while conducting these studies. This section summarizes key lessons derived from the above studies that are relevant for the study presented in this dissertation.

1. In terms of recruitment, it was found that cash compensation was a useful external motivation to encourage people to sign up in the study.
2. From the participants recruited in the time/space and in the pilot study, it was observed that those coming for class credits showed less interest in performing the task. Typical behaviors of this kind of participants included: providing random answers to questionnaires in order to finish quickly, asking the researcher if they could leave if they completed the task fast, stating during the interview that they only participated for class credits and that they were not interested in winning the prize. The latter aspect contrasted with the answers of participants who did not join the study to receive extra credits. These participants showed strong interest in the competition and cash prizes[38].
3. Since participants in collaborative pairs knew each other before joining the study, collaboration was performed naturally. Communication in most cases was fluid as a result of common ground.
4. Audio-based communication channels, while facilitating communication between participants, it was found that they constituted an important source of noise for automatic facial expression recognition systems.
5. Text-based communication, in particular $C3_{Text}$, led pairs to achieve results similar to those reached by pairs in $C4_{Audio}$. While communication

[38] The pilot study was also framed as a competition with cash rewards for the three best performing teams or individuals.

volume was significantly lower in $C3_{Text}$, communication balance in this condition was higher than other communication channels.

6. Communication in $C3_{Text}$, though limited to carry affective information, was still able to transport messages that resulted in participants' affective reactions.
7. Whether affective reactions have a positive or negative classification in information search, it is necessary to evaluate them in context. While it is typically expected that positive affective reactions are linked to satisfaction when relevant information is found, and that negative affective reactions correlate to problems finding relevant information; it is also possible the opposite.
8. Useful affective reactions in information search may go beyond the spectrum of basic emotions. Specific mental states, also expressed through facial expressions and body gestures, could be found more often within search processes.
9. In order to properly explore affective processes in information seeking, whether it is performed individually or in collaboration, it is necessary to design specialized experiments that enable researchers to manipulate affective variables and perform observations from different angles (e.g. subjective experiences, expressive components, and physiological changes).
10. Finally, to investigate affective processes in CIS in the laboratory, it is fundamental to design tasks that promote active collaboration (e.g. non-dividable task) and that enable researchers to perform multiple observations during the experimentation process.

The above list is a simplification of different aspects learned from past studies and also from those conducted in the preparation of this dissertation. Other aspects not listed, but still relevant, include language and age range. These lessons had strong influence in the experimental design introduced in the next chapter.

3.4 Research hypotheses

The theoretical framework and preliminary studies (especially the pilot study) introduced in this chapter led to the formulation of a set of hypotheses that supplement the process for addressing the following research questions:

RQ1: Do initial affective states and their interactions shape the way team members collaborate when searching information, and if so, how?

RQ2: What emotions are typically experienced and expressed (physically, physiologically, and verbally) by team members when collaborating in an information search task?

RQ3: To what extent, if any, do initial positive and negative affective states and those derived from the collaboration of individuals in an information search task influence team performance?

RQ4: To what extent, if any, does the relation 3-to-1 between positive and negative affective states (P/N) (Losada & Heaphy, 2004; Fredrickson & Losada, 2005) apply to CIS?

The hypotheses are related to the influence of positive and negative affective processes in performance, communication, quality of work, and affective processes interactions. This set of hypotheses focuses not only on the products but also on the CIS process.

For example, H_4 operationalizes the relation between positive and negative affective processes in the social interactions of team members while working on a CIS task.

The list of hypotheses is presented below. Figure 3.3, on the other hand, depicts the relations between affective processes as a result of affective stimuli and two measures, namely performance and communication. Note that the first hypothesis is not depicted in this diagram because it is independent of the affective stimuli applied during the experimentation.

Hypothesis 1: The communication volume of pairs in which both members are treated with negative stimuli will be lower than the communication volume of pairs in which at least one of the members received positive stimuli.

Hypothesis 3a: Participants who receive negative stimuli will reach lower quality of work (response precision) than those who are positively treated.

Hypothesis 3b: Participants who receive negative stimuli and collaborate with someone who was also treated with negative stimuli will reach a lower quality of work (response precision) than teams in which one or two members were positively treated.

Hypothesis 4: Based on studies that have successfully evaluated the relation 3-to-1 (P/N) in different domains, it would be expected that teams in a CIS task whose P/N ratio is above this baseline should outperform - in a number of performance measures - those below it.

3.5 Summary

This chapter has introduced the research framework of this dissertation. This comprised four major components, namely, a theoretical framework, evaluation framework, preliminary studies, and a set of hypotheses.

In the theoretical framework, a review of relevant models and theories in information science, psychology, and communication were presented. In particular, the affect infusion model (AIM) and positive psychology provided some ideas of how affective processes could be related to information seeking.

In terms of evaluation framework, a data collection approach and a list of evaluation measures that are used in the experimental design were described. Measures were categorized into four groups, namely, information retrieval, task

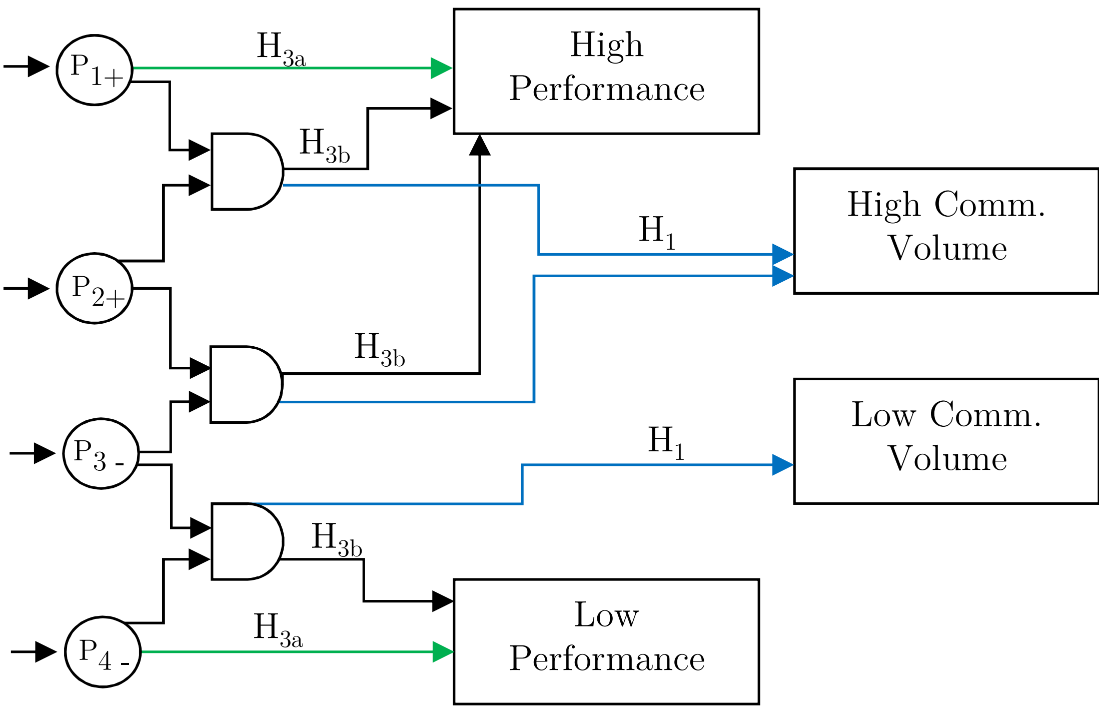


Figure 3.3. Hypotheses diagram. P denotes a participant.

performance, communication, and users' measures.

With regard to preliminary studies, three user studies were presented. Relevant aspects for the design decisions made in the preparation of this dissertation were highlighted. Different lessons were derived from this study, which were later implemented in the research design introduced in the next chapter.

Finally, a set of hypotheses derived from theories as well as from preliminary studies were introduced. These hypotheses covered aspects about collaboration at the level of products (such as performance) and process (e.g. communication, affective states, etc.)

The next chapter describes in detail the methodology derived from this research framework. This includes aspects such as study design, recruitment procedures, instruments, and research protocol.

Chapter 4. Methodology

In order to address the research questions and hypotheses stated in the previous chapter, a mixed-methods approach to investigate causes and effects of users' affective processes in CIS was implemented. This chapter provides a detailed description of the study design, which includes: experimental design, sample, task description, session workflow, laboratory setup, and instrumentation used to collect data. The methodology presented in this chapter is based on the mixed-methods approach depicted in Figure 3.1 in the previous chapter.

4.1 Experimental design

From a quantitative standpoint, a true experimental study based on multiple-group design (Salkind, 2010) was developed. The experiment was designed aiming to evaluate effects of affective processes with positive and negative valence in CIS. Likewise the time-space study described in the previous chapter, this study also investigated effects of affective processes on individual searchers in order to create baseline conditions isolated from the effects of social interactions.

Moreover, in the same way as in preliminary studies, group size was limited to the minimum, that is to say, two members per group (dyads) in order to better control interactions. As noted by Tang et al. (2010) "The number of collaborators working together exponentially increases the complexity of possible interactions, increasing the likelihood of misinterpretation and misunderstanding" (p. 2).

The general structure of the experimental design is derived from the first research question (RQ1), which aims to explore how initial affective states and their interactions shape the way team members collaborate when searching information.

As a result, the following three experimental treatments/conditions were defined:

X $_{++}$: Both team members received stimuli to elicit[39] positive affective states prior starting the task.

X$_{+-}$: One team member received stimuli to elicit positive affective states and the other stimuli to elicit negative affective states prior starting the task.

X$_{--}$: Both team members received stimuli to elicit negative affective states prior starting the task.

With regard to the baseline, two individual conditions with affective stimuli and one control group were added to the overall experimental design. The baseline conditions are:

X$_{+}$: Participants received stimuli to elicit positive affective states prior starting the task.

X$_{-}$: Participants received stimuli to elicit negative affective states prior starting the task.

Control group: No stimuli were applied.

As depicted in Figure 4.1 the overall experimental design consists of two major stages. During the study these two stages were presented to participants simply as Task 1 and Task 2. The first stage, on one hand, corresponds to the

---

[39] Note that elicit when related to stimuli refers to evocation (Coan & Allen, 2009).

Collaborative conditions	$C1^{++}$	R	O_1	PreS	O_2	$\mathbf{X}^{++}$	O_n	PostS	O_{n+1}	MT_{collab}	O_{n+m}
	$C2^{+-}$	R	O_1	PreS	O_2	$\mathbf{X}^{+-}$	O_n	PostS	O_{n+1}	MT_{collab}	O_{n+m}
	$C3^{--}$	R	O_1	PreS	O_2	$\mathbf{X}^{--}$	O_n	PostS	O_{n+1}	MT_{collab}	O_{n+m}
Baseline conditions	$C4_{+}$	R	O_1	PreS	O_2	$\mathbf{X}^{+}$	O_n	PostS	O_{n+1}	MT	O_{n+m}
	$C5^{-}$	R	O_1	PreS	O_2	$\mathbf{X}^{-}$	O_n	PostS	O_{n+1}	MT	O_{n+m}
	$C6^{Control}$	R	O_1	PreS	O_2		O_n	PostS	O_{n+1}	MT	O_{n+m}

Stage 1: short-term effect evaluation (20 min.) Individual evaluations

Stage 2: prolonged effect evaluation (25 min.) Individual/Collaborative evaluations

Figure 4.1. Experimental design summary. (R): Random placement, (PreS): Pre-Stimuli, (PostS): Post-Stimuli, (O_i): observations, (X): treatment/stimuli, (MT): main task, and (collab): collaboration enabled.

evaluation of short-term effects of the affective stimuli in information search. During this stage, all participants (either in collaborative or individual conditions) were treated as individual units. To be precise, participants performed an information search task individually.

The first stage (Task 1), as illustrated below, is divided into three parts: (1) Pre-stimuli evaluation required participants to solve an information problem in no more than five minutes without the influence of affective stimuli; (2) stimuli exposure, in which participants addressed similar information problems for 10 minutes while receiving affective stimuli; and (3) post-stimuli evaluation, which aimed to evaluate short-term effects of affective stimuli. In this final evaluation, participants also had up to five minutes to solve an information problem similar to those addressed in the pre-stimuli evaluation and stimuli stages.

The second stage (Task 2), on the other hand, focused on an evaluation of prolonged effects (during the following 25 minutes) of the affective states derived from the application of affective stimuli. Unlike the first stage, during the prolonged effect evaluation, participants in collaborative conditions were allowed to interact with their teammates, whereas participants in the baseline conditions remained working individually. The task (MT) and topics, however, for both individuals and pairs were the same.

4.2 Recruitment

In order to recruit study participants for this research, a convenience sampling was conducted. Open calls were made available to undergraduate students at Rutgers University, which consist of a population of approximately 43,380 students[40]. Recruitment announcements were posted on campuses' bus stops, Rutgers facilities (e.g. libraries, student centers, and departments' buildings), and also spread through different *listservs*.

The recruitment announcements included general information about the study such as task description, compensation ($10 per participant), risks, IRB protocol approval date[41], and eligibility criteria. Likewise preliminary studies, the study was framed as a competition, hence, the recruitment calls also included information about the possibility to win cash prices per participant based on performance ranking at the end of the study ($50 first place, $25 second place, or $15 third place).

In regard to eligibility criteria, only students who were English native speakers ranging between 18 and 24 years old, and that had intermediate typing and online search skills, were considered as potential participants. For the case of collaborative pairs, it was required that students signed up with someone with

[40] http://www.rutgers.edu/about-rutgers/facts-figures: This number is given only as a reference. Due to the sampling criteria presented below (i.e. language and age range), the population represented by the sample is actually smaller.
[41] IRB study protocol: #E12-619 approved on March 3, 2012.

whom they had previous experience collaborating, this in order to ensure common ground.

The language restriction specified above was made because it has been stated that the use of English by non-native speakers may have implications in the way affective processes as well as other types of messages are communicated and interpreted (Galloway, 1980). Similar effects may occur in terms of interpretation of task instructions and content comprehension while evaluating information to complete the task. In this matter, it has been acknowledged that cognitive effort when reading in a second language may increase (Kirkland & Saunders, 1991).

The registration process took place on a webpage, where students could obtain additional details about the study and reserve a time slot for their participation. Once the registration process was completed, both the recruits and the researcher received an email with the registration information (i.e. name of the participants, contact information, and scheduled date and time for their participation). Registration was subject to evaluation in order to ensure that eligibility criteria were met. Participants that did not meet the requirements were informed about cancellation of their sessions and the particular reasons of these decision (e.g. People who declared that their first language was other than English).

The target sample size was 135 participants. Recruits (individuals and pairs) were randomly assigned at the moment of signing up for the study to one and only one of the six experimental conditions described in the previous section. This way, it was planned to complete 15 samples per condition (45 pairs or 90 participants for collaborative conditions and 45 individuals for baseline conditions). Since the short-term evaluation was conducted with individual participants, the same number of participants could be grouped as follow: 60 samples in the positive and negative conditions and 15 samples in the control group.

While large sample sizes are recommended to achieve greater statistical power, especially in survey-based studies, it has been indicated that increasing the sample size in studies with similar characteristics to the one addressed in this dissertation may involve additional costs as well as risks (Nielsen, 2012). Other studies investigating effects of positive and negative affective states (Davidson & Fox, 1982; Hornik, Risenhoover, & Gunnar, 1987; Zhao, 2006) as well as user studies in information science (Morris & Horvitz, 2007; Capra, Marchionini, Oh,

Table 4.1. Recruitment and sampling specifications.

Target population	Rutgers students
Target sample size (N)	- 45 teams = 90 participants = 15 teams per collaborative condition - 45 individuals = 15 individuals per baseline condition
Sampling method	Convenience sampling
Time frame (recruitment + study)	3 - 4 months
Field	Heterogeneous (science, social science, and humanities)
Online search skills	Intermediate (Google + Firefox)
Typing skills	Average 50 words per minute (Ostrach, 1997).
Number of sessions	1
Session time	~ 60-70 minutes
Native language	English
Age range	[18-24]
Restrictions	Participants can participate once
Compensation for participation	$10 per subject
External motivation	Prize for the three best performing teams/participants (1st place $50, 2nd $25, and 3rd $15)

Stutzman, & Zhang, 2007; Lopatovska, 2009a; Shah & González-Ibáñez, 2010; Shah & González-Ibáñez, 2011; González-Ibáñez. Haseki, & Shah;) have employed similar sample sizes per experimental condition.

The recruitment process was initiated two weeks before the actual study started (June 2012) and it was extended during the experimentation process until September 2012. A summary of the recruitment stage is presented in Table 4.1.

4.3 Task description

As described in the previous chapter, one of the lessons derived from previous work in regard to the study of affective processes in CIS is the need to design tasks that promote active collaboration. Moreover, tasks should enable the researcher to perform multiple observations (from different points of view) of affective processes during information search.

Based on pilot study results, a precision-oriented search task, in particular multiple-step fact finding, was devised. The task comprised a set of questions that were presented sequentially during the sessions. For each question, teams and individual were given a fixed amount of time to find the responses. Additionally, the questions were non-dividable at topic level.

The questions used in the study were obtained from *A Google a Day*[42]. *A Google a Day* is a puzzle-based game implemented by Google to train and evaluate search skills. The puzzle consists of questions that can be answered by searching information online. As stated in the main website, "there is no right way to solve it, but there's only one right answer." One interesting aspect of *A Google a Day*

---

[42] http://agoogleaday.com

is that for each question an ideal search path to find the answer as well as the answer itself are provided for past questions.

A total of 144 questions posted between April 7th and August 31st of 2011 were collected along with answers and their corresponding ideal search paths. After the collection process, the search path of each question was objectively rated based on the number of steps or queries suggested to find the answer. As a result, the number of steps - also referred to as complexity level - for the entire set of questions ranged between 2 and 5. Questions that required non-textual information (e.g. videos, maps, audio, or images) to find the answer were omitted from this study. This decision was made because the complexity criteria used in the classification of questions do not consider cognitive skills required to process non-textual information. Following, an example of a level-2 question that requires textual information is presented:

> ***Question***: You're a detective at a crime scene with no visible evidence. On a hunch, you spray the carpet with a light-emitting solution and it glows, revealing blood. What component of hemoglobin catalyzed the reaction?
>
> ***How to find the answer***: Search [light-emitting solution crime scene] to find that luminol is the active chemical of choice for blood detection. Search [luminol blood] to learn that it reacts with iron to emit a slightly bluish light for about 30 seconds.
>
> ***Answer***: Iron.

Table 4.2. Study questions and corresponding answers. (*) In addition to the answer provided by *A Google a Day*, it was found for some questions alternative answers.

Stage	**QuestionID**	**Question**	**Answer**
Practice	q0	I can run up to 2.5 petaflops. In what city of the world am I located?	Tianjin
PreS	PreS_q1	El Dorado is near a famous hill that's not on Earth. What year was it photographed?	2005
Stimuli	Stim_q2	Contrary to popular belief, the Sahara isn't the biggest desert on Earth. How many square miles is the largest?	5.5 millions
	Stim _q3	What is the molecular formula of the ingredient in dynamite that is made from the oil of a legume that's often mistaken for a nut?	C3H8O3
	Stim _q4	If your doorknob is made with this alloy of copper and zinc, it naturally disinfects itself. How many hours does it take to sanitize?	8
	Stim _q5	Which U.S. vice president was able to read Greek, Latin and the world's second most commonly taught foreign language?	Thomas Jefferson
	Stim _q6	Did Titus Cornelius fight for King George III?	Yes
	Stim _q7	The oldest person to sign the Declaration of Independence criticized the national emblem and suggested what as an alternative?	Wild turkey
	Stim _q8	The Father of Modern Russia taxed wearers of these in 1705. Only Orthodox clergy were exempt from paying to keep their what?	Beards
	Stim _q9*	What is the Latin name of the carnivore with the largest teeth that was found trapped in the area once known as "Los Volcanes de Brea"?	Smilodon or Sabre-toothed cat
PostS	PostS_q10*	Wild salmon are normally born in freshwater and migrate to the sea. In America, there is a variety of salmon that never goes to sea. What was the name of this salmon subspecies, before they got stuck inland?	Sockeye salmon or lacustrine sockeye
Main Task	MT_q1	I am an animal that can grow more than 20,000 teeth in my lifetime. Which of my species, extinct or living, has the largest teeth?	C. megalodon
	MT_q2	During a famous White House séance, witnesses say the president's seat was levitated. Whose spirit was his wife trying to contact?	Willie Lincoln
	MT_q3	In Norse mythology, which god's son will poison Thor at Ragnarök?	Loki or Jormungand
	MT_q4*	An organic compound created a color so fashionable that it inspired a nickname for the decade of the 1890s. What is the name of that organic compound?	Aniline, phenylamine or aminobenzene
	MT_q5	The three-lobed leaves of a tall tree are powdered and used in a traditional New Orleans dish, and the bark is the traditional flavoring for a soft drink. What drink is it?	Root beer
	MT_q6	I was a 19th century lawyer who claimed to have killed 42 men. I was killed while playing dice by a man who was killed over a card game. Who am I?	John Wesley Hardin
	MT_q7	The painter of Starry Night doused his mattress and pillow with what to help him sleep?	Camphor
	MT_q8	Lonesome George's chances of being a dad improved because two ladies moved to his island. Where exactly did they move from?	Island of Española
	MT_q9	Chili pepper plants make the chemical that causes that burning sensation you get from eating the peppers. What are the plants trying to protect?	Seeds
	MT_q10	This invention was initially created as a walking machine and then was tweaked in 1932 to become amphibious. But neither of these are its main purpose now. What is the invention?	Cyclomer
	MT_q11	Before ascending the papal throne, the man Velázquez painted was elevated to cardinal in pectore. How many years was that appointment kept secret?	2
	MT_q12	The author whose father inspired Mr. Micawber slept facing which direction, because he thought it improved his writing?	North
	MT_q13	The second wife of King Henry VIII is said to haunt the grounds where she was executed. What does she supposedly have tucked under her arm?	Her head
	MT_q14*	All cephalopod mollusks with three hearts are carnivorous, but only one type living in temperate waters is deadly to humans. What does this deadly cephalopod normally feed on?	Crabs or/and Shrimps
	MT_q15	If you were in the basin of the Somme River at summer's end in 1918, what language would you have had to speak to understand coded British communications?	Cherokee

Results from the pilot study showed that level-2 questions were more adequate in terms of perceived difficulty, response precision, topic familiarity, and response time (see Appendix J). As a result, a random set of level-2 questions were used for both the stimuli and the main task.

Questions, as discussed in the next chapter, constitute the unit of analysis for different evaluations performed in this dissertation. In this sense, questions were presented in the same order to all participants (solo users and pairs). The list with the questions that participants in the study were able to address during the experimental sessions is presented in Table 4.2. From this list, only the question used in the practice stage was not obtained from *A Google a Day*, instead. this question was especially created so that participants could get familiar with the system and surveys, without using significant amount of the session time.

In terms of time, it was found in the pilot study that solo users and teams spent an average of 3.43 minutes (*SD*=1.57) finding the answers to level-2 questions. However, an important number of questions were not answered because of the lack of time. As a result, both teams and individuals were given a maximum of five minutes for working in each question. As an additional constraint, questions in this study could not be skipped unless an answer was provided.

4.4 Session workflow

Each session was conducted following a standard workflow of activities, which are part of a carefully designed research protocol. The later was the result of lessons derived from the preliminary studies described in the previous chapter. The protocol was devised to ensure consistency in all the session and provide better documentation for possible replications of the study. The overall protocol comprised 56 steps, which includes activities such as: clean browsing history, calibrate sensors, check room temperature, and backup the data collected from each participant. In practice, during each session a tablet device was used to access a webpage containing a checklist with all the steps of the research protocol. It is noteworthy that an important part of the steps were automatically controlled by the experimentation system (i.e. *Coagmento Collaboratory*), which is introduced later in this chapter. The main stages of the session workflow with approximate duration are listed in Table 4.3.

As noted in stage 14 in Table 4.3, participants were exposed to a video with the intention to help them to recover from negative affective states that might result

Table 4.3. Simplified version of the research protocol. (*) Time was fixed for all the participants, which was controlled by the system.

Stage	**Description**	**~Time (min)**
1	Participants were briefly introduced to the study.	1
2	Participants signed consent forms and then were placed in separate rooms (only for collaborative conditions).	1.5
3	Sensors were synchronized with the computer to which participants were assigned and systems' calibrations were performed.	1
4	Participants filled out demographic questionnaire.	1
5	Participants completed positivity test (Appendix B.1)	1
6	Participants read SAM instructions (Appendix B.2) and self-reported their feelings using this instrument for the first time.	2.5
7	Participants watched a brief tutorial that explains how to use the system.	3
8	Participants self-reported their feelings through the SAM questionnaire.	0.25
9	Participants practiced and got familiar with the system while addressing a practice question.	1.5*
10	Short-term effect evaluation: Participants performed a task consisting of a pre-stimuli question, stimuli applied while responding a set of questions, and a post-stimuli question. During the task, feelings were reported through SAM after each question (see explanation in the following sections).	20*
11	Long-term effect evaluation: Participants proceeded with the main task, which comprised several questions similar to the one responded during the short-term evaluation. During the main task, feelings were reported through SAM after each question.	25*
12	Participants completed positivity test with respect to the past 40-50 minutes.	1
13	Participants completed post-task questionnaires: engagement, cognitive load, and system evaluation.	2
14	Participants watched affective recovery video	0.5
15	Participants were briefly interviewed	5
16	Sensors were disconnected	0.25
17	Participants signed compensation receipt	0.25
Total		**~66**

from the application of stimuli or the experimentation process itself. The video was a 2012 Super Bowl announcement, which according to Affdex's results (Affectiva, 2013) for the same year in which the study was conducted, reported an elevated number of smiles from users around the globe.

4.5 Elicitation of affective states

In order to elicit the specific initial affective state in each experimental condition, proper stimuli must be applied. In section 2.1.3.1, different emotion elicitation techniques were described. While these approaches have proven to be effective in eliciting emotions and other affective states, they are non-contextual and difficult to relate to tasks when used in experiments whose focus is not emotion elicitation itself. It was observed that using IAPS or similar approaches to elicit emotions does not make sense to participants, who later are instructed to search information. Therefore, one challenge in the design of the study was to incorporate affective stimuli so that participant could see them as part of the task.

A more flexible approach to elicit affective states that can be integrated with the task described above is game feedback (Martin, 1990), also referred to as false-feedback (Zhao, 2006). This technique consists on providing either positive (e.g. "You are doing great!) or negative (e.g. "Wrong. That was disappointing") feedback to participants regardless of their actual performance when working on a given task.

Side effects of this approach include frustration, disinterest on performing the task, and overconfidence, which can be overcome by balancing the number of positive and negative feedback provided to participants. Following the theory on positive psychology discussed in the previous chapter, the relation 3-to-1 between positive and negative affective states was used to establish the proportion of positive and negative stimuli delivered to participants depending upon the condition to which they were assigned.

Stimuli consisted of a box containing a text message and a blinking emoticon (smiley, frowning, or neutral face) on top of the box. Firstly, text messages contained words from the positive and negative categories found in the Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Francis, & Booth, 2001). Secondly, the blinking effect was implemented to grab the attention of participants while working in the task. Thirdly, the size and position of the color box were adjusted based on observations made on the eye-tracking data, in particular eye fixations collected during the pilot study. Fourthly, boxes and emoticons were presented in three different colors, namely, green for positive stimuli, red for negative stimuli, and yellow for negative-neutral and positive-neutral stimuli. The latter color was used to achieve the relation 3-to-1 by expressing slightly negative message (e.g. "So so. You can do it") for participants in the positive condition and slightly positive messages (e.g. "Just a little better this time") for those in the negative condition. This was implemented in order to avoid cancelling the expected affective response specified by each experimental condition. Finally, color boxes with the corresponding messages and emoticons on top were presented to participants during the stimuli stage on the sidebar panel of the experimental system (section 4.7) for 15 seconds and every 30 seconds. Stimuli were also presented at the moment of submitting the answers to each question. A closer look to the stimuli stage for a positive condition is depicted in Figure 4.2.

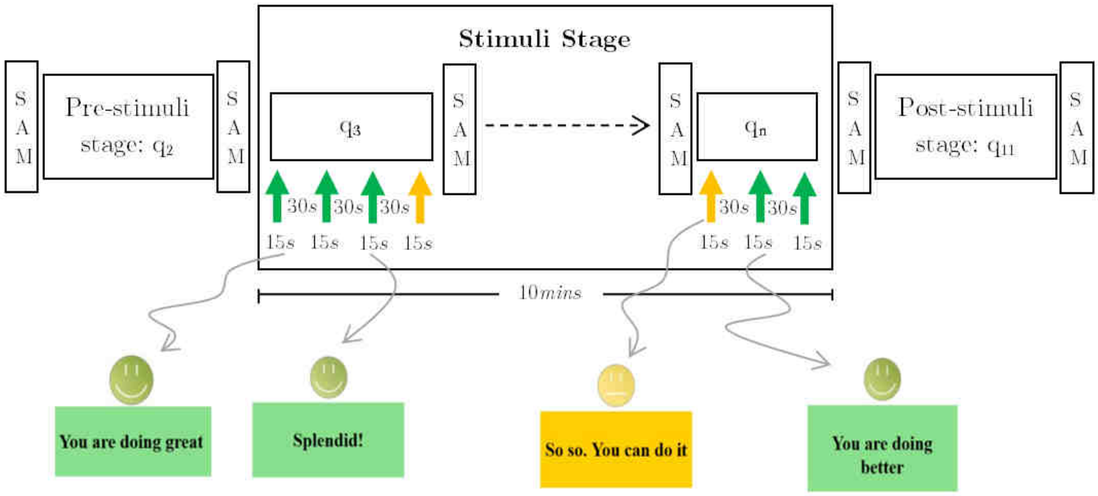


Figure 4.2. A closer look to the stimuli stage of positive condition (X^+).

Participants in the study received stimuli during the short-term evaluation while performing a task similar to the one they had to perform later in the prolonged effect evaluation (Figure 4.1). The evaluation of the emotion elicitation approach chosen for this study was performed in the pilot study following three procedures: (1) self-reports as part of the pretest-posttest experimental design, (2) observations, and (3) interviews. An analyses of the participants' responses to the Self-Assessment Manikin (Bradley & Lang, 1994) during the first 5 minutes of the stimuli stage showed that 66.67% of the participants who were treated with positive stimuli reported feeling happy, 28.20% neutral, and only 5.13% unhappy. On the other hand, 18.75% of those treated with negative stimuli reported feeling happy, 18.75% neutral, and the remaining 62.5% unhappy. Finally, 42.86% of those who did not receive treatment reported feeling happy, 35.71% neutral, and 21.42 unhappy. The latter results were found in a follow up study, which is not reported as part of the pilot study.

Like other approaches, participants exposed to different treatments may react differently, for example, a participant that receives negative feedback may remain in a positive mood. Observations from the pilot study indicate that some participants reacted laughing when receiving negative feedback; however, at the moment of reporting how they felt, they indicated feeling unhappy. On the contrary, participants who were frustrated but engaged with the task being performed, reported feeling happy even after receiving negative feedback. In terms of interviews, when asked about how they felt with respect to feedback provided during the first part of the session, some of the answers of participants who received positive stimuli are: “It was like a boost in confidence”, “I felt pretty good.” On the other hand, examples of the answers from participants who were treated with negative stimuli are: “I started getting less confident”, “It was horrible.”

It is acknowledge that the stimuli is not 100% effective in the elicitation of specific affective states in all the participants; however, as shown in the results presented above, most participants reported feeling in the affective state (i.e. positive or negative) that was intended to evoke with the stimuli applied.

4.6 Laboratory setup

The study was conducted in the Interaction Lab at the School of Communication and Information (SC&I) of Rutgers University. Two rooms with almost identical conditions were used for the experimental sessions. The main difference between these rooms is that one of them (Room-A) has a large window, whereas the other does not (Room-B). Blinds were used in Room-A to mimic the conditions of Room-B. In terms of room temperature, this was adjusted in both rooms to be approximately 70ºF. Figure 4.3 depicts the laboratory setup.

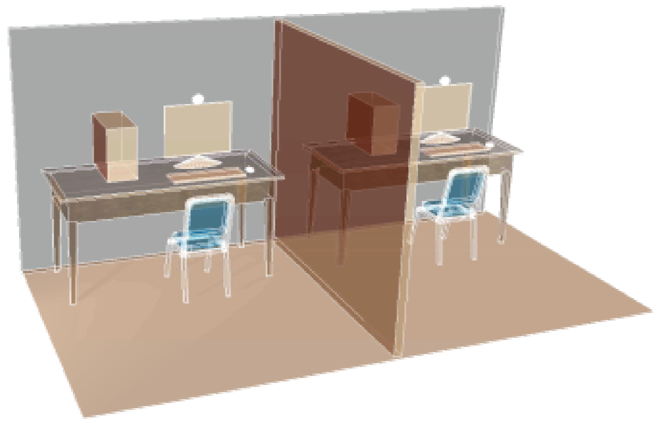


Figure 4.3. Laboratory setup with Room-A (left) and Room-B (right). Note that the division in the middle is a wall.

In addition to these two rooms, a third room (located right outside Room-A and Room-B) was used by the researcher to monitor the session workflow and to perform non-participant observations. The complete setup is illustrated in Figure 4.4.

The above setup was specifically designed to run sessions with pairs of users. For single users, on the other hand, only one room (either Room-A or Room-B) was picked and monitored through the corresponding screen as depicted in

Both rooms were equipped with the identical chairs, desks, data collection instruments, and computers. The three computers used in the study were Dell Optiplex 990 workstations (Intel i7-2600 CPU, 3.4GHz, 64 bits, 8 GB in RAM, 1TB of storage) equipped with identical 19” displays, full size keyboards, mice,

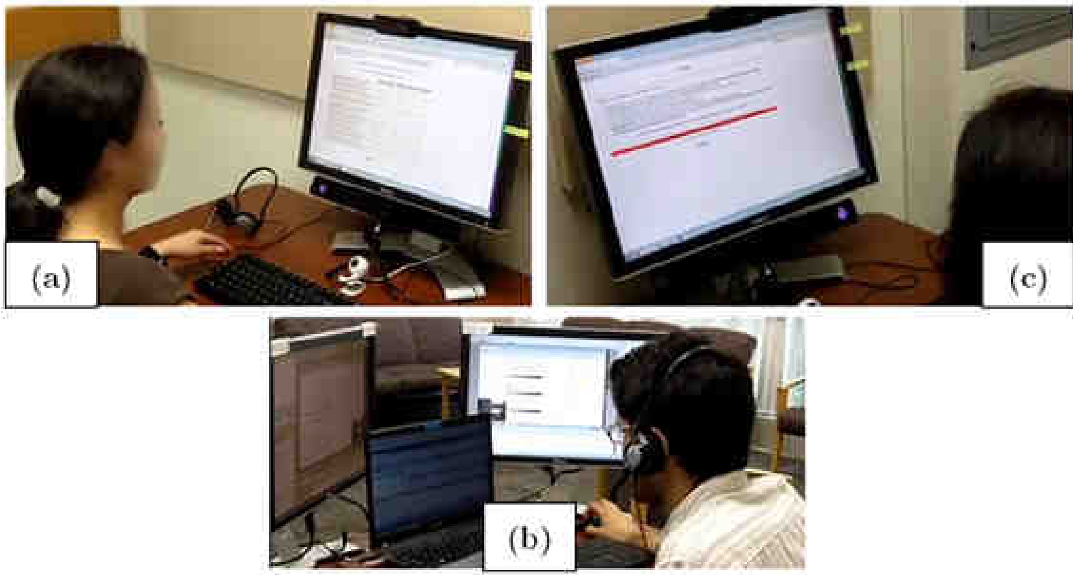


Figure 4.4. Participants and researcher laboratory setup. The participant in (a) was monitored through the display on the left in (b). Likewise, the participant in (c) was monitored through the display on the right in (b). Laptop in the center in (b) was used to monitor the electrodermal signals of both participants.

and headsets. In terms of operating system, the 64-bit version of Windows 7 Enterprise (Service Pack 1) was used in order to ensure compatibility with the hardware and software used in the study. The rest of the equipment and systems used in the study are presented in the following section.

4.7 Instrumentation

In order to collect data, measures, and perform observations during the experimental sessions, multiple devices, instruments, and software were incorporated in the study. These resources are organized into three groups, namely, hardware, questionnaires, and software. Following, brief descriptions of the instruments and systems under each group are presented.

4.7.1 Hardware

At the hardware level, four types of devices were integrated in the laboratory setup, specifically, sensors to measure and capture electrodermal activity (EDA), facial expressions, eye fixations, temperature, and humidity. A brief description of each sensor is provided below.

4.7.1.1 EDA sensor

In order to measure and record EDA, two *Affectiva Q Sensors 2.0 Curve* were used. The participants in both individual and collaborative conditions wore a *Q Sensor* on their non-dominant hand (or the one that was not controlling the mouse) during the entire session. The sensor was set to capture data at its highest rate (i.e. 32 Hz). An illustration of the *Affectiva Q Sensors 2.0 Curve* is presented in Figure 4.5.

Figure 4.5. Affective *Q Sensor* 2.0 Curve.

EDA data captured with *Q Sensors* is timestamped thanks to an onboard clock, which was synchronized with the local time of computers in order to facilitate later event-related analyses. The device also records data regarding movements in a tri-dimensional space and body temperature. While the data is collected, the device controls effects derived from motion artifacts.

Following the recommendations of *Affectiva*, a procedure to activate participants in non-physical demanding tasks was integrated in the research protocol described above. After attaching the sensor to participants' wrists, they were instructed to sit and stand up from a chair several times for approximately one minute. After this activation procedure, if participants' EDA remained weak, a saline solution or a specialized electrode gel recommended by *Affectiva* was applied in participants' wrists where the *Q sensors*' electrodes were placed.

4.7.1.2 Eye tracker

Two *Mirametrix S2 Eye trackers* (one per participant) were used in the experimental sessions in order to log participants' eye fixations on the screen as well as other eye-related data while working in the task. The eye tracking data was captured at 60 Hz and each sample was timestamped internally by the sensor.

Eye trackers were calibrated after the EDA activation protocol described above. Calibration was performed with the *Mirametrix* Tracker software. Additional adjustments were performed based on participants' features such as height and presence of optic glasses.

Eye tracking data in this dissertation had the following three uses: (1) during the pilot study, evaluate participants' attention to components of interests in the system such as task time, notifications, and stimuli; (2) at the interview stage, videos of desktop activity with eye fixations overlaid were used to help participants to recall what they did during their respective sessions; and (3) provide a validation data source to which the researcher can refer to in order to help contextual interpretations of affective data (i.e. EDA, facial expressions, and questionnaires).

4.7.1.3 Webcam

In order to capture and observe (in real time) participants' facial expressions, two webcams were used per participant. One Logitech C910 webcam was placed on top of the computer screen to capture participants' frontal faces. High definition video was captured with this camera at 15fps[43] with a resolution of 920x720. Along with video, audio was also recorded. Video recorded with this camera constituted the main source of data for facial expression analyses.

Additionally, an Ubisoft webcam (equipped with a wide angle lens) was placed on the bottom part of the computers' screens (slightly inclined upwards in approximately 20°) to capture participants' frontal faces even if they were looking down while typing. This camera was set to record video at 15fps with a resolution of 640x480. In order to avoid an impact in the frame rate and video quality of the C910 webcam, the Ubisoft camera was also used to stream video and audio to the researcher so that he could perform real-time observations of participants' facial expressions. The complete setup for the two webcams and the eye tracker is depicted in Figure 4.6.

---

[43] Frames per second

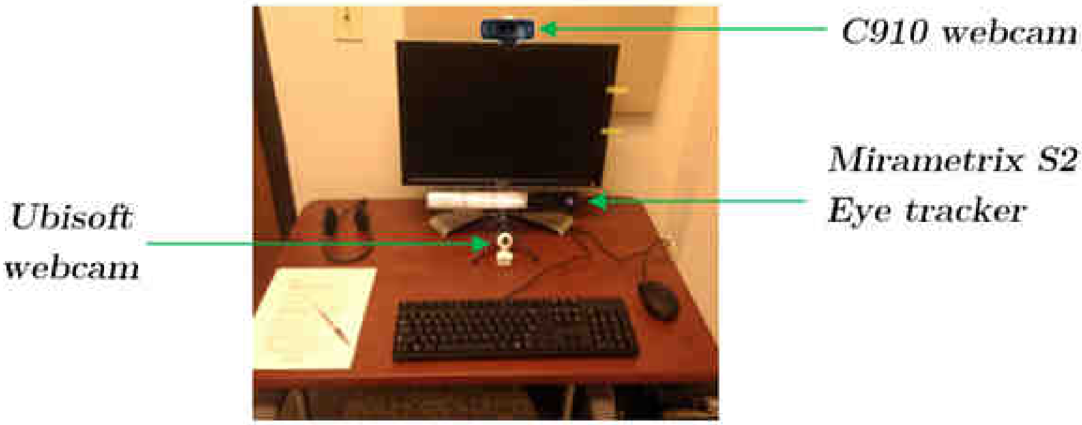


Figure 4.6. Eye tracker and webcams setup.

4.7.1.4 Thermometer and hygrometer

In order to monitor room conditions, an *Acurite* digital humidity and temperature monitor (model 00325) was installed in Room-A and Room-B. This instrument allowed the researcher to register the temperature and humidity of the room before and after each experimental session. Since rooms' doors remained closed during the sessions, a raise in room temperature was expected, which was later stabilized before starting the following sessions. Both temperature and humidity were controlled to avoid possible artifacts in EDA.

4.7.2 Questionnaires and Interview

In order to investigate the perspective of participants in terms of affective processes, cognitive load, and the overall experience during the sessions, questionnaires and a semi-structure interview were used. The following sections describe how these data collection instruments and procedures were included in the study.

4.7.2.1 Self-Assessment Manikin (SAM)

Measures of the affective dimension from the perspective of participants were performed with SAM (Bradley & Lang, 1994). Using SAM, the participants self-reported how they felt immediately before and after working in each question. As described in section 2.1.3.2, SAM is a non-verbal scale that allows individuals to indicate their affective states or reactions in three dimensions, namely, pleasure (happy-unhappy), arousal (excited-calm), and dominance (controlled-in control). There are different variations of the scale; however, in this study the 9-points version of the scale was used. The implementation of SAM in this dissertation is depicted in Appendix B.2.

Due to facial expression and EDA do not reveal how participants feel or experience a given emotion or affective state, this survey allowed individuals to self-report how pleasant or unpleasant their affective states felt at a given moment.

In the pilot study, users read an adapted version of the SAM instructions provided in (Lang et al., 1997) (Appendix B.2). In the pilot study, participants averaged 84.833 seconds (*SD*=49.063) reading the instructions of SAM in an early stage of the sessions. After this stage, although participants had the possibility to revisit the SAM instructions while completing the questionnaire, none of them went back to the instructions. Responding each SAM, on the other hand, took participants approximately 15 seconds.

4.7.2.2 Positivity self test

The positivity ratio (i.e. ratio between positive and negative affective states) was measured using the *positivity self test* (Fredrickson, 2009). As described in section 2.1.3.2, this questionnaire is a modified version of the Differential Emotions Scale (DES) (Izard, 1977), which is referred to as mDES (Fredrickson, Tugade, Waugh, & Larkin, 2003). The *positivity self test*, which consists of 20 items, was used in two instances during the sessions. First, participants were instructed to complete the questionnaire in terms of their affective experiences during the past 24 hours. This was conducted right after the calibration of the devices described in the previous sections. Later, at the end of the session, participants were asked to report how they felt during the past 40 to 50 minutes in regard to their experiences during experimental session.

4.7.2.3 NASA Task Load indeX (TLX)

In addition to measure affective process, cognitive load with respect to the task was measured using a subjective workload self-assessment test (Hart & Staveland, 1988). This questionnaire referred to as NASA TLX (Appendix B.3) consists of six questions and 7-point scales that are used to provide the responses. Each point in the scale is subdivided in three increments representing low, medium, and high estimates. As a result, the overall scale has 21 gradations.

4.7.2.4 Non-Standard Questionnaires

Non-standard questionnaires were designed to evaluate the experience of users in regard to the following aspects: quality of the tutorial, topic familiarity, topic complexity, collaboration experience, and system perception. Answers to these questionnaires were provided in a 5-point Likert scale.

The tutorial survey consisted of two questions, one presented right after watching the tutorial and the other one presented at the very end of the session (Appendix B.6).

A short survey regarding topic familiarity and complexity was presented along with each question. In order to proceed with the task, participants had to first respond the survey. Later, after completing each question or running out of time, another brief survey was showed to participants asking about the perceived difficulty after addressing the question and the level of confidence in their answer (if any). These surveys are presented in Appendix B.4.

For collaborative conditions, each participant had to respond a questionnaire at the end of the session in order to rate their experience collaborating with their teammate during the task. The questionnaire consisted of four items about levels of effort, attention, required concentration, and involvement (Appendix B.5).

Finally, the last questionnaire that participants had to respond consisted of a set of affirmations aimed to evaluate usability factors and user satisfaction with respect to the system (Appendix B.6).

4.7.2.5 Interview

In addition to the above questionnaires, a brief semi-structured interview was conducted at the end of each session. The interview was based on a group of 14 questions that were adapted based on the observations of the researcher during the session. Questions were framed to obtain additional details about affective experiences, affective reactions observed during information search, perception of the stimuli stage, previous experiences participating in similar studies, and motivations to take part in the study. The list of base questions of the interview is presented in Appendix B.4.

Interviews were conducted remotely using Skype. During each interview, the researcher showed the participants a video of their desktop activity with eye fixations overlaid (Figure 4.7). Participants were instructed to use the video and briefly describe aloud (similar to think-aloud protocol) what they did in specific episodes identified by the researcher or that participants recalled to be remarkable.

During the interviews conducted in the pilot study, it was observed that the responses of one of the team members ended up influencing that of their partners. Another side effect of this approach is that responses were provided mostly by one of the team members, while the other remained in silence. In

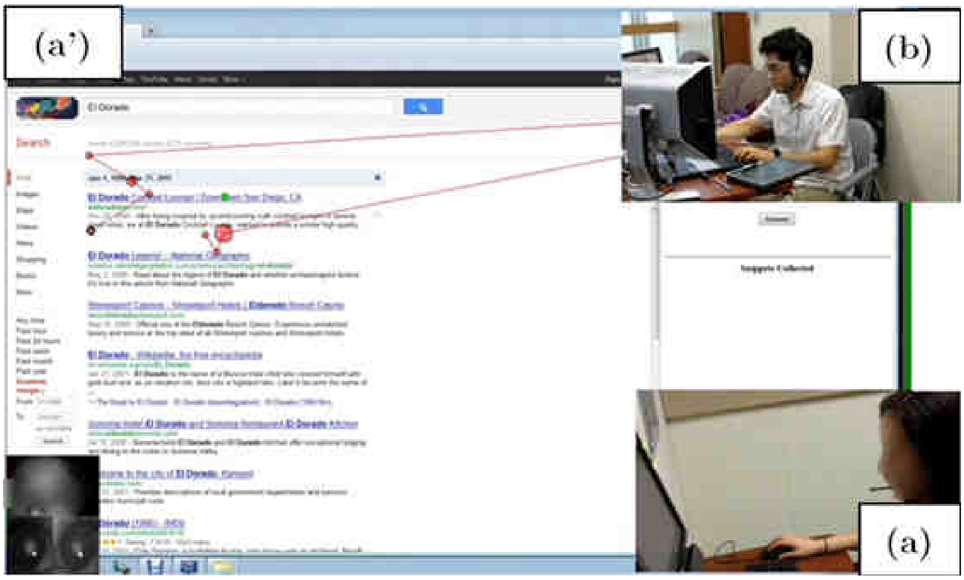


Figure 4.7. Example of an interview with one participant. Both the researcher in (b) and the participant in (a) are watching a video recording of the session with eye fixations overlaid (a').

order to address these shortcomings, both participants in the team were called at the same time; however, they could not hear each other. To avoid confusions, the researcher acted as a moderator by instructing and coordinating participants about when to provide their answers to the questions.

4.7.3 Software

The final group of data collection resources consists of software that helped the researcher to keep track of browsing activity, record interviews, and perform observations. Following, a description of each piece of software and how they were integrated in the study design is presented.

4.7.3.1 Coagmento collaboratory

Coagmento collaboratory is a customized implementation of *Coagmento* (Shah, 2010b; González-Ibáñez & Shah, 2011), a software that supports individual and collaborative information seeking. This new version of *Coagmento* was especially designed to support experimental designs involving affective treatments, timed tasks, multiple-stages sessions, communication, and logging capabilities.

Coagmento Collaboratory comprises three main components: (1) a toolbar, (2) a sidebar, and (3) a server-side Web application. First, the toolbar provided the participants with two buttons, namely, a *search* button to access Google (the search engine enabled during the experiment) and a *snip* button to help them so save snippets of text that along with their sources, which was required during the task. In addition, the toolbar recorded all browsing activity and different users' actions within the browser. In the collaboratory version, this feature was enhanced by providing efficient logging capabilities, robustness, and a wide range of actions. Data was logged along with local timestamps in order to facilitate synchronization with data captured with other devices (e.g. Q Sensor and Mirametrix S2 Eye Tracker).

Second, the sidebar displayed different elements depending upon the stage of the task. For example, during some stages the remaining time to respond questions and the questions themselves were displayed on top of the sidebar. During the stimuli stage, these were displayed on the bottom part of the sidebar. In the long-term evaluation stage and only for collaborative conditions, a chat system was enabled. Finally, snippets collected while working on each question were displayed on the sidebar. When collaboration was enabled, snippets were shared among collaborators. Content in the sidebar was automatically updated with AJAX[44] calls to the served-side Web application. A screenshot of *Coagmento collaboratory* and Firefox 11 during an actual session is presented in Figure 4.8. Note that both the toolbar and the sidebar container were implemented as a Firefox extension.

Finally, the server-side Web application controlled the session workflow and dynamically generated the content displayed in the main container and also that presented in the sidebar container. Workflow properties such as the order in which question were presented, stages duration, and stimuli duration, were parameterized through values stored in a database table. Likewise, the server-side Web application produced content to be displayed in the main container of Firefox, which included surveys, tutorial, instructions, and the questions that participants had to respond as part of the tasks.

---

[44] Asynchronous JavaScript and XML

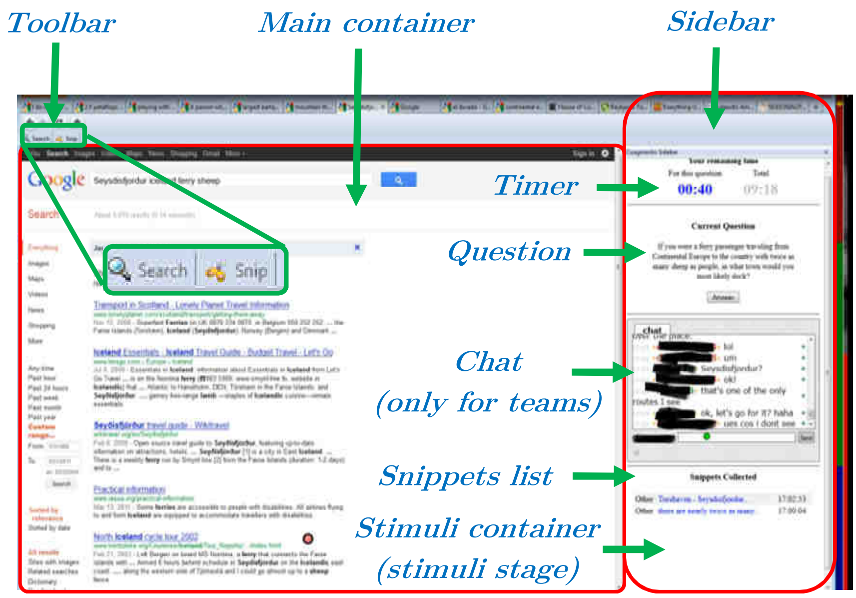


Figure 4.8. Screenshot of *Coagmento Collaboratory* and Firefox 11. Participants' nicknames in the chat box have been blacked out.

Coagmento Collaboratory was installed in Mozilla Firefox 11. As explained in the task description, due to the particular characteristics of the task, searches were restricted to be performed only on *google.com* within a specific period of time. This restriction was controlled by the *Coagmento Collaboratory* toolbar in order to avoid that participants could find the questions and their answers posted on webpages recently indexed by Google.

To reduce effects of external software components available on Firefox, different features were disabled in Firefox. For example, the address bar, bookmarks, other toolbars, and components installed by default in Firefox were disabled.

4.7.3.2 Morae Recorder 3.1

Produced by TechSmith, *Morae* is a software to perform usability studies. This software was specifically used in the study to record (1) desktop activity, (2) users' actions such as mouse clicks and keystrokes, (3) video captured with the

Ubisoft webcam, and (4) system processes. Both desktop activity and webcam were streamed in real-time to *Morae Observer* (see below).

4.7.3.3 Morae Observer 3.1

As part of the Morae suite, the *Observer* is a software that allows researchers to watch in real-time the screen, actions, and webcams of computers running *Morae Recorder*. This software was used by the researcher to perform non-participant observation while participants worked on the task. During the observational process, *Observer* also enabled the researcher to add notes or generate markers to facilitate later analyses.

Morae Observer was installed in the supervisor computer. Depending upon the experimental condition (single users or collaborative pairs), one or two instances of the Observer software were used.

4.7.3.4 NCH Debut

In order to record high resolution video from the webcam at a stable frame rate, the *NCH Debut* (version 1.64) software was used exclusively to record video along with timestamped frames. The latter aspect was implemented to facilitate post-processing procedures such as video segmentation and synchronization. Videos were recorded at a resolution of 920x720 at 15fps.

4.7.3.5 Mirametrix Viewer and Tracker

Viewer is a software designed to capture desktop activity along with eye tracking data. The software was configured to record the screen of users with eye fixations overlaid, so that this could be presented later to the users during the interview stage.

A supplementary software, namely, Mirametrix Tracker was used to perform calibrations of the eye tracker at the beginning of each session. More importantly, this software was also used to deliver eye tracking status to the researcher. This way, the researcher could monitor if participants were within the capturing range of the eye tracker.

4.7.3.6 Affectiva Q Live

The *Q Live* software was used by the researcher to monitor EDA as part of the non-participant observation protocol. Using the Bluetooth capabilities of the sensors, the Q Live software was configured to display the signals of one or two sensors depending upon the experimental condition. The software was also used to remotely add markers in the data when an event of interest was observed.

4.7.3.7 GTalk

In order to communicate with the participants during the experimental sessions, a separate communication channel was enabled. *GTalk* was only used to allow participants to ask questions to the supervisor in case they were experiencing a problem and also to receive instructions from the supervisor (e.g. asking participants to correct their body posture so that the eye tracker could detect their eyes). For participants, two Gmail accounts were created. In each account, only the supervisor's email account was added. Because participants' actions were actively monitored through *Morae Observer*, most interventions were initiated by the supervisor.

4.7.3.8 Teamviewer

In order to assist participants remotely and avoid interventions of the researcher by entering to the laboratory rooms, the researcher used Teamviewer 7 to take control of the participants' computers in case technical problems had to be solved or in the final stage, where the researcher showed the participants the recording of their sessions captured with the *Mirametrix Viewer* software.

4.7.3.9 Google

As indicated in the task description, participants in this study were instructed to use only *google.com* to perform their searches during the task. Because the task was based on a set of questions collected from *A Google a Day*, whose responses can be found few hours after questions are posted, Google was parameterized to display results from documents indexed before March 31st of 2011. The *Coagmento Collaboratory* internally forced these parameters be present in all the participants' searches as an implementation of the notion of wormhole described in section 4.3.

Finally, in order to avoid dynamic changes in the layout of Google as the user typed queries, the Google instant feature was disabled. This design decision was supported by results from an experiment conducted by Shah, Liu, González-Ibáñez, and Belkin (2012), who did not find significant differences in terms of the number of queries issued and pages visited by participants using different Google settings, namely, (1) query suggestion disable, (2) query suggestion enabled but instant feature disabled, and (3) both instant and query suggestion enabled.

4.7.3.10 Skype and Pamela call recorder

The interviews, as indicated in the previous section, were conducted remotely using Skype. Two accounts were specially created to run this study. In order to record the conversations between the researcher and the interviewees, Pamela call recorder was used.

4.8 Summary

This chapter reviewed the methodological approach used to carry out the empirical evaluation of this research. Design decisions were derived from the research framework introduced in Chapter 3.

In terms of experimental design, an experiment comprising three collaborative conditions, two individual conditions, and one control was designed. Experimental conditions were linked to specific affective stimuli to elicit positive and negative affective states in the participants.

The recruitment procedure was conducted through convenience sampling. The target population was students from Rutgers University. With respect to the task, this was designed as a precision-oriented search task comprising several *A Google a Day* questions. In regard to the session workflow, this was conducted following a strict research protocol to ensure consistency in all the experimental sessions.

With respect to the manipulation of affective variables, a procedure to elicit positive and negative affective states was introduced. Finally, the instruments and resources (i.e. software, surveys, and hardware) used to collect data were described. A summary of the data collection instruments and methods is presented in Table 4.4.

Table 4.4. List of instruments for data collection and their purpose in the study.

Type	Data collection instrument/method	Purpose	Data type
Quantitative	Questionnaire: Demographic	Demographic	Categorical, interval, ratio
	Questionnaire: Positivity Self test (Fredrickson, 2009) and SAM (Bradley & Lang, 1994)	Feelings/Affects (Subjective)	Ordinal or interval
	Questionnaire: Engagement	Engagement (Subjective)	Ordinal or interval
	Questionnaire: NASA's TLX (Hart & Staveland, 1988)	Cognitive load (Subjective)	Ordinal or interval
	Coagmento Collaboratory	Browsing activity (pages, queries, snippets)	Categorical
	Coagmento Collaboratory: Answers	Performance	Categorical
	Coagmento Collaboratory: Communication	Communication/Affective processes	Categorical
	Q Sensor 2.0 (EDA)	EDA Affects/Moods (activation only)	Ratio
	Cameras: Facial expression	Basic/Universal emotions	Categorical
	Mirametrix S2 Eye Tracker	Eye fixations/Gaze	Ratio
	Morae Recorder	Desktop activity/Users' actions	Categorical
Qualitative	Morae Observer	Behaviors	Qualitative
	Semi-structured interviews	Experience/Behaviors/Validation	Qualitative

Chapter 5. Results

The previous chapter described the methodology implemented to conduct an experimental study and collect data to address the research questions and hypotheses of this dissertation. This chapter provides a detailed description of the quantitative analyses and results of the study, which includes: sample demographics, data exploration, results for each research question, and hypothesis testing.

5.1 Sample demographics

Overall, 148 students signed up to participate in the study (64 pairs and 84 individuals); however, only 145 (50 pairs and 45 individuals) showed up. A session was cancelled because its participants (both women) could not remove their scarves, which obstructed face regions necessary for facial expression recognition. Other sessions were cancelled because their participants were not English native speakers as they indicated in the recruitment form. In few cases, data collected was discarded due to technical problems with the system or some of the instruments. As a result the data from 135 participants (45 pairs and 45 individuals) were kept. From this sample, 66% of the participants were women. In terms of collaborative pairs, six were men-pairs, 24 women-pairs, and 15 mixed-pairs. The sex distribution across experimental conditions is illustrated in Figure 5.1.

The relationship types of participants in each pair were specified through a multiple-choices question with the following options: couple, friends, coworkers, roommates, siblings, and other. For the “other” category, only two pairs reported that in addition to being friends, they were also labmates and

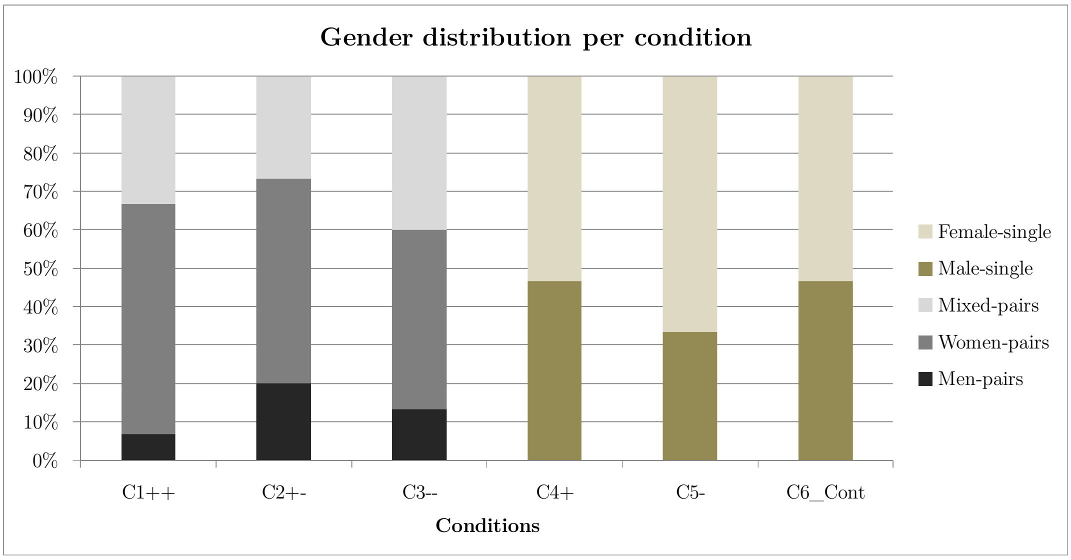


Figure 5.1. Sex distribution per condition.

housemates. A summary of the individual responses for each category is presented in Table 5.1. Note that participants could indicate more than one type of relationship, thus frequencies of individual responses overlap.

In addition to report the types of their relationships, participants also indicated their age, which ranged between 18 and 23 years (*M*=20.13; *SD*=1.36); and for how long they knew each other, which varied between one to 23 years (*M*=4.30,

Table 5.1. Individual responses of participants about their relationship with their partner.

Relationship category	**N (individual responses)**
Friends	42
Roommates	26
Couples	16
Siblings	4
Coworkers	2
Other (labmates and housemates)	2

SD=5.05). During this period of time, participants declared having had previous experience searching information with their teammates, which fluctuated between one and 20 times (*M*=7.17, *SD*=5.64). Participants also indicated that they had searched information in collaboration with others in one or more opportunities (*M*=11.03, *SD*=14.32).

In terms of areas of studies, participants came from 52 programs offered at Rutgers University. Examples of such programs are animal science, biomedical engineering, business, media studies, nursing, information technologies, history, and communication. Only two participants did not have their program majors declared at the moment of taking part in the study. For summarization purposes, programs were categorized into nine areas, specifically, art and humanities, business, engineering, health, management, natural science, social science, technology, and undeclared. The classification criteria were based on the areas of studies specified by the School of Arts and Science of Rutgers[45]. For programs not listed in this source, classification was based on the schools that offered the programs declared by participants (e.g. engineering and business). Table 5.2 summarizes frequencies of individual responses classified into the nine categories listed above. Since some participants reported being enrolled in more than one program, frequencies of individual responses overlap.

At the beginning of each session participants were asked whether they were right-handed, left-handed, ambidextrous and which hand they used to control the mouse. This information was required to decide the location of the Q Sensor

[45] http://sasundergrad.rutgers.edu/academics/requirements/core

Table 5.2. Distribution of participants in different program areas. Note that frequencies in the second columns involve overlapping responses due to participants enrolled in more than one program.

Program area	**N (individual responses)**
Social science	59
Health	28
Natural science	26
Arts and humanities	14
Engineering	12
Management	11
Business	10
Technology	4
Undeclared	2

with the aim of reducing possible effects of motion artifacts in the EDA signal. According to participants' responses, 87.30% were right-handed, 10.32% left-handed, and 2.38% ambidextrous. However, all participants indicated that they used their right hand to control the mouse, hence Q sensors were attached to participants' left hand.

With regard to operating systems preference, 52.59% used Microsoft Windows, 45.93% Mac OS, and 1.48% Linux. In terms of Web browsers, participants' preferences were distributed as follow: 41.48% Google Chrome, 28.15% Safari, 22.96% Mozilla Firefox, and 7.41% Microsoft Internet Explorer. When asked about what search engine they used the most, 96.29% indicated that Google was their preferred option. The remaining 3.71% included answers such as Yahoo, Facebook, and Wikipedia. Finally, search experience was reported by participants using a 5-point Likert scale, where one point indicates very

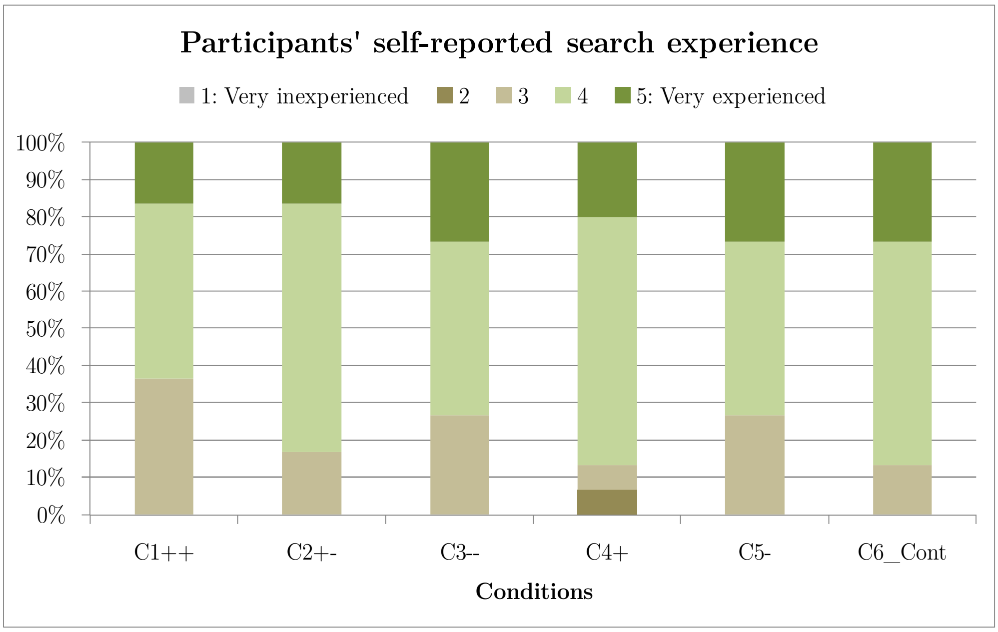


Figure 5.2. Participants' self-reported search experience.

inexperienced and five points very experienced. Responses indicated that 99.25% of the participants felt with intermediate to very advanced search experience (three to five points). Only one participant rated her experience with two points. A summary of participants' self-reported search experience across conditions is depicted in Figure 5.2.

5.2 Data exploration

Before proceeding with data analyses to address the research questions and hypotheses of this dissertation, data were preprocessed and explored. Preprocessing comprised data filtering, computation of measures according to the evaluation framework introduced in section 3.2, and data merging to combine multiple sources of data. On the other hand, data exploration focused on testing statistical assumptions (e.g. normality and homogeneity of variance) and identifying outliers.

5.2.1 Data preprocessing

In a preliminary stage of data preprocessing, participants' answers to each questions were manually evaluated using as a reference the list of answers indicated in Table 4.2. In addition, the search paths were inspected by exploring search logs captured by *Coagmento Collaboratory* and videos of the sessions displaying users' actions and eye tracking overlaid. This process was performed to evaluate whether participants visited banned pages that contained some of the questions of the study with their respective answers. As explained in the previous chapter, *Coagmento Collaboratory* was configured to restrict *google.com* to retrieve information indexed before the date in which the set of questions used in this study were posted on *A Google a Day.* Additionally, the system was set to block access to sites that were known to contain questions and answers from *A Google a Day* (e.g. *wikianswers.com*). In spite of these precautions, few participants accidentally or intentionally found ways to access such pages.

The inspection of participants' search processes provided a way to assign scores to answers that were partially correct or incomplete. For instance, some participants were unable to find the precise answer for some of the questions, however, their queries and the pages they visited indicated that they were in the right direction to find the correct answer. For example, in the practice question (q_0) "I can run up to 2.5 petaflops. In what city of the world am I located?" some participants responded China; however, as stated in Table 4.2, the right answer is *Tianjin*, which is a Chinese city. Few others responded *Tianhe,* which is the name of the supercomputer implicitly referred in the question. In other cases, the answers were spelled wrong (e.g. *Tijan*), however, pages visited and the snippets collected indicated that participants found the right answer.

Derived from this process, a coding scheme was established to evaluate participants' responses to each question. The codes, descriptions, and the number of answers classified under each code are presented in Table 5.3.

Following the evaluation of answers, sessions were segmented using start and end local timestamps for each question according to *Coagmento Collaboratory* logs. This information was used to segment video and electrodermal activity (EDA) data. Samples from these data sources were timestamped with the local time of the computers used by participants during the experimental sessions. As indicated in the protocol presented in Table 4.3, devices and software used to

Table 5.3. Coding scheme to evaluate participants' answers and the corresponding number of answers coded under each category.

Code	Description	# answers
0	No answer was provided.	150
1	The answer is completely incorrect. More importantly, it reflects a divergent search behavior with respect to the information space in which the right answer is located (e.g. answers such as California or Germany in the questions about the supercomputer that can run up to 2.5 petaflops discussed in the text).	342
2	The answer is incorrect, however, it shows convergence toward the right answer (e.g. *China*, while wrong for the question about the supercomputer, it was found in the same information space where the right answer, i.e. *Tianjin,* is located).	274
3	The answer is partially correct. That is to say, it contains part of the full answer for a given question (e.g. in an answer comprising two names, only one was provided by participants)	89
4	The answer is correct, but spelling mistakes are present.	6
5	The answer is completely correct.	820
-1	Discarded because of insufficient time. In some cases, when participants had few seconds left in a particular stage of the session, the system opened a new question with insufficient time to handle it. In such cases the corresponding search logs, if any, were discarded.	32
-2	Discarded because of cheating. Answers obtained primarily from banned sites after participants intentionally or accidentally found ways to access them. In such cases, the corresponding search logs, if any, were discarded.	39

capture facial expressions, EDA, and eye movements were synchronized at the beginning of each session with the respective computers assigned to participant in each session.

Segmented data were later analyzed with specialized tools. Specifically, videos with the face of participants were processed with three different tools to extract features from participants' faces and perform classifications into groups of basic emotions. This process was performed following the method described by González-Ibáñez, Shah, and Córdova-Rubio (2011). EDA data was analyzed with an online service developed by *Affectiva* to extract peaks as well as other features from data collected with the *Q Sensors*. More details about the results of both analyses are provided below in the section devoted to the second research question.

Next, the set of evaluation measures introduced in section 3.2 were computed for each individual participant and team in each question as part of the dependant variables of interest in this study. From this set of measures, it was found that the 30-seconds threshold specified in the literature as an implicit measure of usefulness did not apply to the characteristics of the task assigned to participants, in which a time limit of five minutes was fixed for each question. In order to specify an implicit measure of usefulness in the context of the task of this study, the average dwell time of relevant pages (i.e. pages from where participants collected snippets - as instructed in the task description - that helped them to find the answers to the questions) minus one standard deviation (SD=15) was used as a threshold to distinguish useful from non-useful pages. This new value rounded to 15 seconds was computed excluding outliers comprised of pages in which participants spent more than 70 seconds, which corresponded to 15.93% of the relevant pages. A density plot with the distributions of relevant pages, non-relevant pages, and pages with dwell-time above the 15-seconds threshold (referred to as useful pages) is depicted in Figure 5.3. The number of relevant pages differed significantly from non-relevant pages according to the usefulness criteria, χ^2 (1, N=5416)=407.76, p<.01, thus verifying the new threshold as an implicit measure of usefulness. The contingency table used to compute the χ^2 statistic is presented in Table 5.4.

After completing the cleaning procedures of *Coagmento Collaboratory* data, segmentation, processing facial expressions and EDA data, and the computation of performance measures, all data sources were combined to facilitate integrated analyses such as determining what emotions searchers typically express while visiting relevant or useful pages.

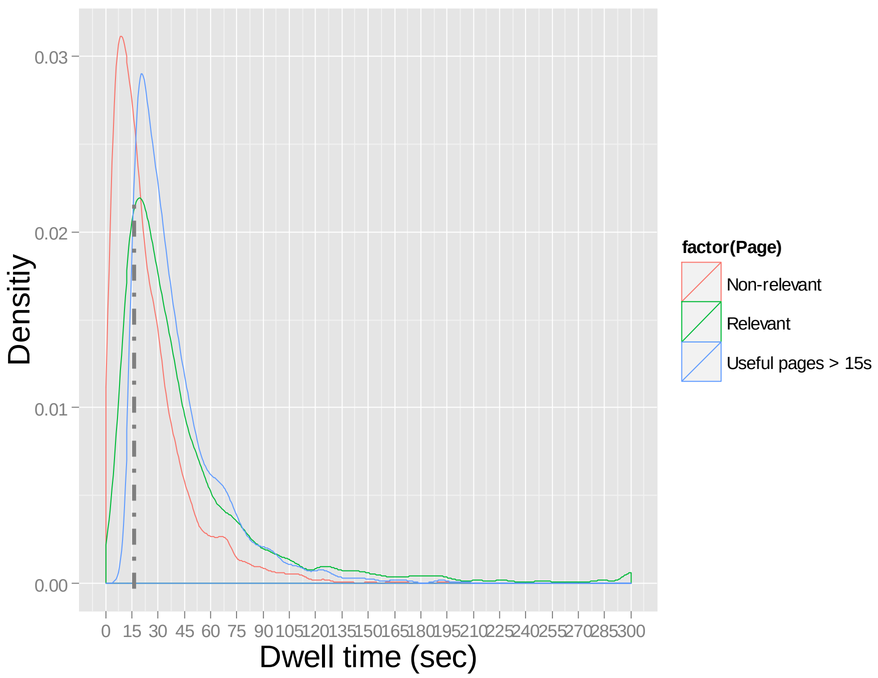


Figure 5.3. Density plot for web pages (non-relevant, relevant, useful) dwell time.

Table 5.4. Contingency table with classified instances according to the usefulness threshold of 15 seconds.

		Explicit measure		
		Relevant	**Non-relevant**	**Total**
Implicit measure	**Non-useful** (dwell time <15 Seconds)	364	1353	**1717**
	Useful pages (dwell time>=15 seconds)	1857	1842	**3699**
	Total	**2221**	**3195**	**5416**

5.2.2 Evaluation

After computing each of the measures as dependent variables, statistical assumptions were evaluated in order to determine whether parametric or non-parametric tests should be applied to perform comparisons within and between groups. Assumptions were evaluated in each question in the evaluation of short-term effects and prolonged effects. All analyses were performed in R[46] and some validations in SPSS[47].

5.2.2.1 Evaluation of short-term effects.

Analyses in this stage were conducted with participants organized into three groups or stimuli-based conditions: positive, negative, control group, this in accordance to the experimental design introduced in section 4.1. Overall, both the positive ($Stim^{+}$) and negative ($Stim^{-}$) groups consist of 60 participants each during the first three questions. On the other hand, the control group ($Stim^{control}$) consists of 15 participants in the same three questions. Unbalanced groups

[46] http://www.r-project.org/

[47] http://www-01.ibm.com/software/analytics/spss/

violate one of the assumptions to run parametric tests such as ANOVA in its standard form.

The distribution of participants per question in each stimuli-based condition is depicted in Figure 5.4. Under this grouping approach, the number of participants fell 13.33% in the fourth question (Stim_q_4) and more than 50% in the fifth one (Stim_q_5), yet the number of samples in $Stim^+$ (n=30) and $Stim^-$ (n=28) was found to be appropriate to perform statistical comparisons between these two conditions. From the sixth (Stim_q_6) to the ninth (Stim_q_9) question, the number of samples was reduced to less than 20%, thus statistical comparisons were not conducted with the measures computed in these three questions due to small sample sizes in each group.

Normality assumption was first visually inspected with Q-Q plots (Appendix D) and evaluated with descriptive measures, specifically kurtosis and skewness. Moreover, the Shapiro-Wilk test was used to evaluate normality in each group,

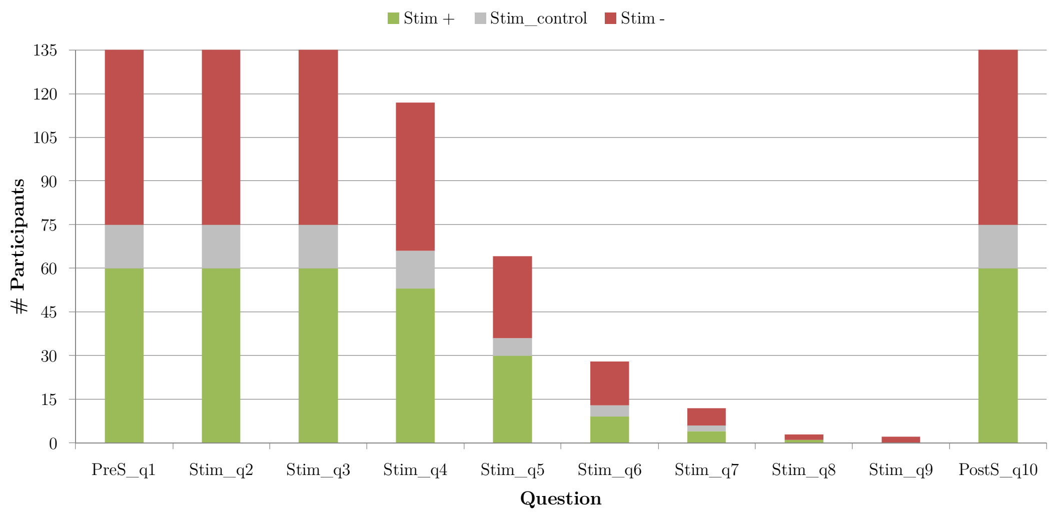


Figure 5.4. Distribution of participants per question in terms of stimuli-based conditions. (PreS) Pre-stimuli question, (Stim) Stimuli, and (PostS) Post-stimuli question.

question, and dependent variable of interest (measures according to the evaluation framework introduced in section 3.2). Results from this test indicated that the distribution of the data was significantly non-normal ($p<.01$) for most variables and data groups according to the three stimuli-based conditions introduced above.

In addition to normality assumption, homogeneity of variance was evaluated. Since most variables were found to be significantly non-normally distributed, Brown-Forsythe's test was used to evaluate the homogeneity of variance for each variable. Results indicated that variance is homogeneous across the three groups ($p>.05$) in most of the questions and for most of the measures. In particular, results of Brow-Forsythe's test were used to determine whether non-parametric test such as the Kruskal-Wallis test and Wilcoxon rank-sum test could be used for between-group comparisons.

5.2.2.2 Evaluation of prolonged effects.

In the second evaluation stage, data were organized into to the six experimental conditions or groups depicted in Figure 4.1 (i.e. $C1^{++}$, $C2^{+-}$, $C3^{--}$, $C4^{+}$, $C5^{-}$, and $C6^{control}$). The distribution of individual participants and pairs per question in each experimental condition is depicted in Figure 5.5. Overall, conditions with pairs of users denoting the interaction of affective states (i.e. $C1^{++}$, $C2^{+-}$, and $C3^{--}$) comprise of 30 participants each during the first five questions of the main task (MT). On the other hand, individual conditions $C4^{+}$, $C5^{-}$, and $C6^{control}$ are comprise of 15 participants in the same five questions. Note that from the experimental design, samples are unbalanced when considering participants in the collaborative conditions as individual units. However, when pairs are treated as units, sample sizes in each group are identical (n=15).

As depicted in Figure 5.5, the number of samples progressively decayed starting in the sixth question of the main task (MT_q_6). In this question, the sample size decreased 9.62%, however, the number of samples in each group was found to be adequate to perform between-group comparisons. Only after the eight question (MT_q_8), the overall sample size was reduced more than 46% with less than 10 individual samples per group.

Normality and homogeneity of variance were analyzed following the same procedures used in the first stage (short-term evaluation). Normality assumption was first visually inspected with Q-Q plots and evaluated with descriptive measures, specifically kurtosis and skewness. Moreover, the Shapiro-Wilk test was used to evaluate normality in each group, question, and dependent variable of interest (measures according to the evaluation framework introduced in

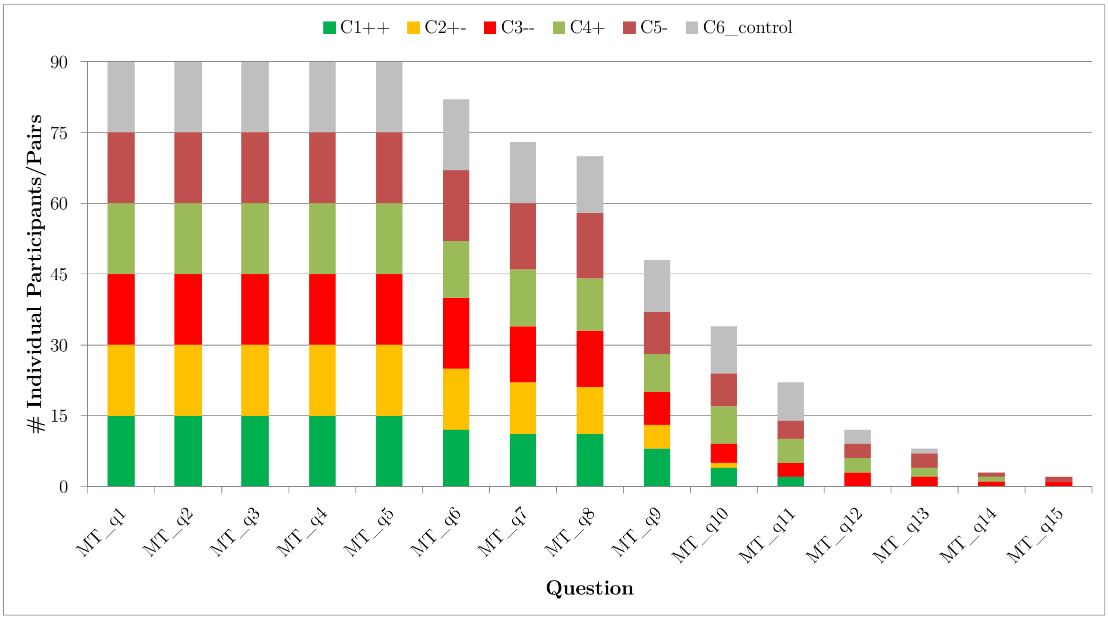


Figure 5.5. Distribution of participants per question in terms of experimental conditions. (MT) Main task.

section 3.2). Results from this test indicated that the distribution of the data was significantly non-normal (p<.01) for most variables and data groups according to the six conditions introduced above.

On the other hand, homogeneity of variance was evaluated with the Brown-Forsythe's test. Results indicate that variance is homogeneous across the three groups (p>.05) in most of the questions and for most of the measures.

5.3 Research question 1: Affective influences in CIS

This section presents the analyses performed to address the first research question (RQ1) of this dissertation, which focuses on effects of initial affective processes and their interactions in searchers' collaborative practices. In particular, this research question was introduced in Chapter 1 as follows:

> **RQ1**: Do initial affective states and their interactions shape the way team members collaborate when searching information, and if so, how?

To address this research question, participants' affective states were manipulated with the aim of establishing initial conditions. Specifically, participants were treated with affective stimuli designed to elicit either positive or negative affective states before the participants were assigned with the search tasks planned for this study. From the literature on emotion elicitation (section 2.1.3) and pilot study results (section 3.3.1), it is implied that stimuli were effective in eliciting the intended affective states in each participant. It is important to note that this approach does not take into account participants' subjective experiences of internal affective processes (i.e. feelings) derived from the exposure to stimuli. However, as discussed in the following research question (RQ2), the affective tone that was intended to be induced with affective stimuli coincided with the valence of self-reported feelings.

The evaluation of interactions of affective processes in pairs during their search processes was approached with respect to the experimental conditions of the study design depicted in Figure 4.1. According to this design, after participants were individually treated with affective stimuli, they were paired with whom they signed up for the study to accomplish a common search task (prolonged effect evaluation). In addition to pairs, this stage also involved three baseline groups. The first one consists of individual participants who were not treated with affective stimuli during the short-term evaluation ($C6^{control}$). The other two baseline groups involve individual participants who received positive ($C4^{+}$) and negative ($C5^{-}$) affective stimuli respectively. These baselines provide a way to contrast individual information behaviors with collaborative ones.

The stage to evaluate prolonged effects was especially designed to address this first research question, which focuses on possible causalities or influences of initial affective processes and their interactions in collaborative practices around information search. Such collaborative practices may comprise multiple activities such as information sharing, communication, group strategy formation, specific information search behaviors expressed in terms of aspects such as information coverage and query formulation, and individual actions.

The following two sections provide detailed results of the analyses conducted to explore (1) short-term effects of affective states in the information search process and (2) prolonged effects of initial affective states and their interactions in the information search process of teams.

5.3.1 Evaluation of short-term effects

After solving a short practice question (q_0), participants were instructed to answer the following question within five minutes:

> *PreS_q_1: El Dorado is near a famous hill that's not on Earth. What year was it photographed?*

This first question was intended to evaluate searchers in different dimensions before being exposed to affective stimuli. Specifically, participants in the three stimuli-based groups (i.e. $Stim^+$, $Stim^-$, and $Stim^{control}$) were evaluated in terms of task-perception measures (i.e. topic familiarity, confidence in answers, and perceived difficulty) and the information retrieval (IR) measures introduced in the evaluation framework presented in section 3.2. As reported in the data exploration section (section 5.2), normality assumption was violated for most measures and groups, however, variance was found to be homogeneous across groups. Therefore, the Kruskal-Wallis test was used to perform statistical comparisons between groups. Overall, results from this test indicated that no significant differences ($p<.05$) across the three conditions were found in any of the IR measures nor in task perception questionnaires. This shows that participants initiated the short-term evaluation stage in roughly similar conditions.

During the following 10 minutes, 120 participants in two groups (i.e. $Stim^+$ and $Stim^-$) were individually exposed to affective stimuli to elicit either positive or negative affective states while addressing a set of *A Google a Day* questions. Note that the 15 participants in the control group ($Stim^{control}$) did not receive stimuli. After the stimuli stage, all participants were instructed to answer the following question within five minutes:

> *PostS_*q_{10}*: Wild salmon are normally born in freshwater and migrate to the sea. In America, there is a variety of salmon that never goes to sea. What was the name of this salmon subspecies, before they got stuck inland?*

This question was intended to investigate the effects of induced initial affective states in individual information search. Note that while the topics of *PreS_q1* and *PostS_*q_{10} were different, the two questions as well as all the other questions in the study were rated with the same level of difficulty (see Section 4.3). Between-group comparisons were conducted in terms of the same set of measures that were used in PreS_q_1. Unlike pre-stimuli evaluation, the Kruskal-Wallis test showed significant differences ($p<.05$) for the average time that participants spent in content pages. Table 5.5 provides descriptive measures for each group and the corresponding test statistic.

Table 5.5. Descriptive measures for average dwell time in content pages and comparison across conditions with Kruskal-Wallis (K-W) test in question PostS_q_{10}.

		Stimuli-based groups		
		Stim^{+}	**Stim^{-}**	**Stim^{control}**
Average dwell time in content pages	**Min**	0	0	0
	Max	141.33	193.43	61.26
	M	40.07	42.27	25.67
	SD	25.27	28.68	15.27
	Mdn	35.88	37.07	23.38
	K-W	$\chi^2(2)=6.20$, $p=0.04$		

Post-hoc comparisons were performed with the Wilcoxon rank-sum test. Results indicate that *average dwell time in content pages* was significantly higher (p<.01) in Stim$^+$ (Mdn=35.88s) and Stim$^-$ (Mdn=37.06s) than that found in the control group (Mdn=23.38s). In addition, the same measure was found to be significantly higher (p<.01) in Stim$^-$ than in Stim$^+$. In other words, Stim^{control} < Stim$^+$< Stim$^-$. Effect sizes in each comparison were found to be small according to Cohen's convention. Wilcoxon rank-sum test results for each pair of groups are summarized in Table 5.6.

Table 5.6. Wilcoxon rank-sum tests and effect sizes. (↑) Significantly higher and (↓) significantly lower.

		Wilcoxon rank-sum test	**Effect size**
Average dwell time in content pages	Stim$^+$(↓) vs Stim$^-$ (↑)	W=1,3685, p<.01	r=0.05
	Stim$^+$ (↑) vs Stim^{control} (↓)	W=5,325, p<.01	r=0.06
	Stim$^-$ (↑) vs Stim^{control}(↓)	W=5,180, p<.01	r=0.06

Within-group analyses as a result of the pretest-posttest experimental design were carried out with the Friedman's test and also the Wilcoxon signed-rank test as non-parametric versions of repeated measures ANOVA. The evaluation focuses on measuring changes between the pre-stimuli evaluation ($PreS\_q_1$) and post-stimuli evaluation ($PostS\_q_{10}$) as a result of the exposure to affective stimuli.

Results from Friedman's test revealed significant changes in participants that were treated with both positive and negative affective stimuli. However, participants in the control group did not show significant variations. Table 5.7 shows results for the Friedman's test and post-hoc analyses for measurement of

Table 5.7. Friedman's test results for significant changes at $p<.05$ from PreS_q$_1$ to PostS_q$_{10}$. (↑) and (↓) denote significant increment and decrement respectively in PostS_q$_{10}$ as a result of the operation (PostS_q$_{10}$ − PreS_q$_1$).

		Stim$^+$		**Stim$^-$**	
		Friedman rank sum test	**Measurement of change** (PostS_q$_{10}$ − PreS_q$_1$)	**Friedman rank sum test**	**Measurement of change** (PostS_q$_{10}$ − PreS_q$_1$)
IR measures and user actions	**Coverage**	χ^2 (1)=13.75, $p<.01$	-1.50 (↓)	χ^2 (1)=11.79, $p<.01$	-1.5 (↓)
	Useful relevant coverage	χ^2 (1)=11.76, $p<.01$	1.00 (↑)		
	Unique useful relevant coverage	χ^2 (1)=4.45, $p<.05$	0.12 (↑)		
	Relevant coverage	χ^2 (1)=5.77, $p<.05$	0.45 (↑)		
	SERPs	χ^2 (1)=13.25, $p<.01$	-2.50 (↓)	χ^2 (1)=9.98, $p<.01$	-3.00 (↓)
	Efficiency	χ^2 (1)=15.52, $p<.01$	0.12 (↑)	χ^2 (1)=14.75, $p<.01$	0.19 (↑)
	Effectiveness	χ^2 (1)=12.00, $p<.01$	0.40 (↑)	χ^2 (1)=11.79, $p<.01$	0.39 (↑)
	Precision	χ^2 (1)=19.10, $p<.01$	0.25 (↑)	χ^2 (1)=7.04, $p<.01$	0.33 (↑)
	Avg. dwell time SERPs (s)	χ^2 (1)=5.40, $p<.05$	2.49 (↑)	χ^2 (1)=9.60, $p<.01$	3.22 (↑)
	Avg. dwell time content pages (s)	χ^2 (1)=9.60, $p<.01$	12.40 (↑)	χ^2 (1)=20.76, $p<.01$	13.11 (↑)
	Snip action			χ^2 (1)=5.49, $p<.05$	0.35 (↑)
	# Queries sequence	χ^2 (1)=10.29, $p<.01$	-2.00 (↓)	χ^2 (1)=9.98, $p<.01$	-2.50 (↓)
	# Distinct queries	χ^2 (1)=13.00, $p<.01$	-2.00 (↓)	χ^2 (1)=11.79, $p<.01$	-2.00 (↓)
	Avg. time to snip			χ^2 (1)=15.08, $p<.01$	14.29 (↑)
Task questionnaires	**Familiarity**	χ^2 (1)=7.12, $p<.01$	0.27 (↑)		
	Pre-perceived difficulty	χ^2 (1)=8.53, $p<.01$	1.00 (↑)	χ^2 (1)=12.74, $p<.01$	1.00 (↑)
	Post-perceived difficulty	χ^2 (1)=11.26, $p<.01$	1.00 (↑)		
	Response confidence	χ^2 (1)=8.33, $p<.01$	-1.00 (↓)	χ^2 (1)=6.72, $p<.01$	-0.50 (↓)

change. The latter is depicted as significant increasing (↑) or decreasing (↓) scores in the differences between post-stimuli (PostS_q$_{10}$) and pre-stimuli (PreS_q$_1$) evaluations with respect to each measure.

The majority of significant changes at $p<.05$ were found in participants who were treated with positive affective stimuli (Stim$^+$). However, particular changes were found to be consistent for participants in both groups (Stim$^+$ and Stim$^-$). For example, significant increments were noted for precision, efficiency, effectiveness, and perception of the challenge (i.e. an indicator of the perceived level of difficulty of questions from the perspective of users before engaging in search processes to find their answers). A raise in precision indicates that participants were able to find more relevant information with respect to their overall information coverage in PostS_q_{10} than in PreS_q_1 (p<.05). Likewise, a raise in efficiency means that participants found more useful pages using less queries in PostS_q_{10} than in PreS_q_1 (p<.05).

This result is consistent with the fact that the number of queries were going down, which was found to be significantly lower in PostS_q_{10} than in PreS_q_1 (p<.05). This shows that participants were able to find an answer (not necessarily the right one) in PostS_q_{10} using less queries and less sources than in PreS_q_1. Similarly, an increasing perception in the level of challenge (Pre-perceived difficulty) indicates that participants who were treated with positive and negative stimuli found PostS_q_{10} to be more challenging than PreS_q_1 even though both questions were rated with the same level of difficulty (section 4.3).

Participants in $Stim^+$ and $Stim^-$ also showed decreasing scores in terms of overall coverage (i.e. the total number of content pages visited while addressing *A Google a Day* questions), exploration of Google search result pages (SERPs), number of sequential queries (i.e. a series of queries including both newly formulated and reused queries), number of distinct queries, and the level of confidence at the moment of providing the answers to both questions. The latter change indicates that participants who were treated with affective stimuli felt less confident of their responses in PostS_q_{10} than what they felt after answering PreS_q_1.

Some significant changes with regard to some measures were observed in either $Stim^+$ or $Stim^-$. In particular, significant changes were observed in $Stim^+$ with regard to relevant coverage, useful relevant coverage, unique useful relevant coverage, topic familiarity, and perceived difficulty at the moment of providing an answer. On the other hand, participants in $Stim^-$ showed significant changes in terms of the number of snippets collected and also with respect to the average time to save the first snippet in each relevant page.

Results from Friedman's test were validated with the Wilcoxon signed-rank test. Table 5.8 provides results from this test, including the test statistic, level of significance, effect size, and measurement of change depicted with the same notation used to represent Friedmans' test results. Results from the Wilcoxon signed-rank test were found to be consistent in most cases with those reported by the Friedmans' test. Specifically, both test showed significant variations for $Stim^+$ and $Stim^-$ in information coverage, SERPs, efficiency, effectiveness, precision, average dwell time

in SERPs and content pages, number of queries, average time to snip, topic familiarity, pre-perceived difficulty, post-perceived difficulty, and response confidence. Wilcoxon signed-rank test differed from Friedmans' test with respect to useful coverage, unique useful coverage, recall, F-measure, and snip action. Additionally, this test informed that participants in the control group ($Stim^{control}$) explored less SERPs in PostS_q_{10} (Mdn=4) than in PreS_q_1 (Mdn=7), p=0.046, r=-.12, however this was not reported by the Friedman's test. Effect sizes for the Wilcoxon signed-rank tests results ranged from small to medium according to Cohen's convention.

Table 5.8. Wilcoxon signed-rank test results for significant changes at $p<.05$ from PreS_q$_{1}$ to PostS_q$_{10}$. (↑) and (↓) denote significant increment and decrement respectively in PostS_q$_{10}$ as a result of the operation (PostS_q$_{10}$ – PreS_q$_{1}$).

		Stim^{+}		Stim^{-}	
		Wilcoxon signed-rank test	**Measurement of change** (PostS_q$_{10}$ - PreS_q$_{1}$)	**Wilcoxon signed-rank test**	**Measurement of change** (PostS_q$_{10}$ - PreS_q$_{1}$)
IR measures and user actions	**Coverage**	$W=1148.5$, $p<.01$, $r=-0.23$	-1.50 (↓)	$W=1191$, $p<.01$, $r=-0.26$	-1.50 (↓)
	Unique useful coverage			$W=48.5$, $p<.05$, $r=-0.13$	0.25 (↑)
	Useful relevant coverage	$W=243$, $p<.05$, $r=-0.12$	0.28 (↑)	$W=113.5$, $p<.01$, $r=-0.21$	1.00 (↑)
	Relevant coverage	$W=204$, $p<.01$, $r=-0.17$	0.45 (↑)		
	Unique Relevant Coverage	$W=51$, $p<.05$, $r=-0.13$	0.27 (↑)		
	SERPs	$W=1317$, $p<.01$, $r=-0.28$	-2.50 (↓)	$W=1194.5$, $p<.01$, $r=-0.26$	-3.00 (↓)
	Efficiency	$W=262.5$, $p<.01$, $r=-0.28$	0.13 (↑)	$W=353.5$, $p<.01$, $r=-0.23$	0.19 (↑)
	Effectiveness	$W=273.5$, $p<.01$, $r=-0.2$	0.40 (↑)	$W=290.5$, $p<.01$, $r=-0.23$	0.39 (↑)
	Precision	$W=211.5$, $p<.01$, $r=-0.3$	0.25 (↑)	$W=269.5$, $p<.01$, $r=-0.18$	0.33 (↑)
	Recall			$W=932$, $p<.05$, $r=-0.14$	-0.01 (↓)
	F-measure			$W=924$, $p<.05$, $r=-0.13$	-0.02 (↓)
	Avg. dwell time SERPs	$W=565$, $p<.01$, $r=-0.16$	2.49 (↑)	$W=506$, $p<.01$, $r=-0.18$	3.22 (↑)
	Avg. dwell time content pages	W=385, p<.01, r=-0.24	12.40 (↑)	$W=304$, $p<.01$, $r=-0.27$	13.11 (↑)
	Snip action	$W=326.5$, $p<.05$, $r=-0.12$	0.45 (↑)	$W=258$, $p<.05$, $r=-0.14$	0.35 (↑)
	Avg. time to snip			$W=256$, $p<.01$, $r=-0.24$	14.29 (↑)
	# Queries sequence	$W=1302$, $p<.01$, $r=-0.25$	-2.00 (↓)	$W=1166$, $p<.01$, $r=-0.24$	-2.50 (↓)
	# Distinct queries	$W=1145.5$, $p<.01$, $r=-0.25$	-2.00 (↓)	$W=1169.5$, $p<.01$, $r=-0.25$	-2.00 (↓)
Task questionnaires	**Familiarity**	$W=25$, $p<.05$, $r=-0.15$	0.27 (↑)		
	Pre-perceived difficulty	$W=226$, $p<.05$, $r=-0.14$	1.00 (↑)	$W=167.5$, $p<.01$, $r=-0.19$	1.00 (↑)
	Post-perceived difficulty	$W=264$, $p<.01$, $r=-0.2$	1.00 (↑)		
	Response confidence	$W=818.5$, $p<.05$, $r=-0.15$	-1.00 (↓)	$W=680$, $p<.01$, $r=-0.16$	-0.50 (↓)

Both between-group and repeated measures analyses within subjects showed that affective stimuli had effects on participants' search behaviors. In particular, between-group analyses showed that in PreS_q$_{1}$ participants in the three groups started in roughly similar conditions. However, in PostS_q$_{10}$ participants who received affective stimuli spent on average more time in content pages than those in the control group. On the other hand, repeated measures analyses in each group showed that affective treatments had significant effects expressed in terms of different IR measures, users' actions, and task questionnaires. More importantly, no significant effects were found for participants in the control group with the exception of borderline significant differences in the number of SERPs as reported by the Wilcoxon signed-rank test.

5.3.2 Evaluation of prolonged effects

Observed changes in the post-stimuli evaluation ($PostS\_q_{10}$) constitute the initial conditions of participants at the moment of starting the evaluation of prolonged-effects. Taking into account the collaborative component of this stage, pairs and individual participants were distributed into six groups defined in terms of the type of affective stimuli to which participants were individually exposed during the short-term evaluation. These six groups (i.e. $C1^{++}$, $C2^{+-}$, $C3^{--}$, $C4^{+}$, $C5^{-}$, and $C6^{control}$) constitute the experimental conditions of the study intended to investigate the effects, if any, of initial positive and negative affective states and their interactions in collaborative practices performed around information search.

In the analyses presented in this dissertation, collaborative practices were studied in terms of the following four factors: information related measures and users' actions, information sharing, communication processes, and participants' perception about the *A Google a Day* questions being solved. First, information related measures consist of the same set of IR measures used in the evaluation of short-term effects and specific participants' actions related to the exploration and collection of information (e.g. snipping text from webpages). Second, information sharing comprises of the specific action of viewing each other's snippets and sending messages that contain information collected from pages. Third, communication processes focuses on specific categories of messages that denote how participants collaborated during the task. Finally, participants' perception of the *A Google a Day* questions presented during the main task includes: level of familiarity with the topics of the questions and the perceived challenge when questions are presented for the first time to participants. Additionally, participants' perception also includes post-question evaluations in order to determine how difficult they found questions after finding an answer or running out of time, and how confident they felt with their answers.

The approach to address the evaluation of prolonged effects followed two levels of comparisons conducted in each question. As depicted in Figure 5.6, the evaluation plan consisted of (1) comparisons across the six experimental

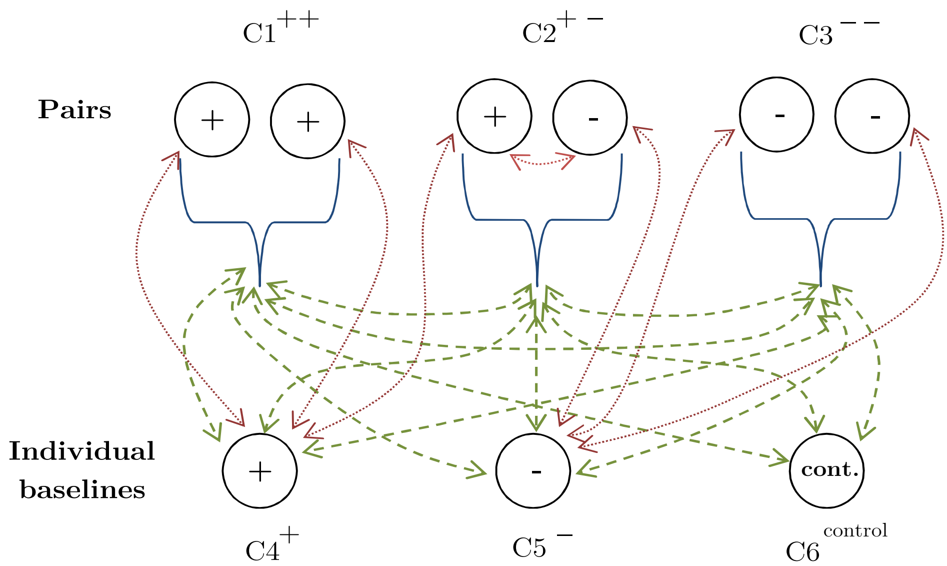


Figure 5.6. Comparison schema for prolonged effect evaluation.

conditions with pairs as unit of analysis (segmented connectors), and (2) individual members in each pair were compared with those in the baseline conditions (dotted connectors). Unlike the evaluation of short-term effects, in this case analyses were performed per question. Note that this evaluation approach was also used to address the other three research questions.

As reported in the previous section, data exploration showed that normality assumption was violated for most measures, groups, and questions. Therefore, comparisons were performed with non-parametric tests. Moreover, as indicated in the data exploration section, due to a reduction in the number of observations in later stages of the experimental sessions (i.e. number of participants in each question, which was reduced as they ran out of time), statistical tests were performed only when 10 or more observations were available in each group. Following, results from each analysis in accordance to the four factors introduced above and the comparison approach depicted in Figure 5.6 are presented.

5.3.2.1 Information-related measures and users' actions

Information related measures consist of the same set of measures used in the short-term stage (i.e. IR measures and specific user actions related to the exploration and collection of information). Analyses were first conducted with pairs as individual units by combining individual measures obtained from each team member with those of their teammates.

Between-group comparisons were conducted with the Kruskal-Wallis test. In addition, post-hoc analyses were performed with the Wilcoxon rank-sum test. Due to the extension of the analyses and comparisons performed in this research question as well as in the following ones, test results are summarized in tables with cells denoting post-hoc comparisons. For additional details, some descriptive measures are provided in Appendix E.

Results for this first analysis are summarized in Table 5.9. The first column contains the set of measures where the Kruskal-Wallis test reported significant results at $p<.05$. On the other hand, the first row corresponds to the questions where significant differences were observed. Due to a progressive reduction in the number of observations after the sixth question of the main task, statistical comparisons were performed between MT_q_1 and MT_q_8. In these questions each group (condition) had 10 or more samples.

Not surprisingly, when comparing the aggregated scores of pairs with those of participants in the baseline conditions, test results consistently showed that the former achieved significant higher values in measures such as number of queries and exploration of SERPs. Since the questions of the main task require at least two queries to find their answers (according to A Google a Day ideal search path presented in section 4.3), the more queries pairs or individual participants formulated to solve questions, the lower was their performance. More discussion in this regard is provided in section 5.5, which focused on the third research question. Conversely, individual participants in the baseline conditions were found to be more efficient than pairs. The latter result in terms of the proportion of effective access to useful pages with respect to the number of queries used to find such pages (Section 3.2.2).

Table 5.9. Summary of information related measures. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with pairs as unit of analysis. Results in cells are significant at $p<.01$ for different IR measures. Due to space restrictions in the table, conditions codes are presented as follow: $1=C1^{++}$, $2=C2^{+-}$, $3=C3^{--}$, $4=C4^{+}$, $5=C5^{-}$, and $6=C6^{control}$.

	Question							
Measures	**MT_q_1**	**MT_q_2**	**MT_q_3**	**MT_q_4**	**MT_q_5**	**MT_q_6**	**MT_q_7**	**MT_q_8**
Coverage		1>2,3,4,5,6 2>3,4,5,6 3>4,5,6 5>4,6 6>4	1>2,3,4,5,6 2>3,4,5,6 3>4,5,6 4>5 6>4,5					1>2,3,4,6 2>6 3>2,4,6
Relevant coverage			1>2,3 5>4					
Useful Coverage		1>4,5,6 2>1,3,4,5,6 3>1,4,5,6 5>6	1>2,3,4,5,6 2>3,4,5,6 3>4,5,6			2>1,3,4 3>1		1>2,3,6
Useful relevant coverage			2>4 3>4 5>4 6>4					
# queries sequence	1>4,5,6 2>1,3,4,5,6 3>1,4,5,6 5>4 6>4,5	1>4,5,6 2>1,4,5,6 3>1,2,4,5,6 5>6	1>2,3,4,5,6 2>4,5,6 3>2,4,5,6 6>4,5	1>3,4,5,6 2>1,3,4,5,6 3>4,5,6 4>5 6>4,5	1>3,4,5,6 2>1,3,4,5,6 3>4,5,6 5>4,6 6>4		1>2,3,4,5,6 2>4,5,6 3>2,4,5,6 4>5,6 6>5	1>2,3,4,6 2>3,4,6 3>4,6 6>4
# unique queries	1>4,5,6 2>1,3,4,5,6 3>1,4,5,6 5>4 6>4,5	1>4,5,6 2>1,4,5,6 3>1,2,4,5,6 6>5	1>2,3,4,5,6 2>4,5,6 3>2,4,5,6 4>5 6>4,5	1>2,3,4,5,6 2>3,4,5,6 3>4,5,6 4>5,6 6>5	1>3,4,5,6 2>1,3,4,5,6 3>4,5,6 5>4,6 6>4	1>4,5,6 2>1,3,4,5,6 3>1,4,5,6 5>4 6>4,5	1>2,3,4,5,6 2>4,5,6 3>2,4,5,6 4>5,6 6>5	1>2,3,4,6 2>3,4,6 3>4,6 6>4
Team query diversity		1>2 3>1,2		2>1,3 3>1			1>2 3>1,2	
SERPs	1>4,5,6 2>1,3,4,5,6 3>1,4,5,6 5>4 6>4,5	1>4,5,6 2>1,4,5,6 3>1,2,4,5,6 5>6	1>2,3,4,5,6 2>4,5,6 3>4,5,6 4>5 6>4,5	1>3,4,5,6 2>1,3,4,5,6 3>4,5,6 4>5 6>4,5	1>3,4,5,6 2>1,3,4,5,6 3>4,5,6 5>4,6 6>4	1>3,4,5,6 2>1,3,4,5,6 3>4,5,6 5>4 6>4,5		1>2,3,4,6 2>3,4,6 3>4,6 6>4
Recall			1>2,3,4,5,6 2>3,4,5,6 3>4,5,6 5>4 6>4,5					
Precision								1>3,4 2>1,3,4,6 3>4 5>1,2,3,4,6 6>1,3,4
Effectiveness								2>1,3,4 3>1,4 4>1 5>1,2,3,4,6 6>1,2,3,4
Efficiency	1>2,3 2>3 4>1,2,3,5,6 5>1,2,3,6 6>1,2,3	1>3 2>1,3 4>1,2,3,5,6 5>1,2,3 6>1,2,3,5	2>1,3 3>1 4>1,2,3,6 5>1,2,3,4,6 6>1,2,3		1>2,6 2>6 3>1,2,6 4>1,2,3,5,6 5>1,2,3,6			2>1,3 3>1 4>1,2,3 5>1,2,3,4,6 6>1,2,3,4
Snip		1>2 4>6 5>4,6	1>2,3,5,6 2>3,5,6 3>5,6	1>2,3,5 2>3,5				

Results also show that pairs did not necessarily cover more pages than individuals. As shown in Table 5.9, coverage was significantly higher at $p<.01$ for pairs than for individuals in three questions: MT_q_2, MT_q_3, and MT_q_8. A closer look to collaborative pairs indicates that those in $C1^{++}$ covered more pages than those in $C2^{+-}$ and $C3^{--}$ while working in these three questions. Moreover, pairs in $C2^{+-}$ covered more pages than those in $C3^{--}$ in MT_q_2 and MT_q_3.

Despite the objective level of difficulty assigned to all questions (section 4.3), observational research and interviews showed that MT_q_2 and MT_q_3 generated interesting discussions around their topics. For example, in MT_q_2 ("During a famous White House séance, witnesses say the president's seat was levitated. Whose spirit was his wife trying to contact?") team members generally tackled the question individually. In first place they investigated to which president this question was referring to. Following, participants exchanged their individual findings; however, discrepancies with respect to the name of the president led them to pursue additional searches and further discussions before moving to the next search step. After reaching consensus, team members focused on finding the name of the spirit that was being contacted. The spirit was one of President Lincoln's sons, but since more than one son passed away by the time Ms. Lincoln was participating in the séance, team members had to explore additional webpages to discuss the answer.

During the interview stage of the study, it was rather common that participants indicated that the hardest time they had in the session was figuring out where to start to find the answer. For example, participant S473 said "in some questions I was not sure where to start ... it was hard to pick the right answer because there is lot of different sources I could rely on." Similarly, when asked about what was the most difficult part during the session, S651 stated "Umm just knowing what to google."

In MT_q_3 ("In Norse mythology, which god's son will poison Thor at Ragnarök?") participants found possible answers very quickly, however, disagreement within pairs were observed due to different interpretations of this question. Although all participants in the study were English native speakers, in general they interpreted MT_q_3 in different ways. For example, in the interview stage some participants reported that they thought that the question was about the name of the father (Loki), whereas other participants indicated that the question was oriented in finding the name of the son (Jörmungandr). This situation led participants in pairs to explore more pages in order to reach consensus. Conversely, participants in the baseline conditions addressed this question without having to match their own interpretation with someone else's.

A distinctive difference found in this question is that pairs in $C1^{++}$ not only explored more pages than those in the other two collaborative conditions, but also they covered more relevant and useful pages. Nevertheless, as discussed in the results for the third research question, despite these significant differences, not all pairs in $C1^{++}$ responded MT_q_3 correctly.

Immediate effects of the interactions of affective states can be observed in the first question (MT_q_1: "I am an animal that can grow more than 20,000 teeth in my lifetime. Which of my species, extinct or living, has the largest teeth?") of the evaluation of prolonged effects. Table 5.9 shows that pairs in $C2^{+-}$ used significantly more queries (*Mdn*=6) to find an answer to this question than pairs in the other two collaborative conditions. Consistently, pairs in $C2^{+-}$ explored significantly more SERPs (*Mdn*=6) at p<.01 than pairs in $C1^{++}$ (*Mdn*=5) and $C3^{--}$ (*Mdn*=5). On the other hand, pairs in $C1^{++}$ used significantly less queries (*Mdn*=5) at p<.01 than pairs in $C2^{+-}$ (*Mdn*=6) and $C3^{--}$ (*Mdn*=5). At the same time, efficiency to discover useful pages was significantly higher at p<.01 for pairs in $C1^{++}$ (*Mdn*=.17) than for pairs in $C2^{+-}$ (*Mdn*=0.13) and $C3^{--}$ (*Mdn*=0.12). Details for performance in this question are provided in section 5.5.

In addition to investigate interactions of affective states with pairs as units of analyses, participants were also treated as individual units to perform comparisons across experimental conditions. Since participants in $C2^{+-}$ started the main task in opposite affective states, this condition was split into two groups: $C2^{+}$ and $C2^{-}$. As presented in Table 5.10, significant differences at p<.05 according to the Kruskal-Wallis test were scattered in different questions. For example, differences observed in MT_q_1 are not consistent with those found in the following questions. Additionally, differences that were found with pairs as unit of analyses (e.g. number of queries, recall, and efficiency, among others) are not present as significant differences in the comparisons with individual participants.

Analyses in MT_q_1 show that the coverage of pages that were not only useful and relevant, but also unique (i.e. only one participants was able to access such pages) was significantly higher at p<.01 for participants in $C2^{+}$ (*Mdn*=0) and $C4^{+}$ (*Mdn*=0) than for those in the other conditions. Conversely, it is observed that the coverage of such pages was significantly lower for individual searchers in the baseline condition $C5^{-}$ (*Mdn*=0) than for participants in the other

Table 5.10. Summary of information related measures. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with individuals as unit of analysis. Results in cells are significant at $p<.01$ for different IR measures. Due to space restrictions in the tables, conditions codes are presented as follow: $1=C1^{++}$, $2^+=C2^+$, $2^-=C2^-$, $3=C3^{--}$, $4=C4^+$, $5=C5^-$, $6=C6^{control}$.

	Question			
Measure	**MT_q_1**	**MT_q_2**	**MT_q_3**	**MT_q_4**
Coverage				$1>2^+,2^-,3,6$ $2^+>2^-,3$ $4>1,2^+,2^-,3,5,6$ $5>1,2^+,2^-,6$ $6>2^-$
Unique coverage				$1>2^-$ $2^+>1,2^-$ $3>1,2^-$ $4>1,2^+,2^-,3,5,6$ $5>1,2^+,2^-,3,6$ $6>1,2^+,2^-,3$
Unique useful relevant coverage	$1>5$ $2^+>1,2^-,3,5,6$ $2^->5$ $3>5$ $4>1,2^-,3,5,6$ $6>1,2^-,3,5$			
Avg. dwell time SERPs			$1>5,6$ $2^+>1,5,6$ $2^->1,2^+,3,4,5,6$ $3>1,2^+,4,5,6$ $4>1,2^+,5,6$ $5>6$	
Average time to snip in relevant pages		$1>2^+,3$ $2^+>3$ $2^->1,2^+,6$ $4>1,2^+,2^-,3,6$ $5>1,2^+,2^-,3,4,6$ $6>1,2^+$		$1>2^+,2^-,3,4,5,6$ $2^+>2–,3,4,5,6$ $2^->3,6$ $4>2^-,3,6$ $5>2^-,3,4,6$ $6>3$

	Question			
Measure	**MT_q_5**	**MT_q_6**	**MT_q_7**	**MT_q_8**
Useful Coverage				$1>3$ $2^+>3$ $2^->2^+,3$ $5>3$
Unique useful relevant coverage	$1>2^+$ $2^->1,2^+,3$ $3>2^+$ $4>1,2^+,2^-,3,6$ $5>1,2^+,2^-,3,6$ $6>1,2^+,3$			
User Query diversity			$1>2^+,2^-,3,4,6$ $2^+>2^-,4$ $2^->3,4$ $3>4$	
Snip		$1>2^-,3$ $2^+>1,2^-,3,4$ $3>2^-$ $4>2^-,3$ $6>1,2^+,2^-,3,4$		
Average time to snip in relevant pages		$1>3$ $2^+>3,6$ $4>1,2^+,2^-,3,5,6$ $5>1,2^+,2^-,3,6$		

experimental conditions. In fact, nobody in $C5^{-}$ visited this kind of pages in the same question. A closer look to the data shows that participants who started in negative affective states (i.e. $C2^{-}$, $C3^{--}$, and $C5^{-}$) were less likely to visit this kind of pages. Specifically, only two participants in $C3^{-}$ (*Mdn*=0) and one in $C2^{-}$ (*Mdn*=0) accessed one of such pages while working in MT_q_1. Note that while the reported medians for this measure are equal to zero, significant differences found for $C2^{+}$ and $C4^{+}$ are attributed to five and four participants respectively that were able to access one or two of these pages.

Although participants in $C1^{++}$ also started MT_q_1 in positive affective states, their coverage of pages that were unique, relevant, and also useful, was significantly lower ($p<.01$) than that of participants in $C2^{+}$ and $C4^{+}$. Perhaps working with someone in a congruent positive affective state led team members to reinforce each other's overconfidence, which was derived from their individual positive experiences in the evaluation of short-term effects.

On the other hand, the fact that participants in $C2^{+}$ incurred in extra efforts to compensate that of their partners who started MT_q_1 in negative affective states ($C2^{-}$), might have contributed to the exploration of pages with distinctive characteristics (i.e. useful, relevant, and unique). As discussed in the evaluation of communication processes, this difference in effort was in part expressed through higher volume of communication coming from participants in $C2^{+}$ than that of participants in the other collaborative conditions while working on the first questions. With regard to participants in the baseline condition $C4^{+}$, even though they worked individually, they behaved similarly to those in $C2^{+}$ in terms of information coverage. Nevertheless, unlike participants in collaborative conditions, participants in $C4^{+}$ were not subject to behavioral influences of collaborators.

Significant differences found in MT_q_1 for the control condition ($C6^{control}$) serve as a threshold to distinguish effects of initial affective states and their interactions in the coverage of a specific group of pages. In fact, participants in the control group were able to cover more pages that were relevant, useful, and also unique than participants who were previously treated with negative affective stimuli (i.e. those in $C2^{-}$, $C3^{-}$, and $C5^{-}$). Conversely, the coverage of such pages was significantly lower for participants in $C6^{control}$ compared to that of participants in $C2^{+}$ and $C4^{+}$.

As discussed later in section 5.5, despite the significant differences found in the first question of the evaluation of prolonged effects, participants who achieved higher or lower coverage of pages that were unique, relevant, and also useful, did not necessarily find the right answer to MT_q_1.

In MT_q_3 the only observed difference was in terms of the average time that participants spent in SERPs. Results show that participants in the control group ($C6^{control}$) spent significantly less time in SERPs (*Mdn*=7.06 seconds) at $p<.01$ than those in the other conditions. On the other hand, participants in $C2^{-}$ remained significantly more time in SERPs (*Mdn*=21.35 seconds) at $p<.01$ than participants in all the other conditions. Similarly, participants in $C3^{--}$, who were also treated with negative stimuli, spent more time in SERPs (*Mdn*=15.37 seconds) at $p<.01$ than participants in the other conditions, except $C2^{-}$. Despite this significant results observed in collaborative conditions, the opposite was observed in individual conditions. In particular, the time spent in SERPS was significantly higher in $C4^+$ (*Mdn*=12.95 seconds) at $p<.01$ than that found in the collaborative conditions in which the participants were previously treated with positive stimuli (i.e. $C1^{++}$ and $C2^+$). On the other hand, dwell time in SERPS in $C5^-$ (*Mdn*=10.12 seconds) was only significantly higher at $p<.01$ than that found in the control group.

The major number of significant differences between individuals were found in MT_q_4 ("An organic compound created a color so fashionable that it inspired a nickname for the decade of the 1890s. What is the name of that organic compound?"). First of all, it was found that information coverage in $C2^-$ was significantly lower (*Mdn*=2) at $p<.01$ than that found in all the other conditions. Second, the coverage of unique pages in $C2^-$ (*Mdn*=0) was significantly lower at $p<.01$ than that achieved in the other conditions. On the other hand, results for $C4^+$ showed that information coverage (*Mdn*=3) and unique coverage (*Mdn*=1) were significantly greater at $p<.05$ than that found in the other conditions. Similar results were also observed in $C5^-$. As discussed later in section 5.5, information coverage was not related with finding the right answer for MT_q_4.

Results for MT_q_5 show that the coverage of pages that were not only useful and relevant, but also unique, was significantly lower at $p<.01$ in $C2^+$ than that found in other conditions. In fact, nobody in $C2^+$ visited this kind of pages. Participants in other conditions visited one or two of such pages. These results differ from those observed in MT_q_1, where participants in $C2^+$ actually visited more pages with these characteristics.

Another interesting finding as a result of comparative analyses with individuals was found in MT_q_2, MT_q_4, and MT_q_6. It was observed that it took participants in C3^{--} significantly less time to collect the first snippet from pages they found to be relevant to answer these questions (*Mdn*=8.54, 8.86, and 6.18 seconds respectively). However, this behavioral aspect was not observed in participants in the corresponding baseline condition (C5^{-}), who actually required significantly more time to collect the first snippet in the pages they visited to respond these questions (*Mdn*=38.17, 14.78, and 17.82 seconds respectively).

Contrary to what was described above for participants in C1^{++} while working in MT_q_1, this observed difference between participants in C3^{--} and those in the respective baseline (C5^{-}) is attributed to reinforced feelings of frustration and rush in pairs where both participants were initially exposed to negative affective stimuli. This interpretation derives from the interview stage in which participants who were treated with negative affective stimuli reported feelings of frustration and rush. Moreover, it was observed in C3^{--} sessions that participants, in general, worked almost independently with little to no communication. At some point during the sessions, participants in this condition engaged in sort of within-pair competitions. That is to say, whoever found the answers to questions, submitted them without coordinating this action with their teammates. This situation in some cases provoked frustration and irritation in team members who were not able to give their opinion before answers were submitted. For example, while working in MT_q_4, participant S596 screamed and pounded the computer desk when her teammate responded the question without waiting for her. In the interview stage, when asked about this particular moment at the same time that a video of her search session was being displayed to help her to relive the situation, she said "I did not know if the *aniline dye* was actually the answer, so I did not want to answer before I was sure it was the answer. So I was frustrated because I was not positive [about the answer]," on the other hand, her teammate (participant S597), said "I asked her opinion on it [the answer], she never got back to me, so well I couldn't do more searching, so I guessed I'll go with the best thing we have."

Beyond the scope of qualitative observations, when looking at participants' responses to the NASA TLX (Hart & Staveland, 1988) questionnaire, in particular to the question "How hurried or rushed was the pace of the task?" no significant differences were found between groups. Another aspect to consider is that participants started working on MT_q_4 on average 12.60 minutes (*SD*=4.17) after completing the evaluation of short-term effects (i.e. right after responding $PostS\_q_{10}$). At this point, it was expected that participants' affective states changed as a result of the interactions with participants in similar or different affective states or due to a number of internal factors such as tiredness and boredom. As discussed later in section 5.4, participants' affective states - from the perspective of participants (i.e. self-reported feelings) - were found to converge progressively to neutrality as they completed questions.

Analyses in MT_q_6 ("I was a 19th century lawyer who claimed to have killed 42 men. I was killed while playing dice by a man who was killed over a card game. Who am I?"), in addition to show differences with regard to the average time that participants spent in SERPs, it also reported significant differences with respect to the number of snippets that participants collected from pages that they found to be relevant. In this case, the number of snippets was significantly lower at $p<.01$ for participants in $C2^-$ (*Mdn*=1) than in any other condition. In this case, only seven participants contributed each with one snippet. On the contrary, $C6^{control}$ was the group that concentrated the majority of the snippets collected (*Mdn*=1) for MT_q_6. In this condition, nine out of 15 participants collected between one and four snippets. Note that participants started working on MT_q_6 on average 18.27 minutes (*SD*=4.71) after completing the evaluation of short-term effects. Therefore, they had less than seven minutes (on average) to answer as much questions as they could. Perhaps, this situation might have led participants to prioritize their time usage by focusing on finding and submitting answers in the shortest time possible rather than collecting snippets.

Further analyses within each condition in the context of MT_q_6 showed that participants in $C6^{control}$ and also in $C4^{+}$ started this question on average 15.50 minutes (*SD*=4.14) from the beginning of the evaluation of prolonged effects. Therefore, their remaining time to respond other questions was on average close to 10 minutes, which left them with more time to collect snippets. However, regardless of time constraints, participants in $C2^{+}$ collected more snippet than participants in the other collaborative conditions and also than those in their respective baseline ($C4^{+}$). This is attributed to additional efforts from participants in $C2^{+}$ to compensate the reduced number of snippets collected by their teammates in $C2^{-}$.

Question MT_q_7 ("The painter of Starry Night doused his mattress and pillow with what to help him sleep?") was initiated on average 19.90 minutes (*SD*=4.85) after the completion of the evaluation of short-term effects. Analyses in this question show significant differences only with respect to query diversity within participants. This measure computed with the Levenshtein edit distance, represents how different were the queries that participants formulated in order to find relevant information to respond questions. As shown in Table 5.10, participants who formulated the major spectrum of queries were those in C1^{++}(*Mdn*=22). Moreover, as discussed later in section 5.5, participants in C1^{++} also achieved one of the highest levels of query precision (a measure that compared participant's queries with those provided by *A Google a Day* as part of the ideal search path to respond each question) in this question.

The last question in which between-group analyses were conducted is MT_q$_8$. Results for this question show significant differences only in terms of useful coverage. As illustrated in Table 5.10, useful coverage was significantly lower at $p<.01$ for participants in C3^{--} (*Mdn*=1) than that of participants in the other collaborative conditions and also than those in their respective baseline (C5^{-}). Usefulness, as explained in section 3.2, is an implicit measure based on the time people spent in webpages (in section 5.2 the usefulness threshold for this study was set in 15 seconds). Since participants started working on MT_q$_8$ on average 21.24 minutes (SD=4.97) after completing the evaluation of short-term effects, their time for other questions was reduced to less than four minutes. Furthermore, as discussed above, observations and interviews showed that feelings of rush and frustration were reinforced when two participants initiated the evaluation of prolonged effects in negative affective states. Perhaps, the combination of these two factors (i.e. time constraints and negative feelings) led participants in C3^{--} to spend less than 15 seconds on the webpages they visited.

Analyses of the collaboration process of searchers are presented in the following sections. In particular, information sharing behaviors, communication processes, and task perception are investigated to further understand and interpret the observed differences across experimental conditions presented above.

5.3.2.2 Information sharing

Collaborative information seeking provides searchers with unique capabilities that are not possible in individual settings. For example, collaboration enable searchers to exchange information with others working toward common goals, provide and receive feedback on each other's' information findings, and share queries to tackle the information problem being addressed. These and other similar actions are executed relying on different levels of awareness, which may or may not be supported by systems.

Since the interest in this section is on information sharing, analyses presented here were performed only with data collected from pairs in $C1^{++}$, $C2^{+-}$, and $C3^{--}$. In the context of this study, information sharing was evaluated following two different approaches. The first one focused on an explicit action of searchers for viewing snippets collected by their teammates. The second one was based on the exchange of information through the chat system enabled during the study. This section presents results for the first approach. Results for the second one are presented in the following section, which focuses on communication processes.

The system used to support CIS in this study (i.e. *Coagmento Collaboratory*) provided two resources for information sharing: the snip button and chat system. In particular, the snip button enabled participants to save fragments of text that were found to be relevant to respond the questions of the study. After saving snippets, these were made available as short links on the sidebar for both the participant who collected them and his/her teammate. To view the content of a snippet, participants had to click on the corresponding link provided on the sidebar. This action opened a popup window with the snippet and information about it (e.g. when it was collected, who collected it, and from where it was collected). Overall, participants in all the conditions performed this action on average 1.90 times (*SD*=1.25) per question. This shows that participants, in general, used this feature of *Coagmento Collaboratory* at least one time during the task to share information with others and to obtain information from others.

Figure 5.7 provides a detailed view of the average number of times that the "view other's snippet" action was performed in each question with individuals as unit of analysis in each condition. In this chart it is possible to observe that during the first two questions, participants in C3¯¯ viewed each other's snippets, on average, more times than participants in other conditions. To some extent,

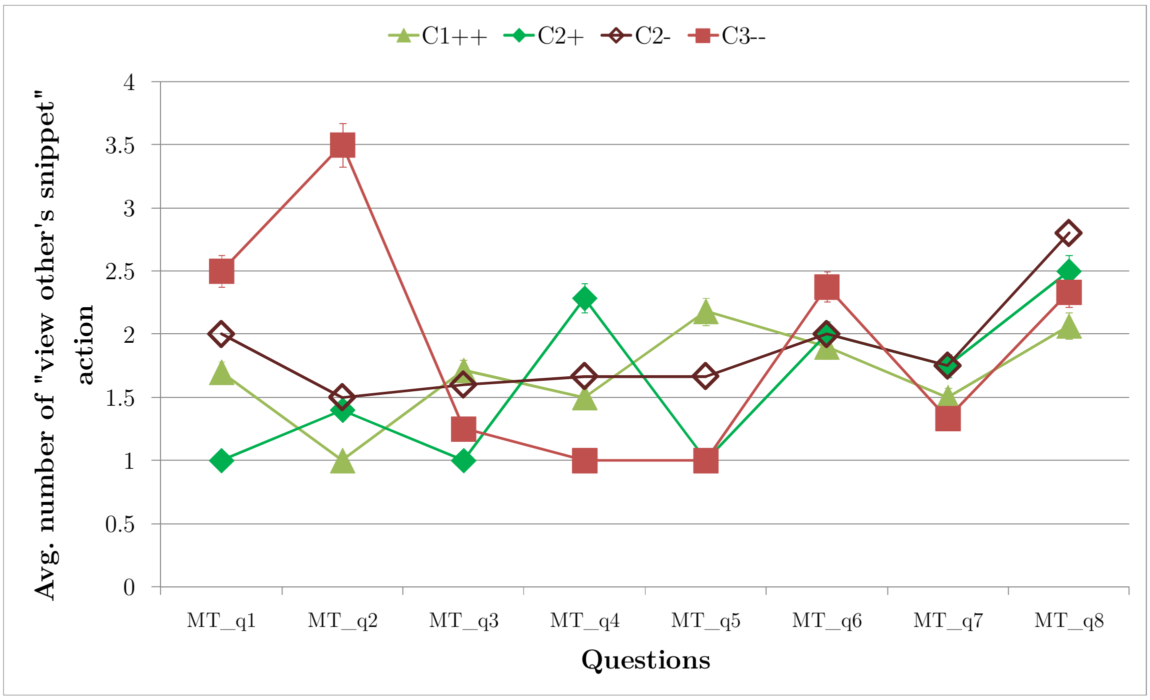


Figure 5.7. Average number of times participants viewed each other's snippets in each question.

this explains the lower communication volume in $C3^{--}$, especially in question MT_q_1, which is discussed in more detail in the following section. Despite the independent working style of participants in $C3^{--}$, as described in the previous section, they still used each other's snippets as a resource for awareness. However, as illustrated with the situation experienced by participant S596 and S597 in $C3^{--}$, this awareness resource was not sufficient to satisfy coordination needs within the team, which later resulted in feelings of frustration and irritation in S596.

A similar situation is observed for participants in $C2^{-}$ who, on average, viewed the snippets collected by their teammates in $C2^{+}$ more times, with the exception of MT_q_4. On the contrary, participants in $C1^{++}$ used less frequently each other's snippets in the first two questions. Perhaps this was compensated with a higher communication volume in those questions, which is discussed in more detail in the following section. Regardless the differences depicted in Figure 5.7; comparisons across conditions with the Kruskall-Wallis test did not show significant differences at $p<.05$ in questions MT_q_1 to MT_q_8.

In addition to studying information sharing behaviors within pairs, the individual use of snippets collected while searching for answers (i.e. participants viewing their own snippets) was also investigated. Unlike the analyses presented above, participants in collaborative conditions were also contrasted with those in the baseline conditions, who also had the possibility to save snippets and access them later. On average, participants viewed their own snippets 1.76 times (*SD*=1.41) per question. Figure 5.8 shows the average number of times participants performed this action in each question. It is observed during the

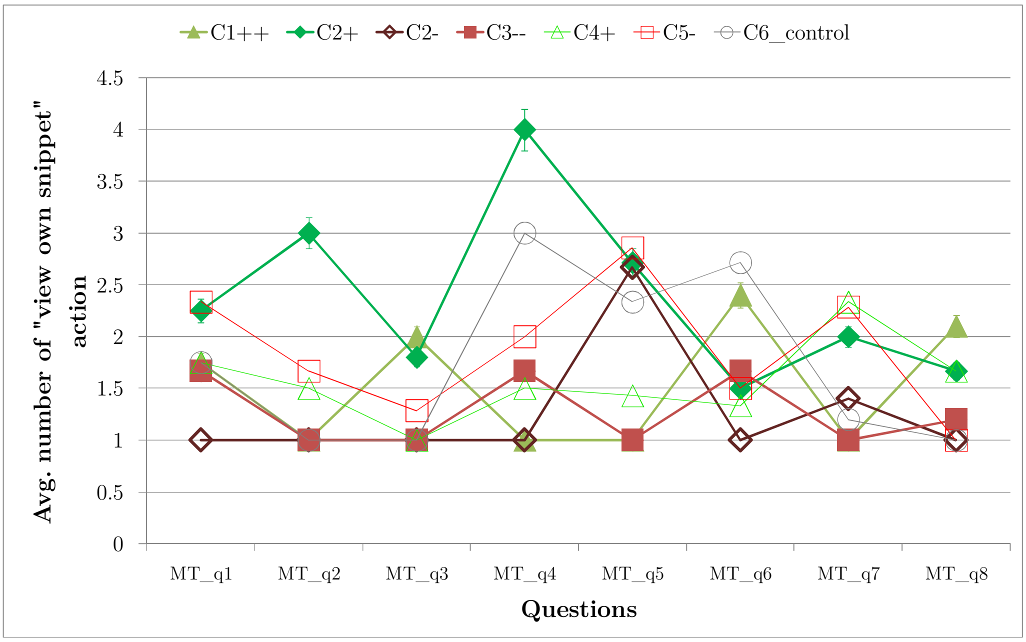


Figure 5.8. Average number of times participants viewed their own snippets in each question.

first questions that participants in C2⁺ accessed the snippets collected by themselves, on average, in more opportunities than those in other conditions. This situation is attributed to the fact that participants in C2⁻ were less active in the collection of snippets during the first questions, thus most of the snippets were collected by their teammates in C2⁺. Despite differences depicted in Figure 5.8, these were not found to be significantly different at $p<.05$.

The following section focuses on communication processes, providing additional details about information sharing as well as other information-related behaviors.

5.3.2.3 Communication processes

A distinctive aspect in CIS is the ability of participants to communicate with each other. Communication as discussed in section 2.2 is one corner stone of CIS, which enables participants to coordinate, contribute, and cooperate as part of the collaborative process. In this study communication was analyzed in order to extract different features that can be related to information search and affective dimension. Overall, 2489 messages were produced by participants in collaborative conditions (i.e. $C1^{++}$, $C2^{+-}$, and $C3^{--}$). Each message was analyzed following two different approaches. First, chat messages were classified by two independent coders in accordance to an extended version of the coding scheme introduced by González-Ibáñez, Haseki, and Shah (2012a, 2012b). The second approach to analyze communication processes was based on the use of natural language processing techniques. In particular, messages were analyzed in terms of the types of words used by participants in the messages exchanged.

As indicated above, the first approach to analyze communication involved the use of a coding scheme to classify messages in accordance to predefined categories established in a coding book (see Appendix H). Two independent coders were trained to classify all messages. This process was conducted in three iterations. In the first iteration coders coded 20% of the messages. Following, intercoder reliability was performed. Due to low score achieved in the first iteration, corrective actions and improvements in the coding book were implemented. In the second iteration, the two coders were instructed to recode the original 20% plus an additional 10% of the messages. After completing this iteration, intercoder reliability was found to be $Kappa$=.71, p<.01. Finally, in the third iteration, the coding process of the remaining 70% was performed by one of the coders.

The original coding scheme used by González-Ibáñez, Haseki, and Shah (2012a, 2012b) comprises of general categories of messages, specifically: task coordination

(TC), task content (TN), task social (TS), non-task related (NT), and non-codable (NC). The extended version of the coding scheme implemented for this study incorporates complementary codes aiming to capture additional aspects found in messages such as: collaborative support, information sharing, control, conflict resolution, strategy definition, awareness, search process, uncertainty, success in finding the answers, and affective tone. Unlike the general coding category, which consists of five possible codes, all the complementary coding categories were binary. For example codes denoting collaborative support, control, strategy definition, awareness, search process, uncertainly, and success consisted of two possible values: presence (1) and absence (0). Others such as conflict resolution were expressed as agreement (A) or disagreement (D).

Analyses were conducted in first place with pairs as units of analyses. Communication was found to be prominent between questions MT_q_1 and MT_q_5. As participants were more restricted in time in later questions, communication had a negative trend. Figure 5.9 depicts the average

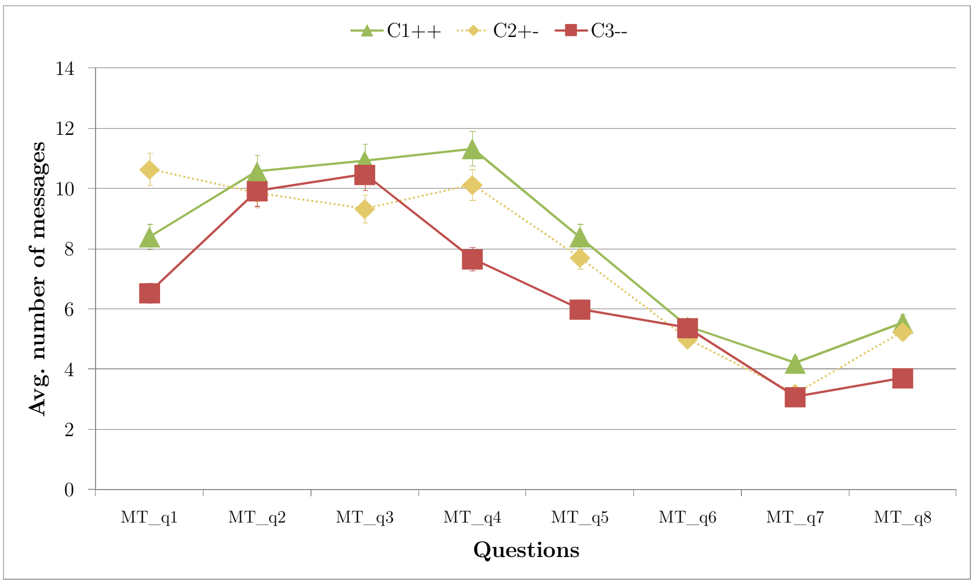


Figure 5.9. Communication volume of pairs in each question.

communication volume (number of messages) per condition in each question. According to this figure, pairs in conditions in which at least one participant was previously treated with positive stimuli (i.e. $C1^{++}$ and $C2^{+-}$) exchanged more messages during the evaluation of prolonged effects with the exception of question MT_q_3. Despite these differences in terms of averages, statistical analyses did not show significant variations at $p<.05$ with respect to the overall communication volume of pairs in each question.

Comparisons across conditions with regard to particular communication processes according to the coding categories introduced above were conducted with the Kruskall-Wallis test. Post-hoc analyses for significant results at $p<.05$ were performed with the Wilcoxon rank-sum test. Between-group comparisons were carried out with individual codes aggregated per pair in each question. Results are summarized in Table 5.11, which shows significant differences at

Table 5.11. Summary of communication analyses. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with pairs as unit of analysis. Results in cells are significant at $p<.01$ for different categories of communication processes. Numbers 1, 2, and 3 in cells, represent conditions $C1^{++}$, $C2^{+-}$, and $C3^{--}$ respectively.

	Question		
	MT_q_1	**MT_q_4**	**MT_q_5**
Task content	$C1^{++}>C3^{--}$ $C2^{+-}>C1^{++},C3^{--}$	$C1^{++}>C3^{--}$ $C2^{+-}>C1^{++},C3^{--}$	
Non-task related			$C1^{++}> C2^{+-},C3^{--}$
Search process		$C2^{+-}>C1^{++},C3^{--}$	
Disagreement		$C2^{+-}>C1^{++},C3^{--}$	
Questions		$C1^{++}>C3^{--}$, $C2^{+-}$	
Answers	$C1^{++}>C3^{--}$ $C2^{+-}>C3^{--}$		

$p<.01$ in six codes that characterize the type of communication in which participants engaged. In particular, differences were found with respect to communication about task content, non-task related, search process, disagreements, questions, and answers.

As indicated in the previous analyses, an immediate evaluation of prolonged effects of initial affective states and their interactions was performed by looking at observed differences in the first question (MT_q_1). Results for this question show significant differences at $p<.01$ with respect to the number of messages in which the participants discussed the content of the questions (TN) (i.e. discovering the name of animal that can grow more than 20,000 teeth its lifetime). This was higher in $C2^{+-}$ (*Mdn*=7) than in $C1^{++}$ (*Mdn*=6). In turn, the volume of TN messages was also higher in C1++ than in $C3^{--}$ (*Mdn*=2). Table 5.11 also shows differences in the number of answers provided as part of question-answering processes that took place within pairs. Note that no differences were found with respect to the number of questions. This is due to the fact that while participants in the three conditions sent questions to their teammates, not every question received an answer. In this case, the number of answers was significantly lower at $p<.01$ for participants in $C3^{--}$ (Mdn=0) than in $C1^{++}$ (*Mdn*=1) and $C2^{+-}$ (*Mdn*=1). Differences in the number of answers in each condition implicitly indicate the level of collaboration within pairs in which participants could expect support from their partners.

As shown in Table 5.11, after MT_q_1 no significant differences are observed until MT_q_4. In the fourth question, which focused on finding the name of an organic compound very popular in the 1890s, the communication volume of TC messages was also significantly higher at $p<.01$ in $C2^{+-}$ (*Mdn*=8) than in $C1^{++}$ (*Mdn*=7) and $C3^{--}$ (*Mdn*=5). This kind of messages was also significantly more frequent in $C1^{++}$ than in and $C3^{--}$. Additionally, communication volume was significantly higher at $p<.01$ in $C2^{+-}$ than in the other two conditions when it came to discuss while searching information. This included messages about search results, findings, and queries being used. Disagreements within pairs were also found to be higher in $C2^{+-}$ than in the other two conditions. The medians for the three conditions in these two measures was found to be zero. Finally, the number of questions exchanged in $C1^{++}$ (*Mdn*=3) was higher than in $C2^{+-}$ (*Mdn*=2), but no differences were observed in the number of answers.

In the following question (MT_q_5), results show that non-task related (NT) communication was significantly higher at $p<.01$ in $C1^{++}$ than in the other two conditions. NT messages were characterized by participants talking about topics outside the context of the study. For example, "i love you" and "lol love u too." As shown in the preliminary studies discussed in section 3.3, NT communication was found to be rather unusual in similar laboratory studies where a text-chat was used to support communication. In fact, 0.84% of the messages exchanged in the entire study were non-task related. An inspection of the data showed that the difference reported above was attributed to only four participants in $C1^{++}$ who exchanged NT messages.

Analyses with respect to other communication measures according to the evaluation framework introduced in section 3.2 such as communication effort, balance in the exchange of messages, and balance in communication effort, did not report significant differences.

In addition to investigating differences among pairs in each condition, analyses were conducted with individuals members based on the affective stimuli they received during the previous evaluation stage. Similar to the previous analyses, participants originally in $C2^{+-}$ were reassigned to two individual conditions: $C2^{+}$ and $C2^{-}$. Figure 5.10 depicts the average communication volume (also referred to as number of messages) in each question of individuals in the four groups. The chart indicates that communication volume was on average higher in $C2^{+}$ than in the other conditions in the first two questions than. Before question MT_q_6, the plot shows that in general communication volume was lower for participants who were originally treated with negative stimuli (i.e. $C2^{-}$ and $C3^{--}$).

Unlike the analyses with pairs, the Kruskal-Wallis test showed significant differences at $p<.05$ with respect to the overall communication volume, communication effort, number of messages discussing the questions (TN), socialization around the topic of the question (TS), non-task related

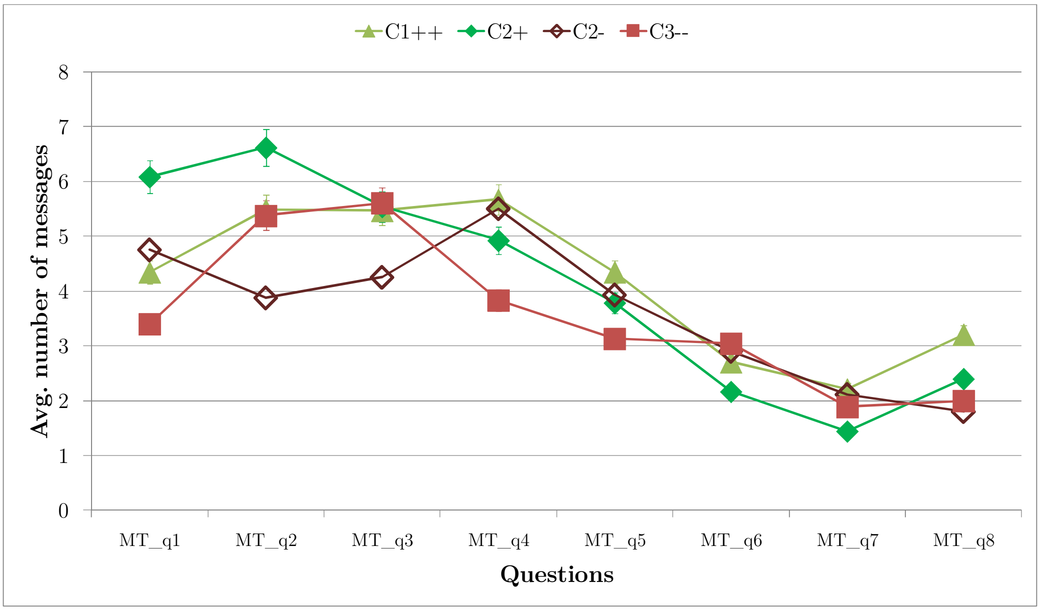


Figure 5.10. Communication volume of individuals in each question.

communication (NT), disagreements, questions, and answers. Table 5.12 summarizes significant differences at $p<.01$ as reported by post-hoc analyses with the Wilcoxon rank-sum test. As illustrated in this table, $C3^{--}$ is in most cases on the right side of the comparisons, which means that measures in this condition were found to be significantly lower than in the other conditions.

For example, in MT_q_1 (i.e. immediate effects of the interactions of positive and negative affective states) results showed that communication volume was significantly lower at $p<.01$ in $C3^{--}$ (*Mdn*=2) than in $C1^{++}$ (*Mdn*=4), $C2^{+}$ (*Mdn*=5) and $C2^{-}$ (*Mdn*=5). In turn, the number of messages in $C2^{+}$ was significantly higher at $p<.01$ than in the other three groups. Additionally, communication effort (i.e. proportion between the length of messages and the average number of words per minute. See section 3.2.2 for more details) was significantly lower in $C3^{--}$ (*Mdn*=45.27) than in $C1^{++}$ (*Mdn*=56.18), $C2^{+}$ (*Mdn*=118.91), and $C2^{-}$ (*Mdn*=92.73). On the other hand, communication effort was found to be significantly higher at $p<.01$ in $C2^{+}$ than in $C1^{++}$, $C2^{-}$, and $C3^{--}$. Note that the average typing speed of participants in the six groups ranged between 45 and 50 words per minutes. Results for this particular analysis are illustrated in Appendix D.2.

As illustrated in Table 5.12, same results were found with respect to TN communication (e.g. “i looked up which animal has 20000 teeth in a lifetime”, “read latest snippet”, and “called Leviathan melvillei”). In other words, the volume of TN messages was significantly lower at $p<.01$ in $C3^{--}$ (*Mdn*=1) than in the other three groups. On the other hand, TN communication was

Table 5.12. Summary of communication analyses. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with individuals as unit of analysis. Results in cells are significant at $p<.01$ for different categories of communication processes.

	Question			
	MT_q₁	**MT_q₄**	**MT_q₅**	**MT_q₈**
Volume	$C1^{++}>C3^{--}$ $C2^{+}>C1^{++},C2^{-},C3^{--}$ $C2^{-}>C1^{++},C3^{--}$			$C1^{++}>C2^{+},C2^{-}$
Effort	$C1^{++}>C3^{--}$ $C2^{+}>C1^{++},C2^{-},C3^{--}$ $C2^{-}>C1^{++},C3^{--}$	$C1^{++}>C2^{+},C3^{--}$ $C2^{+}>C3^{--}$ $C2^{-}>C1^{++},C2^{+}, C3^{--}$	$C1^{++}>C2^{-},C3^{--}$ $C2^{+}>C1^{++},C2^{-},C3^{--}$ $C2^{-}>C3^{--}$	
Task content	$C1^{++}>C3^{--}$ $C2^{+}>C1^{++},C2^{-},C3^{--}$ $C2^{-}>C1^{++},C3^{--}$	$C1^{++}>C3^{--}$ $C2^{+}>C1^{++}, C3^{--}$ $C2^{-}>C1^{++},2^{+}, C3^{--}$	$C1^{++}>C2^{+},C2^{-},C3^{--}$ $C2^{+}>C2^{-}$	$C2^{+}>C3^{--}$ $C2^{-}>C3^{--}$
Task Social	$C1^{++}>C2^{+},C2^{-},C3^{--}$ $C2^{+}>C2^{-}$ $C3^{--}>C2^{+},C2^{-}$			
Non-task related			$C1^{++}>C2^{+},C2^{-},C3^{--}$	
Disagreement		$C2^{+}>C1^{++},C2^{-},C3^{--}$ $C2^{-}>C1^{++},C3^{--}$		
Questions				$C1^{++}>C2^{-},C3^{--}$ $C2^{+}>C1^{++},C2^{-},C3^{--}$ $C2^{-}>C3^{--}$
Answers				$C1^{++}>C2^{+},C2^{-}, C3^{--}$ $C2^{+}>C2^{-},C3^{--}$ $C3^{--}>C2^{-}$

significantly higher at $p<.01$ in $C2^{+}$ (*Mdn*=4) than in $C1^{++}$ (*Mdn*=3), $C2^{-}$ (*Mdn*=3) and $C3^{--}$.

Socialization around the questions (e.g. "i am very confident", "i saw the movie", and "record time"), referred to as TS communication, was significantly higher at $p<.01$ in $C1^{++}$ (*Mdn*=1) than in the other three groups. This type of messages were also more frequent in $C3^{--}$ (*Mdn*=0.5) than in $C2^{+}$ (*Mdn*=0) and $C2^{-}$ (*Mdn*=0). For the particular case of condition $C2^{+-}$, TS communication was significantly higher in $C2^{+}$ than in $C2^{-}$.

No significant differences at $p<.05$ were found in MT_q_2 and MT_q_3. However, in MT_q_4 significant differences with regard to effort, TN communication, and disagreements also place $C3^{--}$ on the right side of all the inequalities. Unlike

MT_q1, in this case communication effort was found to be higher for participants in $C2^{-}$ (*Mdn*=81.82) than in $C1^{++}$ (*Mdn*=68.73), $C2^{+}$ (*Mdn*=56.73), and $C3^{--}$ (*Mdn*=39.82). This measure was also found to be higher in $C1^{++}$ than in $C2^{+}$ and $C3^{--}$.

Table 5.12 also shows that TN communication was significantly higher at $p<.01$ in $C2^{-}$ (*Mdn*=4) than in $C1^{++}$ (*Mdn*=3), $C2^{+}$ (*Mdn*=4), and $C3^{--}$ (*Mdn*=2). Additionally, results show that TN messages were more frequent in $C2^{+}$ than in $C1^{++}$, and the later also higher than $C3^{--}$.

With respect to discrepancies, participants in $C2^{+}$, followed by those in $C2^{-}$, showed significantly more disagreements with their teammates than that found in the other groups. Medians were found to be zero in the four groups for this particular measure.

In MT_q5, results also show that measures for $C3^{--}$ were found to be significantly lower than in the other groups. In particular, communication effort was significantly higher at $p<.01$ in $C2^{+}$ (*Mdn*=75.27) than in $C1^{++}$ (*Mdn*=63.27), $C2^{-}$ (*Mdn*=53.45), and $C3^{--}$ (*Mdn*=22.91). Moreover, effort was higher in $C1^{++}$ than in $C2^{-}$, and in turn communication effort in the latter was higher than in $C3^{--}$.

Contrary to results in MT_q_4, in the fifth question the volume of TN messages was significantly higher at $p<.01$ in $C1^{++}$ (*Mdn*=3) than in $C2^{+}$ (*Mdn*=3), $C2^{-}$ (*Mdn*=2), and $C3^{--}$ (*Mdn*=1). Within $C2^{+-}$, this type of message was more frequent in participants in $C2^{+}$ than for those in $C2^{-}$. Moreover, in accordance with the results presented in Table 5.11, non-task related communication was found to be higher in $C1^{++}$ than in the other three groups. However, as discussed above these messages were isolated cases attributed to only four participants in this condition.

The last group of significant differences was found in MT_q_8. In this question, communication volume was significantly higher at $p<.01$ in $C1^{++}$ (*Mdn*=2) than in $C2^{+}$ (*Mdn*=2) and $C2^{-}$ (*Mdn*=1.5). Volume of TN messages was significantly higher at $p<.01$ in $C2^{+}$ (*Mdn*=1) and $C2^{-}$ (*Mdn*=1) than in $C3^{--}$ (*Mdn*=0). The number of questions was significantly higher in $C2^{+}$ (*Mdn*=0.5) than in the other three groups. In contrast, participants in $C3^{--}$ (*Mdn*=0) issued less questions than those in $C1^{++}$ (*Mdn*=0), $C2^{-}$ (*Mdn*=0), and $C2^{+}$. On the other hand, the number of answers was significantly higher in $C1^{++}$ than in the other groups. Also, answers were more frequently found in $C3^{--}$ than in $C2^{-}$. Note that in this question, as illustrated in Figure 5.10, the average number of messages was quite low compared to that in the first questions. Perhaps this reduction in communication volume is explained in part by time pressure. In fact, pairs initiated MT_q_8 on average 23.73 (*SD*=4.46) after the evaluation of short-term effects. Therefore, with less than two minutes (on average) to complete the task, participants opted to focus on finding the answers to more questions rather than in communicating.

The analyses presented above with pairs and individual team members point to confirm the first hypothesis introduced in the research framework of this dissertation (section 3.4), which states:

> **Hypothesis 1**: The communication volume of pairs in which both members are treated with negative stimuli will be lower than the

communication volume of pairs in which at least one of the members received positive stimuli.

As shown in Table 5.11 and Table 5.12 communication volume for particular categories of messages in MT_q_1 (i.e. evaluation of immediate effects of affective states interactions) was significantly lower at $p<.01$ in $C3^{--}$ than in conditions where at least one team member started the task in a positive affective state. Moreover, an inspection of communication processes of individual team members (Table 5.12) also showed that, in general, communication volume (total and also of specific categories) was significantly higher in $C1^{++}$ and/or in $C2^{+-}$ than in $C3^{--}$. Even though differences were not observed in all the questions, significant results found in later questions were consistent in showing that communication volume, in general, is lower when both participants start the task in a negative affective state. While abundance of communication does not correlate with high performance, lack of communication may affect collaboration sub processes such as coordination and awareness.

As pointed out at the beginning of this section, communication was also investigated with the support of natural language processing techniques. In particular, each message was analyzed in terms of the types of words used by participants in each condition. In order to accomplish this, messages were decomposed into single words and then compared with words in an extended version of the LIWC dictionary (Pennebaker et al., 2001), which also includes categories and words of the WordNet affect (Strapparava & Valitutti, 2004) as well as a list of interjections, punctuation, and emoticons. The evaluation approach was based on the bag-of-feature framework. Therefore, rather than creating a frequency vector with all the words in the dictionary (bag-of-words approach), a simplified vector with the major categories of the dictionary as features was constructed for each message.

Results from this analysis (with pairs and also with individuals as units of analyses) showed significant differences at the level of linguistic processes, which were scattered between questions MT_q_1 and MT_q_8. Linguistic processes in the LIWC dictionary include general categories of words such as pronouns, conjunctions, verbs, and adverbs. For instance, in MT_q_1 the use of the pronoun "I" was significantly higher in C2^{-} (*Mdn*=4) than in C1^{++} (*Mdn*=2), C2^{+} (*Mdn*=3), and C3^{--} (*Mdn*=1). The use of this pronoun indicates self-references in messages like: "i got elephants for largest teeth in the world", "i got his dead son willie", "i am very confident", and "i searched the 1890s nickname." On the other hand, significant differences at $p<.01$ with regard to the use of the pronoun "you" were found in MT_q_5. In this case participants in C1^{++} (*Mdn*=1) were more likely to refer to their partners than participants in the other groups. Examples of such messages are: "if your confident lets answer!", "you wanna answer?", and "how are you doing?" No significant differences were found with respect to the use of the pronoun "we".

Significant differences were also found in terms of the use of affective words (e.g. great, dear, jeez, phew, and wow). In question MT_q_5 these were significantly lower at $p<.01$ in C3^{--} (*Mdn*=0) than in the other conditions. At the level of cognitive processes, it was found that participants in C2^{+} and C2^{-} used significantly more words at $p<.01$ in the discrepancy category than those in the other conditions. Finally, significant differences were also found with respect to words about biological processes, time, space, perception, and personal concerns. A complete list of the significant differences found with this approach to analyze communication is provided in Appendix E.

5.3.2.4 Task perception

The last factor investigated to address this first research question focuses on the perception of participants before and after working in each question. This was measured through short questionnaires presented to participants before searching information to respond the *A Google a Day* questions and at the moment of providing the answers. The first questionnaire focused on topic familiarity and perceived challenge, whereas the second questionnaire focused on perceived difficulty and confidence in the answer provided. Responses to both questionnaires were provided with a 5-point Likert scale. Both questionnaires are provided in Appendix B.4.

Between-group comparisons were conducted with the Kruskal-Wallis test and

Table 5.13. Summary for task perception. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with individuals as unit of analysis. Results in cells are significant at p<.01 for different IR measures.

		Question	
		MT_q_4	**MT_q_5**
Before question	**Perceived challenge**	$C1^{++}>C2^{-},C4^{+},C5^{-}$ $C2^{+}>C1^{++},C2^{-},C4^{+},C5^{-}$ $C2^{-}>C4^{+},C5^{-}$ $C3^{--}>C1^{++},C2^{+},C2^{-},C4^{+},C5^{-}$ $C4^{+}>C5^{-}$ $C6^{control}> C1^{++},C2^{+},C2^{-},C3^{--},C4^{+},C5^{-}$	
After question	**Confidence Answer**		$C2^{+}>C1^{++},C2^{-}, C3^{--},C4^{+},C5^{-}$ $C2^{-}>C1^{++}, C3^{--},C4^{+},C5^{-}$ $C3^{--}>C1^{++},C4^{+},C5^{-}$ $C4^{+}>C1^{++},C5^{-}$ $C5^{-}>C1^{++}$ $6>C1^{++},C2^{+},C2^{-},C3^{--},C4^{+},C5^{-}$

post-hoc analyses were performed with the Wilcoxon rank-sum test. Results for the latter analyses are summarized in Table 5.13. Note that since questionnaires were individually responded, analyses were conducted only with individual participants as units of analysis. The evaluation of immediate effects (MT_q_1) of the interactions of affective states did not show significant differences at $p<.05$. This indicates that regardless the collaborative component introduced in three conditions in this stage, this did not affect the perception of the task in early stages of the evaluation of prolonged effects.

Later, significant differences at $p<.05$ were found for perceived challenge and confidence in the answer in the fourth and fifth questions respectively. For the case of perceived challenge (indicated before engaging in a search process), it was discovered that participants in $C5^{-}$ found the fourth question (MT_q_4) less challenging (*Mdn*=3), $p<.01$. than what participants in the other conditions reported. Conversely, those in the control group ($C6^{control}$) found this question significantly more challenging (*Mdn*=4) at $p<.01$ than what participants in the other conditions indicated.

With respect to confidence in the answer (indicated at the moment of providing answers), it was found that confidence levels in $C1^{++}$ (*Mdn*=2) with respect to the answers provided to MT_q_5 were significantly lower at $p<.01$ than that found in all the other conditions. Conversely, confidence levels in $C6^{control}$ (*Mdn*=4) were significantly higher at $p<.01$ than that reported in the other groups.

It is important to note that no differences were found with respect to topic familiarly, therefore differences in perceived challenge, confidence, and also performance in the response are not attributed to prior knowledge in the topic of MT_q4 and MT_q5. As explained in the development of the following research questions, variations in task perceptions in later stages of the task can be linked to affective variability during the session.

5.4 Research question 2: Affective experience and expression in CIS

This section presents the analyses performed to address the second research question (RQ2) of this dissertation, which focuses on the experience and expressivity of affective processes during the information search process of teams. The research question was introduced in Chapter 1 as follow:

> **RQ2**: What affective processes are typically experienced and expressed (physically, physiologically, and verbally) by team members when collaborating in an information search task?

This question is concerned with possible differences in the expressivity and experience of affective processes in information search as a result of initial affective states and their interactions in collaborative settings. In other words, is it the case that the participants who start a search task in positive affective states express and/or experience affective states triggered during information search in a different way than participants who start in negative affective states? Moreover, how do social factors as a result of collaboration influence expressivity and experience of affective states triggered during information search?

To address this second research question, affective processes in the context of this study were analyzed from four different perspectives: facial expressions, physiological responses, communication, and self-reports.

The following two sections provide descriptive analyses of expressed and experienced affective processes during the evaluation of (1) short-term effects and (2) prolonged effects.

5.4.1 Evaluation of short-term effects

As explained in previous sections, the evaluation of short-term effects of affective processes was performed through a pretest-posttest experimental design. In this experiment participants were assigned to three groups: positive stimuli group ($Stim^{+}$), negative stimuli group ($Stim^{-}$), and control group ($Stim^{control}$). Participants in these groups had to solve a common information search task that consisted in finding the answers to a set of *A Google a Day* questions. During the experimental sessions, both sensors and questionnaires were employed in order to capture different types of affective data. The analyses presented in the following sections focus on facial expressions, electrodermal activity (EDA) data, and self-reported feelings.

5.4.1.1 Affective processes expressed through face

The first approach to analyze affective processes of participants focuses on facial expressions. As explained in section 2.1.3.2, facial expressions are the expressive component of emotions. To analyze participants' faces, a modified version of the procedure described by González-Ibáñez, Shah, and Córdova-Rubio (2011) was used. The method involves the use of two or more software-based solutions to extract features from faces and classify them into a group of basic emotions.

While this type of software typically link facial expressions to specific human emotions (e.g. happiness), the procedure used in this dissertation refers to facial expressions (e.g. smile) as indicators of the presence of particular emotions. It is important to mention that due to technical limitations of current approaches to automatically detect facial expressions, this dissertation focuses primarily on smile, which is typically reported with high accuracy rates in its detection. As noted below, the three tools used in this study are able to detect smiles with an accuracy above 85%.

Similar to González-Ibáñez, Shah, and Córdova-Rubio (2011), three software were used to process videos containing participants' faces. These three software are *FaceDetect* (Kueblbeck and Ernst, 2006; Face and Object Detection Webpage, 2011), *eMotion* (Sebe et al., 2007), and *BMERS* (González-Ibáñez, 2006). First, *FaceDetect* was developed at the Fraunhofer Institute of Integrated Circuits (Kueblbeck and Ernst, 2006; Face and Object Detection Webpage, 2011). This software is capable of detecting facial expressions associated to happiness, anger, sadness, and surprise. According to the authors, the happiness (i.e. smiles) recognition rate on an specific database of faces is close to 95%. Second, *eMotion* was developed at the University of Amsterdam. It became later a commercial solution and today is commercialized by ThirdSight. *eMotion* can detect facial expressions related to six basic emotions (i.e. happiness, anger, disgust, fear, sadness, surprise) and neutrality. According to the authors, the recognition rate of happiness (i.e. smiles) was around 92% on a specific dataset (Sebe et al., 2007). This software was used by Arapakis and Gray (2008) and also by Lopatovska (2009b) in the evaluation of affective processes in the information search process of individuals. Finally, BMERS (Basic Moods and

Emotions Recognition System) was developed by the author of this dissertation to investigate affective expressions in collaborative contexts (González-Ibáñez, 2006). This software is capable of recognizing facial expressions of six basic emotions (i.e. happiness, anger, tenderness, fear, sadness, surprise) and neutrality. The recognition rate of smiles (or happiness) was reported by the author to be higher than 85%.

As described in González-Ibáñez, Shah, and Córdova-Rubio (2011), the three tools were evaluated with a video containing images of people's faces extracted from standard databases. Overall 187 images with smiles and 187 with other facial expressions were collected from four databases: Japanese Female Database (JAFFE) (Lyons et al, 1998), Indian Female Face Database, Indian Male Face Database (Jain and Mukherjee, 2002), and CVL Face DB (Peer, 2011; Solina et al., 2003). Each image was replicated 15 times in order to create one second of facial expression in a video recorded at 15 frames per second (fps). This evaluation procedure reported different recognition rates for smiles than those reported in the corresponding articles. Since the evaluation procedure involved the use of fixed images, it is acknowledged that recognition rates could be affected due to the particular implementations of each tool.

Despite the differences detected in terms of recognition rates, the three tools performed quite good with a recognition threshold of 80%. This threshold indicates that facial expressions recognized with a probability that is equal or higher to 0.8 were considered as valid detections. Although previously authors such as Arapakis and Gray (2008) and also Lopatovska (2009b) have used a threshold of 90%, higher thresholds are useful with prominent smiles. As pointed out by González-Ibáñez, Shah, and Córdova-Rubio (2011) such smiles "are less likely to appear in information seeking settings" (p.6). By setting a threshold over 80%, less prominent forms of smiles, which sometimes last a fraction of a second, are not recognized.

Since the three tools have different characteristics (i.e. detection with rotated faces, and detections with fixed faces) that make them useful in the context of this study, participants' faces were processed with each software independently.

Prior to processing the videos, these were segmented, normalized, and synchronized with other sources of data. First, segmentation was performed by manually inspecting each video in order to locate the exact moment when participants initiated and terminated their sessions. Valid segments were obtained from *Coagmento Collaboratory* logs using the local time saved along with the pages and other actions performed by the participants. Moreover each frame in the video was timestamped with the local time of the corresponding machine, thus facilitating the identification of segments.

Second, synchronization comprised the combination of the different data collected during the study. This includes: *Coagmento Collaboratory* logs, EDA data, and videos with facial expressions. Synchronization was performed with respect to the local time of the computer through which data was collected.

Finally, normalization was conducted by adding a five-seconds segment at the beginning of each video, which contained a moment in which participants expressed neutrality. To accomplish this, each video was manually inspected in order to locate moments in which participants expressed neutrality. As a result of this inspection, one frame per participant containing a neutral face was replicated 75 times in order to create a five-seconds sequence (in a video recorded at 15 fps). This normalization procedure was conducted in first place to processes videos with *eMotion*, which as indicated in the documentation of the software, needs to be calibrated with participants' neutral faces in order to achieve better results. Moreover, normalization also provided a way to compare the classification of the three software during the first five seconds.

In addition to the procedures described above, facial markers for each participant were manually created for the particular case of *eMotion*. This procedure consisted on placing markers on key features used in the detection of facial expressions. Examples of such features are: head contours, eyes, nose, mouth, and neck. As indicated by the authors of *eMotion*, facial markers contribute to better classification results.

Overall, 56.25 hours of video were processed. Results of *BMERS*, *FaceDetect*, and *eMotion* during the evaluation of short-term effects are summarized in Figure 5.11. As illustrated in these plots, the average duration of smiles accounted for less than 30% of the facial expressions detected by the three tools. Two out of three tools (*BMERS* and *eMotion*) showed that participants who received positive stimuli ($Stim^{+}$) smiled more than those who received negative ones ($Stim^{-}$). In turn, the latter smiled on average more than participants who were not exposed to stimuli ($Stim^{control}$). This result, which is attributed to the different affective stimuli used in e ach group, indicates that exposing participants to false feedback had effects on their expression of emotions. Observations during the study showed that participants treated with positive

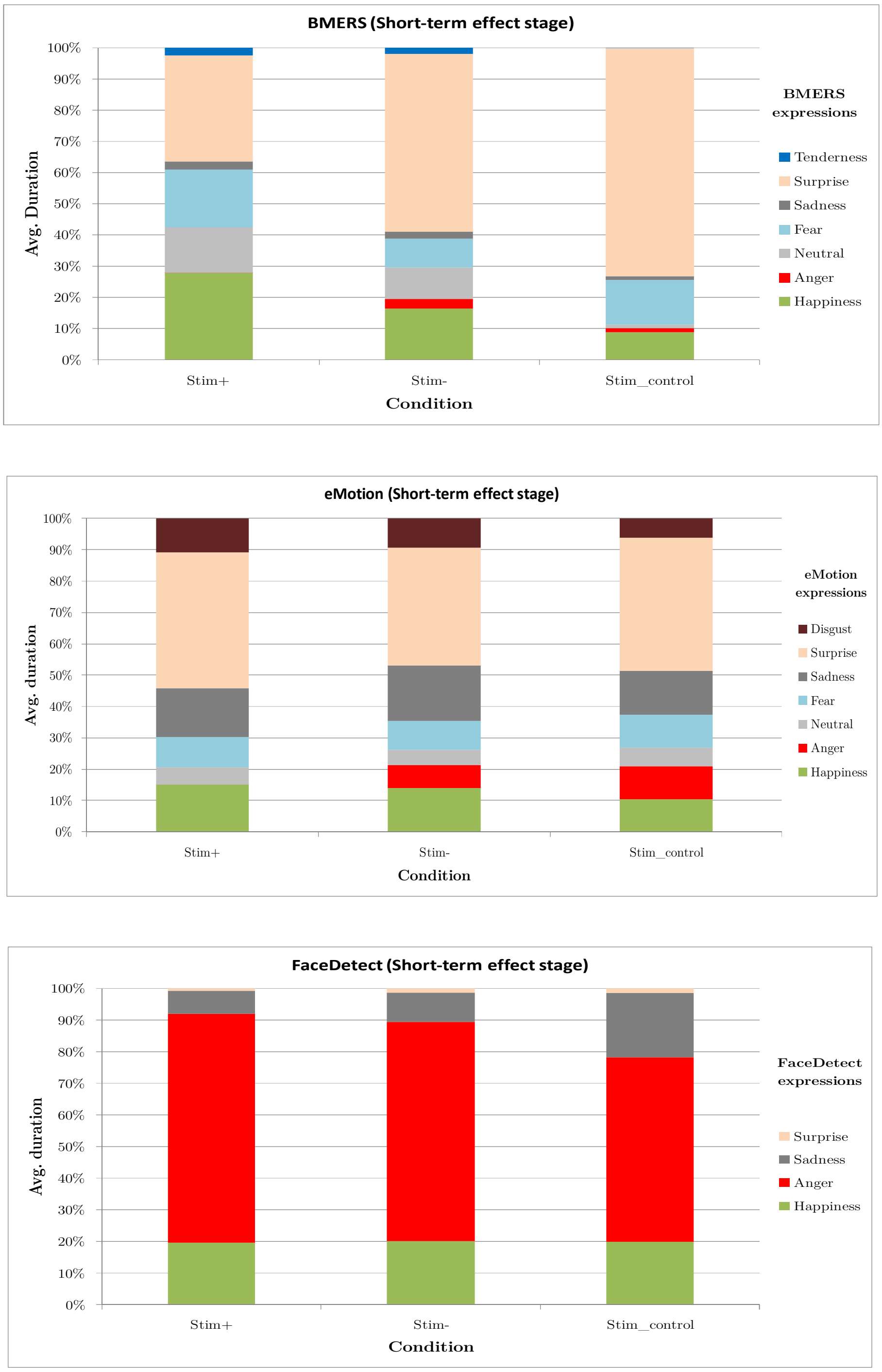


Figure 5.11. Average facial expressions duration during the evaluation of short-term effects. Results from three different software: *BMERS*, *eMotion*, and *FaceDetect*.

stimuli typically reacted positively when receiving encouraging messages and smiley faces at the moment of searching information and responding questions. On the other hand, there were a few cases in which participants treated with negative stimuli reacted smiling after receiving discouraging messages and frowning faces. In other cases smiles were observed when participants in negative conditions received slightly positive feedback (e.g. “Not bad this time”). As explained in section 4.5, the latter kind of feedback was intended to avoid extreme frustration and disinterest in participants. When asked about their experiences during the stimuli stage, some participants in $Stim^-$ indicated that they could not believe how poorly they were doing, so they smiled because they found the situation funny.

To explore more in detail when smiles were expressed by the participants during this stage, smiling episodes were aggregated in each question with respect to its normalized elapsed time. Normalization in this case was achieved by dividing the actual elapsed time of questions by the overall time spent in them. This resulted in a time scale from zero to one for all questions. This analysis was performed in the pre-stimuli evaluation (PreS), stimuli stage, and post-stimuli evaluation (PostS). Results for each stage are presented in Figure 5.12, which combines results from *BMERS*, *eMotion*, and *FaceDetect*. This combination was possible by selecting points in each unit of time with the highest similarity. Similarity in this case was calculated by computing the distance between the vectors with the smile percentages obtained from each tool. The average distance for pairs of points with the minimal distance was 1.24% (SD=1.23).

Bearing in mind that each question has its own search process, Figure 5.12 shows that during the pre-stimuli evaluation, smiles were concentrated mainly in later stages of the participants’ search processes. This is particularly clear in $Stim^+$ and $Stim^-$.

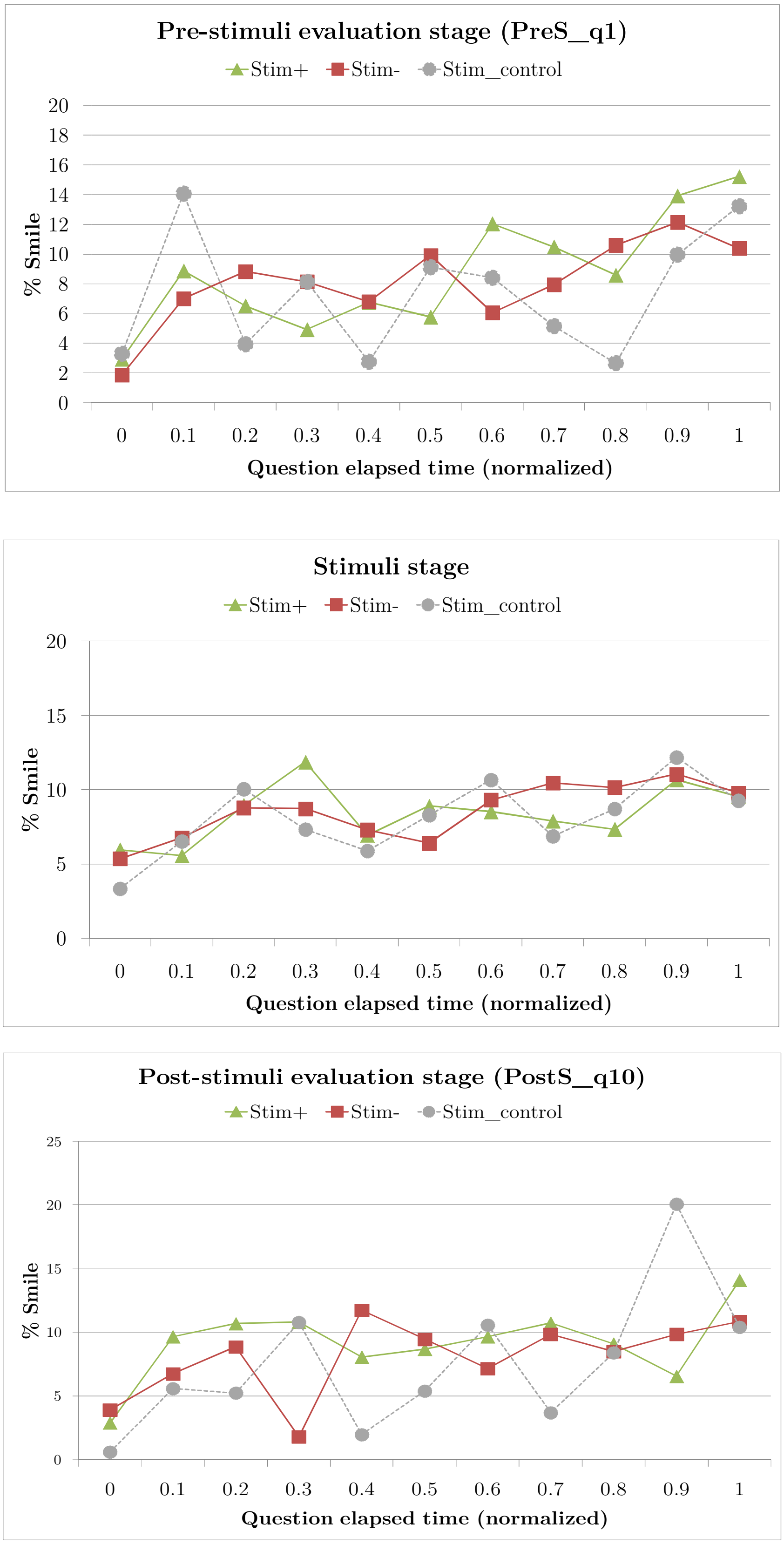


Figure 5.12. Smiles recognized during the development of questions in different stages of the evaluation of short-term effects. Combined results for *BMERS*, *eMotion*, and *FaceDetect* based on similarity in each unit of time.

However, in $Stim^{control}$, similar concentrations of smiles were found in both early stages and toward the end of the participants' search processes (14.06% and 13.22% respectively). Note that the pre-stimuli evaluation comprised only one question ($PreS\_q_1$).

Similar results were found for $Stim^-$ and $Stim^{control}$ during the application of stimuli (stimuli stage). On the other hand, $Stim^+$ presents a peak in 0.3 with a concentration of 11.86% of smiles identified during the session. This distinctive peak is attributed to participants being exposed to positive stimuli.

Figure 5.12 also shows that in the post-stimuli evaluation stage, smiles in $Stim^{control}$ were mostly concentrated (30.47%) in the last two segments of the development of $PostS\_q_{10}$. Similarly, in $Stim^+$ participants smiled more times (14.01%) than in early stages of the development of $PostS\_q_{10}$. On the other hand, similar frequencies of smiles were found for participants in $Stim^-$ in intermediate and last segments of $PostS\_q_{10}$. Yet, a moderate increment is observed toward the end of the development of this question.

The fact that smiles were typically concentrated in later stages of the search processes associated to questions can be attributed to feelings of satisfaction, relief, and success experienced by the participants after answering each question. Feelings such as satisfaction or success can be related to the levels of confidence that participants reported at the moment of providing their answers. As explained in section 4.5, participants in $Stim^+$ and $Stim^-$ also received affective feedback after providing their answers. However, although such stimuli could be negative or positive contributing factors in the expressivity of smiles, results for the control group ($Stim^{control}$) show that regardless of the exposure to affective stimuli, smiles are more likely to be expressed toward the end of the participants' search processes.

Since the expressive component of emotions may be altered by particular characteristics of the participants (e.g. personality or psychological disorders), it is acknowledged that the tools used to detect smiles might have failed in capturing hidden affective and mental states that can be expressed through smiles (e.g. happiness, joy, success, satisfaction, and confidence). To tackle this shortcoming, the next section focuses on the physiological component of affective processes in order to access internal affective changes that could inform participants' reactions that are not expressed through their faces.

5.4.1.2 Affective processes expressed through electrodermal activity

The second approach to investigate affective processes in the information search process focused on their physiological component. As explained in section 2.1.3, one way to do this is through electrodermal activity (EDA), which is a response of the human body that relates to arousal levels or physiological activation. Unlike facial expressions, which are mainly corresponded to categorical or discrete approaches in the study of affective processes, EDA is mostly related to dimensional approaches.

The study of EDA was achieved in this study with the use of *Q Sensors*, which are specialized devices capable of collecting this type of physiological signal. The participants in the study wore this device in their left wrists during the entire session. This device was the only one that the participants had to wear during the sessions; however, during the interview they reported being aware of this and the other sensors (i.e. webcam and eye tracker) only at the beginning of the sessions. Later participants focused mainly on the tasks they had to perform. For example S474 said “Yea it [the Q sensor] didn’t bother me at all.” S461 added “I mean [pause] at first I felt kind of aware of them [the sensors] so I was kind of [pause] just like looking at them and the stuff but then when I was doing the study I kind of just forgot about them.” Similarly, S450 said “Umm, I didn't really paid much attention to them [the sensors] it didn't bother me.”

EDA data was first smoothed. Then it was segmented, synchronized, and combined with other data sources following the same procedures described in the previous section for videos with participants’ faces. Synchronization was possible since *Q Sensors* were synchronized at the beginning of sessions with the local time of the corresponding computers assigned to participants.

Data was then processed with the *Affectiva Q Analytics*[48] software in order to extract peaks and related measures from the EDA signals. The set of peak-related measures includes: height, start time, duration, attack, magnitude, area under the curve, decay. lability[49], and *Q Score*. In particular, *Q Score*, as explained in the documentation of the software[50], is a “measure designed to give an indication of engagement/arousal” (p.7). This measure is the result of a combination of different features such as labile regions and area under the curve. Although *Q Score* can be an implicit indicator of engagement, the documentation explains that this measure is context-dependent. That is to say, high *Q Scores* could be linked to excitement, stress, or engagement, whereas

---

[48] http://www.qsensortech.com/q-analytics-beta/

[49] Measure related to regions with multiple peaks per minute. Lability characterizes individuals in terms of their high frequencies of peaks under particular conditions.

[50] http://www-assets.affectiva.com/assets/Q-Analytics-Guide.pdf

lower *Q Scores* could indicate affective and mental states such as disinterest and calm.

EDA analyses presented here focused on the occurrence of peaks as indicators of participants' responses to particular events such as finding relevant pages, finding the answers to the questions, or being exposed to affective stimuli. Following a procedure similar to that performed with facial expressions, the factor of interest (in this case peaks) was aggregated in each unit of time of the normalized duration of the questions in which they were detected. Figure 5.13 depicts the cumulative frequencies of peaks (in percentages) in each unit of time during the development of questions in different stages of the evaluation of short-term effects.

As noted in previous section, each question can be referred to as an independent search process. Generally speaking, it is possible to observe that the lowest concentration of peaks in all the cases occurred in the very first segment of the search processes. On the other hand, the highest concentrations were typically found in the last segments. In the context of this study, the positive trend of cumulative peaks observed throughout the search processes can be explained as increasing levels of engagement and/or stress. The latter could be a result of time pressure as a consequence of the participants being constantly aware of their remaining time to complete the questions and the overall task (presented on the top region of the sidebar as depicted in Figure 4.8).

In addition to distinctive peaks at later stages of the development of questions, it is also possible to observe salient points in early stages (between 0.1 and 0.3). Perhaps such concentrations of peaks indicate (1) frustration or anxiety during early stages of information search as a result of uncertainty, or (2) excitement after finding the first piece (or pieces) of information that helped the participants to later find the answer to the questions.

As illustrated in Figure 5.13, it is interesting to note that during the stimuli stage the concentration of peaks in the three conditions ($Stim^{+}$, $Stim^{-}$, and $Stim^{control}$) was very similar regardless of the type of affective stimuli or their absence in the control group. Note that unlike the pre-stimuli and post-stimuli evaluations, the stimuli stage comprised multiple questions that participants addressed in a period of 10 minutes. During this time participants in $Stim^{+}$ and $Stim^{-}$ were frequently exposed to stimuli.

As stated in the literature, multiple trials or a frequent exposure to stimuli may result in anticipatory affective reactions (Figner & Murphy, 2011). This aspect may explain similarities in the distribution of peaks during the stimuli and post-stimuli evaluation stages for $Stim^{+}$ and $Stim^{-}$, but not for $Stim^{control}$. Lower concentration of peaks in later stages of the search processes of $Stim^{control}$ can be attributed to participants feeling more relaxed as a result of being aware that that $PostS\_q_{10}$ was the last question of the first task[51]. For additional details, a summary of the peaks identified in the context of different types of pages, snip action, and the occurrence of smiles is provided in Appendix F.

It is important to recall that the dimensional approach to study affective processes comprises aspects such as activation (arousal) and tone (valence).

---

[51] As explained in section 4.1 the entire the evaluation of short-term effects (i.e. pre-stimuli, stimuli, and post-stimuli evaluation) was presented to participants simply as Task 1.

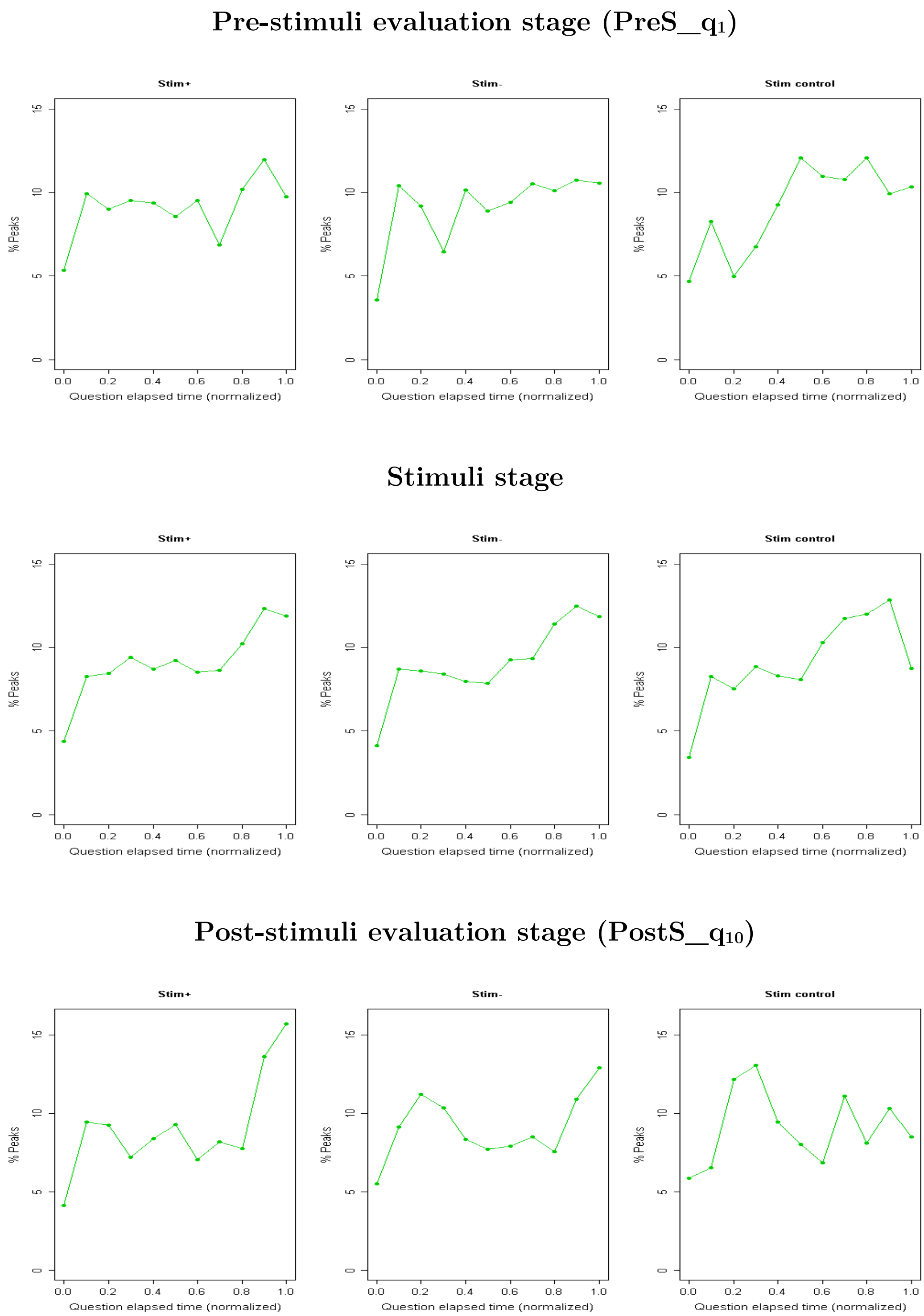


Figure 5.13. Peaks detected in EDA at different stages of the development of questions responded during the evaluation of short-term effects. The plot on top corresponds to results in the pre-stimuli evaluation. The plot in the center summarizes the activity during the stimuli stage. Finally, the plot on the bottom part depicts peaks detected in the post-stimuli evaluation.

However, as explained early in this section and also in section 2.1.3, EDA relates to physiological activation. Therefore, a key limitation of EDA is its inability to indicate whether responses have positive or negative valence. This explains opposite interpretations (e.g. engagement or frustration) provided above. To elucidate this ambiguity, facial expressions could help to interpret whether EDA relates to pleasant (positive) or unpleasant (negative) affective experiences. For example, the fact that both peaks and smiles were mostly concentrated in later stages of the search processes sheds light that EDA in such stages could have a positive connotation. However, as illustrated in Appendix F, not every smile has a correspondence with the occurrence of peaks in EDA. This aspect relates to another limitation of EDA, which is the fact that either instruments are not sensitive enough to capture this type of signal in some participants or that simply some people are less labile[52] than others. Indeed, during the study there were 41 participants (17 in $Stim^{+}$, 17 in $Stim^{-}$, and 7 in $Stim^{control}$) who did not present peaks during the execution of their sessions. Such participants presented small to no changes in their EDA during the activation procedure (explained in section 4.7.1.1) performed at the very beginning of the sessions.

Although both facial expressions and physiological signals provide an objective way to investigate affective processes, participants may experience these processes in different ways. This subjective perception is not captured by the evaluation approaches used so far. To address this, the next section investigates

---

[52] In the context of this study and based on the measure conducted with *Affectiva Q Analytics*, a person is considered to be labile when his/her affective states are easily altered. This is expressed in EDA through multiple peaks in a given region.

participants' self-reported feelings as a subjective component of the affective dimension.

5.4.1.3 Affective processes expressed through self-reports

The third approach to investigate affective processes focused on self-reported feelings. Previous sections focused on the results of objective approaches using both observer-based (specifically facial expressions) and neurophysiological (in particular EDA) methods. However these approaches fail in capturing participants' subjective experiences, which in some case could differ from what their bodies express.

Feelings were studied by asking the participants to report how they felt right before being exposed to each question and right after providing their answers. To facilitate self-reporting, the Self-Assessment Manikin (SAM) questionnaire (Bradley & Lang, 1994) (more details in section 0) was presented to the participants between questions and also between particular stages of the study (e.g. tutorial, practice question, task instructions). Note that the use of SAM between questions served a dual purpose: (1) collect data about participants' feelings as a post-condition of the question responded, and (2) collect data about participants' feelings as a pre-condition for the following question. Instructions of this questionnaire, which were provided to the participants in early stages of the session, indicate that responses should be provided within 15 seconds. On average, the participants in this study responded the SAM questionnaires in 10.08 seconds (SD=6.33).

SAM allowed the participants to report their feelings in three different, but complementary dimensions: valence/tone (i.e. happy-unhappy), activation/arousal (i.e. excited – calm), and dominance/control (i.e. in control – controlled). These dimensions were presented to the participants as three independent 9-point scales. Note that the scales use different terms to express aspects such as positive and negative (happy and unhappy respectively) and high activation and low activation (excited and calm respectively). This alternate terminology, which is more familiar to people than terms such as activation, is used in the original SAM questionnaire.

With the data collected, the first analyses were conducted with regard to valence. The average scores of valence computed before and after each question in $Stim^{+}$, $Stim^{-}$, and $Stim^{control}$ are summarized in Figure 5.14. In this plot points connected with thin lines represent responses before questions (Pre). On the

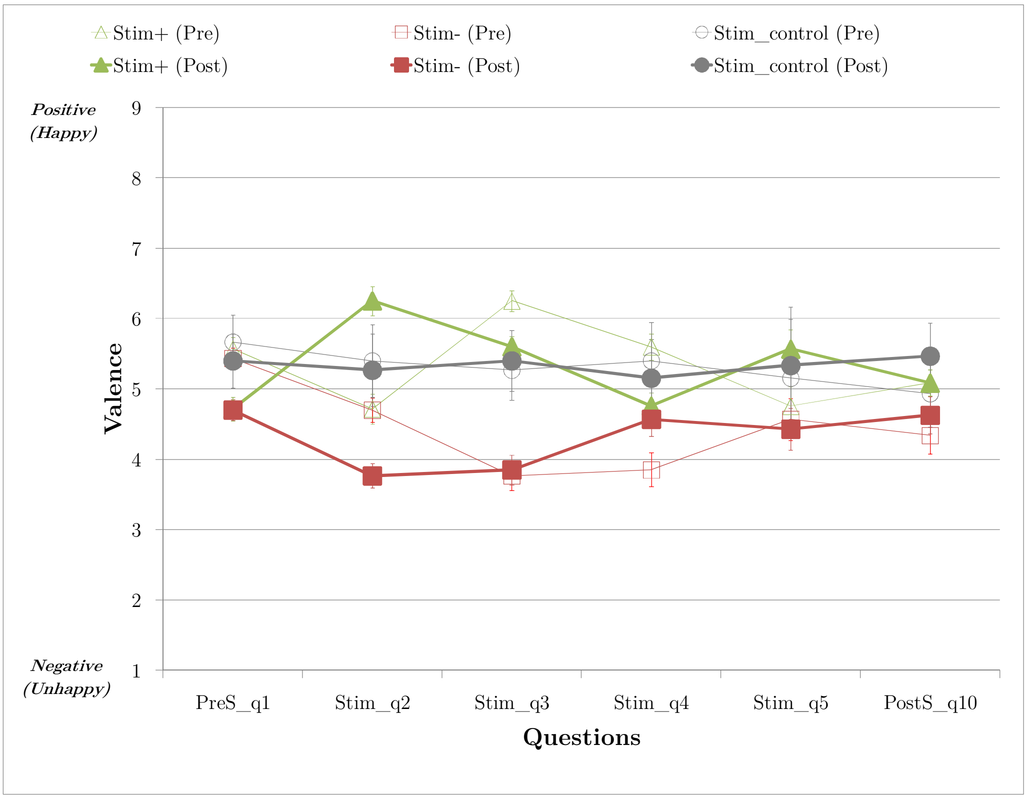


Figure 5.14. Average scores for self-reported valence (i.e. positive versus negative) in each question.

other hand, points connected with thick lines correspond to self-reported valence right after providing answers to the questions (Post).

Statistical comparisons between groups were performed with the Kruskall-Wallis tests and post-hoc analyses were conducted with the Wilcoxon rank-sum test. Results from this test did not show significant differences with regard to valence reported before being exposed to question PreS_q_1. This aspect is particularly important because it indicates that the participants started the first evaluation stage (i.e. evaluation of short-terms effects) in roughly similar conditions. The test also did not show significant differences after the completion of PreS_q_1. Differences became significant at $p<.01$ after completing the first question of the stimuli stage (Stim_q_2). In particular results showed that valence, according to the 9-point scale, was significantly higher at $p<.01$ in $Stim^+$ (*Mdn*=6) than in $Stim^{control}$ (*Mdn*=5). In turn, valence in the latter was significantly higher at $p<.01$ than $Stim^-$ (*Mdn*=4). Note that values greater than five points denote positive valence, whereas those lower than five points indicate negative valence. Values equal to five are interpreted as neutral. These results have particular implications because it shows the effectiveness of the stimuli as experienced by the participants. In fact, it was found that stimuli intended to elicit positive affective states reached an effectiveness of 71.67% in the first questions. On the other hand, stimuli planned to evoke negative affective states reached an effectiveness of 65% in the same questions.

Similar results were found after participants responded the following question (Stim_q_3). While means varied as illustrated in Figure 5.14, medians were almost the same as in Stim_q_2. In this case the Wilcoxon rank-sum test also showed that valence was significantly higher at $p<.01$ in $Stim^{+}$ (Mdn=5.5) than in $Stim^{control}$ (Mdn=5), and in turn, valence in the latter was significantly higher at $p<.01$ than $Stim^{-}$ (Mdn=4).

Both questions ($Stim\_q_2$ and $Stim\_q_3$) are particularly important since these are the only two questions that all 135 participants responded. As illustrated in Figure 5.4, after $Stim\_q_3$ the number of observations was reduced because the majority of the participants used the 10 minutes assigned to respond $Stim\_q_2$ and $Stim\_q_3$. Regardless this reduction in the number of observations, significant differences with respect to valence were also observed after participants completed $Stim\_q_5$. More importantly, such differences were consistent with respect to the type of stimuli used in each group. As reported after this question, valence was significantly higher at $p<.01$ in $Stim^{+}$ (Mdn=5) than in $Stim^{control}$ (Mdn=5). Moreover, valence in the latter was also higher than in $Stim^{-}$ (Mdn=5).

Differences found during the stimuli stage, specially self-reported valence by the participants in $Stim^{+}$ and $Stim^{-}$, are attributed to the two types of stimuli (i.e. positive and negative) to which participants in both groups were exposed and not due to their experiences during information search. In fact, participants in the control group typically placed their affective valence in the center of the unhappy-happy scale, thus denoting neutrality.

The main result of the analyses with respect to valence was found at the beginning of the post-stimuli evaluation ($PostS\_q_{10}$). In this case the Wilcoxon rank-sum test showed that participants in the three groups started this question in different levels of affective states. In particular, valence was found to be significantly higher at $p<.01$ in Stim^{+} (*Mdn*=5) than in Stim^{control} (*Mdn*=5). Furthermore, valence in the latter was also higher than in Stim^{-} (*Mdn*=4). As discussed in the development of the first research question (RQ1), the analyses performed with the different groups were conducted under the assumption that stimuli were effective in eliciting internal affective states without taking into account the subjective experiences of the participants. Therefore, the importance of this result lies in the consistency between the elicitation of internal affective states and the subjective experience of participants as pre-conditions for PostS_q_{10}, which focused on the short-term effects of positive and negative affective processes in the information search process.

As indicated above, in addition to valence, the participants also reported their levels of activation. This dimension was presented to the participants on a 9-point scale with two extremes: excited and calm. The average scores of activation computed before and after each question in Stim^{+}, Stim^{-}, and Stim^{control} are summarized in Figure 5.15. In this plot points connected with thin lines represent responses before questions (Pre). On the other hand, points connected with thick lines correspond to self-reported valence right after providing answers to the questions (Post).

Statistical comparisons between groups were performed with the Kruskall-Wallis tests and post-hoc analyses were conducted with the Wilcoxon rank-sum test. Significant differences at $p<.01$ were found right before starting PreS_q_1, with participants in Stim^{-} reporting feeling more activated (excited in the terminology presented to the participants) (*Mdn*=6) than those in Stim^{+}(*Mdn*=5) and

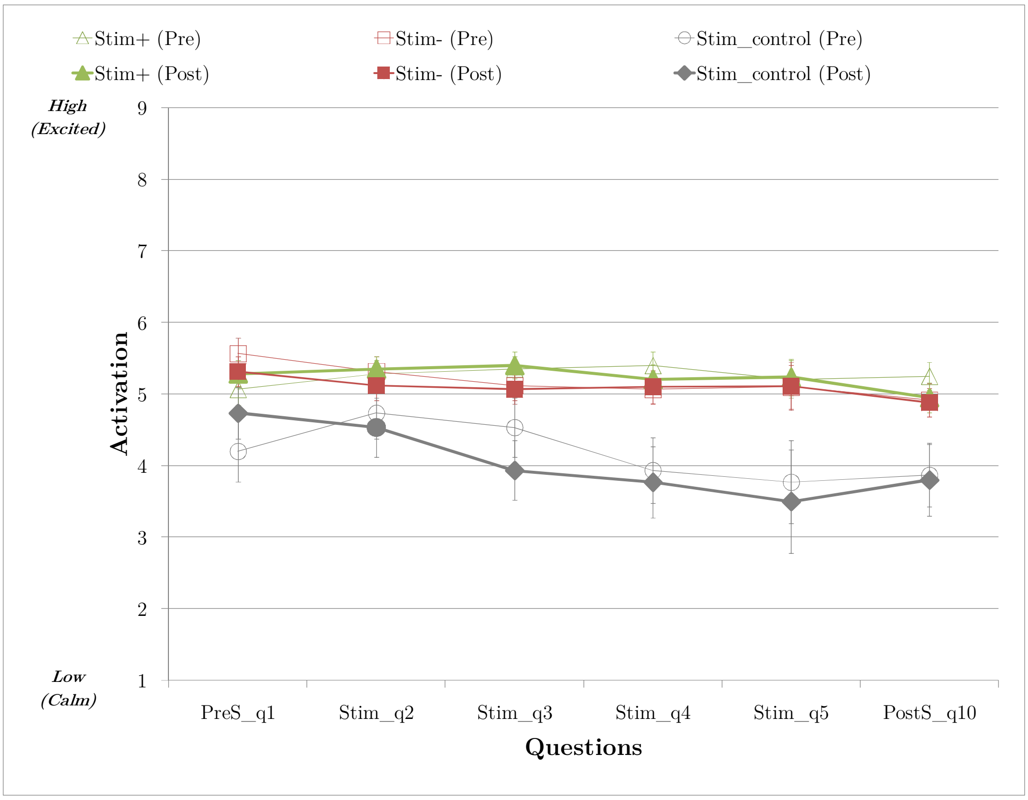


Figure 5.15. Average scores for self-reported activation (i.e. high versus low).

Stim^{control}(Mdn=5). Regardless of this unexpected dissimilarity, no significant differences were found after PreS_q_1 was completed.

Differences in terms of activation as a result of the stimuli stage were only observed right after Stim_q_3. In this case the levels of activation were found to be significantly higher at p<.01 in Stim$^+$ (Mdn=5) than in Stim$^-$ (Mdn=5) and the latter also higher than Stim^{control} (Mdn=4). Note that activation has positive and negative interpretations depending upon self-reported valence. Since valence was mostly positive in Stim$^+$, high activation levels have positive connotations (e.g. engagement). On the other hand, because of the valence in Stim$^-$ was mainly negative, high activation levels would have a negative interpretations (e.g. frustration).

The same medians and significant differences between the three groups were found in Stim_q$_4$. As noted in the analyses of valence, the number of observations in this question was lower than in Stim_q$_3$, nevertheless, results showed a consistent pattern in the elicitation of affective states during the stimuli stage. In fact, although some participants were able to address more than two questions in this stage, it was found that self-reported activation right before starting PostS_q10 followed the same relations found in Stim_q$_3$ and Stim_q$_4$. In other words, Stim$^+$ > Stim$^-$ > Stim^{control}.

High activation levels found in Stim$^+$ and Stim$^-$ are attributed to the exposure to affective stimuli. Conversely, as depicted in Figure 5.15, Stim^{control} showed a trend toward low activation levels (e.g. calm). Similar to the results for valence, the significant differences between the three groups reported above provides additional evidence about the effectiveness of the affective stimuli used to elicit positive and negative affective states.

Finally, self-reported feelings were also investigated with respect to dominance. This dimension indicated whether the participants felt in control (high dominance) or controlled (low dominance) during the execution of the study. The average scores of dominance computed before and after each question in Stim$^+$, Stim$^-$, and Stim^{control} are summarized in Figure 5.16. In this plot points connected with thin lines represent responses before questions (Pre). On the other hand, points connected with thick lines correspond to self-reported valence right after providing answers to the questions (Post).

Statistical comparisons between groups were performed with the Kruskall-Wallis tests and post-hoc analyses were conducted with the Wilcoxon rank-sum test.

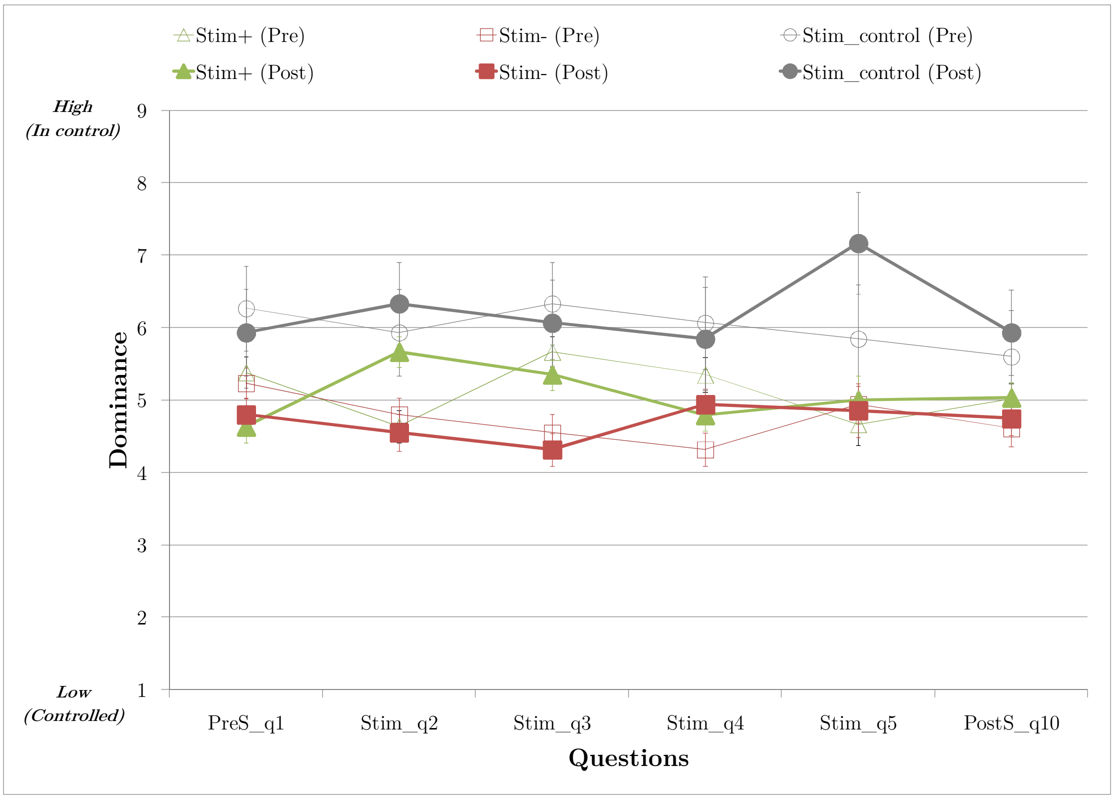


Figure 5.16. Average scores for self-reported dominance (i.e. high versus low).

Results from this test did not show significant differences with regard to dominance levels reported before being exposed to question PreS_q_1. This aspect is particularly important because it indicates that the participants started the first evaluation stage (i.e. evaluation of short-terms effects) in roughly similar conditions. The test also did not show significant differences after the completion of PreS_q_1.

Significant differences at $p<.01$ were found right after completing the first question of the stimuli stage (Stim_q_2). In particular results showed that dominance levels, according to the 9-point scale, were significantly higher at $p<.01$ in $Stim^{control}$ (Mdn=7) than in $Stim^{+}$ (Mdn=5). In turn, dominance levels in the latter were significantly higher at $p<.01$ than $Stim^{-}$ (Mdn=5). The same relation between the three groups was found right after Stim_q_3 and Stim_q_5.

This result is consistent with some comments provided by the participants during the interview conducted at the end of the sessions. For example, when asked about their reception of the feedback provided during Task 1 (evaluation of short-term effects) participant S449 in $Stim^-$ said “I felt like I wasn't in control of what I was doing.” On the other hand, participants S592 in $Stim^+$ said “I guess they [stimuli] were like confidence boosters.” Despite this common pattern found in these three questions, after $Stim\_q_4$ dominance levels were found to be significantly higher at p<.01 in $Stim^-$ (*Mdn*=5) than in $Stim^+$ (*Mdn*=5). Yet, self-reported dominance was still significantly higher in $Stim^{control}$ (*Mdn*=7) than in the other two conditions.

Feeling more controlled (low dominance) during the stimuli stage is in part attributed to the exposure to affective stimuli. Since stimuli were contextualized as positive or negative feedback with regard to the performance of the task, participants might have related such feedback to intentional manipulations. Indeed, one participant in $Stim^-$ said “I felt that they [feedback/stimuli] were made on purpose.” Despite participants' impressions in terms of their levels of control (dominance), no differences were found at the moment of starting the post-stimuli evaluation ($PostS\_q_{10}$). This result shows that pre-conditions to $PostS\_q_{10}$ were mostly attributed to valence and activation levels and not to different levels of dominance.

A summary of the results described in this section for valence, activation, and dominance is provided in Table 5.14. Overall, results for the evaluation of self-reported feelings demonstrated the effectiveness of the procedure implemented to elicit affective states.

Table 5.14. Summary for self-reported feelings. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with individuals as unit of analysis. Results in cells are significant at $p<.01$. Due to space restrictions in the tables, conditions codes are presented as follow: $S^+=$ $Stim^+$, $S^-=$ $Stim^-$, and $S^c=$ $Stim^{control}$.

		Question					
		PreS_q1	**Stim_q2**	**Stim_q3**	**Stim_q4**	**Stim_q5**	**PostS_q10**
Before (Pre)	**Valence**			$S^+>S^-,S^c$ $S^c>S^-$	$S^+>S^-,S^c$ $S^c>S^-$		$S^+>S^-,S^c$ $S^c>S^-$
	Activation	$S^+>S^c$ $S^->S^+,S^c$			$S^+>S^-,S^c$ $S^->S^c$	$S^+>S^-,S^c$ $S^->S^c$	$S^+>S^-,S^c$ $S^->S^c$
	Dominance			$S^+>S^-$ $S^c>S^+,S^-$	$S^+>S^-$ $S^c>S^+,S^-$	$S^-> S^+$ $S^c>S^+,S^-$	
After (Post)	**Valence**		$S^+>S^-,S^c$ $S^c>S^-$	$S^+>S^-,S^c$ $S^c>S^-$		$S^+>S^-,S^c$ $S^c>S^-$	
	Activation			$S^+>S^-,S^c$ $S^->S^c$	$S^+>S^-,S^c$ $S^->S^c$		
	Dominance		$S^+>S^-$ $S^c>S^+,S^-$	$S^+>S^-$ $S^c>S^+,S^-$	$S^-> S^+$ $S^c>S^+,S^-$	$S^+>S^-$ $S^c>S^+,S^-$	

This section focused on investigating the expressivity and experience of affective processes in information search during the evaluation of short-term effects. As indicated in the analyses of facial expressions and EDA, these approaches have intrinsic limitations, especially when it comes to interpreting their contextual meaning. Self-reports provide additional details of the experience of the participants at particular moments and they were provided right after each question, peaks in EDA that were found in the last segments of the questions can be interpreted with respect to responses in the valence scale. Therefore, based on aggregated results presented in this and previous sections it is possible to say that, in general, the peaks found in the last segments of the questions that participants in $Stim^+$ and $Stim^{control}$ responded have positive connotations (e.g. excitement as a result of success). On the other hand, peaks found in $Stim^-$ would be related with negative states such as frustration as a result of failure.

The following section focuses on further exploration of the prolonged effects of the affective states resulted from this stage. In addition to address this investigation with individuals, the analyses presented below also focus on the interaction of affective processes as a result of collaboration between pairs of participants.

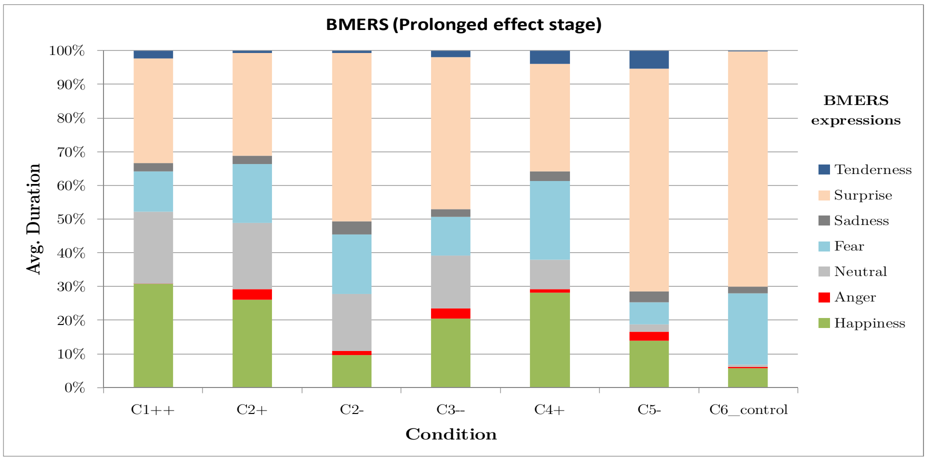
5.4.2 Evaluation of prolonged effects

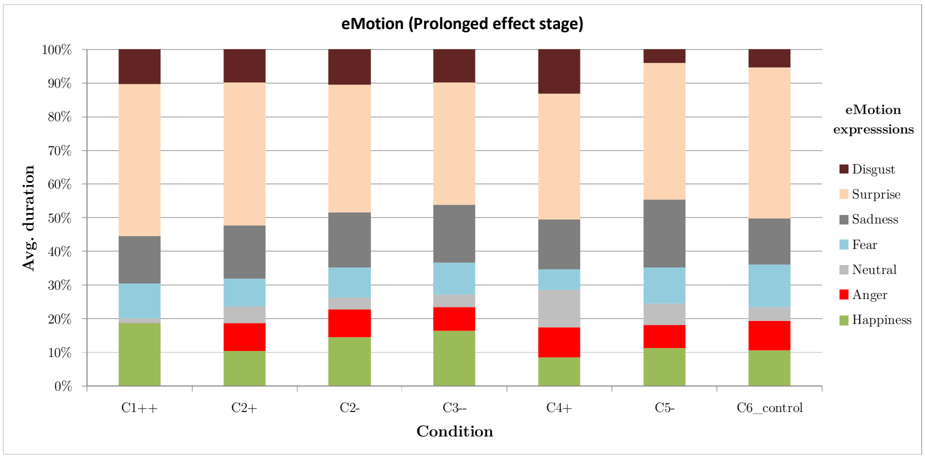
This section presents results for the analyses conducted to investigate expressiveness and experience of affective states during the information search process of teams. Results presented in the previous section showed that regardless of initial affective states, individual searchers present similar expressive patterns in terms of smiles and EDA in particular stages of information search. This section, on the other hand, focuses on studying whether interactions of affective processes - as a result of collaboration between team members - influence searchers' expressiveness and experience of affective states in the long run.

Below, results for facial expressions, EDA, and self-reports are presented. Additionally, taking into account the social component introduced in this stage, this section also shows results from communication analyses, which focus on verbally expressed affective states that were observed during the interactions of team members while working on the search tasks.

5.4.2.1 Affective processes expressed through face

Following the same procedure described in the previous section, analysis of facial expressions were performed with *BMERS*, *eMotion*, and *FaceDetect*. Overall, 67.5 hours of videos were processed with each system in order to extract features from participants' faces and classify them into basic emotions. A summary of the results obtained with each software is presented in Figure 5.17.





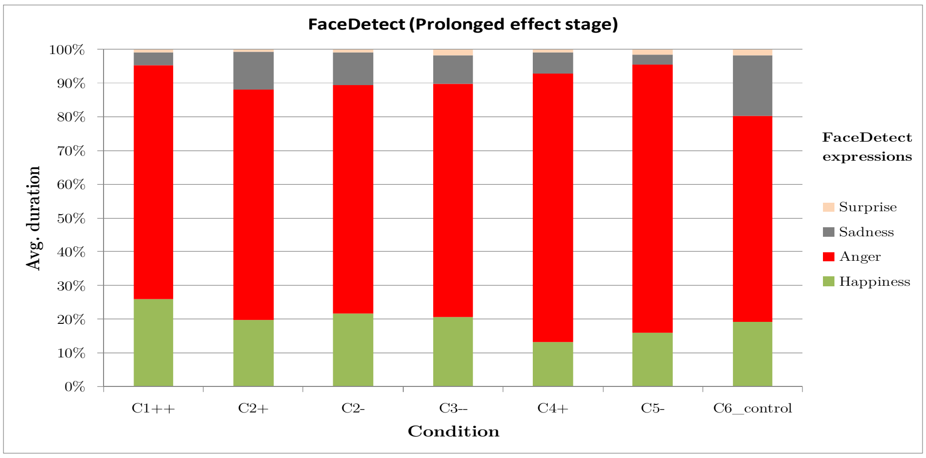


Figure 5.17. Average facial expressions duration during the evaluation of prolonged effects. Results from three different software: *BMERS*, *eMotion*, and *FaceDetect*.

As noted above, the three software classify facial features into different sets of facial expressions related to basic emotions. The recognition rates of these facial expressions differ from tool to tool. A comparison with respect to the common set of facial expressions labeled as: happiness, anger, surprise, and sadness, shows noticeable differences with respect to surprise, anger, and sadness. For example, *FaceDetect* reported that the average duration of surprise in all six conditions was under 5%. On the other hand, both *BMERS* and *eMotion* showed that surprise was a predominant expressions in most conditions. Similarly, *BMERS* reported low presence of sadness compared to results obtained with *FaceDetect* and *eMotion*. With respect to anger, *FaceDectect* showed that this expressions accounted for more than 70% of participants' facial expressions in all six conditions, whereas the same expressions in *BMERS* and *eMotion* was under 10%.

Among the four facial expressions that are common in the three tools, the only one that presented some degree of similarity was that linked to happiness (i.e. smiles). This general comparison shows reliability issues in the detection of facial expressions other than smiles. This is the main reason behind the analyses of facial expressions presented in the previous section and also here focus on smiles.

An evaluation of smiles with respect to their occurrences at different stages of the information search processes associated with the development of the *A Google a Day* questions is depicted in Figure 5.18. As described in the previous section, this combination was possible by selecting points in each unit of time with the highest similarity. Similarity was calculated by computing the distance between the vectors with the smile percentages obtained from each tool. The

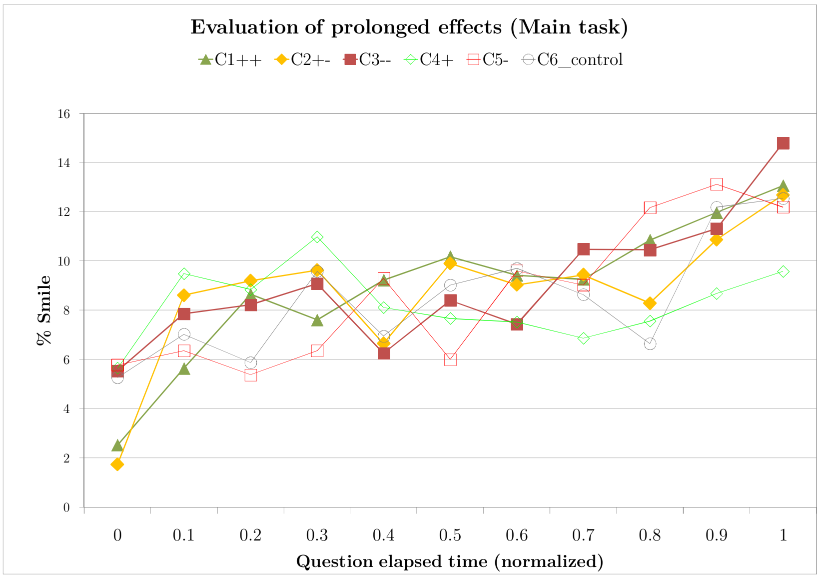


Figure 5.18. Smiles recognized in different stages during the development of questions in the evaluation of prolonged effects. Combined results for *BMERS*, *eMotion*, and *FaceDetect* based on similarity in each unit of time.

average distance for pairs of points with the minimal distance was in this case 0.93% (SD=0.74).

Consistently with the results presented in the previous section, Figure 5.18 shows that higher concentration of smiles are typically found in later stages of the participants' search processes. From all the conditions analyzed, only $C4^+$ presented a higher concentration of smiles (10.98%) in an early stage (in 0.3) than that found in the last segments.

In addition to investigate smiles with respect to the overall search process, a closer examination was carried out in order to study the expression of smiles in different types of webpages. Note that these analyses were not conducted in the previous stage due to the presence of stimuli that could make certain facial expressions, such as smiles, more noticeable when conducting fine-grain analyses.

Analyses in this respect focused on comparing the expressivity of smiles and other facial expressions (grouped into one category) in relevant and non-relevant pages and also in useful and non-useful pages. For the case of relevant pages, results in all six conditions showed that the majority of smiles in such pages lasted on average less than 1.36 seconds, that is to say, less than 3% of the average dwell time found in relevant pages (*M*=45.22s, *SD*=48.03). The density of smile duration in non-relevant pages was similar to that found in relevant pages; however, since the average dwell time in non-relevant pages was 17.50 seconds (*SD*=18.73), smiles in such pages typically lasted a fraction of a second. On the other hand, the duration of other facial expressions accounts for up to 30% of the time participants spent in both relevant and non-relevant pages. These results are summarized with density plots in Figure 5.19 based on the facial expression analyses conducted with *FaceDetect*. Plots obtained with the

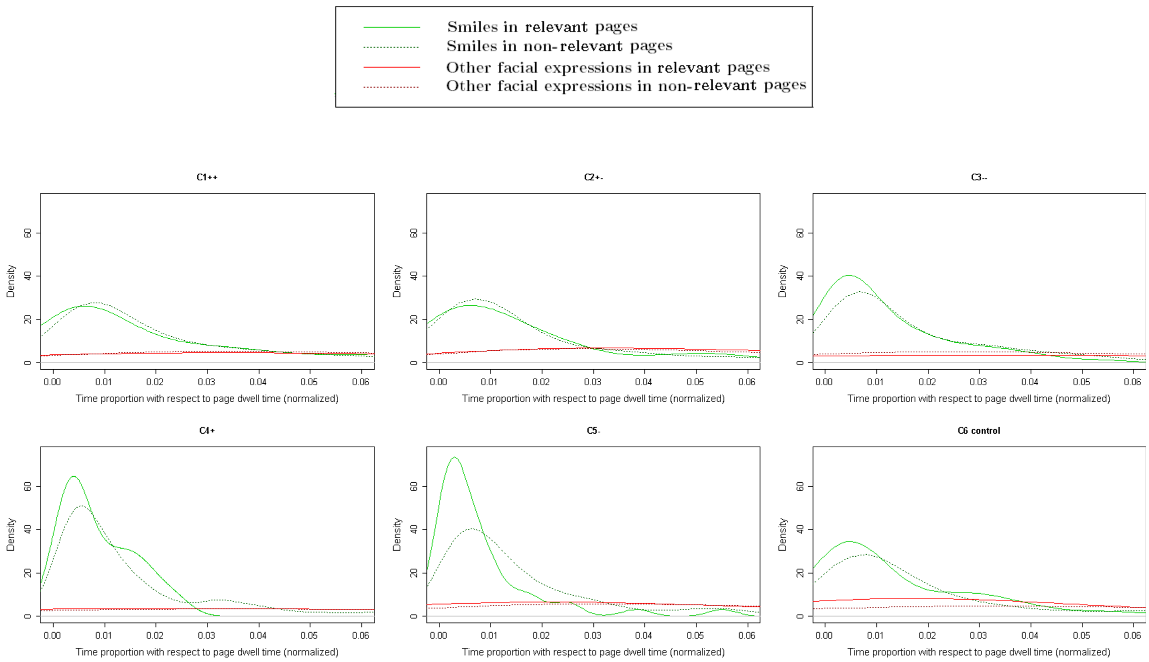


Figure 5.19. Proportional duration of smiles and other facial expressions with respect to dwell time (normalized) in relevant and non-relevant pages.

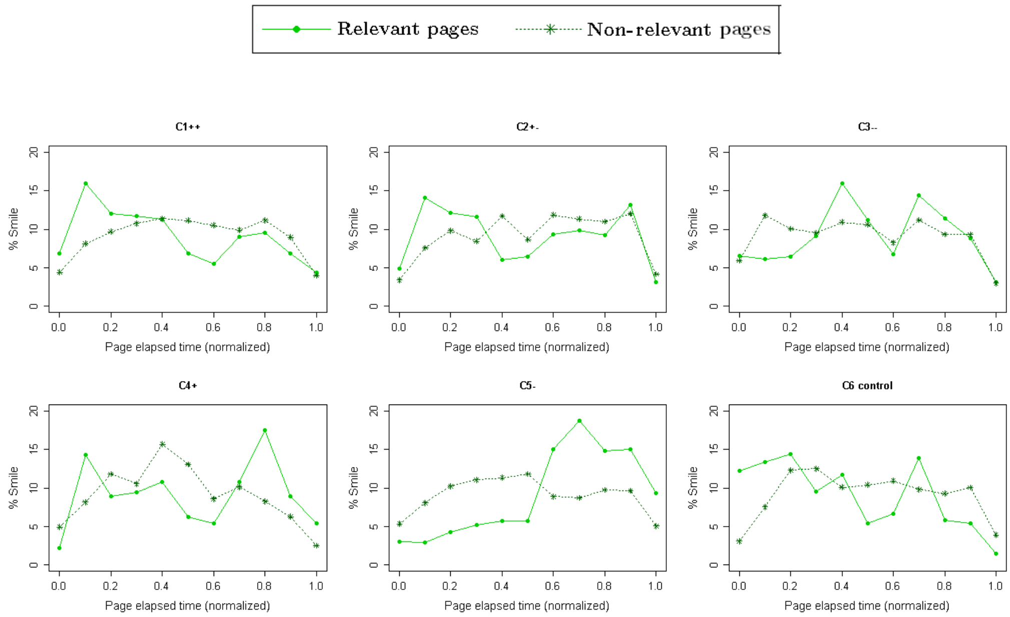


Figure 5.20. Smiles recognized with *FaceDetect* during the exposure to the content of relevant and non-relevant webpages.

other two tools are provided in Appendix F.

While smiles duration in relevant and non-relevant pages were found to be proportionally similar with respect to dwell time in such webpages, the actual difference (estimated in one extra second in relevant pages) could serve as an indicator of searchers' satisfaction.

In addition to studying the duration of smiles and other facial expressions in relevant and non-relevant pages, an analysis was conducted to investigate when the participants smiled during the exposure to the content of such webpages. This examination was carried out by aggregating smiles with respect to the elapsed time (normalized) of webpages. Analyses were conducted with facial expressions detected with the three software. Results based on *FaceDetect* are presented in Figure 5.20, whereas those based on *eMotion* and *BMERS* are provided in Appendix F. Both *FaceDetect* and *BMERS* showed that smiles in conditions other than $C5^-$ were concentrated in two moments during the exposure to the content of webpages. Perhaps, smiles in initial segments indicate a positive first impression of relevant pages from the participants' point of view. On the other hand, smiles detected in later segments could be related with the moment in which the participants found information that helped them to find the answers or the answers themselves, thus denoting satisfaction or success. This specific moment was explicitly indicated by the participants by saving snippets from such pages. Indeed, analyses conducted with respect to this action showed that the participants smiled more than 80% of the times at the exact moment in which snippets were saved. These results increased slightly in most conditions when the analyses were carried out with a window of 10 seconds around the snip action (five seconds before and after). Results for these analyses are summarized in Table 5.15.

In condition $C5^-$ smiles were mostly concentrated in later segments of the exposure to relevant pages. The absence of the first peak in this group, which was found to be a common aspect in the other five conditions, could be attributed to the influence of negative affective states that were induced on participants in the previous stage. This aspect can be related with the affect infusion model (AIM) introduced in section 3.1.2 since negative affective states could have influenced the evaluation criteria and perception of the participants. In other words, the first impression of the participants in $C5^-$ at the moment of visiting relevant webpages was not necessarily positive. Only after carefully reviewing the content of webpages and finding the information they were looking for, they smiled.

Table 5.15. Results for smile analyses conducted with respect to the snip action.

	Smiles			
Condition	**At snip**	**5 sec before snip**	**5 sec after snip**	**At snip ± 5 sec**
$C1^{++}$	85.04%	85.41%	85.33%	85.34%
$C2^{+-}$	82.17%	82.14%	82.19%	82.16%
$C3^{--}$	82.15%	82.12%	82.30%	82.19%
$C4^{+}$	91.78%	91.79%	91.82%	91.83%
$C5^{-}$	93.23%	93.27%	93.26%	93.27%
$C6^{control}$	90.14%	90.11%	90.11%	90.11%

Something interesting to note in this regard is that while participants in $C2^{-}$ (as part of $C2^{+-}$) also started the evaluation of prolonged effects in negative affective states, their smiles were concentrated in early and late segments during the exposure to the content of webpages. In fact, an analysis conducted with participants in $C2^{-}$ showed the same pattern (i.e. two salient segments of smiles) found in positive conditions. Possibly the presence of social interactions during the collaboration process contributed to participants' affective recovery, thus reducing the effects of initial negative affective states. In the case of $C3^{--}$, although two concentration peaks of smiles were found in this condition, the first one was located almost in the middle, which suggests a late reaction to the content or also reactions derived from communication with their partners when sharing their individual findings.

It is noteworthy that smiles were not absent in non-relevant pages, however, their distributions were found to be roughly uniform during the exposure to the content of pages. Note that the high concentration of smiles in $C4^{+}$ in segment 0.4 in non-relevant pages, as shown in Figure 5.20, was considered a recognition error of *FaceDetect* since the other two tools (*BMERS* and *eMotion*) did not show this salient point.

In addition to investigate smiles in relevant and non-relevant pages, analyses were carried out to study this facial expression in the context of useful and non-useful pages. Recall that the usefulness of pages was determined based on the time the participants spent in pages. Pages in which the participants spent 15 or more seconds were considered as useful. Conversely, pages with a dwell time under this threshold were classified as non-useful.

Results for these analyses are depicted in Figure 5.21. Unlike results described above for relevant pages, useful pages does not present distinctive peaks of smiles compared to non-useful pages. Note that the high concentration of smiles in $C4^{+}$ in the segment 0.4 correspond to a detection error found in *FaceDetect*, which was described above. Some of the small peaks of smiles found in specific segments of the different conditions can be related to the overlap between useful and relevant pages as shown in Table 5.4.

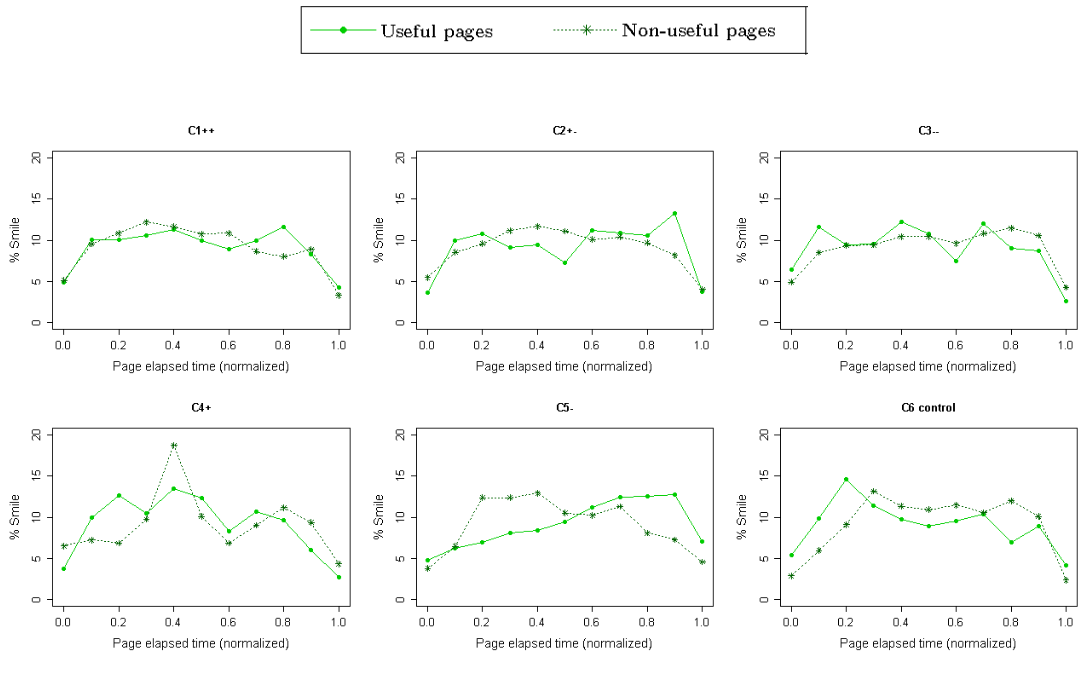


Figure 5.21. Smiles recognized with *FaceDetect* during the exposure to the content of useful and non-useful webpages.

As explained in the evaluation of short-term effects, it is acknowledged that the tools used to analyze facial expressions might have failed in detecting smiles. In some cases this could be attributed to the participants being out of the capturing range of the camera or in a position in which key features such as the mouth were not properly captured. To address this limitation, the following section focuses on electrodermal activity (EDA) in order to access internal affective changes that could inform participants' reactions that were not detected through their faces.

5.4.2.2 Affective processes expressed through electrodermal activity

This section presents results for the analyses conducted with respect to EDA. The purpose of these analyses was to examine participants' affective states at the physiological level. As described in the evaluation of short-term effects, participants presented a common pattern of peaks at later stages of their search processes regardless of the exposure to affective stimuli.

EDA analyses were first conducted with respect to the information search process linked to the development of each question. This was done by aggregating the peaks found in the electrodermal signal in each unit of time (normalized) of the duration of questions. Similar to the results presented in the previous evaluation stage, Figure 5.22 shows high concentration of peaks in EDA in the last segments of the search processes.

While most conditions displayed progressive positive trends of aggregated peaks toward the last minutes of the development of questions, $C5^-$ presented an abrupt change with salient points in the last two segments. Note that in this stage the participants were not exposed to stimuli, however, they were aware of

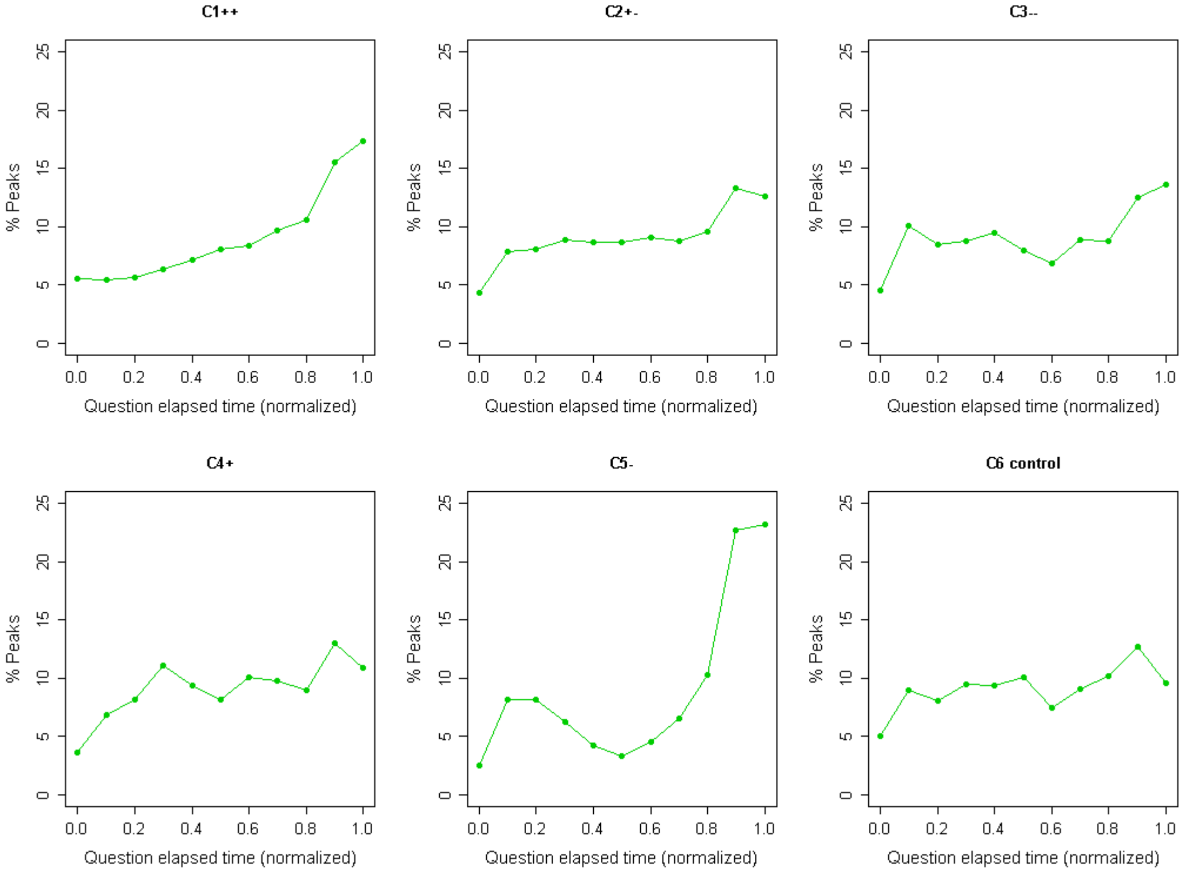


Figure 5.22. Peaks detected in EDA at different stages of the development of questions responded during the evaluation of prolonged effects.

their remaining time for each question and for the overall task. Although the participants in $C2^{-}$ (as part of $C2^{+-}$) and $C3^{--}$ also started this stage under negative affective states, it could be possible that social interactions helped the participants in these conditions to mitigate the effects of time pressure. Another explanation of this prominent concentration of peaks in $C5^{-}$ could be related to the participants' feelings of satisfaction or distress as a result of finding or not finding the necessary information to answer the questions.

In a similar fashion to the analyses conducted with respect to smiles, EDA was also analyzed in the context of relevant and non-relevant page. Results for this analysis are presented in Figure 5.23. Unlike the results obtained for smiles, peaks in EDA presented similar distributions for both relevant and non-relevant pages in conditions $C2^{+-}$, $C3^{--}$, and $C4^{+}$. For the particular case of $C1^{++}$, the concentration of peaks increased progressively in relevant pages and decreased progressively in non-relevant pages. Yet, the distributions of peaks in this condition was very similar to those found in $C2^{+-}$, $C3^{--}$, and $C4^{+}$.

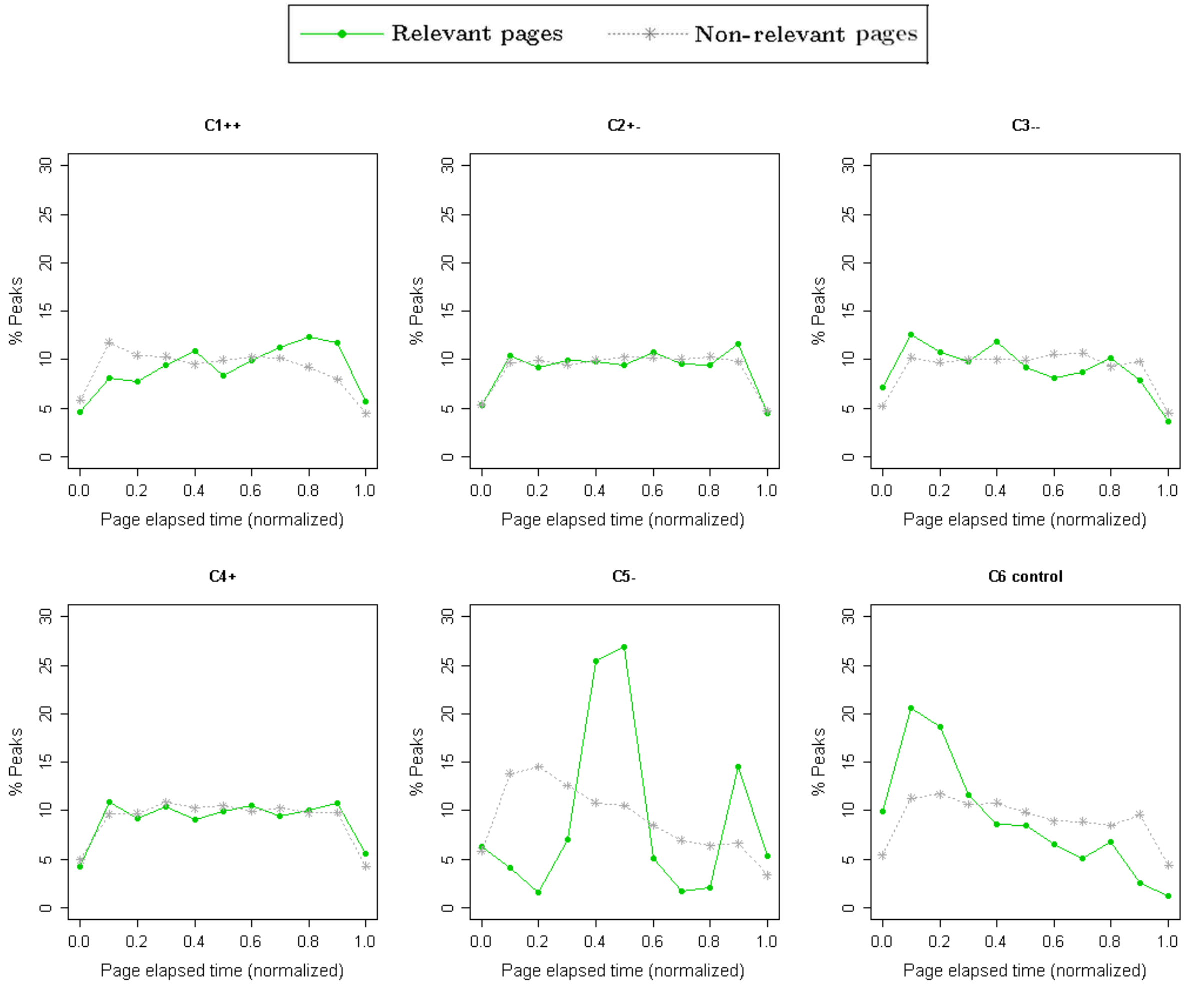


Figure 5.23. Peaks detected in EDA during the exposure to the content of relevant and non-relevant pages.

As shown in Figure 5.23, only $C5^{-}$ and $C6^{control}$ presented different patterns in the concentration of peaks found in EDA. Specifically, $C5^{-}$ displayed prominent agglomeration of peaks in the middle and also in the last moments of exposure to the content of relevant pages. To examine the source of this pattern, additional analyses were carried out with data collected during the evaluation of short-term effects. Results showed that the participants in $Stim^{-}$ (who were later assigned to $C2^{-}$, $C3^{--}$, and $C5^{-}$) presented a pattern of EDA peaks similar to those found in $Stim^{+}$ and $Stim^{control}$ in the pre-stimuli evaluation ($PreS\_q_1$) and during the stimuli stage. However, results for the post-stimuli evaluation ($PostS\_q_{10}$) revealed that the pattern of EDA peaks in $Stim^{-}$ (Figure 5.24) was similar to that found in $C5^{-}$. This indicates that either affective stimuli or induced negative affective states had an effect in the affective reactions expressed through EDA. The fact that a similar pattern was found later in $C5^{-}$, but not in $C3^{--}$ and $C2^{-}$ could be explained by the absence of social interactions in $C5^{-}$. Similar to the observations made with regard to smiles, it is possible that the initial negative affective states influenced the way people processed and

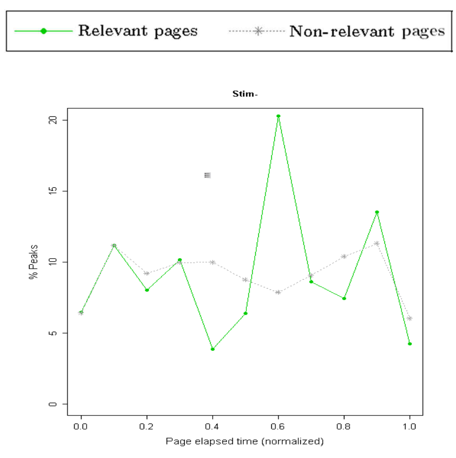


Figure 5.24. Peaks detected in EDA during the exposure to the content of relevant and non-relevant pages in the post-stimuli evaluation ($PostS\_q_{10}$).

reacted to information, which was also expressed at the physiological level.

On the other hand, in $C6^{control}$ the concentration of peaks was high in the first seconds of exposure to the content of relevant pages, but then it abruptly decayed. Perhaps, initial peaks in EDA could indicate the first reactions to the content of relevant pages, which were not observed in non-relevant pages. Yet, it is not clear why the participants in this condition presented this particular pattern that differed from those found in the other individual and also collaborative conditions. One possible explanation for such a difference is that regardless of the exposure to affective stimuli in the previous stage (i.e. evaluation of short-term effects), the participants in $C6^{control}$ started this stage in either positive or negative affective states, thus aggregated results displayed combined patterns.

As shown in Figure 5.23, with the exception of $C1^{++}$ and $C5^{-}$, peaks were not concentrated primarily in the last segments of exposure to the content of relevant pages. This indicates that peaks in EDA were not necessarily attributed to local findings of relevant content, which were explicitly indicated by snipping text from webpages. Indeed, additional analyses were conducted in order to investigate if the snip action was preceded (five seconds before), accompanied (at the exact moment), or followed (five seconds after) by peaks in EDA. Unlike the results for smiles, peaks in EDA occurred very few times (less than 2%) around the snip action. Results for these analyses are summarized in Table 5.16.

Table 5.16. Results for peaks in EDA conducted with respect to the snip action.

	Peaks in EDA	
Condition	**At snip**	**At snip ± 5 sec**
$C1^{++}$	0.79%	0.74%
$C2^{+-}$	1.14%	1.15%
$C3^{--}$	1.90%	1.65%
$C4^{+}$	1.98%	1.79%
$C5^{-}$	0.14%	0.09%
$C6^{control}$	1.39%	1.25%

Analyses similar to those conducted with relevant and non-relevant pages were carried out with useful and non-useful pages. As shown in Figure 5.25, results in this case were rather similar to those described above. These results show that EDA, unlike smiles, does not seem to be a reliable indicator of the relevancy and usefulness of the content of webpages (i.e. local level). Conversely, results obtained at the question level showed that this physiological signal could be a

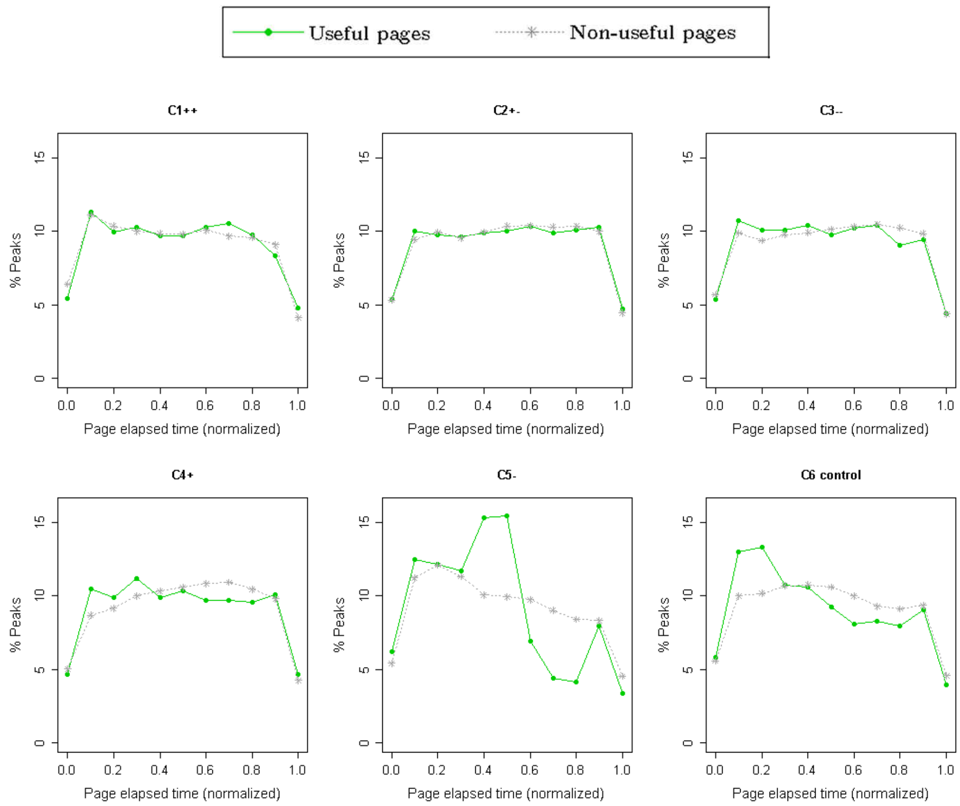


Figure 5.25. Peaks detected in EDA during the exposure to the content of useful and non- useful

good indicator of success (along with smiles) or failure of the information search process (i.e. global level) in both collaborative and individual settings.

As discussed in the evaluation of short-term effects, neither facial expressions nor EDA account for the subjective experience of affective processes. To address this limitation, the next section investigates participants' self-reported feelings as a subjective component of the affective dimension.

5.4.2.3 Affective processes expressed through self-reports

Analyses with regard to the subjective component of participants' affective states were conducted with data collected through the Self-Assessment Manikin (SAM) questionnaire. Likewise in the evaluation of short-term affects, self-reported feelings were studied with regard to valence, activation, and dominance. Between-group comparisons were conducted with the Kruskal-Wallis test and post-hoc analyses were carried out with the Wilcoxon rank-sum test.

Results showed that valence was different only at the moment of starting the evaluation of prolonged effects (i.e. right before starting MT_q_1). In particular, it was found that valence was significantly higher at p<.01 in $C1^{++}$ (*Mdn*=5.5) than the other conditions, except $C4^{+}$. Indeed, valence in $C4^{+}$ (*Mdn*=6) was significantly higher at p<.01 than in all the other conditions. Moreover, valence in $C2^{+}$ (*Mdn*=5) was significantly higher at p<.01 than that found in conditions in which the participants were previously treated with negative stimuli (i.e. $C2^{-}$, $C3^{--}$, and $C5^{-}$). Similarly, valence in the control group ($C6^{control}$) was found to be significantly higher (*Mdn*=5) at p<.01 than that found in $C2^{+}$, $C2^{-}$, $C3^{--}$, and $C5^{-}$. With the exception of $C2^{+}$, this results places $C6^{control}$ almost between positive and negative groups. Comparisons conducted with regard to groups previously treated with negative stimuli showed that the valence in $C2^{-}$ (*Mdn*=4) was significantly lower at $p<.01$ than that found in $C3^{--}$ (*Mdn*=5) and $C5^{-}$ (*Mdn*=5). With respect to $C3^{--}$ and $C5^{-}$, no significant differences were found. The fact that valence was found to be significantly higher in positive groups and also in the control group than in negative groups confirms the efficacy (from the participants' point of view) of the stimuli stage in inducing initial affective states.

After completing MT_q_1 and the following questions in the evaluation of prolonged effects, no significant differences were found with regard to self-reported valence. While it may be possible that participants' affective states could have reached a stable point in the long run (either because of the absence of stimuli, the presence of social interactions, or participants getting used to the task), it is also plausible that the participants have become used to the SAM questionnaire or tired of having to answer it after each question. Either way, this could have caused that at some point of this stage most of the participants' answers were placed in the center of the scales for the three affective dimensions (i.e. valence, activation, and dominance).

Between-group comparisons with respect to activation showed that this was significantly higher at $p<.01$ in $C2^{-}$ (*Mdn*=5) than in all the other conditions right before starting MT_q_1. Conversely, activation in $C6^{control}$ (*Mdn*=4) was significantly lower at $p<.01$ than in the other groups. Significant differences at $p<.01$ with respect to the other groups were organized as follow: $C2^{+}>C1^{++}>C3^{--}>C5^{-}>C4^{+}$ (*Mdn*=5 in all five groups).

Following MT_q_1, differences with respect to activation were found only after MT_q_4 was completed. In this case, activation in $C2^{-}$ (*Mdn*=6) was found to be significantly higher at $p<.01$ than that found in the other conditions. On the contrary, activation levels in $C6^{control}$ (*Mdn*=4) were found to be significantly lower at $p<.01$ than in all the other conditions. The relations between the other conditions (*Mdn*=5 in all five groups) are the following: $C2^{+}>C1^{++}>C3^{--}>C4^{+}>C5^{-}$. Activation levels after MT_q_4 constituted the initial levels of activations for MT_q_5.

With regard to dominance, no significant differences were observed during the entire evaluation of prolonged effects. The overall results are summarized in Table 5.17. Plots for each dimension with the means and standard errors in each question are provided in Appendix F.

Table 5.17. Summary for self-reported feelings. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with individuals as unit of analysis. Results in cells are significant at $p<.01$.

		Question		
		MT_q_1	**MT _q_4**	**MT _q_5**
Before (Pre)	**Valence**	$C1^{++}>C2^{+},C2^{-},C3^{--},C5^{-},C6^{control}$ $C2^{+}>C2^{-},C3^{--},C5^{-}$ $C3^{--}>C2^{-}$ $C4^{+}>C1^{++},C2^{+},C2^{-},C3^{--},C5^{-},C6^{control}$ $C5^{-}>C2^{-}$ $C6^{control}>C2^{+},C2^{-},C3^{--},C5^{-}$		
	Activation	$C1^{++}>C3^{--},C4^{+},C5^{-},C6^{control}$ $C2^{+}>C1^{++},C3^{--},C4^{+},C5^{-},C6^{control}$ $C2^{-}>C1^{++},C2^{+},C3^{--},C4^{+},C5^{-},C6^{control}$ $C3^{--}>C4^{+},C5^{-},C6^{control}$ $C4^{+}>C6^{control}$ $C5^{-}>C4^{+},C6^{control}$		$C1^{++}>C3^{--},C4^{+},C5^{-},C6^{control}$ $C2^{+}>C1^{++},C3^{--},C4^{+},C5^{-},6^{control}$ $C2^{-}>C1^{++},C2^{+},C3^{--},C4^{+},C5^{-},C6^{control}$ $C3^{--}>C4^{+},C5^{-},C6^{control}$ $C4^{+}>C5^{-},C6^{control}$ $C5^{-}>C6^{control}$
	Dominance			
After (Post)	**Valence**			
	Activation		$C1^{++}>C3^{--},C4^{+},C5^{-},C6^{control}$ $C2^{+}>C1^{++},C3^{--},C4^{+},C5^{-},6^{control}$ $C2^{-}>C1^{++},C2^{+},C3^{--},C4^{+},C5^{-},C6^{control}$ $C3^{--}>C4^{+},C5^{-},C6^{control}$ $C4^{+}>C5^{-},C6^{control}$ $C5^{-}>C6^{control}$	
	Dominance			

Finally, in addition to investigating responses to SAM, the overall frustration levels were measured at the end of the evaluation of prolonged effects based on responses to the sixth question of the NASA-TLX questionnaire (Hart & Staveland, 1988) (i.e. "How insecure, discouraged, irritated, stressed, and annoyed were you?"). Between-group comparisons showed that frustration levels were significantly higher at $p<.01$ in $C2^{+}$ (*Mdn*=15) than in all the other groups. On the other hand, the lowest levels of frustration were found in $C1^{++}$ (Mdn=7.5), followed by $C4^{+}$ (*Mdn*=9) and $C6^{control}$ (*Mdn*=11).

In general, conditions in which the participant were initially treated with negative stimuli showed higher levels of frustration than in positive groups, except $C2^{+}$. The relation between negative conditions was as follow: $C2^{-}$ (*Mdn*=13) > $C3^{--}$ (*Mdn*=12) > $C5^{-}$ (*Mdn*=11). The fact that frustration levels were significantly higher at $p<.01$ in $C2^{+-}$ and $C3^{--}$ than in the baseline conditions, shows the negative implications of having pairs in which at least one of the team members started working on the main task in negative affective states. It is also interesting to note the implications for participants who originally started in positive affective states ($C2^{+}$), which after having to work with someone in negative affective states ($C2^{-}$) increased their levels of frustration. On the contrary, having pairs in which both members commenced the main task in positive affective states contributed to keep or reduce the frustration levels. The complete list of results for frustration is presented in Table 5.18.

With respect to the other dimensions measured through the NASA-TLX questionnaire (i.e. mental effort, physical effort, time pressure, performance, and

Table 5.18. Summary for frustration levels. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with individuals as unit of analysis. Results in cells are significant at *p*<.01.

Frustration levels	$C2^{+}>C1^{++},C2^{-},C3^{--},C4^{+},C5^{-},C6^{control}$ $C2^{-}>C1^{++},C3^{--},C4^{+},C5^{-},C6^{control}$ $C3^{--}>C1^{++},C4^{+},C5^{-},C6^{control}$ $C4^{+}>C1^{++}$ $C5^{-}>C1^{++},C4^{+},C6^{control}$ $C6^{control}>C1^{++},C4^{+}$

task difficulty), no significant differences were reported by the Kruskal-Wallis test.

Finally, the affective load (AL) of participants was estimated in accordance to Equation 3.16 introduced in section 3.2.2.

$$AL = Uncertainty * TimePressure$$

Since uncertainty in this formulation corresponds to levels of irritation as a result of anxiety, frustration, or rage experienced during a search task, this component was expressed by answers to the same question of the NASA-TLX analyzed above for frustration levels. On the other hand, time pressure was expressed by the answers to the third question of this questionnaire (i.e. "How hurried or rushed was the pace of the task?"). Participants' responses to these two questions were multiplied and then compared across groups with the Kruskal-Wallis test. Significant results at $p<.01$ found with the Wilcoxon rank-sum test as a post-hoc test are presented in Table 5.19. Unlike results for the previous analysis, AL was found to be significantly higher at $p<.01$ in $C1^{++}$ (*Mdn*=89) than that found in the corresponding baseline $C4^{+}$ (*Mdn*=80). This particular difference can be attributed to the presence of social factors in the former. Also, AL was significantly higher at $p<.01$ in $C2^{-}$ (*Mdn*=228) than in

Table 5.19. Summary for affective load. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with individuals as unit of analysis. Results in cells are significant at p<.01.

Affective load	$C1^{++}$>$C4^{+}$ $C2^{+}$>$C1^{++}$,$C3^{--}$,$C4^{+}$,$C5^{-}$,$C6^{control}$ $C2^{-}$>$C1^{++}$,$C2^{+}$,$C3^{--}$,$C4^{+}$,$C5^{-}$,$C6^{control}$ $C3^{--}$>$C1^{++}$,$C4^{+}$,$C6^{control}$ $C5^{-}$>$C1^{++}$,$C3^{--}$,$C4^{+}$,$C6^{control}$ $C6^{control}$>$C1^{++}$,$C4^{+}$

$C2^{+}$ (*Mdn*=180), which shows that despite the high levels of frustration reported for $C2^{+}$, participants in this condition were less affected by the pressure of time than their partners. Another interesting result is that AL was found to be significantly higher at p<.01 in $C5^{-}$ (*Mdn*=165) than in $C3^{--}$ (*Mdn*=162). As discussed earlier, although participants in both conditions initiated the task in negative affective states, it seems to be that social factors contributed to alleviate their impact in the long run.

5.4.2.4 Affective processes expressed through verbal communication

Unlike the evaluation of short-term effects, the incorporation of collaboration in this stage provided an alternative way to investigate expressed affective states. This approach focused on examining communication with the aim of evaluating whether initial affective states influenced the verbal expressivity of the participants when interacting with their teammates. Since communication was only present in collaborative conditions, these analyses were carried out with $C1^{++}$, $C2^{+-}$, and $C3^{--}$.

The results presented below expand those reported earlier in the first research question. Specifically, this section focuses on the tone (or valence) of messages - which was coded by human coders as positive, negative, or neutral - and also on specific types of words contained in the extended version of the LIWC dictionary (Pennebaker et al., 2001).

Although the first analysis did not show significant differences at $p<.05$, this section provides additional details about the way searchers in different affective states express while collaborating. Regardless of the type of initial affective states, it was found that messages were mostly neutral (objective). Moreover, participants expressed similar number of positive and negative messages during their search processes. In general, the participants sent one or two positive messages (e.g. “shit u put it down lol”, “sounds good”, “oh you did it? lol”) and one with a negative tone (e.g. “no idiot”, “wtf is lonesome george”, “rarest animal on earth :(”) while working in each question. Only in question MT_q_6 (“I was a 19th century lawyer who claimed to have killed 42 men. I was killed while playing dice by a man who was killed over a card game. Who am I?”) and MT_q_7 (“The painter of Starry Night doused his mattress and pillow with what to help him sleep?”) none of the participants sent negative messages. As discussed in the first research question, participants in these two questions were able to find the answers very quickly. Moreover, for the particular case of MT_q_7, several participants reported being familiar with the topic of the question.

In general the exchange of positive and negative messages was balanced within each pair of participants. However, in $C2^{+-}$ it was interesting to observe that most of the negative message came from participants in $C2^{+}$. Participants in $C2^{-}$ only expressed negative messages during the first two questions. Note that communication volume in $C2^{+}$ and $C2^{-}$ was roughly similar, thus this finding can be related with the high levels of frustration reported in the previous section by the participants in $C2^{+}$. The distributions of neutral, positive, and negative messages in each condition are provided in Appendix F.

The second approach to analyzing affective processes through communication was based on the type of words the participants used in their messages. This analysis was conducted with particular categories of the extended version of the LIWC, which includes categories and words of the WordNet affect (Strapparava & Valitutti, 2004) as well as a list of interjections, punctuation, and emoticons.

Comparisons between groups showed that the participants in $C1^{++}$ used more swear words (e.g. “damn”, “piss”) than in the other two conditions in MT_q_1 and MT_q_3. However, an inspection of the messages showed that the participants used this kind of words to intend positive meaning (e.g. “good shit”, “shit u put it down lol”, “avengers shit dude”). Results for this analysis are summarized in Table 5.20.

As shown in preliminary studies, text-based communication and time constraints limit the exchange of messages and especially affective expressivity (for more details see section I.2). As described in the first research question, communication in this study was primarily task oriented. It is possible that

Table 5.20. Comparison between users in collaborative conditions

	Question	
	MT_q_1	**MT_q_3**
Swear_22	$C1^{++}>C2^{+-},C3^{--}$	$C1^{++}>C2^{+-},C3^{--}$

richer communication channels (e.g. audio, face-to-face, or video) that enable searchers to communicate and search information at the same time, would facilitate the expressivity of participants.

5.5 Research question 3: Affective processes and performance

This section presents the analyses performed to address the third research question (RQ3) of this dissertation, which focuses on the implications, if any, of initial affective states and those derived from collaboration on performance. The research question was introduced in Chapter 1 as follow:

> **RQ3**: To what extent, if any, do initial positive and negative affective states and those derived from the collaboration of individuals in an information search task influence team performance?

So far the analyses presented in the previous chapters have not paid attention to the overall performance of participants. The first research question focused on general aspects that can be related to performance such as number of queries, information coverage, precision, recall, efficiency, and effectiveness; however, none of these measures really indicate how well or bad the participants did in the tasks. To address this research question, performance was investigated in terms of the following three measures:

1. *Response time*: This measure corresponds to the time that it took to the participants to find the answers to the *A Google a Day* questions. The lower the response time the higher the performance. Note that this measure is independent of the quality of the answer.

2. *Query precision*: This measure shows to what extent the participants used queries similar or equal to those specified in the baseline search path (or ideal search path) established by *A Google a Day* for each question. The higher query precision the higher the performance. Similar to response time, this measure is independent of the quality of the answer.
3. *Response precision*: This measure focuses in the overall quality of the work done by pairs and individual. It is expressed by the proportion between correct answers and the overall number of questions addressed by each pair or individual. The higher the response precision the higher the performance. Unlike the previous two measures, this one takes into account the quality of the answer.

The three analyses were carried out in the evaluation of both short-term and prolonged effects.

5.5.1 Evaluation of short-term effects

The evaluation of short term effects in this case focuses on investigating to what extent initial affective states had an impact on the performance of the participants. The following sections presents the results for the three types of analyses introduced above.

5.5.1.1 Response time

Response time was defined as the time that it took the participants to find the answer to the *A Google a Day* questions from the moment they were exposed to the question for the first time until the moment they submitted their answers. The analysis in this case was conducted with the questions in the following stages: pre-stimuli evaluation, stimuli stage, and post-stimuli evaluation. On

Table 5.21. Descriptive measures for response time in the three conditions. Values in cells are expressed in seconds.

Condition	**Mdn**	**M**	**SD**
$Stim^{+}$	165.44	179.07	88.99
$Stim^{-}$	158.44	171.92	90.61
$Stim^{control}$	172.41	179.07	96.78

average, the participants spent less than three minutes in each question. Moreover, as the session progressed, the participants spent less time in questions with the aim of responding more questions. This behavior was common across all the three conditions. Indeed, analyses with the Kruskal-Wallis test did not show significant difference between the three groups in any of the questions of this stage. These results show that neither affective stimuli nor the affective states induced on the participants had effects in response time. A summary of descriptive measures for this analysis is provided in Table 5.21.

While response time indicates how efficient the participants were in providing the answers, this does not show how well they used their time or if they actually provided the right answers. The next sections examine these two aspects in detail.

5.5.1.2 Query precision

The analysis presented in this section focuses on the ability of the participants to formulate precise queries that lead them to find the right answer. As described in section 4.3 the *A Google a Day* questions used in this study are characterized by ideal search processes that involve two queries. Based on the narrative of these ideal search paths, the two queries required in each question denote two major steps required to find the answers (this under the assumption that searchers are unfamiliar with the topics of the questions). On one hand, the first query gives access to partial information that helps searchers to formulate the following query. On the other hand, the second query lead searchers to the answers themselves.

With regard to the assumption about topic familiarity, it was found in this study that 96.76% of the time the participants reported being very unfamiliar (less than four points in the familiarity scale presented to the participants at the beginning of each question) with the topic of the questions presented in the evaluation of short-term effects. Moreover, as shown in the first research question, no significant differences were found between conditions with respect to topic familiarity. Because of this aspect in addition to specific settings in Google (for more details, see section 4.7.3), it was presumed that the participants would not be able to find the answer, in most cases, with less than two queries.

To investigate query precision, participants' queries in each question were first contrasted with those found in the corresponding ideal search paths (also referred to as baseline queries). Comparisons were carried out with a string similarity function based on the Levenshtein edit distance[53], whose results are expressed in a range from 0 (very dissimilar) to 1 (exact match). Then, results from this comparison were added up and dived by the total number of queries. This ratio, referred to as query precision (a task-dependent measure), was formalized as follow:

---

[53] *levenshteinSim:* http://www.inside-r.org/packages/cran/RecordLinkage/docs/winkler

$$Query\ precision = \frac{\sum levenshteinSim(participant's\ queries, baseline\ queries)}{n\{participant's\ queries\}} * 100 \qquad \text{(Eq. 5.1)}$$

The outcomes for query precision not only indicate if the participants used queries similar to those in the ideal search path, but it also accounts for the number of distinct queries. For example, in a given question if participants issued two different queries that matched perfectly with the baseline queries, then their query match for that question was 100%. Any additional query that differs, even slightly, from the baseline ones will impact negatively query precision.

Results for query precision in this stage are summarized in Figure 5.26. As shown in this plot, high query precision (greater than 90%) in this stage was unusual (less than 1.2% of the time) among the participants. Regardless of the stimuli to which the participants were exposed during this stage, in most cases the similarity between participants' queries and those in the ideal search path ranged between 30% and 40%. This low score indicates that the participants' queries were not well aligned with the ideal search paths associated with each question. Note that between-group comparisons with regard to this measure

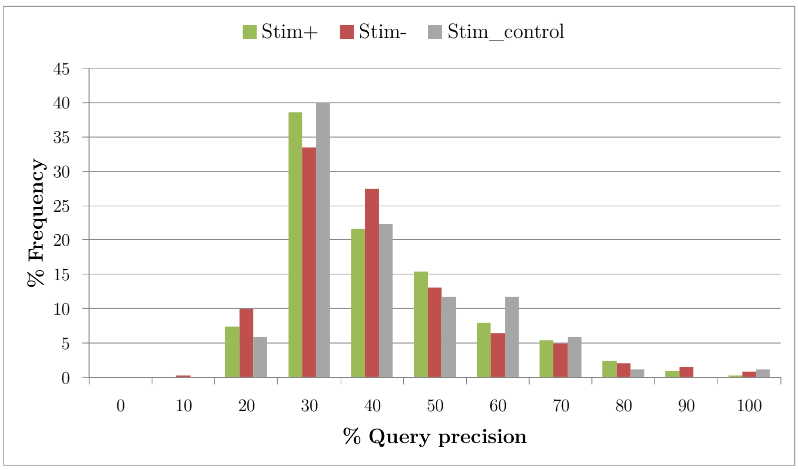


Figure 5.26. Query precision during the evaluation of short-term effects.

(performed with the Kruskal-Wallis test) in each question, did not report significant differences at $p<.05$.

As shown in the results for the first research question, the number of queries in Stim+ and Stim⁻ was significantly reduced at $p<.01$ from the pre-stimuli to the post-stimuli evaluation (*Mdn*=3 in both groups). This indicates that although query precision was low during the evaluation of short-term effects, participants were able to find alternative search paths, which in some cases were more efficient and effective than those established in the corresponding baselines (i.e. ideal search paths). In fact, three out of the four observations with 100% query precision corresponded to participants who were not able to find an answer. Answers in the remaining observation were completely wrong. On the other hand, 37.63% of the observations with query precision between 30% and 40% were linked to answers partially or completely correct. It was also found that two observations with query precision below 17% were related to right answers. Figure 5.27 summarizes query precision with respect to the quality of the answers. Assuming that values 0 to 5 in the coding scheme were equidistant as

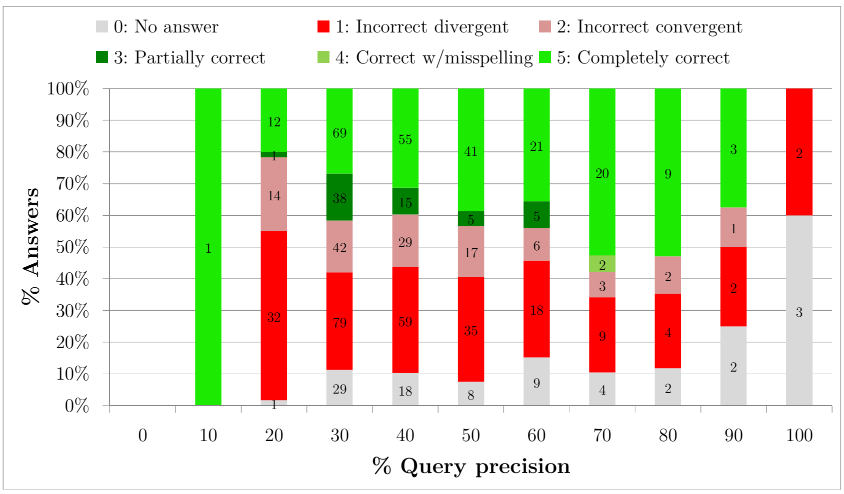


Figure 5.27. Query precision with respect to answer quality during the evaluation of short-term effects. Values in bars correspond to the actual number of answers.

part of a rating scale, query precision and quality of the answers in the three conditions were found to be not correlated ($0<r<.12$). This indicates that high levels of query precision did not ensure finding the right answers. Instead, the participants were more effective in finding the answers through search heuristics involving queries that partially matched the baseline ones. The following section provides further analyses with regard to the quality of the participants' answers in the evaluation of short-term effects.

5.5.1.3 Response precision

The third analysis conducted to evaluate the performance of the participants focused on the quality of the work. To study this aspect, participants' answers were compared with those provided by *A Google a Day* (for additional details about questions and the corresponding answers, see Table 4.2). Moreover, participants' search processes in each question were inspected with regard to the queries formulated and the pages visited. These two analyses derived in the coding scheme presented early in this chapter (see Table 5.3). The number of observations during the first task (evaluation of short term effects) that fell under each coding category is provided in Table 5.22.

In accordance to the results of this first analysis, participants' response precision was computed. Response precision (also a task-dependent measure) in this study was defined as the ratio between the total number of correct answers (partially correct, correct with spelling mistakes, and completely correct answers) divided by the total number of answers provided by each participant.

Table 5.22. Scale to rate the quality of answers and the corresponding number of observations during the evaluation of short-term effects.

Code	**Description**	**Observations short-term**
5	Correct answer	231
4	Correct with spelling mistakes	2
3	Partially correct (e.g. in a list of 3 names, only 2 are provided)	64
2	Incorrect search answer with convergent search process	114
1	Incorrect search answer with divergent search process	243
0	No answer	79
-1	Discarded because of insufficient time	2
-2	Discarded because of cheating	37

$$Response\ precision = \frac{Correct\ Answers}{Total\ number\ of\ answers} \qquad \text{(Eq. 5.2)}$$

Results from this analysis did not show significant differences among the three groups. As illustrated in the box plots in Figure 5.28, participants in the three groups responded on average the same number of questions (four questions). With regard to response precision, this was roughly the same (*Mdn*=0.33, M=0.34, SD=0.21) in both Stim^{+} and Stim^{-}. On the other hand, response

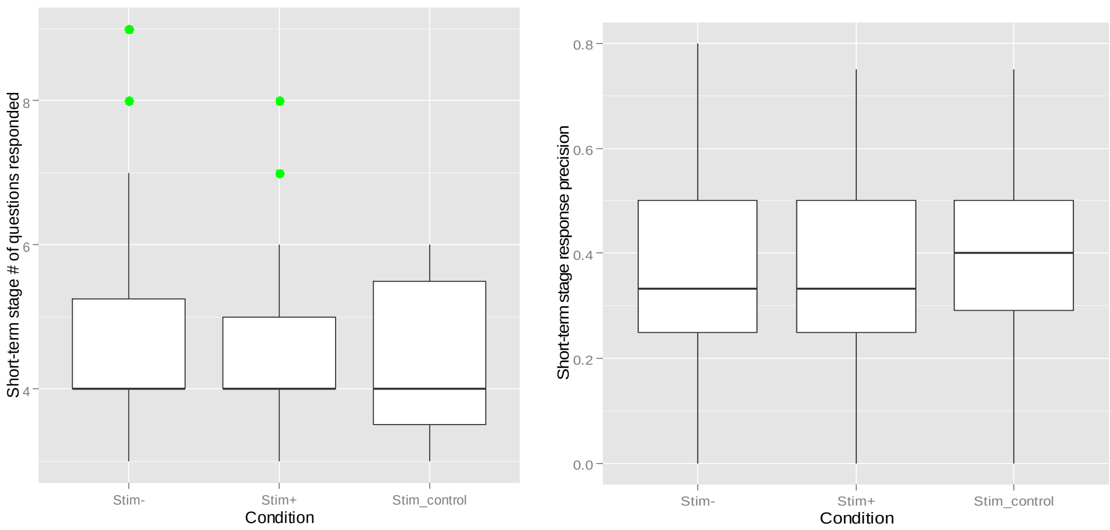


Figure 5.28. Box plots for number of questions responded (left) and response precision (right) in each condition during the evaluation of short-term effects.

precision in $Stim^{control}$ was slightly higher, but not enough to be significantly different from the other two conditions. Note that this evaluation considered aggregated values for questions responded in the entire stage and not just $PostS\_q_{10}$ as a post-stimuli evaluation. Therefore, these results indicate that stimuli did not have short-term effects in the number of questions addressed by the participants or in their response precision.

The following section expands the analyses presented above by exploring prolonged effects of initial affective states in collaborative and non-collaborative conditions.

5.5.2 Evaluation of prolonged effects

This section presents results for the same performance measures introduced above. However, in this case the goal is to investigate prolonged effects, if any, of initial affective states and those derived from collaboration in the performance of teams. The following sections present the results for response time, query precision, and response precision.

5.5.2.1 Response time

As explained above, response time was defined as the time that it took participants to find the answer to the *A Google a Day* questions from the moment they were exposed to the question for the first time until the moment they submitted their answers. On average, participants and pairs spent less than 190 seconds working on these questions (Table 5.23). In general terms, results per question showed that individuals and pairs responded questions in roughly the same amount of time; however, in MT_q_3 ("In Norse mythology, which god's son will poison Thor at Ragnarök?"), response time was significantly

Table 5.23. Descriptive measures for response time in the six conditions. Values in cells are expressed in seconds.

Condition	Mdn	M	SD
$C1^{++}$	168.71	172.72	82.64
$C2^{+-}$	174.67	185.16	88.93
$C3^{--}$	146.75	155.07	88.46
$C4^{+}$	140.58	163.65	89.40
$C5^{-}$	145.74	151.25	77.82
$C6^{control}$	134.07	151.54	80.14

higher for pairs than for individuals. As discussed in the first research question, an interesting aspect about MT_q_3 was the interpretation of the question. For the case of pairs, different interpretations led the participants to explore more pages in order to reach consensus, which resulted in an increment in response time. Since participants in individual conditions did not have to match their own interpretation with someone else's, they were able to respond the questions quicker. It is also interesting to note that participants who initiated the second task in negative affective states spent less time working on this question. Specifically, response time in $C3^{--}$ (*Mdn*=202.63s) was significantly lower at $p<.01$ than in the other two collaborative conditions. On the other hand, response time in $C1^{++}$ (*Mdn*=220.24s) was significantly higher at $p<.01$ than in the other two collaborative conditions.

At the individual level, participants who were treated with negative stimuli in the previous stage ($C5^{-}$) spent significantly less time (*Mdn*=147.38s), $p<.01$, than in all the other conditions. Similar to collaborative conditions, response time in $C4^{+}$ (*Mdn*=188.41s) was significantly higher at $p<.01$ than that found in $C5^{-}$ and control group (*Mdn*=159.32s). According to the affect infusion model

Table 5.24. Summary for response time. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with pairs as unit of analysis. Results in cells are significant at p<.01.

	Question
	MT_q3
Response time	$C1^{++}>C2^{+-},C3^{--},C4^{+},C5^{-},C6^{control}$
	$C2^{+-}>C3^{--},C4^{+},C5^{-},C6^{control}$
	$C3^{--}>C4^{+},C5^{-},C6^{control}$
	$C4^{+}>C5^{-},C6^{control}$
	$C6^{control}>C5^{-}$

(see section 3.1.2), it may be possible that initial affective states accounted for specific information processing strategies chosen by the participants to tackle MT_q3. Although individual participants did not have to agree with others in the interpretation of this question, it is possible that they were still confused by the ambiguity of this question. Perhaps, while the participants in $C4^{+}$ dedicated more time deciding how to interpret and address MT_q3, those in $C5^{-}$ were biased by the principle of least effort, thus dedicating less time. A summary of the results for this analysis is provided in Table 5.24.

5.5.2.2 Query precision

The analysis presented in this section focuses on the ability of the participants and pairs to formulate precise queries to find the right answer. As described above, this analysis was carried out in each question by comparing participants' queries with those in the search path provided by *A Google a Day*.

Results for query precision in the evaluation of prolonged effects are summarized in Figure 5.29. Similar to the results obtained in the evaluation of short-term effects, high query precision (greater than 90%) was also unusual among the

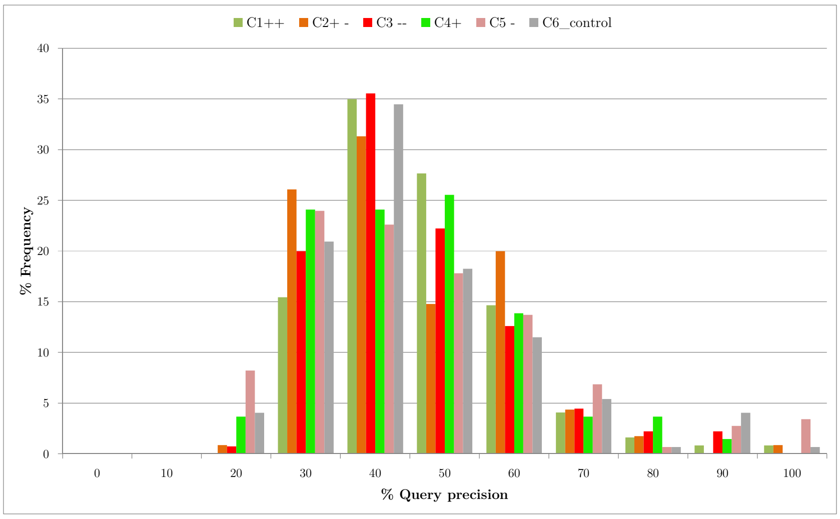


Figure 5.29. Query precision during the evaluation of prolonged effects.

participants in this stage (less than 5% of the time). However, unlike the previous stage it was found that 62.5% (five out of eight) of the observations with high query precision were linked correct answers. Figure 5.30 provides detailed results for query precision with respect to the quality of the answers. Assuming that values in the coding scheme were equidistant as part of a rating scale, query precision and quality of the answers in the six conditions were found to be not correlated ($0<r<.1$). This indicates that high levels of query precision did not ensure finding the right answers.

As shown in Figure 5.29, the major concentrations of query precision ranged between 30% and 60%. At the same time, the majority of correct answers were associated to participants and pairs whose query precision fell in this range (Figure 5.30). Between-group comparisons with the Kruskal-Wallis test reported significant differences for query precision only in the seventh question (MT_q_7: "The painter of Starry Night doused his mattress and pillow with what to help him sleep?"). The two baseline queries for this question according to the search path suggested by *A Google a Day*

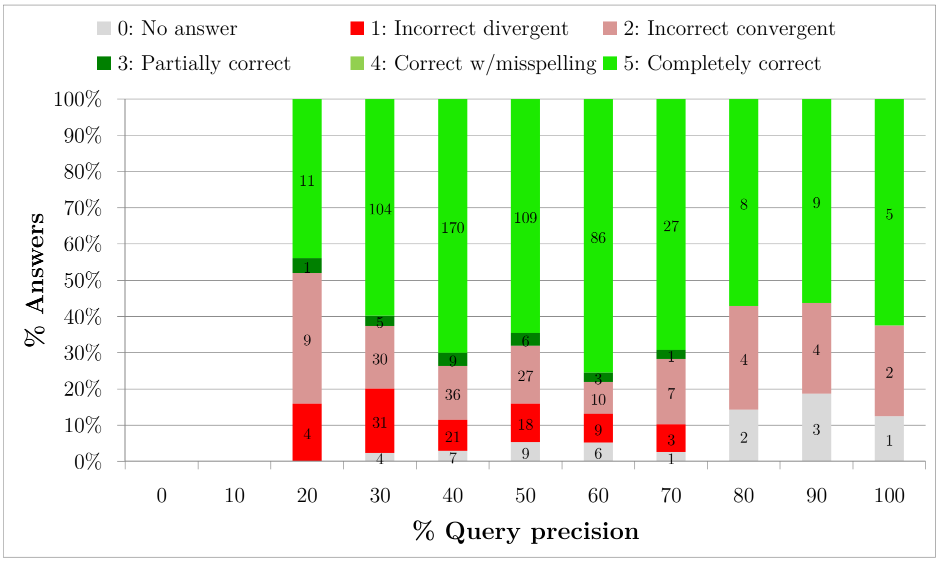


Figure 5.30. Query precision during the evaluation of prolonged effects. Values in bars correspond to the actual number of answers.

are: (1) "painter Starry Night" and (2) "van Gogh doused his mattress." Post-hoc analyses conducted with the Wilcoxon rank-sum test showed that query precision in $C5^-$ (*Mdn*=56.85%) was significantly higher at $p<.01$ than in the other conditions. On the other hand, results for this measure were significantly lower at $p<.01$ in $C4^+$ (*Mdn*=38.39%) than in all the other groups. With respect to the relation between query precision and the quality of the response, it was found that all the participants in $C5^-$ who reached this question (N=14) found the right answer. For the case of $C4^+$, 10 participants out of 12 found the right answer. Interestingly, the participant with the highest query precision (78.26%) did not provide an answer. Overall 66.45% of the participants found the right answer and 3.06% provided answers that were partially correct. As mentioned in the development of the first research question, the topic of MT_q_7 was one of the few ones that participants reported being familiar with. Therefore, regardless of query precision, topic familiarity could be the main reason of the high percentage of correct answers found in this particular question. Other significant

Table 5.25. Summary for query precision. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with pairs as unit of analysis. Results in cells are significant at $p<.01$.

	Question
	MT_q$_7$
Query precision	$C1^{++}>C2^{+-},C3^{--},C4^{+},C6^{control}$ $C2^{+-}>C3^{--},C4^{+},C6^{control}$ $C4^{+}>C3^{--}$ $C5^{-}>C1^{++},C2^{+-},C3^{--},C4^{+},C6^{control}$ $C6^{control}>C3^{--},C4^{+}$

results found with the Wilcoxon rank sum test in MT_q$_7$ are provided in Table 5.25.

5.5.2.3 Response precision

This section presents performance analyses during the evaluation of prolonged effects with regard to the quality of the answers provided by the participants. Unlike results presented in the previous stage, the proportion of correct answers and wrong answers increased from 0.8 to 2.5. A summary with the number of observations coded under each quality code is provided in Table 5.26.

Table 5.26. Scale to rate the quality of answers and the corresponding number of observations during the evaluation of prolonged effects.

Code	**Description**	**Observations short-term**
5	Correct answer	529
4	Correct with spelling mistakes	0
3	Partially correct (e.g. in a list of 3 names, only 2 are provided)	25
2	Incorrect answer with convergent search process	129
1	Incorrect answer with divergent search process	87
0	No answer	43
-1	Discarded because of insufficient time	0
-2	Discarded because of cheating	2

Results showed that on average individual participants were able to address more questions than pairs. However, no significant differences at $p<.05$ were found in this regard. Moreover, no differences were found with respect to the number of correct answers, however, the overall number of wrong answers was found to be significantly higher at $p<.01$ in $C6^{control}$ (*Mdn*=4) than in other conditions but $C1^{++}$ (*Mdn*=1). A comparison between groups and individuals showed that the latter provided significantly more wrong answers at $p<.01$ than pairs. A closer inspection to incorrect answers showed that $C3^{--}$ (*Mdn*=0) provided significantly less wrong answers with divergent search processes than all other conditions. Regarding response precision (ratio between correct answers and the total number of answers provided) it was found that pairs did significantly better at $p<.01$ than individuals. Moreover, results showed that response precision in $C3^{--}$ (*Mdn*=0.85) was significantly higher at $p<.01$ than all other conditions. On the other hand, response precision in $C4^{+}$ (*Mdn*=0.6) was significantly lower than that found in all the other groups. A list with significant results for the measurements described above is provided in Table 5.27.

Table 5.27. Comparison summary for number of answers and incorrect answers. Results for Wilcoxon rank-sum test (as post-hoc analyses for Kruskal-Wallis) with pairs as unit of analysis. Results in cells are significant at $p<.01$.

	Comparison
# Overall Incorrect answers	$C4^{+}>C2^{+-},C3^{--}$ $C5^{-}>C1^{++},C2^{+-},C3^{--},C4^{+}$ $C6^{control}>C2^{+-},C3^{--},C4^{+},C5^{-}$
# Incorrect answers with divergent search process	$C1^{++}>C3^{--}$ $C2^{+-}>C1^{++},C3^{--}$ $C4^{+}>C1^{++},C2^{+-},C3^{--}$ $C5^{-}>C1^{++},C2^{+-},C3^{--}$ $C6^{control}>C1^{++},C2^{+-},C3^{--}$
Response precision	$C1^{++}>C2^{+-},C4^{+},C5^{-},C6^{control}$ $C2^{+-}>C4^{+},C5^{-}, C6^{control}$ $C3^{--}>C1^{++},C2^{+-},C4^{+},C5^{-},C6^{control}$ $C5^{-}>C4^{+}$

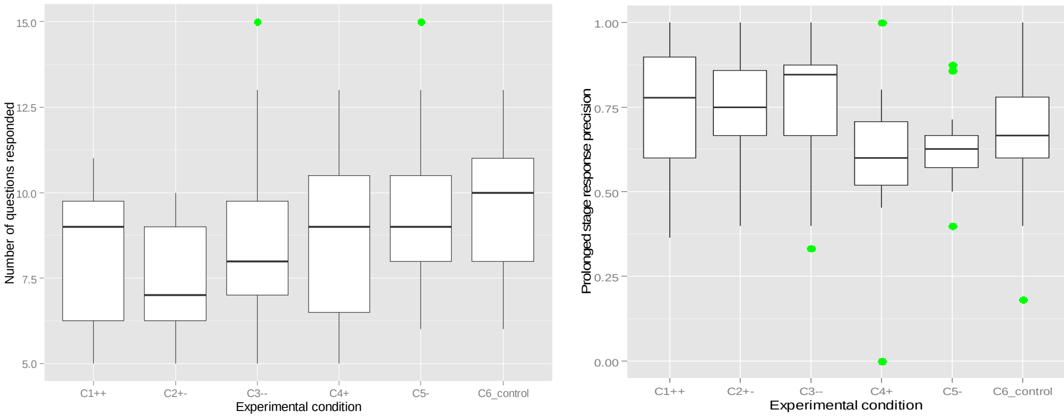


Figure 5.31. Box plots for number of questions responded (left) and response precision (right) in each condition during the evaluation of prolonged effects.

Additionally box plots for response precision and number of questions responded in each condition are provided in Figure 5.31.

Results for response precision indicate that collaboration, regardless of participants' initial affective states, contributed to achieve a higher quality of work, but not quantity. While this could be attributed to pairs spending more time in questions as a result of additional factors such as switching between communication and search, results for response time did not show significant differences among the groups with the exception of MT_q_3. Therefore, instead of extra time, it is possible that participants in pairs were able to compare their individual findings with their teammates', which helped pairs to achieve a better quality of work. As discussed in previous questions, social interactions in questions such as MT_q_2 and MT_q_3 helped participants to make careful decisions when choosing among multiple options. On the other hand, individuals were limited by their personal evaluation criteria and confidence levels, which in some cases was a determining factor of the quality of their answers.

With respect to affective states derived from collaboration, results obtained in the second research question did not show significant differences at the level of positive and negative affective states during the collaboration stage. Significant changes were only observed with regard to activation after completing MT_q_4 and at the same time right before starting MT_q_5. However, as reported above, none of the performance analyses showed significant variations in these two question. Analyses in each question did not show correlation between self-reported valence, activation, or dominance and the quality of answers, query precision, or response time.

At the level of physiological changes, for the particular case of collaborative conditions, it was found a weak correlation (r=.34) between the average *Q Score* and the number of correct answers. Recall that *Q Score* is defined as an implicit measure of engagement or arousal, which is determined by combining features such as number of peaks in electrodermal activity. Figure 5.32 depicts a linear correlation analysis between average *Q score* and the number of correct answers.

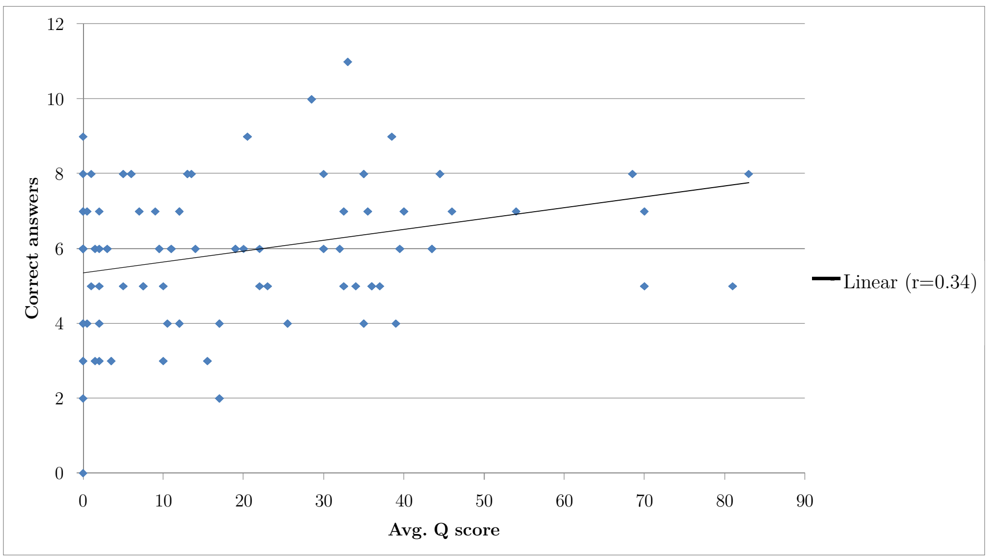


Figure 5.32. Linear correlation between average *Q score* and correct answers.

The results presented above with pairs and individual team members, in particular those for response precision, provide evidence to reject the two hypotheses linked to the third research question introduced in the research framework of this dissertation (section 3.4). These two hypotheses were formulated as follow:

> **Hypothesis 3a:** Participants who receive negative stimuli will reach lower quality of work (response precision) than those who are positively treated.
>
> **Hypothesis 3b:** Participants who receive negative stimuli and collaborate with someone who was also treated with negative stimuli will reach a lower quality of work (response precision) than teams in which one or two members were positively treated.

The first of these hypothesis (3a) states that induced negative affective states will contribute negatively to the quality of work achieved by the participants. However, results presented above showed exactly the opposite, that is to say, it was $C5^{-}$ (participants treated with negative stimuli in the previous stage) the group in which response precision was significantly higher than in the other two individual conditions (i.e. $C4^{+}$ and $C6^{control}$).

Hypothesis 3b states that the quality of work should be better in collaborative pairs in which at least one of the team members was treated with positive stimuli. However, results presented above showed that the quality of the work (response precision) was indeed significantly better in $C3^{--}$ than in the other two collaborative conditions in which both or one team member was previously treated with positive stimuli. Moreover, results also showed that response precision was significantly better in $C1^{++}$ than in $C2^{+-}$, which suggests that mixed initial affective states may affect negatively the quality of the work.

5.6 Research question 4: Positivity ratio in CIS

The analyses and results presented in this section focus on the fourth research question (RQ4) of this dissertation. This research question aims to investigate if the 3-to-1 ratio between positive and negative affective states, which has been evaluated in different domains (including collaborative ones), applies in the context of collaborative information seeking. This question was introduced in Chapter 1 as follows:

> **RQ4**: To what extent, if any, does the relation 3-to-1 between positive and negative affective states (P/N) (Losada & Heaphy, 2004; Fredrickson & Losada, 2005) apply to CIS?

Theory and empirical research in positive psychology have shown that positivity (which encompasses not only positive affective states, but also attitudes and social relations, among others) is a contributing factor in different domains of life such as health, collaboration, and relationships, to name a few. It has been also shown that positivity should be accompanied, in the right proportions, by negativity. In the context of teamwork, Losada and Heaphy (2004) and Fredrickson and Losada (2005) showed that ideally the proportions of positivity and negativity for high performance teams should fall in what is call Losada's zone, which is characterized by a lower limit of 2.9013 and an upper limit of 11.6346. Roughly speaking the lower limit, referred to as 3-to-1 ratio (Fredrickson, 2009), indicates that for every three positive affective states, judgments, attitudes, and so on, there should be one negative. On the other hand, the upper limit indicates that there should not be extreme abundance of positivity with respect to negativity. Beyond the scope of teamwork, it has been shown that the 3-to-1 ratio has practical applications in other contexts such as health and relationships; however, there are no studies in this regard in information seeking or CIS.

To address this research question, the positivity ratio of teams was estimated with two different approaches. First, globally by combining positivity and negative scores derived from the *positivity self test* (see section 4.7.2.2) that the participants individually responded at the beginning and at the end of the sessions. Second, locally by combining levels of positivity and negativity individually reported before and after each question through the SAM questionnaire.

The estimated positivity ratio was analyzed with respect to the global and local performance of teams. Global performance was defined in terms of response precision (i.e. ratio between correct answers and the total number of questions responded). Local performance, on the other hand, was defined in terms of the quality of the answers explained in section 5.2.1 (i.e. completely correct, partially correct, incorrect, and no answer).

In both cases, pairs were classified into three groups: high performance, medium performance, and low performance. This classification was conducted in accordance to the criteria specified in Table 5.28. The following sections present results obtained with each approach.

Table 5.28. Classification of teams based on their local and global performance

Group	Local performance (quality of answer)	Global performance (response precision)
High performance pairs	Correct	0.80 – 1.00
Medium performance pairs	Partially correct	0.60 – 0.79
Low performance pairs	Incorrect and no answers	0.00 – 0.59

5.6.1 Global approach

To estimate the positivity ratio from responses to this questionnaire, Waugh and Fredrickson's (2006) procedure was used, which consists on calculating the ratio between positivity and negativity scores. On one hand, positivity score consists of the number of positive affective terms stated in the questionnaire (e.g. grateful, inspired, and interested) to which participants responded with two or more points (on a scale from 0 to 4). On the other hand, negativity score consists of the number of negative affective terms stated in the questionnaire (e.g. disgust, hate, and embarrassed) to which participants responded with one or more points (on a scale from 0 to 4). Since the analyses were conducted with pairs as unit of analysis, the positivity and negativity scores of both participants in each pair were added and then divided.

The distribution of positivity ratios estimated at the beginning of the session (i.e. with regard to the past 24 hours prior the starting the session) and at the end of the session (i.e. with regard to the affective experiences during the session) is depicted in Figure 5.33. As shown in this figure, very few participants had a positivity ratio equal or higher than 2.9 at the moment of starting the sessions. Similar results were found at the moment of completing the sessions. The major concentration of positivity ratio found in both instances ranged between 1.0 and 2.0, which means that the positivity score of most participants

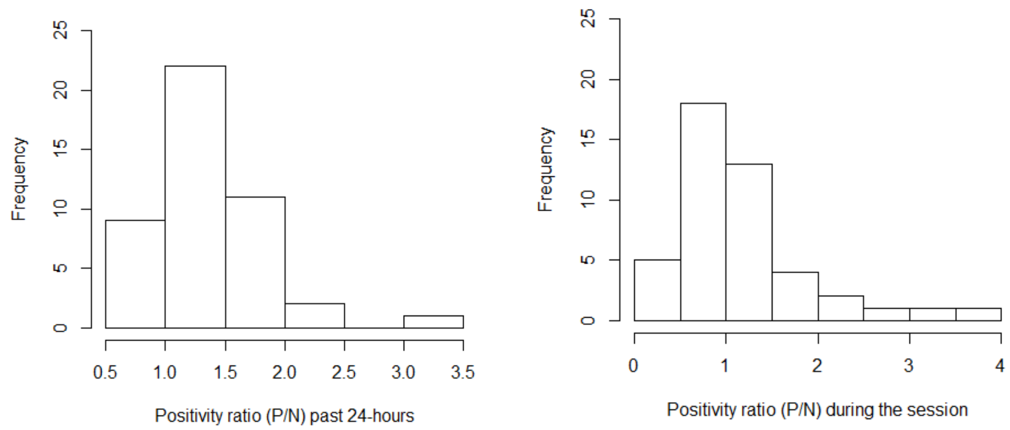


Figure 5.33. Histogram of positivity and negativity scores reported through the *positivity self test* (Fredrickson et al., 2003) in regard to affective states experienced by participants 24 hours before their sessions (left) and during the session (right).

was equal (1:1) or greater (2:1) than their negativity scores. Note that the positivity ratio measured at the beginning of the session is considered to be the result of long-term effects of participants' experiences during the past 24 hours before sessions were carried out. This aspect was not directly manipulated, however, within-subject analyses with the Wilcoxon signed-rank test showed significant decrements of the positivity ratios measured at the end of the session in $Stim^{+}$ (W=112, p<.01, r=-.36) and $Stim^{-}$ (W=109.5, p<.01, r=-.35), but not in the control group. Comparisons performed between groups did not show significant differences.

Correlation analyses were conducted between the positivity ratio of pairs measured at the beginning and also at the end of sessions with respect to their overall response precision and also with IR measures reported in the first research question. Results from these analyses showed that positivity ratio was not correlated with performance nor with IR measures. Additionally, based on the classification of teams introduced in Table 5.28, between-group comparisons were conducted with the Kruskal-Wallis test in order to evaluate differences with regard to the positivity ratios of teams in the three groups. Results from this test did not show significant differences at $p<.05$ neither for the initial positivity ratios nor for the end-session positivity ratios. These results show that low, medium, and high performance teams presented similar positivity ratios.

A closer examination to the data showed that none of the high performance pairs started the session with a positivity ratio equal or higher than 2.9. However, 80.95% of the pairs (17 out of 21) started with a positivity score equal or higher than their respective negativity scores. At the end of the session, the percentage of high performance pairs with positivity higher than negativity fell to 57.14%. In this group, only one pair had a positivity ratio of 3.33. Moreover, 8 out of 9 pairs with a positivity ratio lower than 1.0 were in conditions in which at least one participant was previously treated with negative stimuli (i.e. $C2^{+-}$ and $C3^{--}$). Average positivity ratios measured at the beginning and at the end of the sessions are provided in Table 5.29.

With regard to medium performance pairs, 85.33% (15 out of 18) initiated their sessions with a positivity ratio greater than 1.0 (i.e. positivity greater or equal than negativity). After completing the search tasks, positivity was higher than negativity in 14 pairs (77.78%). In this group, one pair in $C1^{++}$ that started with

Table 5.29. Average positivity ratios measured at the beginning and at the end of sessions.

Group	Positivity ratio at the beginning of the session (past 24-hours)	Positivity ratio at the end of the session (as a result of the session)
High performance pairs	M=1.64, SD=1.78	M=1.07, SD=0.74
Medium performance pairs	M=1.54, SD=0.83	M=1.29, SD=0.18
Low performance pairs	M=1.25, SD=0.72	M=0.97, SD=0.79

a positivity ratio equal to 0.64 end up with a positivity ratio equal to 3.0. Furthermore, in this stage the four pairs who presented positivity ratios under 1.0 were in condition $C2^{+-}$ and $C3^{--}$. Averages of positivity ratios measured at the beginning and at the end of the sessions are provided in Table 5.29.

Finally, all low performance pairs (six in total) initiated the task with a positivity ratio higher than 1.0. Interestingly, only one of these pairs started the session with a positivity ratio equal to 3.33. Moreover, during the session, both participants in this pair were treated with positive stimuli (i.e. $C1^{++}$), yet their overall response precision was equal to 40%. At the end of the sessions, only one pair in $C3^{--}$ reported a positivity ratio under 1.0. The positivity ratio of the remaining five pairs ranged between 1.0 and 1.29. Averages positivity ratios measured at the beginning and at the end of the sessions are provided in Table 5.29.

As indicated above, it is acknowledged that individuals' positivity ratios could have changed as a result of affective stimuli during the evaluation of short-term effects. Note that as explained in section 4.5, such stimuli were applied with proportions 3-to-1 (i.e. three positive and one negative stimuli) in positive conditions and 1-to-3 (i.e. one positive and three negative stimuli) in negative conditions. It is also possible that other confounding factors that were not measured (e.g. boredom or tiredness) could have caused variations in positivity and negativity scores. To address the shortcomings of this global approach to investigate the relation between positivity ratios and performance, the following section provides results from a local approach that investigate positivity ratios and performance in each question.

Table 5.30. Classification of teams based on their local and global performance

Description	SAM valence (low to high)	Recoded valence (low to high)
Positivity score	4 to 1	1 to 4
Neutral	5	0
Negativity score	6 to 9	-1 to -4

5.6.2 Local approach

The second approach to investigate relations between positivity ratio and performance was conducted by combining positivity and negativity levels individually reported by the participant before and after each question through the valence scale in the SAM questionnaire. For analysis purposes, since this questionnaire is based on a 9-points scale (from 1 to 9), values were re-coded in accordance to the scale presented in Table 5.30. Then, positivity ratio of each pair of participants (p_1,p_2) before each question was computed as follows:

$$positivityRatioSAM(p1, p2) = \left|\frac{positivityScore(p_1) + positivityScore(p_2)}{negativityScore(p_1) + negativityScore(p_2)}\right| \quad \text{(Eq. 5.3)}$$

Similar to the results obtained with the global approach, between-group comparisons were carried out with the Kruskal-Wallis test in order to evaluate if the positivity ratio of high performance pairs differed significantly from that of medium and low performance pairs. Results from this test did not show significant differences at $p<.05$. Additionally, no correlations were found between positivity ratio and different IR measures computed in each question.

According to this approach, 85.19% of the questions responded correctly (i.e. high performance) were addressed by pairs who initiated such questions with a positivity ratio greater or equal than 1.0. In this group, positivity ratio was greater or equal than 2.9 in 38.43% of the questions that were responded correctly by pairs. The average positivity ratio in questions responded correctly was 2.14 (*SD*=1.31).

With regard to medium performance pairs, only eight observations in MT_q_2 corresponded to participants responding questions partially correct. In this question, six pairs (75%) began with a positivity ratio higher than 1.0, but only 2 of them started with a positivity ration equal or greater than 2.9. The average positivity ratio in questions responded partially correct was 1.59 (*SD*=1.31).

Finally, 87.04% of the questions that were not responded correctly were initiated by pairs whose positivity ratio before starting such questions was greater than 1.0. From this group, 51.85% began these questions with a positivity ratio equal or greater than 2.9. The average positivity ratio in questions responded incorrectly was 2.22 (*SD*=1.48). The average positivity ratio measured before each question is illustrated in Figure 5.34.

Both global and local evaluations showed that the relation 3-to-1 between

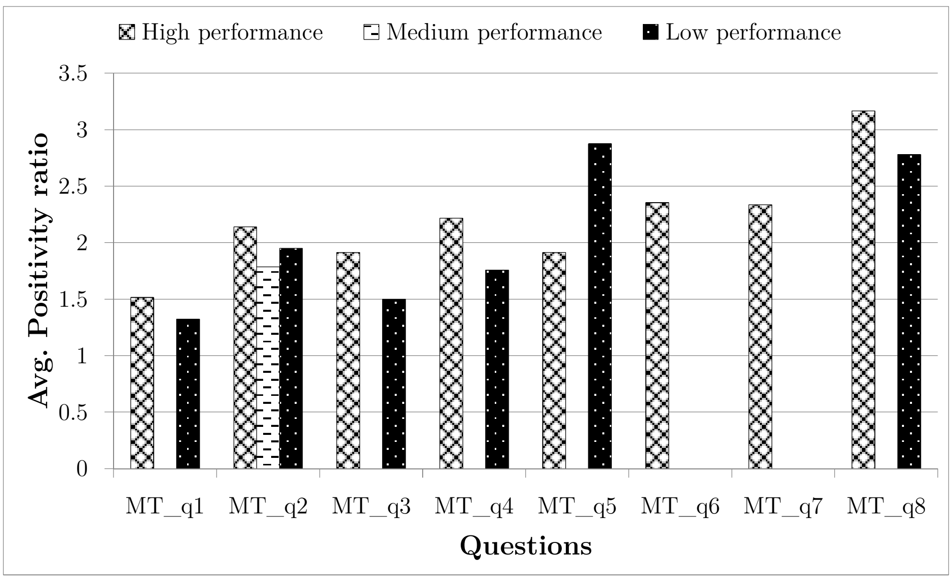


Figure 5.34. Average positivity ratio of pairs measured before each question.

positivity and negativity was not consistently found in high performance pairs. In fact, several pairs with positivity ratios under this threshold and even lower than 1.0 (i.e. negativity greater than positivity) were able to reach higher response precision (i.e. global evaluation) or respond questions correctly (i.e. local evaluation). Since the global approach was limited due to lack of samples with positivity ratios above 2.9 as informed by the positivity self-test, these results only provide initial evidence to reject the fourth hypothesis of this dissertation, which was introduced in the research framework as follows:

> **Hypothesis 4:** Based on studies that have successfully evaluated the relation 3-to-1 (P/N) in different domains, it would be expected that teams in a CIS task whose P/N ratio is above this baseline should outperform - in a number of performance measures - those below it.

It is noteworthy that communication analyses were planned in order to investigate the positivity ratio at the level of speech acts following an approach similar to that used by Losada and Heaphy (2004). However, as shown in the results provided in the first and second research questions, messages with positive and negative tones as well as agreement and disagreements were rather scarce. This situation restricted the possibility to carry out an exploration of the positivity ratio and its relation with the performance of pairs at the process level.

5.7 Summary

This chapter provided a detailed explanation of the analyses conducted to address the four research questions and hypotheses of this dissertation. Results obtained from these analyses were presented along with local interpretations, which will be expanded an integrated in the following chapter.

In the first part of this chapter, data exploration and preprocessing procedures were presented with the purpose of justifying the statistical analyses used to address the research questions and hypotheses. Then, in accordance to the structure of the experimental design of the study, analyses and results for the first three research questions were split into two major parts: evaluation of short-term and prolonged effects.

The evaluation of short-term effects showed that affective stimuli had effects on participants' search practices as well as in their affective experiences. The former results were expressed in terms of information coverage, precision, and number of queries, to name a few. With regard to affective experiences, results showed that affective stimuli were able to elicit the intended affective state, which later became the initial affective condition for the evaluation of prolonged effects.

On the other hand, the evaluation of prolonged effects was designed to investigate how initial affective states influence information practices, task perception, affective expressiveness, affective experiences, and the overall performance of individual searchers and collaborative pairs in the long run. For the particular case of pairs, these factors were studied along with communication processes with respect to the interactions of the affective states of participants. Note that while the focus of this dissertation is CIS, individual information seeking was studied in order to define baselines to perform comparisons and interpret results obtained with pairs.

Results for the first research question, which focused on effects of initial affective processes and their interactions in searchers' collaborative practices, showed isolated differences in the development of *A Google a Day* question. However, differences that could be attributed to initial affective states were not always consistent and could be the result in some cases of randomness. Not surprisingly one of the few consistent results was with regard to the number of distinct queries formulated in each question. These results were significantly higher in pairs than in the baseline conditions in the majority of the questions analyzed. However, significant results in this case are a direct consequence of the combination of queries issued by each team member and not as a result of affective states or their interactions.

It was interesting to find that information coverage was not necessarily higher for collaborative pairs. Indeed, significant differences were found in only two questions (MT_q_2 and MT_q_3), which were later characterized by discussions that emerged as a result of discrepancies within pairs with regard to the answers or the interpretation of questions. In these two cases, differences were attributed to influences of initial affective states and their interactions.

At the level of communication processes, immediate effects of initial affective states and their interactions were observed in the first question (MT_q_1) of the evaluation of prolonged effects. Although these differences were not found in all questions, communication processes were typically more frequent in pairs in which at least one member started the task with positive affective states. These results provided evidence to support the first research hypothesis of this dissertation, which stated that the communication volume of pairs in which both members are treated with negative stimuli will be lower than the communication volume of pairs in which at least one of the members received positive stimuli.

In the second research question, which focuses on the experience and expressivity of affective processes during the information search process, it was found that smiles can be a good indicators of the relevance of documents (local) and also of satisfaction or relief at the moment of completing information search tasks (global). At the global level, it was found that smiles increase progressively following similar patterns in collaborative and individual conditions. At the local level, smiles were concentrated at the beginning and at the end of the exposure to the content of relevant pages in five out of six conditions. A distinctive pattern in the concentration of smiles was found in the condition were individual participants started the second task in negative affective states ($C5^{-}$). This aspect was attributed to a different information processing strategy influenced by negative affective states. Perhaps, individual participants in negative affective states were more reluctant or pessimistic to view a relevant document as potentially relevant in the first seconds of exposure to the content of such pages.

Affective processes expressed through electrodermal activity were found to be a good global indicator of engagement and success (if accompanied by smiles) or frustration in the development of information search tasks. However, at the level of pages (local) distinctive patterns were found only in the control group ($C6^{control}$) and also in $C5^{-}$.

Analyses with regard to self-reported affective experiences did not show consistent differences after the completion of the first question. It was confirmed, however, that participants started the evaluation of prolonged effects with different perception of their internal affective processes that were consistent with the intended affective states that were elicited in the previous stage.

Results for the third research question, which focuses on the implications of initial affective states and those derived from collaboration on performance, showed isolated differences with regard to response time and query precision. However, overall analyses in terms of response precision revealed that collaborative conditions reached higher performance in terms of the quality of their work. More importantly, collaborative conditions in which both members started the evaluation of prolonged effects in negative affective states ($C3^{--}$) outperformed those in which at least one team member was treated with positive affective states ($C1^{++}$ and $C2^{+-}$). This result was consistent with those found in the baseline conditions, where $C5^{-}$ outperformed $C4^{+}$. As pointed out above, perhaps initial negative affective states influenced the use of particular information processing strategies (this along the lines of affect infusion) that made them more critical about the content and the quality of documents. These results provided evidence to reject the two hypotheses linked to the third research question of this dissertation, which stated that positive affective states should lead searchers to achieve higher response precision.

Finally, results for the fourth research question, which focused on positivity and its relation to performance in CIS, showed that proportions between positivity and negativity (positivity ratio) did not correlate with the performance of pairs in the context of this study. The positivity ratio was analyzed with two different approaches (i.e. global and local) and then compared with respect to the performance of pairs. Results from both analyses did not show significant differences in the positivity ratio of high performance, medium performance, and low performance pairs. As stated in the fourth hypothesis introduced in the research framework of this dissertation, it was expected (based on literature in positive psychology) that pairs with a positivity ratio higher than 2.9 (roughly speaking positivity three times higher than negativity or 3-to-1 ratio), but lower than 11.6, should outperform pairs that are outside this optimal zone. However, results showed that pairs with such positivity ratio were distributed among the three performance categories stated above. Moreover, it was found that cases in which negativity was higher than positivity (positivity ratio lower than 1.0) were also classified as high performance pairs. These results provided evidence to reject this hypothesis.

The following chapter integrates and expands the discussion of the results obtained in each research question. Additionally, implications for theory, practical applications, limitations, and future work are discussed in detail.

Chapter 6. Discussion and Conclusions

The previous chapters have provided a comprehensive review of the objectives of this research, literature, research framework, methodology, and results obtained in a laboratory study. As stated earlier in this dissertation, the main motivation and objective of this work was to understand what role, if any, initial affective processes (in particular positive and negative affective states) and their interactions play in collaborative information seeking (CIS).

As described in the research framework of this dissertation (Chapter 3), studies in psychology and other fields have shown that initial affective processes take an active role in complex processes such as decision-making, social relations, and perception. As noted in the literature review of this dissertation (Chapter 2), although there exist research about the affective dimension in the context of individual and collaborative information seeking, most of these works focus on affective states as byproducts of information search. In other words, they indicate that searchers experience and express emotions, moods, affects, or feelings as a result of different stages that take place in information search. For instance — returning to the examples of John ($S_{John[-]}$) and April ($S_{April[+]}$) used in the first chapters — if John fails in finding information to solve a particular problem, it is likely that during and after his search process, he experiences different negative affective states such as frustration, sadness, and anger. On the other hand, if April successfully finds information to solve an important part of her project, it is likely that during and after her search process, she experiences positive affective states such as enthusiasm, confidence, and joy. Note that in both cases it is indicated that the resulting affective processes are not guaranteed, but are likely to occur. This aspect, which is a limitation of past work and also of specific parts of this dissertation, is discussed in detail in section 6.2.

Rather than focusing on affective states as a result of information search, this dissertation aims to understand how affective states (in particular positive and negative affective states) that precede information search processes influence information behaviors and affective states that take place in CIS. The following section expands upon the discussion of the research questions of this study along with integrated interpretations based on the results presented in Chapter 5, followed by limitations of this dissertation, theoretical and practical implications, future directions, and final remarks.

6.1 Role of initial affective states and their interactions in CIS

This dissertation was framed to investigate what role initial affective processes (in particular positive and negative affective states) as well as their interactions play in CIS. The following questions were presented in determining this role:

> **RQ1**: Do initial affective states and their interactions shape the way team members collaborate when searching information, and if so, how?
>
> **RQ2**: What affective processes are typically experienced and expressed (physically, physiologically, and verbally) by team members when collaborating in an information search task?
>
> **RQ3**: To what extent, if any, do initial positive and negative affective states and those derived from the collaboration of individuals in an information search task influence team performance?

RQ4: To what extent, if any, does the relation 3-to-1 between positive and negative affective states (P/N) (Losada & Heaphy, 2004; Fredrickson & Losada, 2005) apply to CIS?

To address these research questions, a laboratory study was conducted in order to investigate short-term and prolonged effects of initial affective states and their interactions in CIS. Note that although the focus of this dissertation is on CIS, the experimental evaluation also involved individual searchers who were used to establish baselines that helped in the interpretation of results obtained in collaborative contexts. The following sections present discussions about each research question with respect to short-term and prolonged implications of initial affective processes and their interactions in information search.

6.1.1 Short-term effects

The evaluation of short-term effects in this study focused on immediate implications as a result of the elicitation of affective states with the false feedback technique. Although the evaluation of short-term effects did not involve collaboration between participants, results from this stage were necessary to confirm the initial conditions of the participants before they were evaluated in the long run with or without collaboration.

As part of the analyses conducted to address RQ1, it was shown that the participants who received affective stimuli displayed significant variations in their search practices (e.g. less number of queries, less information coverage, and higher efficiency). More importantly, such variations were not observed in the control group. This result not only showed the efficacy of the stimuli used to elicit particular affective states, but also that affective changes had immediate effects in the participants' search processes.

For example, between-group comparisons in the post-stimuli question ($PostS\_q_{10}$) showed that the average time participants spent on content pages was significantly higher in conditions where the participants received stimuli (i.e. $Stim^{+}$ and $Stim^{-}$) than those in the control group ($Stim^{control}$). In turn, results for this measure were significantly higher in the negative condition ($Stim^{-}$) than in the positive condition ($Stim^{+}$). Regardless of these differences, response time — as part of short-term performance analyses in RQ3 — was not found to differ significantly across the three conditions in $PostS\_q_{10}$. This suggests that effects derived from the stimuli stage did not have global implications in the search processes (i.e. question level), but did have local effects (e.g. pages, queries, SERPs).

Perhaps, differences in the time spent on content pages suggest that the participants who started $PostS\_q_{10}$ in negative affective states ($Stim^{-}$) were more careful when reading the content of pages than their peers in $Stim^{+}$. To be more precise, this behavioral change in $Stim^{-}$ could be attributed to a reduction in the levels of confidence of participants (i.e. negative affective state) as a result of the false-feedback technique used in the stimuli stage. Recall that for the case of $Stim^{-}$, this technique involved displaying discouraging messages to the participants as they searched for and provided information to answer the *A Google a Day* questions regardless of their actual performance. To some extent, participants in $Stim^{-}$ might have developed a feeling of insecurity that made them more cautious at the moment of deciding what information was relevant and what was not. In this regard, within-group analyses (with both Friedman rank sum test and Wilcoxon signed-rank test) showed a significant increment in the average time to snip (explicit relevance judgment in this study) in $Stim^{-}$, which confirms that the participants in this condition typically needed more time to decide what information was relevant in $PostS\_q_{10}$ than that reported for the same group in $PreS\_q_{1}$.

Conversely, it may be possible that participants who started $PostS\_q_{10}$ in positive affective states ($Stim^{+}$) were more relaxed than their peers in $Stim^{-}$ while reviewing content pages. This behavior could be attributed to an increment in their levels of confidence, which were derived from the stimuli stage. For example, participant S592 reported that the feedback provided by the system in that stage "were like confidence boosters." Likewise, participant S640 stated that "it [(receiving positive feedback)] helped boost [his] confidence. It made [him] happier." Note that unlike $Stim^{-}$, participants in $Stim^{+}$ received positive feedback regardless of their actual performance.

It is noteworthy that in $Stim^{+}$ and $Stim^{-}$, stimuli were designed to avoid side effects such as disinterest, overconfidence, and excessive levels of frustration (Zhao, 2006). As explained in section 4.5, this was achieved by implementing 3-to-1 and 1-to-3 ratios between positive and negative feedback in $Stim^{+}$ and $Stim^{-}$ respectively.

Although between-group analyses did not show significant differences for other measures, within-group analyses (with both Friedman rank sum test and Wilcoxon signed-rank test) showed significant variations from $PreS\_q_{1}$ to $PostS\_q_{10}$ in $Stim^{+}$ and $Stim^{-}$, but not in $Stim^{control}$. Specifically, results for both tests were consistent for information coverage, relevant coverage, SERPs, efficiency, effectiveness, precision, average dwell time in SERPs and content pages, number of queries, average time to snip, levels of familiarity, perceived challenge (when exposed for the first time to the question), and response confidence. Most of these variations were found in both $Stim^+$ and $Stim^-$, for example a decreasing level of information coverage was found in both groups, which is attributed in part to the participants spending more time on content pages. Likewise, a consistent reduction in the number of queries and SERPs in both conditions may be related to the participants spending more time in SERPs. Reduced information coverage contributed to higher precision, whereas reduced number of queries contributed to higher efficiency.

There were only three measures that changed exclusively in $Stim^+$ from $PreS\_q_1$ to $PostS\_q_{10}$: relevant coverage, topic familiarity, and perceived difficulty (when providing the answer). Relevant coverage increased significantly in $PostS\_q_{10}$, which also contributed to higher precision levels. As indicated above, it is believed that the participants in $Stim^+$ were more relaxed during the exploration of pages in $PostS\_q_{10}$, but at the same time were less critical or less hesitant – but not necessarily less careful – when deciding if a page was relevant or not. Perhaps feelings of confidence, security, and certainty derived from the stimuli stage contributed to lighten the decision making process. Regarding topic familiarity, although this was found to increase in the post-stimuli evaluation ($PostS\_q_{10}$), the perceived challenge and also difficulty were higher than that reported in the pre-stimuli evaluation ($PreS\_q_1$). This change in the perception of familiarity may also be attributed to increased levels of confidence as a result of exposure to positive feedback.

As noted in the Affect Infusion Model (AIM) (Forgas, 1995; Forgas & George, 2001), the influence of affective states (in particular mood) would depend on information processing strategies, which may be selected due to additional factors such as familiarity with the situation, target complexity, specific motivations, cognitive capacity, and situational pragmatics (Forgas, 2009). Since the participants were not very familiar with the topic of the task and the complexity of the task required the participants to look for different pieces of information before responding the *A Google a Day* questions, information processing strategies were probably beyond the scope of direct access. On the other hand, because of the precision-oriented nature of the task, heuristic strategy was not suitable. This strategy and also substantial information processing consider the absence of motivations; however, as explained in section 4.2, the participants in this study were given extrinsic motivations that consisted of financial compensations. When asked about their main motivation to participate in the study, several participants pointed out their interest in the money that they could obtain (i.e. up to $50 per participant) if they were among the three best performance pairs or individuals. Some quotes from the interview in this regard are: "To see what the whole study is about and try to win the prizes of course" [S616], "Umm…main motivation was money, honestly (haha)" [S617], "Umm to help [(a friend)] and to get money" [652], "we're getting paid obviously" [S653], "the cash prizes" [S654], and "just the monetary compensation" [S489]. Although it has been indicated in the literature that it is difficult to find motivated processing strategies in experimental studies (Forgas, 2009), based on participants' comments, the extrinsic motivation used in this study proved effective in most cases. Therefore, it may be possible that motivated processing was the main strategy used across the three conditions. As stated by Forgas, with motivated processing strategy “people are likely to engage in highly selective, guided, and targeted information search strategies to support a motivational objective” (Forgas, 2009, p. 104).

Since motivated processing is less likely to be infused by affective states (Forgas, 2009), the behavioral differences found between the three groups and also within $Stim^{+}$ and $Stim^{-}$ could be attributed to underlying or additional information processing strategies. For example, Sinclair and Mark (1983) showed through experimental studies (also using emotion elicitation procedures) that people in positive affective states (in particular referring to happy participants) employed information processing strategies that are “relatively passive or nonsystematic, [and] less detailed [(accurate)]” (p. 417). Conversely, people in negative affective states (in particular referring to unhappy participants) were found to be “more active or systematic, [and] detailed” (p. 417). Along the same lines, Isen and Means (1983) found that decision making in a group of participants with induced positive affective states was faster than it was in a control group. As the authors pointed out, this aspect contributed to achieve higher efficiency. With regard to information processing strategies, the authors noted that the participants in positive affective states typically did not revisit information already seen and they “were more likely to ignore information considered unimportant” (p. 18). The authors related this behavior to a strategy called “elimination by aspects”, which consists of “eliminating from further consideration alternatives that did not meet a criterion on a selected important dimension.”

In addition to exploring effects of initial affective states in information-related behaviors, the analyses performed to address RQ2 aimed to investigate the expression and experience of affective states during the information search process, taking into account potential effects of both positive and negative affective states as preconditions. Results for the evaluation of short-term effects showed interesting patterns at the level of smiles as well as electrodermal activity (EDA). In particular, it was observed that smiles presented slightly positive trends in the three groups (i.e. $Stim^{+}$, $Stim^{-}$, and $Stim^{control}$) before, during, and after the stimuli stage. In all three cases, smiles were generally concentrated during the last segments (based on normalized time of questions' elapsed time) or stages of the participants' search processes. It was noted that such high concentrations of smiles could be attributed to feelings of success, relief, or satisfaction. In turn, feelings of satisfaction or success can be related to increased levels of confidence at the moment of providing answers.

Results for EDA showed high concentration of peaks during the last segments of the participants' search process (based on normalized time of questions elapsed time). Recall that peaks are attributed to affective reactions that could derive from increased levels of physiological activation. Variations in the levels of activation (under a dimensional approach) have been related to affective states (e.g. levels of excitement) and also cognitive states (e.g. levels of attention and engagement). It is important to mention that although results for $Stim^{+}$ and $Stim^{-}$ may be confounded with extrinsic factors such as affective stimuli, results for $Stim^{control}$ presented similar patterns before ($PreS\_q_1$) and during the stimuli stage. Differences across conditions were only observed during the post-stimuli evaluation ($PostS\_q_{10}$), where the concentration of EDA peaks in $Stim^{control}$

presented a negative trend. This difference was attributed to a feeling of relaxation experienced by the participants in $Stim^{control}$ as a result of being aware that PostS_q_{10} was the last question of the first task. Conversely, the concentration of peaks in PostS_q_{10} for $Stim^{+}$ and $Stim^{-}$ displayed patterns similar to those found in PreS_q_1 and also during the stimuli stage. Considering that participants in $Stim^{+}$ and $Stim^{-}$ were also aware that PostS_q_{10} was the last question of the first task, this common pattern was attributed to anticipatory affective reactions (Figner & Murphy, 2011) derived from multiple trials involving affective stimuli.

Although facial expressions and EDA provide objective measures of participants' affective processes from discrete and dimensional approaches, it was acknowledged that they both have limitations. To partially address possible shortcomings associated with these methods, participants' subjective experiences of their affective states (i.e. feelings) were also evaluated. This was addressed through the self-assessment manikin (SAM) (Bradley & Lang, 1994), a pictorial instrument that allowed the participants to self-report their affective states through three dimensions: valence (i.e. positive-negative), activation (i.e. excited-calm), and dominance (i.e. control-in control). Results derived from this questionnaire showed that in general, the participants in $Stim^{+}$ and $Stim^{-}$ experienced positive and negative affective states respectively as a result of their exposure to affective stimuli. Additionally, it was found that activation levels were typically higher in $Stim^{+}$ and $Stim^{-}$ than in the control group.

With regard to dominance levels, it was found that the participants in the control group felt typically more in control of the situation (i,e. task or session) and those who received stimuli felt more controlled. This particular result corresponds to one of the limitations of this study, which will be discussed in section 6.2. In addition to showing that stimuli were effective during the stimuli stage, the most important result obtained from self-reports was the confirmation (from the participants' point of view) that the participants started the post-stimuli evaluation (PostS_q_{10}) in the intended affective states. Specifically, it was found that valence in $Stim^{+}$ was significantly higher (i.e. positive) and valence in $Stim^{-}$ was significantly lower (i.e. negative) than that found in the control group.

The evaluation of short-term effects showed that initial affective states had immediate implications in the information behaviors of individual searchers. It was presumed that such differences would be related to different information processing strategies that the participants used to address the *A Google a Day* questions. Bearing in mind that this evaluation stage was specifically designed to test the effectiveness of the stimuli used to elicit positive and negative affective states, the following objective of this study (evaluation of prolonged effects) was to investigate whether the observed effects in PostS_q_{10} endure in time and also if interactions of affective states as a result of collaboration produce significant changes.

6.1.2 Prolonged effects

In the evaluation of prolonged effects the participants were redistributed into six experimental conditions (i.e. $C1^{++}$, $C2^{+-}$, $C3^{--}$, $C4^{+}$, $C5^{-}$, and $C6^{control}$). Such distribution was performed depending upon the stimuli that the participants received in the previous stage and also whether they signed up for the study with someone else. Conditions C1^{++}, C2^{+-}, and C3^{--} correspond to collaborative conditions, whereas C4^{+}, C5^{-}, and C6^{control} were used as baselines for comparisons and interpretation purposes. As part of the analyses conducted to address RQ2, it was shown that there were significant differences in the affective states of the participants — specifically for valence and activation — at the moment of starting the evaluation of prolonged effects. More importantly, such differences were consistent with the stimuli that the participants received in the previous stage. The major levels of initial positive affective states (self-reported positive valence) were concentrated mainly in C4^{+}, whereas the major levels of initial negative affective states (self-reported negative valence) were found mostly in C2^{-} (i.e. participants who received negative stimuli and later were assigned to work with their partners who were treated with positive stimuli in C2^{+-}).

Despite observed differences in initial affective states in the six experimental conditions, results for RQ1 did not show that the average time the participants spent in content pages varied significantly among the baseline and also collaborative conditions. Recall that this particular aspect was found to be significantly different across the three groups in the evaluation of short-term effects. Additionally, such differences were attributed to varying levels of confidence of the participants and also to specific information processing strategies.

Although results for RQ1 showed significant differences across the six conditions with respect to different measures, such differences were not consistent in all the questions. While this indicates that significant differences could be derived from random factors, it is possible that differences in specific cases may be associated to particular aspects of the questions. For example, in MT_q_2 ("During a famous White House séance, witnesses say the president's seat was levitated. Whose spirit was his wife trying to contact?") and MT_q_3 ("In Norse mythology, which god's son will poison Thor at Ragnarök?") participants in collaborative conditions had to deal with conflict resolution derived from disagreements in their findings (MT_q_2) or in their interpretation of the question (MT_q_3).

While particular differences found between collaborative and baseline conditions such as the number of queries derived from the combination of individual results, it was interesting to find that information coverage was in most cases roughly similar for pairs and individuals. This result suggests that pairs did not necessarily incur in additional efforts in terms of information coverage. The only two exceptions in this regard (with pairs as unit of analysis) were found in MT_q_2 and MT_q_3, which as indicated above were described as having distinctive features that were heightened in collaborative conditions.

Other dimensions associated with CIS, such as communication and information sharing, were also explored. However, similar to information-related measures, significant differences were typically scattered in different questions with little consistency from question to question. A key finding in this regard was that communication volume was typically higher in conditions where at least one of the participants started the task in initial positive affective states (i.e. $C1^{++}$ and $C2^{+-}$) as opposed to communication volume found in $C3^{--}$. This particular result provided evidence to support the fourth hypothesis introduced in the research framework of this dissertation:

> **Hypothesis 4**: The communication volume of pairs in which both members are treated with negative stimuli will be lower than the communication volume of pairs in which at least one of the members received positive stimuli.

Further analyses in this regard did not show correlations between communication volume and the performance of pairs. However, it was indicated that lack of communication or low communication volume may have negative implications for fundamental aspects of collaboration such as coordination and awareness.

Results for RQ1 showed that significant differences observed in the evaluation of short-term effects, in particular those referred to the average time that the participants spent in content pages, did not persist in the long run. However, analyses conducted to address RQ2 showed that distinctive expressive patterns associated with initial affective conditions could act as implicit indicators of the information processing strategies used by the participants. In particular it was found that individual participants who started in negative affective states ($C5^{-}$) smiled typically in the last seconds of exposure to the content of relevant pages (pages from where they collected snippets). Perhaps, negative affective states made the participants more reluctant or pessimistic to view a page as potentially relevant in the first seconds of exposure to the content of such pages. This particular result relates to the discussion presented in the previous section, where it was stated that participants in initial negative affective states would use information processing strategies that made them more systematic, critical, and meticulous in the evaluation of the content of pages.

This result contrasts with the expressive patterns found in other individual conditions and also in collaborative contexts. In conditions $C1^{++}$, $C2^{+-}$, $C4^{+}$, and $C6^{control}$, the participants typically smiled during the first and last few seconds of exposure to the content of relevant pages. Early smiles would be attributed to the participants' first impression of the content pages, whereas late smiles would be an indication of satisfaction after finding the information of interest. For instance, at the moment of a page visit, particular features (e.g. title, images, and highlights, among others) may have indicated to the searchers that they were in the right place, thus resulting in smiles. Later, smiles expressed at the moment of making explicit judgments of relevancy would indicate satisfaction or success when finding relevant information.

Condition $C3^{--}$ (i.e. pairs of participants who initiated the task in negative affective states), also presented two peaks with high concentration of smiles during the exposure to the content of relevant pages, however, the first peak was found almost in the middle of the dwell time of relevant pages. Such smiles could be associated with delayed reactions of participants to the content of pages in combination with sharing and commenting episodes with their partners in the context of collaboration. Similar to other conditions, the second peak of smiles would indicate satisfaction or success as a result of finding relevant information. A delayed reaction to the content of relevant pages suggests that participants in $C3^{--}$, similar to those in $C5^{-}$, used systematic information processing strategies.

In the absence of affective stimuli during the evaluation of prolonged effects, some participants indicated that collaboration provided an alternative way to confirm that they were on the right track. For instance, S476 said that "working with a partner was similar [(in relation to affective stimuli)] because if we got the same answer I was more confident." Having two participants collaborating toward common goals would provide confirmation mechanisms that increased confidence levels and precision of the answers provided. It is suggested that this collaboration-based feedback would contribute to the overall performance of pairs. In fact, based on the results obtained for RQ3, pairs were able to achieve higher levels of response precision than individual participants in the baseline conditions.

Regarding performance, results for RQ3 showed that response precision (i.e. ratio between correct answers and the total number of questions addressed) was significantly higher in $C3^{--}$ than in all the other five conditions. Consistently, response precision in the corresponding baseline condition $C5^{-}$ was significantly higher than that found in its counterpart $C4^{+}$. These results suggest that initial negative affective states, which might have triggered the use of systematic information processing strategies, would have implications in the long run with regard to the quality of work.

An important aspect to consider in collaborative contexts is that frustration levels and overall affective load may increase when participants in opposed affective states (i.e. positive and negative affective states) interact. Based on the results presented for RQ2, the levels of frustration of participants were significantly higher in $C2^{+-}$ than in any other condition. Interestingly, a closer examination of this condition showed that frustration levels were significantly higher in $C2^{+}$ than those found in $C2^{-}$. Affective load, on the other hand, was found to be significantly higher in $C2^{-}$ than in $C2^{+}$. It is speculated that such increased levels of frustrations in $C2^{+}$ would be a side effect of elevated communication efforts for $C2^{+}$ when interacting with partners in $C2^{-}$. Perhaps high affective load and frustration explain why response precision was significantly lower in $C2^{+-}$ than in the other two collaborative conditions.

Overall, results found in this study suggest that initial affective states may define or shape information processing strategies. Additionally, in collaborative settings, results indicate that the interplay of similar or different affective processes could change the way searchers interact with each other, their frustration levels, affective load, and also the quality of their work. Below, concise answers for each research question are provided.

6.1.3 Summary

Based on the results presented in the previous chapter and the integrated discussion provided above, the research questions of this dissertation can be responded to as follows:

RQ1: Do initial affective states and their interactions shape the way team members collaborate when searching information, and if so, how?

> Results suggest that initial affective processes would not have persistent effects on information search practices such as information coverage, precision, recall, or in the number of queries used to find information. However, initial affective states would influence information processing strategies and also social interactions between team members including communication volume and communication effort during the coordination, exchange, and evaluation of information.

RQ2: What affective processes are typically experienced and expressed (physically, physiologically, and verbally) by team members when collaborating in an information search task?

Results suggest that while team members may experience and express a wide spectrum of affective states, smiles (physical/expressive) and peaks in EDA (physiological) seem to act as local (relevance judgments) and global (search process) indicators of satisfaction, success, confidence, engagement, interest, and relief. Moreover the expression of affective processes may be influenced by initial affective states and also by particular information processing strategies.

With regard to the experience of affective processes, results indicate that the interaction of opposed affective states in collaborative search may lead team members to experience high levels of frustration and affective load. Conversely, the interaction of positive affective states would contribute to lessen the levels of frustration and affective load.

RQ3: To what extent, if any, do initial positive and negative affective states and those derived from the collaboration of individuals in an information search task influence team performance?

Results suggest that initial affective states would have prolonged implications in the quality of work. Specifically, initial negative affective states and also their interactions would lead collaborative searchers to achieve higher quality of work than those who start in positive affective states. Conversely, the interaction of opposed affective states (i.e. positive and negative) would lead searchers to achieve low quality in their work.

RQ4: To what extent, if any, does the relation 3-to-1 between positive and negative affective states (P/N) (Losada & Heaphy, 2004; Fredrickson & Losada, 2005) apply to CIS?

Based on the theory and empirical evidence on positive psychology, it was expected that positive affective states would have positive effects in the local and/or global performance of the participants. However, results for this question showed that the performance of pairs was not related to the positivity ratio (ratio between positivity and negativity). Specifically, the relation 3-to-1 between positive and negative affective states was not associated to high-performance pairs. One possible reason for this result is that affective states in this study showed mood-congruent influence in the way people approached the search tasks.

6.2 Limitations

While the results of this study suggest that initial affective states would influence the way people search, evaluate, and use information, it is necessary to consider underlying limitations such as sample, stimuli, search task, group size, and assumptions. Although these aspects provide high internal validity, they also restrict the generalization (external validity) of the results. Below, a discussion of possible limitations of this study is presented.

First of all, a sample of convenience was used in this study. The target population was composed of college students from Rutgers University who came from different fields of study. Additionally, the following recruitment restrictions were established: English as a native language, participants' age ranging from 18 to 24 years old, average typing speed of 50 words per minute, intermediate search skills, and for the particular case of collaborative conditions participants signed up with someone with whom they had previous experience working together. Such constraints limit the generalization of the results to a very specific population.

Second, the stimuli used in the study were contextual to the search task used in the study. This was done in order to avoid participants' awareness of the purpose of the feedback provided during the stimuli stage. Unlike the examples of John ($S_{John[-]}$) and April ($S_{April[+]}$), in which initial affective processes were derived from personal situations independent of the context in which their information search processes took place (at work), in this study participants might have created direct associations between their performance (as a result of the stimuli received) and the induced affective states. This aspect could have caused that participants treated with negative stimuli to experience not only negative affective states but also feelings of self-improvement, which result in mood-incongruent behaviors.

Additionally, stimuli were designed to elicit negative or positive affective states without specifying particular categories such as sadness, enthusiasm, frustration, confidence, or happiness, among others. By targeting general affective states, it is possible that in negative conditions, some participants may have experienced frustration while others pessimism. Conversely, in positive conditions, some participants could have experienced happiness, while others, confidence. Perhaps different affective states under a discrete approach and also the decontextualization of stimuli may have different implications in information search that were not addressed in this study.

Regarding the stimuli and experimental evaluation, it is necessary to consider the perception of the participants beyond the scope of the assigned tasks including whether the sensors affected the participants' behaviors during the study or if the participants realized that the feedback provided during the stimuli stage was inaccurate. With regard to instrumentation, most participants reported feeling comfortable with the devices used during the study. On the other hand, as reported in section 5.4.1.3, participants in the positive and negative conditions during the evaluation of short-term effects felt significantly more controlled than their peers in the control group.

Third, only precision-oriented search tasks with fixed level of difficulty were used in this study. As explained in section 4.3, this type of task was selected because it allowed the performing of multiple observations of affective states in short periods of time. Additionally, this type of tasks promoted active collaboration by hindering particular group strategies such as division of labor. In this regard it is acknowledged that other types of search tasks (e.g. recall-oriented) with different levels of difficulty may produce different results.

Fourth, the study design to investigate CIS involved only dyads. This decision was made in order to achieve better control of the complexity of interactions. It is acknowledged, however, that varying group sizes may have different implications in information search and also in the interaction of affective states.

Fifth, this study relied on the assumption that successful completion of a search task derives in positive affective states. Although this relation has been stated in the literature and in particular models such as Kuhlthau's ISP, it may be the case that affective states experienced at the successful completion of a search task are positive, negative, or mixed. For example, consider a situation in which a person searches information about his/her father's symptoms for a disease. If the searcher discovers that the symptoms can be easily addressed with natural medicine, he/she may end up experiencing feelings such as satisfaction, relief, and happiness. On the other hand, if the outcomes of her search process suggest a terminal health condition, then affective states derived from it are likely to have a negative tone (e.g. sadness, worry, and frustration). In this study, it was assumed that the participants had none or little personal involvement with the topic of the questions addressed. Yet, depending on particular aspects of the participants such as personality or past experiences, it is possible that some participants may have experienced negative affective states at the end of their search processes, even after finding the right answers. For example, during the interview stage, one participant reported feeling scared while addressing the question about the White House séance (MT_q2) because she did not like ghosts. In the same question, S600 reported to her partner that she felt sad after discovering that the séance was intended to contact a dead son.

Finally, limitations derived from focusing on one particular facial expression (i.e. smiles) are acknowledged. As explained in the previous chapter, one of the reasons of focusing on smiles was the inaccuracy of *BMERS*, *eMotion*, and *FaceDetect* in recognizing other facial expressions. While smiles were found to be potential global and local indicators of satisfaction and success in information search, it is worth considering a deeper exploration with other facial expressions and gestures. In this regard it is also important to consider facial expressions beyond the scope of emotions. Examples of this are mental states such as interest, confusion, attention, and boredom. Note that some of these mental states were partially covered (from a dimensional approach) in this dissertation with the incorporation of EDA as an indicator of physiological activation.

6.3 Theoretical implications

This dissertation explored the role of initial affective processes and their interactions in CIS. Unlike previous studies, this research places affective states as preconditions of information search, which may or may not influence the way people search, process, evaluate, and make sense of information either individually or in collaboration with others.

Research on affective dimension in individual and collaborative information seeking typically places affective states within information search. In other words, searchers experience and express affective states at different stages of their search processes. For instance, based on Kuhlthau's ISP, when people face an information problem, they first experience uncertainty. While exploring different sources they may feel frustrated or confused. As they start collecting information, confidence may arise. Toward the end of their search process and depending upon the outcomes, searchers may feel relieved, satisfied, or disappointed.

The major theoretical implication of the results presented in this dissertation is that initial affective states may be determining factors of the way search processes are carried out. As discussed above, initial affective states would play a central role in the definition or selection of information processing strategies. Initial positive affective states would lead searchers to employ less systematic and detailed information processing strategies than those used by searchers who commence their search processes in negative affective states. In collaborative settings, initial affective states and their interactions would influence communication processes, frustration levels, affective load, and performance. Note that since the study presented in this dissertation may be limited by design decisions (e.g. search task, sample, group size, and experimental setting), it is necessary to expand this research by investigating if initial affective states are influenced or act in combination with other factors.

It is noteworthy that the empirical evidence derived from this study provided additional support to previous findings that suggest that different stages of information search would be linked to particular affective processes. Furthermore, at the local level, results showed that the evaluation of information based on different information processing strategies would be linked with distinctive expressive patterns that denote whether the information is relevant or not.

The theoretical implications of this study are expressed in the proposed model illustrated in Figure 6.1. Note that instead of defining a new model, this proposal could serve to extend existing models such as Kuhlthaus' ISP by incorporating affective states as preconditions of information search processes.

As stated in this model, initial affective states may affect task perception when facing an information need. They also act as modifiers or triggers of different information processing strategies. The latter in turn would have an active role in

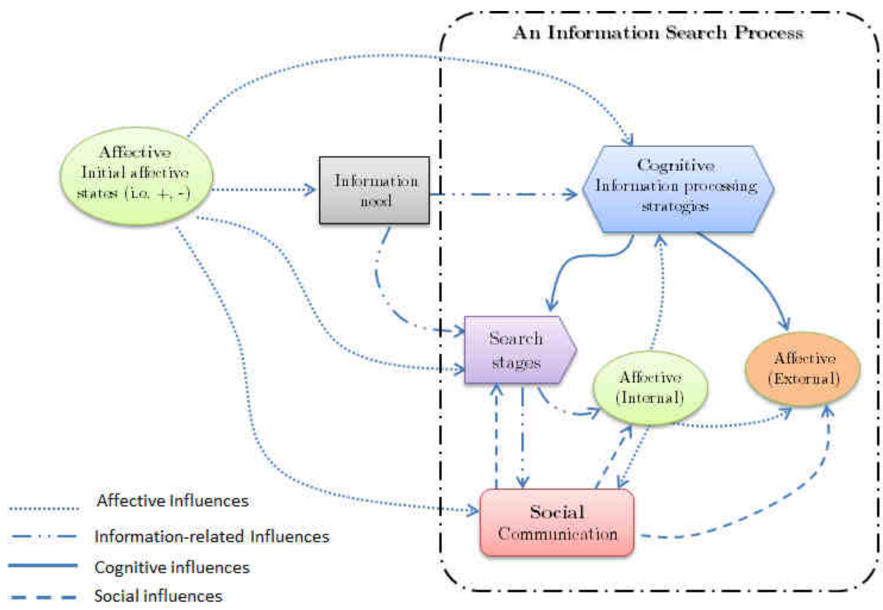


Figure 6.1. A model of influences between affective, cognitive, social, and information-related processes.

information search, influencing mainly how information is perceived and evaluated. Moreover, as shown in previous sections, it is speculated that initial affective states could be decisive factors in the quality of work achieved in the long run.

Note that the proposed model refers only to positive and negative affective states (i.e. dimensional approach) as initial conditions; however, as discussed in the previous section, affective states under a discrete approach may also be included in the model if future studies provide evidence in this regard.

The model also specifies that affective states would have effects on the expressive component of the affective states of searchers (external). Such effects are presented as direct or indirect because the study results did not provide enough evidence to determine with certainty if changes in the expressive component were due to initial affective states (direct) or if they depended purely on the information processing strategy selected as a result of initial affective states (i.e. indirect).

In the case of CIS, initial affective processes would have an effect on the communication process taking place during information search, evaluation, and information exchange. As evidenced in this dissertation, initial affective processes had an effect on the volume of communication and communication effort. For the case of two collaborators, communication volume would be lower when both partners begin in negative affective states than when both partners begin in positive affective states. When team members begin a search task in opposite affective states, communication volume may be high but at the same time more complex requiring greater communication effort in one or both group members.

Beyond the scope of the theoretical implications of this work, the results presented above may also have various practical implications. The next section expands the discussion in this regard.

6.4 Practical implications

As discussed above, the results of this study may not be generalized; however, these results and theoretical implications outlined in the previous section could have different practical implications. By way of example, the results of this work can be related to research issues such as relevance feedback and the design of systems to aid searh processes of individual and groups.

If future studies validate the results of this work in different experimental settings (e.g. different sample, different search tasks, and different systems), then it could be established with a greater degree of certainty that the initial affective states play a central role in individual and collaborative information seeking.

There have been considerable efforts to bridge the gap between searchers and IR systems; however, such attempts have failed in incorporating the affective dimension. According to the results presented in this dissertation, affective processes could help to gain access to hidden cognitive aspects (e.g. information processing strategies). For example, determining how people evaluate information (beyond the collective spectrum) would allow systems to properly deliver and present the information to their users. However, anticipating or identifying what information processing strategies are used by searchers can be unpractical, especially if searchers are required to explicitly declare their information processing strategies.

In this regard, initial affective states could help to predict the type of information processing strategies that searchers are likely to use. In accordance to the results of this study, if it is determined that the initial affective states of a searcher are negative, it would be possible to infer that the information processing strategies used by this searcher are likely to be systematic. Conversely, if the initial affective states are positive, then the information processing strategies are likely to be non-systematic.

Although identifying affective states is as difficult as determining information processing strategies, to date different disciplines have contributed to the development of methods and techniques that enable technology to capture, process, and interpret affective signals. For instance, affective computing (Picard, 1997, 2003) and related disciplines have worked actively in bridging the gap between technology and people. One of the goals in this matter is providing systems or machines with resources for recognizing, understanding, and reacting to affective processes of their users. Affective computing is an interdisciplinary field that combines aspects from other areas such as computer science, electronics, engineering, psychology, and cognitive science. Some contributions of affective computing are advanced algorithms for processing and identifying facial expressions and sophisticated instruments to keep track unobtrusively of people's affective variations during long periods of time. More interestingly is the fact that these technologies are becoming increasingly popular and within the reach of lay people. In this sense it is not difficult to imagine that within a few years, technology will progress in such a way that it will be possible to accurately determine the affective states of people at any time. More importantly, this information will be available to IR systems so that they can better assist their users.

In collaborative scenarios, an important practical implication of this research has to do with group design. For example, after determining initial affective states of individuals, it would be possible to optimize group performance based on affective interactions that could arise during information search. For example, if the objective of the work is similar to that of search tasks used in this study (precision-oriented task), then it would be desirable to have teams whose members employ systematic information processing strategies (i.e. teams of two members whose initial affective states are negative). In other types of search tasks it may be desirable to promote discussions among group members with different viewpoints and attitudes at the time of interacting with information (i.e. teams of two members with opposite initial affective states).

With regard to technological support for CIS, the identification of affective states that group members experience before and during information search could provide better mechanisms to mediate within-group interactions. For example, it would be possible to provide resources for emotional awareness (Garcia, Favela, & Machorro, 1999); that is to say, searchers being aware of their own and their partners' affective states with the aim of supporting empathy and pertinence in the social interactions. For instance, in this study the participants in collaborative conditions were not aware of their partners' affective states, hence, in the case of group members with different affective states (i.e. $C2^{+-}$), searchers in positive affective states were likely to feel frustrated and experience high affective load after infructuous attempts to interact with their partners in negative affective states.

The results obtained in relation to the expressive component of affective states and its connection with the evaluation of the information would have direct implications in studies of relevance feedback. Unlike Arapakis' et al. (2008) work, findings from this study showed that smiles could be a good local indicator of the relevance of the documents and that the expressions of smiles could be affected by initial affective states. Additionally, it was found that both smiles and EDA would be good global indicators of successful completion (from the point of view of searchers) of search tasks.

It is important to note that since more evidence is required to validate the results of this research, the practical implications discussed above are based on some speculations. Yet, they provide some insights for new research on this topic. In this regard, the following section describes future research directions derived from this work.

6.5 Future directions

While the results of this work addressed a set of research questions and led to the evaluation of different hypotheses, the limitations of this study raise new questions and problems to be addressed. Both theoretical and practical implications of this work require further studies to validate and generalize the results presented above. Below, theoretical and practical issues to be addressed in future studies are listed.

1. To what extent can the results obtained in this study be replicated with different samples? Is it the case that varying demographic aspects or sampling procedures have similar effects?
2. As noted in the limitations of this work, initial affective states were induced through contextual stimuli. Therefore, it was speculated that with this type of affective induction, participants could have related their affective states with their performance, thus producing a mood-incongruent behavior. In this regard, would the results be the same if non-contextual stimuli are applied instead?
3. This dissertation was framed around positive and negative initial affective states; however, further research is necessary to explore particular categories of affective processes and evaluate their implications in information search. For example, does feeling angry have different implications for information search than feeling sad? Similarly, does

feeling happy have different implications for information search than feeling confident?

4. To what extent can the results obtained in this study be replicated with other types of search tasks? This study was conducted using precision-oriented search tasks; however, other types such as recall-oriented search task may have similar or different effects.
5. To what extent can the research findings of this work be found in natural settings? What are effective ways to investigate this research topic outside the lab?
6. As noted above the study presented in this dissertation focused on dyads; hence, as part of the future directions of this work it will be necessary to investigate if the research findings for dyads could be applied to larger group sizes. For instance, how does a group of three individuals in initial positive affective states and one individual in negative affective states contrast with a group of three individuals in negative affective states and one individual in positive affective states?
7. In a search task with affectively loaded information (e.g. sensible topics), what would be good affective indicators of the relevancy of documents? Likewise, what would be good affective indicators of successful completion of a search task?
8. How can IR systems integrate affective components to enhance the way information is retrieved and presented to searchers?
9. How can systems take advantage of initial affective processes to better mediate the interactions of group members in CIS?

10. What are technical and ethical considerations that should be taken into account when designing systems that use affective signals to enhance the interaction with IR systems and other searchers?

Addressing the above research problems involves different challenges for researchers. Some of these problems are fundamental to validate the results of this dissertation and thus support the theoretical implications discussed above. While this list is not exhaustive, it provides general ideas to further explore this line of research.

6.6 Final remarks

This dissertation has provided interesting insight about the potential role of affective states in CIS. While the major focus of this research was on CIS, research findings derived from this work also have theoretical and practical implications for individual settings. As discussed in previous sections, the study design addressed important research questions, however unaddressed issues still require attention in order to expand the interpretations and implications of this work.

As a result of this work, it is now hypothesized that affective states as preconditions of search processes have implications in the way information is searched and processed by individuals and groups. Based on the proposed model depicted in Figure 6.1, this hypothesis aims to expand upon existing models that consider the affective component only as an internal resource or byproduct of different stages of information search.

In addition to providing initial evidence for this hypothesis, this dissertation has contributed through extensive literature review toward affective research and its relation with information seeking and CIS. Moreover, the research framework introduced in Chapter 3 (i.e. theoretical framework, evaluation framework, and preliminary studies) provides a valuable foundation to address future research in this topic. Regarding the methodological approach, in spite of its intrinsic limitations, the study design, instruments, and analyses procedures have provided a rigorous methodological approach to investigate affective states in information seeking. Nonetheless, future research using these resources as a reference should first address the limitations discussed in section 6.2 in order to improve the external validity of their findings.

References

Affectiva Inc. (2013). Affectiva Affdex. http://www.affectiva.com, Retrieved March 6, 2013.

Agarwal, N. (2012). Happiness: Exploring relationship with information and place in information science research and education. Poster in *Proceedings of ASIST 2012*. Baltimore, MD, USA. October 26 - 30, 2012.

Allen, B. (1997). Information needs: A person-in-situation approach. In: Vakkari, P., Savolainen, R. & Dervin, B. (eds.), *Information seeking in context.* London: Taylor Graham: 111-122.

Arapakis, I., Jose, J.M., & Gray, P.D. (2008). Affective feedback: An investigation into the role of emotions in the information seeking process. *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA*, 395–402.

Baeza-Yates, R., & Pino, J. A. (1997). A first step to formally evaluate collaborative work. In *Proceedings of ACM Group Conference*, pp. 56–60, Phoenix, AR.

Bartsch, A., & Hübner, S. (2005). Towards a theory of emotional communication. *CLCWeb: Comparative Literature and Culture, 7*(4), http://docs.lib.purdue.edu/clcweb/vol7/iss4/2

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*, 407-424.

Bechara, A., & Damasio, H. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and economic behavior, 52*, 336-372.

Bloch, S. (1988). Rejoinder to comments on "Effector patterns of basic emotions". *J. Social & Biological Structures, 11*, 213-214

Bloch, S., Lemeignan, M., & Aguilera-T, N. (1991). Specific respiratory patterns distinguish between basic emotions . *International Journal of Psychophysiology*, 11: 141-154.

Bloch, S. (2002). *Alba Emoting: Bases científicas del Emocionar*. Santiago de Chile: Editorial Universidad de Santiago.

Boucsein, W. (1992). *Electrodermal activity.* New York: Plenum Press.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25 (1)*, 49-59.

Bradley, M. M., & Lang, P. J. (1999). International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida

Bradley, M.M., & Lang, P.J. (2007). The international affective picture system (IAPS) in the study of emotion and attention. In J.A. Coan, & J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 29-46). Oxford University Press.

Buckminster Fuller, R. (1975). Synergetics: Explorations in the Geometry of Thinking. http://www.rwgrayprojects.com/synergetics/synergetics.html. Retrieved March 1, 2013.

Cannon, W.B. (1922). Bodily changes in pain, hunger, fear and rage: An account of recent researches into the function of emotional excitement. New York and London. D. Appleton and Company.

Cannon, W.B. (1927). The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *The American Journal of Psychology 39*, 106–124.

Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., & Zhang, Y. (2007). Effects of structure and interactions style on distinct search tasks. In *Proceedings of JCDL'07*, pp. 442-451.

Chou, S. W., & Min, H. T. (2009). The impact of media on collaborative learning in virtual settings: The perspective of social construction. *Computers and Education, 52*, 417-431.

Churchill, E.F., Sullivan, J.W., Golovchinsky, G., & Snowdon, D. (1998). Collaborative and co-operative information seeking: CSCW'98 workshop report. ACM SIGGROUP Bulletin, 20(1), pp. 56-59.

Ciarrochi, J., & Forgas, J.P. (2000). The pleasure of possessions: affective influences and personality in the evaluation of consumer items. *European journal of social psychology, 30*(5), 631-649.

Clark, H. H., & Brennan S. E. (1991). Grounding in communication. In L. B. Resnick, R. M. Levine, and S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). The American Psychological Association.

Clark, D.M., & Teasdale, J.D. (1985). Constraints of the effects of mood on memory. *Journal of Personality and Social Psychology*, 48, 1595-1608.

Coan, J.A., & Allen, J.J.B. (Eds.) (2007). Handbook of emotion elicitation and assessment. Oxford University Press.

Cohn, J.F., Ambadar, Z., & Ekman, P. (2007). The observer-based measurement of facial expression with the facial action coding system. In J.A. Coan, &

J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 203-221). Oxford University Press.

Cohn, J.F., & Kanade, T. (2007). Use of automated facial image analysis for measurement of emotion expression. In J.A. Coan, & J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 222-238). Oxford University Press.

Daft, R. & Lengel, R. (1984). Information richness: A new approach to managerial behavior and organizational design. In B. Straw, & L. Cummings (Eds.), *Research in organizational behavior*. Greenwich, CT: JAI Press.

Damasio, A. R. (1994). *Descartes error: Emotion, reason, and the human brain.* New York, NY, USA: Putnam/Grosset Press.

Darwin, C. A. (1872). *The expression of the emotion in man and animals.* London: John Murray.

Davidson, R.J., & Fox, N.A. (1982). Asymmetrical brain activity discriminates positive and negative affective stimuli in human infants. *Science, 218*(4578), 1235-1237.

Denning, P.J., & Yaholkovsky, P. (2008). Getting to "We". *Communications of the ACM*, 51(4):19-24.

Derks, D., Bos, A. E., & Grumbkow, J. V. (2007). Emoticons and social interaction on the Internet: the importance of social context. *Computers in Human Behavior, 23*(1), 842-849. doi: 10.1016/j.chb.2004.11.013.

Derks, D., Bos, A. E. R., & Grumbkow, J. V. (2008). Emoticons and Online Message Interpretation. *Soc. Sci. Comput. Rev., 26*(3), 379-388.

Dervin, B. (1983). An Overview of Sense-making Research: Concepts, Methods, and Results to Date. *Annual meeting of International Communication Association*. Dallas, Texas, USA.

Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative–quantitative methodology. In J. D. Glazer & R. R. Powell (Eds.), Qualitative research in information management (pp. 61–82). Englewood, CO: Libraries Unlimited, Inc.

Dervin, B. (1998). Sense-Making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, *2* (2), 36-46.

Eich, E., Ng, J.T.W., Macaulay, D., Percy, A.D., & Grebneva, I. (2007). Emotion elicitation using films. In J.A. Coan, & J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 124-136). Oxford University Press.

Ekman, P. (1977). Biological and cultural contributions to body and facial movement. In J. Blacking (Ed.), *Anthropology of the body* (pp. 34-84). London: Academic Press.

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

Ekman, P. (2006). Darwin and facial expression: A century of research in review. Malor books. Los Altos, CA.

Ekman, P. (2007). The directed facial action task: Emotional response without appraisal. In J.A. Coan, & J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 47-53). Oxford University Press.

Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (Eds.). (2002). Facial Action Coding System [E-book]. Salt Lake City, UT: Research Nexus.

Ekman, P., & Friesen, W. V. (2003). Unmasking the face: A Guide to recognizing emotions from facial expressions. Malor books. Los Altos, CA.

El Kaliouby, R., & Robinson, P. (2004). Mind Reading Machines: Automated Inference of Cognitive Mental States from Video. In *Proceedings of The IEEE International Conference on Systems, Man and Cybernetics*.

Evans, D. (2001). *Emotion: the Science of Sentiment*, Oxford University Press.

Evans, D. (2002). The search hypothesis of emotion. *The British Journal for the Philosophy of Science, 53(40)*, 497-509.

Face and Object Detection (2011). Official Website. http://www.iis.fraunhofer.de/en/bf/bv/ks/gpe/index.jsp, Retrieved May 20, 2011.

Barrett, L.F., & Russell, J.A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology 74* (4), 967–984.

Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), A handbook of process tracing methods for decision research (pp. 163-184). New York, NY: Psychology Press.

Fidel, R., Pejtersen, A. M., Cleal, B., & Bruce, H. (2004). A multidimensional approach to the study of human information interaction: A case study of collaborative information retrieval. *Journal of the American Society for Information Science and Technology, 55(11)*, 939-953.

Forgas J.P. (1991). Mood effects on partner choice: Role of affect in social decisions. *Journal of Personality and Social Psychology 61,* 708-720.

Forgas, J.P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117*, 39-66.

Forgas, J.P. (2009). The affect infusion model (AIM): An integrative theory of mood effects on cognition and judgments. In L.L. Martin, & G.L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 101-136). Mahwah, N.J., Lawrence Erlbaum Associates, Inc.

Forgas, J., Dunn, E., & Granland, S. (2008). Are you being served…? An unobtrusive experiment of affective influences on helping in a department store. *European Journal of Social Psychology, 38*(2), 333-342.

Forgas, J.P., & George, J.M. (2001). Affective influences on judgments and behavior in organizations: An information processing perspective. *Organizational behavior and human decision processes*, *86*(1), 3-34.

Foster, J. (2006). Collaborative information seeking and retrieval. Annual Review of Information Science and Technology (ARIST), 40, 329-356.

Fox, S., Karnawat, K., Mydland, M., Dumais, S., & White, T. (2005). Evaluating implicit measures to improve web search. ACM Transactions on Information Systems (TOIS). 23(2): 147-168.

Fredrickson, B.L. (2009). Positivity: Top-notch research reveals the 3-to-1 ratio that will change your life. Crown publishers, New York, USA.

Fredrickson, B. L., & Levenson, R. W. (1998). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition and Emotion*, 12, 191-220.

Fredrickson, B.L. & Losada, M.F. (2005). The positive affect and the complex dynamics of human flourishing. *American Psychologist, 60(7)*, 678-686.

Fredrickson, B. L., Mancuso, R. A., Branigan, C., & Tugade, M. M. (2000). The undoing effect of positive emotions. *Motivation and Emotion. 24,* 237-258.

Fredrickson, B.L., Tugade, M.M., Waugh, C.E., Larkin, G. (2003). What good are positive emotions in crises?: A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology*, 84, 365-376.

Friedman, R., Anderson, C., Brett, J., Olekalns, M., Goates, N., & Lisco, C.C. (2004). The positive and negative effects of anger on dispute resolution: Evidence from electronically mediated disputes. Journal of Applied Psychology, 59(2), 369-376.

Fussell, S. (2003). *The verbal communication of emotions*. Mahwah, NJ:Lawrence Erlbaum Associates.

Gable, S., & Haidt, J. (2005). What (and why) is Positive psychology? *Review of General Psychology, 9*(2), 103–110.

Galloway, V.B. (1980). Perceptions of the communicative efforts of American students of Spanish. *The modern language journal, 64*(4), 428-433.

García, O., Favela, J., & Machorro, R. (1999). Emotional awareness in collaborative systems. In *Proceedings of String Processing and*

Information Retrieval Symposium & International Workshop on Groupware, pp. 296-303.

Gobl, C., & Chasaide, A.N. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication, 40* (1–2), 189-212.

Golovchinsky, G., Dunnigan, A. & Diriye, A. (2012). Designing a tool for exploratory information seeking. Proc. CHI, 1799–1804.

González-Ibáñez, R. (2006). Evaluación de la Integración del Darse-Cuenta Emocional en una Aplicación Colaborativa. Master's Thesis, Universidad de Santiago de Chile. Santiago, Chile.

González-Ibáñez, R. (2010). A Proposal for Studying Users' Behaviors in Collaborative Information Seeking through a Convergence Map. In *Proceedings of Collaborative Information Seeking Workshop at GROUP 2010*. Sanibel Island, FL, USA. November 7, 2010.

González-Ibáñez, R., (2011). Skin Conductance response for information science research. Manuscript prepared for the Google Research Award project "Implicit Detection of Relevance Decisions and Affect in Web Search" lead by Dr. Jacek Gwizdka.

González-Ibáñez, R., Haseki, M., & Shah, C. (2011). A Method to Evaluate the Synergic Effect in Collaborative Information Seeking. Workshop on Human-Computer Interaction and Information Retrieval (HCIR). October 20. Google Main Campus, Mountain View, CA.

González-Ibáñez, R., Haseki, M., & Shah, C. (2012a). Understanding Effects of Time and Proximity on Collaboration: Implications for Technologies to Support Collaborative Information Seeking. Work in progress in Proceedings of CHI'12, Austin, TX, USA, May 5-10, 2012.

González-Ibáñez, R., Haseki, M., & Shah, C. (2012b). Time and Space in Collaborative Information Seeking: The Clash of Effectiveness and Uniqueness. In *Proceedings of ASIS&T 2012*, Baltimore, MD, USA, October 26-30, 2012.

González-Ibáñez, R., Haseki, M., & Shah, C. (2013). Let's work together, but not too close! An analysis of communication and performance in collaborative information seeking. *Information Processing and Management, 49,* 1165-1179.

González-Ibáñez, R., & Shah, C. (2010). Group's affective relevance: a proposal for studying affective relevance in collaborative information seeking. Poster in *Proceedings of GROUP 2010*. Sanibel Island, FL, USA. November 6 - 10, 2010.

González-Ibáñez, R., & Shah, C. (2011). Coagmento: A System for Supporting Collaborative Information Seeking. *Integrated demo at ASIST'2011*, New Orleans, LA, USA, October 9-12, 2011.

González-Ibáñez, R., Shah, C., & Córdova-Rubio, N. R. (2011). Smile! Studying expressivity of happiness as a synergic factor in collaborative information seeking. In *Proceedings of American Society of Information Science & Technology (ASIST'2011) Annual Meeting*, New Orleans, LA, USA, October 9-12, 2011.

González-Ibáñez, R., Shah, C., & White, R. (2012). Pseudo-Collaboration as a Method to Perform Selective Algorithmic Mediation in Collaborative IR Systems. In *Proceedings of ASIS&T 2012*, Baltimore, MD, USA, October 26-30, 2012.

Gray, E.K., & Watson, D. (2007). The international affective picture system (IAPS) in the study of emotion and attention. In J.A. Coan, & J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 171-183). Oxford University Press.

Grice, H. P. (1975). Logic and conversation. In *Studies in the Way of Words*, H. P. Grice (Ed.), pp. 22–40. Cambridge, MA: Harvard University Press (1989).

Gross, J.J., & Levenson, R.W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. Journal of abnormal psychology, 106(1), 95-103.

Gupta, R., Duff, M., & Tranel, D. (2011). Bilateral amygdale damage impairs the acquisition and use of common ground in social interaction. *Neuropsychology, 25(2)*, 137-146.

Gwizdka, J., & Lopatovska, I. (2009). The role of subjective factors in the information search process. *JASIST. 60* (12), 2452–2464.

Hansen, P., & Järvelin, K. (2004). Collaborative information searching in an information-intensive work domain: Preliminary results. *Journal of Digital Information* Management*, 2,* 26-29.

Hart, S. G., & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.

Haseki, M., Shah, C., González-Ibáñez, R. (2012). Time as a trigger of interaction and collaboration in research teams: A diary study. In *Proceedings of CSCW'12.* ACM DL.

Hertzum, M. (2008). Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing & Management*, 44(2), p. 957-962.

Hornik, R., Risenhoover, N., & Gunnar, M. (1987). The effects of maternal positive, neutral, and negative affective communications on infant responses to new toys. *Child development, 58*(4), 937-944.

Hyldegård, J. (2006a). Collaborative Information Behaviour - Exploring Kuhlthau's Information Search Process Model in a Group-based Educational Setting. *Information Processing and Management, 42(1)*, 276–298.

Hyldegård, J. (2006b). Using diaries in group based information behavior research - a methodological study. In *Proceedings of Information Interaction in Context, IIiX*, Copenhagen Denmark, pp.153-161.

Hyldegård, J. (2006c). *Between individual and group – exploring group members' information behavior in context.* Doctoral dissertation, Department of Information Studies. Royal School of Library and Information Science, Denmark.

Hyldegård, J. (2009). Beyond the search process: exploring group members' information behavior in context. *Information Processing and Management, 45*, 142-158.

Ingwersen, P. (2005). Integrative framework for information seeking and interactive information retrieval. In K. E. Fisher, S. Erdelez, & E. F. McKechnie (Eds.), *Theories of Information Behavior* (pp. 215-220). Medford, NJ: Information Today.

Irtel, H. (2007). *PXLab: The Psychological Experiments Laboratory* (Vers. 2.1.11). Online im Internet: URL: http://www.pxlab.de (Stand 19. 6. 2007).

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

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

Izard, C.E. (1983). Maximally discriminative facial movement coding system (MAX). Unpublished manuscript, University of Delaware, Newark.

Jaffe, J. M., Lee, Y., Huang, L., & Oshagan, H. (1995). Gender, Pseudonyms, and CMC: Masking Identities and Baring Souls. Paper presented at the 45th Annual Conference of the International Communication Association (ICA) in Albuquerque, New Mexico, USA; May 1995. http://research.haifa.ac.il/~jmjaffe/genderpseudocmc/

Jain, V., and Mukherjee, A. (2002). The Indian Face Database. http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/. Last time visited on May 20, 2011.

Jäverlin, K., & Ingwesen, P. (2004). Information seeking research needs extension towards tasks and technology. *Information research, 10* (1), October. http://informationr.net/ir/10-1/paper212.html. Retrieved March 7, 2013.

Karunakaran, A., Spence, P. R., & Reddy, M. (2010). Towards a model of Collaborative Information Behavior. Workshop on Collaborative Information Retrieval. ACM Conference on Computer Supported Cooperative Work (CSCW 2010). Savannah, GA. Feb. 7, 2010.

Kelly, G. A. (1963). *A theory of personality: The psychology of personal constructs*. New York: Norton.

Kelly, D., and Belkin, N. J. 2001. Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback. In *Proceedings of the 24th Annual international ACM SIGIR '01*. 408-409.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy Enlisted Personnel. Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station Memphis.

Kirkland, M.R., & Saunders, M.A.P. (1991). Maximizing student performance in summary writing: Managing cognitive load. TESOL Quarterly, 25(1), 105-121.

Klein, G., & Adelman, L. (2005). A collaboration evaluation framework. MITRE Corporation.

Kueblbeck, C. & Ernst, A. (2006). Face detection and tracking in video sequences using the modified census transformation. *Journal on Image and Vision Computing*, 24, (6), 564-572, 2006, ISSN 0262-8856.

Kuhlthau, C. (1991). Inside the Search Process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42 (6)*, 361-371.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Larsen, R.J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. *Review of Personality and Social Psychology, 13*, 25-59.

Lerner, J.S., & Keltner, D. (2002). Beyond valence: Toward a model of emotion-specific influence on judgment and choice. *Cognition and emotion, 14*(4), 473-493.

London, S. (1995). Collaboration and community. An essay prepared for Pew Partnership for Civic Change. http://www.scottlondon.com/reports/ppcc.html. Retrieved March 1, 2013.

Lopatovska, I. (2009a). Does the mood matter? Proceedings of the 2009 International Conference on Affective Computing & Intelligent Interaction ACII 2009, pp. 237-240, Amsterdam, The Netherlands.

Lopatovska, I. (2009b). *Emotional Aspects of the Online Information Retrieval Process*. Doctoral dissertation, School of Communication, Information, and Library Studies, Rutgers University.

Lopatovska, I., & Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction. *Information processing and management, 47(4)*, 575-592.

Losada, M., & Heaphy, E. (2004). The Role of Positivity and Connectivity in the Performance of Business Teams: A Nonlinear Dynamics Model. *American Behavioral Scientist, 47(6)*, 740-765.

Lyons, M., Akamatsu, A., Kamachi, M., and Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. In *Proceedings of the third IEEE International Conference on Automatic Face and Gesture Recognition.* April 14-16, 1998, Nara, Japan, pp. 200-205.

Martin, M. (1990). On the induction of mood. *Clinical Psychology Review*, 10, 669-697.

Mennecke, B. E., Valacich, J. S., & Wheeler, B.C. (2000). The effects of media and task on user performance: A test of the task-media fit hypothesis. *Group Decision and Negotiation, 9*, 507-529.

Merriam-Webster (2013). Merriam-Webster: An encyclopedia Britannica company. http://www.meriam-webster.com.

Morris, M.R. (2007). Collaborating Alone and Together: Investigating Persistent and Multi-User Web Search Activities, Microsoft Research Technical Report #MSR-TR-2007-11, January 2007.

Morris, M. R. (2008). A Survey of Collaborative Web Search Practices. In Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 1657-1660, Florence, Italy.

Morris, M. R., & Horvitz, E. (2007). SearchTogether: An interface for collaborative web search. *ACM Symposium on User Interface Software and Technology (UIST)*, pp. 3–12, Newport, RI.

Nahl, D. (2005). Affective load theory (ALT). In K. E. Fisher, S. Erdelez, & E. F. McKechnie (Eds.), *Theories of information behavior* (pp. 39-43). Medford, NJ: Information Today.

Nahl, D. (2009). The centrality of affective in information behavior. In D. Nahl & D. Bilal (Eds.), *Information and emotion: The emergent affective paradigm in information behavior research and theory* (pp 3-37). Medford, NJ: Information Today, Inc.

Nahl, D., & Bilal, D. (Eds.) (2009). *Information and emotion: The emergent affective paradigm in information behavior research and theory* (pp 3-37). Medford, NJ: Information Today, Inc.

Nahl, D., & Tenopir, C. (1996). Affective and cognitive searching behavior of novice end-users of a full-text database. *Journal of the American Society for Information Science, 47(4)*, 276–286.

Nakasone, A., Prendinger, H., & Ishizuka, M. (2005). Emotion recognition from electromyography and skin conductance. In *Proceedings of The Fifth International Workshop on Biosignal Interpretation (BSI-05).*

Neal, D.M., & McKenzie, P.J. (2011). Putting the pieces together: endometriosis blogs, cognitive authority, and collaborative information behavior. *J Med Libr Assoc 99*(2), 127-134.

Neale, D.C., Carroll, J.M, & Rosson, B. (2004). Evaluating computer-supported cooperative work: Models and frameworks. In *Proceedings of the 2004 ACM conference on computer supported cooperative work*. Chicago, Illinois, USA. pp. 112-121.

Nielsen, J. (2012). How many test users in a usability study? Nielsen Norman Group. http://www.nngroup.com/articles/how-many-test-users/. Retrieved March 26, 2013.

Ostrach, T. R. (1997). Typing Speed: How Fast is Average. Five Star Staffing, Inc., Orlando, FL.

Owren, M.J., & Bachorowski, J. (2007). Measuring emotion-related vocal acoustics. In J.A. Coan, & J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 239-266). Oxford University Press.

Oxford (2013). Oxford dictionaries: The world's most trusted dictionaries. http://oxforddictionairies.com.

Pantic, M., & Rothkrantz, L. (2000), Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on pattern analysis and machine intelligence, 22* (12), 1424-1445.

Paul, S.A, & Morris, M.R. (2009). CoSense: Enhancing sensemaking for collaborative web search. In *Proceedings of CHI'2009*. Boston, MA, USA.

Palmero, F., Guerrero, C., Gómez, C., & Carpi, A. (2006). Certezas y controversia en el estudio de la emoción. *Revista electrónica de motivación y emoción (R.E.M.E), 9*, 23-24.

Peer, P., (2011). CVL Face Database: http://www.lrv.fri.uni-lj.si/facedb.html on May 20.

Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). Linguistic Inquiry and Word Count (LIWC): LIWC2001. Mahwah, NJ: Erlbaum Publishers.

Pereira, C. (2000). Dimensions of emotional meaning in speech. In *ISCA Workshop on Speech and Emotion*.

Peter, C., & Herbon, A. (2006). Emotion representation and physiology assignments in digital systems. *Interacting with computers*, *18(2),* 139-170.

Pettigrew, K., Fidel, R., & Bruce, H. (2001). Conceptual Frameworks in Information Behavior. In M.E. Williams (Ed.), Annual Review of Information Science and Technology, vol. 36 (pp. 43–78). Medford, NJ: Information Today.

Picard, R. W. (1997). Affective Computing. Cambridge, MA: The MIT Press.

Picard, R. W. (2003). Affective computing: challenges. International Journal of Human-Computer Studies, 59(1-2), 55-64.

Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., & Back, M. (2008). Algorithmic Mediation for Collaborative Exploratory Search. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR'*2008). ACM, New York, NY, pp. 315-322.

Plutchik, R. (1970). Emotions, evolution, and adaptive process. In M. B. Arnold (Ed.), *Feelings and emotions* (pp. 3-24). New York: Academic Press.

Plutchik, R., (1980). A general psychoevolutionary theory of emotion. In: Plutchik, R., Kellerman, H. (Eds.), Emotion: Theory, Research, and Experience: vol. 1. Theories of Emotion (pp. 3–33). Academic Press, New York.

Provine, R. R., Spencer, R. J., & Mandell, D. L. (2007). Emotional Expression Online: Emoticons Punctuate Website Text Messages. *Journal of Language and Social Psychology, 26*(3), 299-307. doi: 10.1177/0261927X06303481.

Reeve, J.M. (1994). *Motivación y emoción.* McGraw-Hill / Interamericana de España, S.A. ISBN: 8448118251 ISBN-13: 9788448118259

Reddy, M., & Dourish, P. (2002). A finger on the pulse: Temporal rhythms and information seeking in medical work. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work CSCW 2002, Chicago, IL.

Reddy, M., & Jansen, J.A. (2008). Model for understanding collaborative information behavior in context: A study of two healthcare teams. *Information Processing and Management*, 44(1): p. 256-273.

Rottenberg, J., Ray, R.D., & Gross, J.J. (2007). Emotion elicitation using films. In J.A. Coan, & J.J.B. Allen (Eds.), Handbook of emotion elicitation and assessment (pp. 9-28). Oxford University Press.

Russell, J.A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology 45*(6), 1281–1288.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57, 493–502.

Salkind, N J. (2010). *Encyclopedia of research design.* SAGE Publications. ISBN: 9781412961271.

Savolainen, R. (2007). Information behavior and information practice: Reviewing the “umbrella concepts” of information-seeking studies. *The library quarterly, 77* (2), 109-132.

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

Scherer, K.R., & Shorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: theory, methods, research.* Canary, NC: Oxford University Press.

Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proceedings of Eurospeech 2001*, Aalborg, vol. 1, pp. 87–90.

SearchTogether (2008), Microsoft SearchTogether (Beta). http://research.microsoft.com/en-us/um/redmond/projects/searchtogether/ Retrieved March 1, 2013.

Sebe, N., Lew, M. S., Sun, Y., Cohen, I., Gevers, T., Huang, T. (2007). Authentic facial expression analysis. *Image Vision Comput.* 25(12): 1856-1863.

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress empirical validation of interventions. *American Psychologist, 60(5)*, 410–421.

Shah, C. (2008). Toward collaborative information seeking (CIS). In Proceedings of JCDL Workshop on Collaborative Information Retrieval, Pittsburgh, PA, USA.

Shah, C. (2010a). Collaborative information seeking: A Literature Review. *Advances in Librarianship*, 32, 3-33.

Shah, C. (2010b). Coagmento - A Collaborative information seeking, synthesis and sense-making framework. *Integrated demo at CSCW 2010.* Savannah, GA: February 6-11, 2010.

Shah, C. (2010c). A framework for supporting user-centric collaborative information seeking. Doctoral dissertation, University of North Carolina at Chapel Hill.

Shah, C. (2012). Collaborative information seeking: The art and science of making the whole greater than the sum of all. The Information Retrieval Series, Vol. 34. Springer. ISBN 978-3-642-28812-8.

Shah, C., & González-Ibáñez, R. (2010). Exploring Information Seeking Processes in Collaborative Search Tasks. In Proceedings of American Society of Information Science and Technology (ASIST). Pittsburgh, PA. October 22-27, 2010.

Shah, C., & González-Ibáñez, R. (2011). Evaluating the Synergic Effect of Collaboration in Information Seeking. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR'2011),* pp. 913-922, Beijing, China.

Shah, C., & González-Ibáñez, R. (2012). Spatial Context in Collaborative Information Seeking. *Journal of Information Science (JIS)*, April 27, 2012.

Shah, C., Liu, J. González-Ibáñez, R., & Belkin, N. (2012). Exploration of dynamic query suggestions and dynamic search results for their effects on search behaviors. In *Proceedings of ASIST 2012*, Baltimore, MD, USA, October 26-30, 2012.

Shaw, M. E. (1981). Group dynamics: The psychology of small group behavior. New York: McGraw Hill.

Shepherd, P. (2001). The GSR Meter Course: Biofeedback monitoring skills in the context of transformational psychotherapy. Tools for Transformation, http://www.freezoneearth.org/allmeters/index.htm.

Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. London: Wiley Press.

Sinclair, R.C., & Mark, M.M. (1995). The effects of mood state on judgemental accuracy: Processing strategy as a mechanism. *Cognition and Emotion, 9(5)*, 417-438.

Smeaton, A.F., Lee, H., Foley, C., & McGivney, S. (2007). Collaborative video searching on a tabletop. *Multimedia Systems, 12*(4-5), 375-391

Smyth, B., Balfe, E., Boydell, O., Bradley, K., Briggs, P., Coyle, M., & Freyne, J. (2005). A live-user evaluation of collaborative web search. Paper presented at *International Joint Conference on Artificial Intelligence (IJCAI)*, Edinburgh, Scotland.

Solomon, P. (1999) Information mosaics: Patterns of action that structure. In T.D. Wilson & D.K. Allen (Eds.), Exploring the Contexts of Information behaviour: Proceedings of the Second International Conference on Research in Information Needs, Seeking and Use in Different Contexts (pp. 150 –175). London: Taylor Graham.

Sonnentag, S., & Sparr, J.L. (2007). Affective traits. In S.G. Rogelberg (Ed.), Encyclopedia of industrial and organizational psychology (pp. 14-16), SAGE Publications, Inc.

Sonnenwald, D. H., & Pierce, L. G. (2000). Information behavior in dynamic group work contexts: Interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing & Management, 36*, 461-479.

Strauss, M., Reynolds, C., Hughes, S., Park, K., McDarby, G. & Picard, R. (2005). The Handwave Bluetooth Skin Conductance Sensor. Springer. pp. 699-706.

Strapparava, C. and Valitutti, A. (2004). Wordnet-Affect: An Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon.

Strijbos, J. W., Martens, R. L., Jochems, W., & Broers, N. J. (2004). The effect of functional roles on group efficiency: Using multilevel modelling and content analysis to investigate computer-supported collaboration in small groups. Small Group Research, 35, 195-229

Sutherland, G. Newman, B., & Rachman, S. (1982). Experimental investigations of the relations between mood and intrusive unwanted cognitions. *British Journal of Medical Pshychology*. 55, 127-138.

Tafrate, R.C., Kassinove, H., & Dundin, L. (2002). Anger episodes in high- and low-trait-anger community adults. *Journal of clinical psychology, 58* (12), 1573-1590.

Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J. C., & Buxton, V. (2010). Three's company: Understanding communication channels in three-way distributed collaboration. In *Proceedings of CSCW 2010*. Savannah, GA: February 6-11, 2010

Taylor, G.J. (1984). Alexithymia: Concept, measurement, and implications for treatment. *The American Journal of Psychiatry, 141*(6), 725-732.

Taylor-Powell, E.,Rossing, B., & Geran, J. (1998). Evaluating collaboratives: reaching the potential. Technical report, University of Wisconsin-Extension, Madison, Wisconsin.

Tenopir, C. (2003). Information metrics and user studies. *Aslib proceedings, 55*(1/2) 13-17.

Twidale, M. B. T., Nichols, D. M. N., & Paice, C. D. (1997). Browsing is a Collaborative Process. Information Processing and Management, 33(6), 761-783.

Watson, D., Clark, Lee A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54(6)*, 1063-1070.

Watson, J.B. (1924). *Behaviorism.* New York: Norton.

Waugh, C.E., & Fredrickson, B.L. (2006). Nice to know you: Positive emotions, self-other overlap, and complex understanding in the formation of a new relationship*. The Journal of Positive Psychology, 1(2)*, 93–106.

White, R. W., & Huang, J. (2010). Assessing the scenic route: Measuring the value of search trails in web logs. Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR). Geneva, Switzerland.

Wilson P. (1983). *Second hand knowledge: an inquiry into cognitive authority.* Westport, CT: Greenwood.

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, 37(1), 3–15.

Wilson, M.L., & Wilson, M. (2010). Social anxieties and collaborative information seeking. In *Proceedings of Collaborative Information Seeking Workshop at GROUP 2010*. Sanibel Island, FL, USA. November 7, 2010.

Yue, Z., & He, D. (2009). A model for understanding collaborative information behavior in e-discovery. In *Proceedings of TREC* 2009.

Yuen, K.S.L., & Lee, T.M.C. (2003). Could mood state affect risk-taking decisions? *Journal of affective disorders, 75*, 11-18.

Zhao, J. (2006). The effects of induced positive and negative emotions on risky decision making. Talk presented at the *28th Annual Psychological Society of Ireland Student Congress*, Maynooth, Ireland.

Zhou, N., & Stahl, G. (2007). Towards Building a Math Discourse Community: Investigating Collaborative Information Behavior. *Online Communities and Social Computing. Lecture Notes in Computer Science 4564*, 509-518.

Appendix A. Glossary

1. *Affect*: “affect represents the essence of behavior, this understood from the most elemental formulation of approaching what likes, gratifies, or pleases, and estranging from what produces opposite consequences” (Palmero et al., 2005, p. 17). Long lasting affective process.
2. Affective processes/states/dimension: These terms are used in this dissertation when referring in a general sense to the following specific concepts: affect, mood, emotion, and feeling.
3. *Arousal*: Arousal refers to physiological activation or response (e.g. excited, relaxed, calm, etc.)
4. *CIS*: Collaborative information seeking
5. *Elicit*: Evoke, induce
6. *Emotion*: Multidimensional response of individuals to external or internal stimuli. Emotions last less than affects and moods.
7. *Feeling*: Subjective component of emotions.
8. *Information behavior*: According to Reddy and Jansen (2008), information behavior is "philosophy of seeking and use" (p.266)
9. *Labile*: In the context of this study and based on the measure conducted with *Affectiva Q Analytics*, a person is considered to be labile when his/her affective states are easily altered. This is expressed in EDA through multiple peaks in a given region.
10. *Lability*: In the context of this study is a measure related to regions with multiple peaks per minute. Lability characterizes individuals in terms of their high frequencies of peaks under particular conditions.

11. *Mood*: “[mood] denotes the existence of a set of beliefs about the probability of a subject to experience pleasure or pain in the future; that is, of experiencing positive or negative affect” (Palmero et al., 2005, p. 17). Long lasting affective process.
12. *Q score*: A “measure designed to give an indication of engagement/arousal” (http://www-assets.affectiva.com/assets/Q-Analytics-Guide.pdf, p.7)
13. *Search*: According to Reddy and Jansen (2008), searching is "strategic maneuvering" (p.266)
14. *Seeking*: According to Reddy and Jansen (2008), seeking is "tactical maneuvering" (p.266)
15. *SERP*: Search result page.
16. *Valence*: Refers to the tone of an affective process, that is to say, positive or negative.
17. *Trait affect*: “[I]s defined as a tendency to respond to specific classes of stimuli in a predetermined, affect-based manner. Therefore, an affective trait is considered a relatively stable characteristic of personality” (Sonnentag & Sparr, 2007)
18. *Mood-congruency*: Indicates that actions, decision making, judgments, and expressions are all consistent with one’s mood. For example, novice Target customer representatives in a positive mood are more helpful with customers (Forgas, Dunn, & Granland, 2008).
19. *Mood-incongruence*: Indicates that behavior and cognitive processing occurs in an opposite direction to mood. For instance, a person in a negative mood sees the outcomes of a risky decision making process as more favorable than a person in a positive mood.

Appendix B. Questionnaires and other instruments

Below, questionnaires to collect data about affective and cognitive aspects are provided.

B.1 Positivity Self Test

This test has been obtained from Fredrickson (2009) with permission of the author to be used in this study.

> How have you felt in the past twenty-four hours? Look back over the past day (i.e., from this time yesterday up to right now). Using the scale below, indicate the greatest degree that you've experienced of each of the following feelings. Take two minutes to complete this questionnaire.
>
> -- 0: Not at all | 1: A little bit | 2: Moderately | 3: Quite a bit | 4: Extremely --

1. What is the most amused, fun-loving, or silly you felt?
2. What is the most angry, irritated, or annoyed you felt?
3. What is the most ashamed, humiliated, or disgraced you felt?
4. What is the most awe, wonder, or amazement you felt?
5. What is the most contemptuous, scornful, or disdainful you felt?
6. What is the most disgust, distaste, or revulsion you felt?
7. What is the most embarrassed, self-conscious, or blushing you felt?
8. What is the most grateful, appreciative, or thankful you felt?
9. What is the most guilty, repentant, or blameworthy you felt?
10. What is the most hate, distrust, or suspicion you felt?
11. What is the most hopeful, optimistic, or encouraged you felt?
12. What is the most inspired, uplifted, or elevated you felt?
13. What is the most interested, alert, or curious you felt?
14. What is the most joyful, glad, or happy you felt?
15. What is the most love, closeness, or trust you felt?
16. What is the most proud, confident, or self-assured you felt?
17. What is the most sad, downhearted, or unhappy you felt?
18. What is the most scared, fearful, or afraid you felt?
19. What is the most serene, content, or peaceful you felt?
20. What is the most stressed, nervous, or overwhelmed you felt?

B.2 Self-assesment manikin (SAM)

Using the SAM scale, take the following 15 seconds to indicate how you feel. Respond as honestly as you can. Your responses will NOT affect your final score.

The instructions below were obtained and slightly modified from Bradley and Lang (1999)

Following you will see 3 sets of 5 figures, each arranged along a continuum. We call this set of figures SAM, and you will be using these figures to rate how you feel at different stages in this session. SAM shows three different kinds of feelings: Happy vs. Unhappy, Excited vs. Calm, and Controlled vs. In-Control.

Happy vs. Unhappy Scale

You can see that each SAM varies along each scale. In the illustration above, the first SAM scale is the happy-unhappy scale, which ranges from a smile to a frown. At one extreme (LEFT) of the happy vs. unhappy scale, you feel, pleased, satisfied, contented, hopeful. If you feel completely happy at the moment of completing this questionnaire, you can indicate this by selecting the option below the figure at the LEFT.

The other end of the scale (RIGHT) is when you feel completely, unhappy, annoyed, unsatisfied, melancholic, despaired, bored. You can indicate feeling completely unhappy by selecting the option below the figure at the RIGHT.

The figures also allow you to describe intermediate feelings of pleasure, by selecting the option below any of the other figures.

If you feel completely neutral, neither happy nor sad, select the option in the middle (RED ARROW).

If, in your judgment, your feeling of pleasure or displeasure falls between two of the pictures, then select an option BETWEEN the figures. This permits you to make more finely graded ratings of how you feel.

Excited vs. Calm Scale

The excited vs. calm dimension is the second type of feeling displayed here. At one extreme of the scale (LEFT) you feel stimulated, excited, frenzied, jittery, wide-awake, aroused. If you feel completely aroused at the moment of completing this questionnaire, select the option below the figure at the LEFT.

On the other hand, at the other end of the scale (RIGHT), you feel completely relaxed, calm, sluggish, dull, sleepy, unaroused. You can indicate you feel completely calm by selecting the option below the figure at the RIGHT.

As with the happy-unhappy scale, you can represent intermediate levels by selection options below any of the other figures.

If you are not at all excited nor at all calm, select the option in the middle (RED ARROW).

Again, if you wish to make a more finely tuned rating of how excited or calm you feel, select an option BETWEEN the figures.

Controlled vs. In-Control

The last scale of feeling that you will rate is the dimension of controlled vs. in-control. At one end of the scale (LEFT) you have feelings characterized as completely controlled, influenced, cared-for, awed, submissive, guided.

Please indicate feeling completely controlled by selecting the option below the figure at the LEFT.

At the other extreme of this scale (RIGHT), you feel completely controlling, influential, in control, important, dominant, autonomous. You can indicate that you feel dominant by selecting the option below the figure at the RIGHT.

Note that when the figure in this scale is large, you feel important and influential, and that it will be very small when you feel controlled and guided.

If you feel neither in control nor controlled you should select the option in the middle (RED ARROW).

Remember you can also represent your feelings BETWEEN these endpoints.

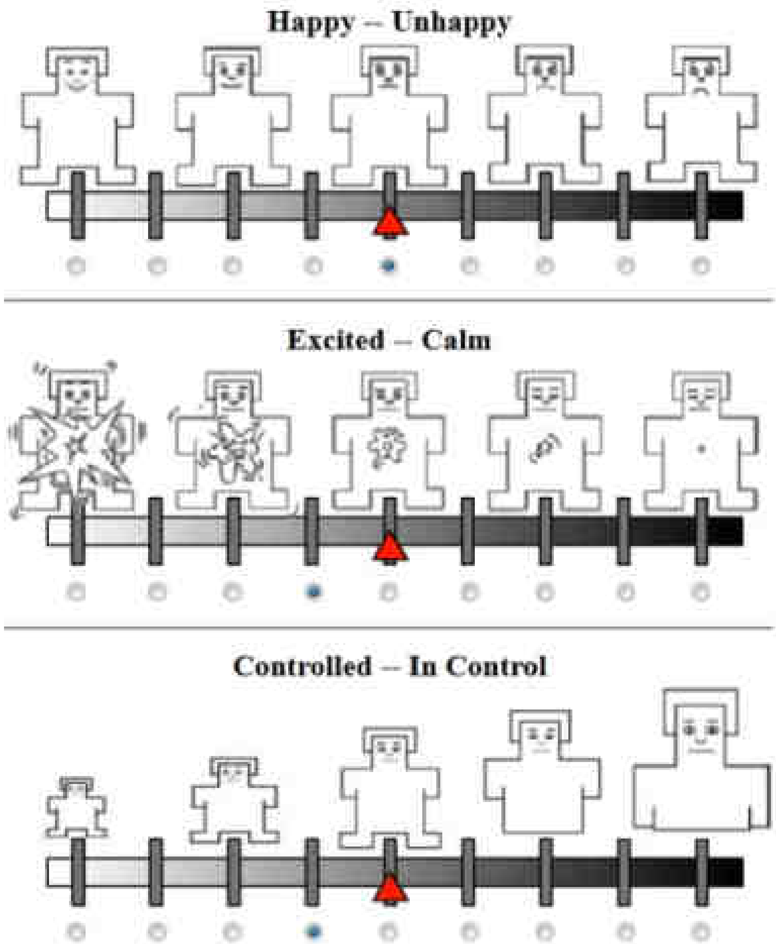


Figure B.1. Self Assessment Manikin (SAM). Adapted version from (Irtel, 2007). Some pictures were changed and scales were inverted for valence and arousal in order to match with the original SAM scale by Bradley and Lang (1994).

B.3 NASA Task Load indeX(TLX)

This questionnaire is a subjective workload self assessment test (Hart & Staveland, 1988). The test consists of six questions and 7-point scales that are used to provide the responses. Each point in the scale is subdivided in three increments representing low, medium, and high estimates. As a result, the overall scale has 21 gradations.

1. How mentally demanding was the task?
2. How physically demanding was the task?
3. How hurried or rushed was the pace of the task?
4. How successful were you in accomplishing what you were asked to do?
5. How hard did you have to work to accomplish your level of performance?
6. How insecure, discouraged, irritated, stressed, and annoyed were you?

B.4 Task perception questionnaires

Before questions

- Please answer the following questions on the scale of 1 to 5.
    1. How familiar are you with the topic of this question?

       (Not familiar at all) 1 2 3 4 5 (Very familiar)
    2. How challenging do you think this question will be for you?

       (Not challenging at all) 1 2 3 4 5 (Very challenging)

After questions

- Please answer the following questions on the scale of 1 to 5.
    1. How confident are you with the answer you found?

       (Not Confident at All) 1 2 3 4 5 (Very Confident)
    2. How difficult was to find the answer to this question?

       (Very Easy) 1 2 3 4 5 (Very Difficult)

B.5 Collaboration questionnaire

- Please answer the following questions on the scale of 1 to 5.

    How did you feel while collaborating?

    (Not absorbed intensely) 1 2 3 4 5 (Absorbed intensely)

    (Attention was not focused) 1 2 3 4 5 (Attention was focused)

    (Did not concentrate fully) 1 2 3 4 5 (Concentrated fully)

    (Not deeply engrossed/involved) 1 2 3 4 5 (Deeply engrossed/involved)

B.6 Supplementary questionnaires

Tutorial questionnaire

1. Was this tutorial clear?

   (Not clear at all) 1 2 3 4 5 (Very Clear)

End-session questionnaire

- Please rate the following statements on the scale of 1 (Not at all) to 5 (Very much).

  Overall, I am satisfied with how easy it is to use this system

  It was easy to learn to use this system

  The tutorial was helpful to use the system

  Overall, I am satisfied with this system

- Please rate the following statements on the scale of 1 to 5

  Using the system was:

  (Uninteresting) 1 2 3 4 5 (Interesting)

  (Not enjoyable) 1 2 3 4 5 (Enjoyable)

  (Dull) 1 2 3 4 5 (Exciting)

  (Not Fun) 1 2 3 4 5 (Fun)

B.7 Interview

The following questions were used by the researcher as a reference to guide the discussion during the interview stage.

1. How did you feel during the session? how different were your feelings between the first and second task?
2. What was the most difficult part and why?

3. Is there any particular moment that you can recall gave you the hardest time?
4. What was the easiest part and why?
5. Any question that made you feel good or bad?
6. How did you feel with respect to the feedback that the system provided you while working on the first task?
7. Have you ever participated in a study like this?
8. Any motivation?
9. What was the part that you liked most (if any) and why?
10. How did you feel collaborating with each other today?
11. How did you feel with respect to all the sensors we used in this study? is there any particular elemenet that made you feel uncomfortable?
12. How did you hear from this study?
13. Have you ever collaborated with other people to search information online?
14. Did you take, drink, or use anything during the past 12 hours that may affect your nervous system? like cafeine, nicotine, medicine or something else? You only need to reply yes or no.

Appendix C. Supplementary Figures and Tables

C.1 Affective dimension

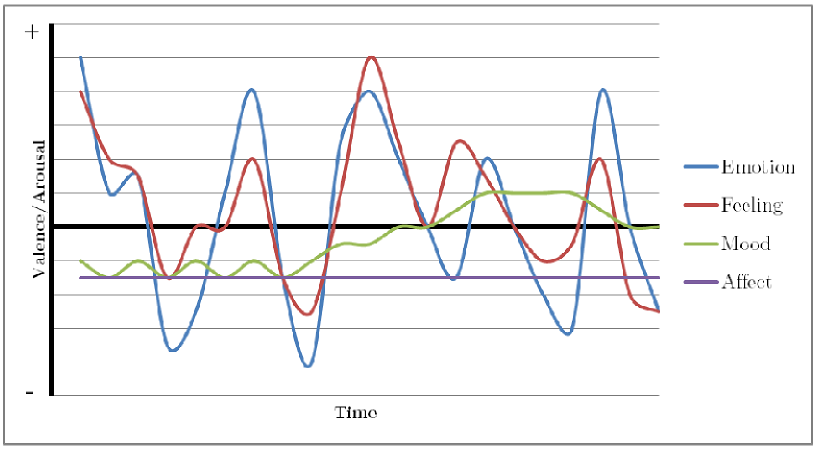


Figure C.1. Sketch of affective processes over time with respect to valence and arousal.

C.2 Group member in context (GMIC) model

Activities					
⇨ Task stages	Construction		Performance		Completion
⇨ Work task stages	Task initiation Topic selection	Prefocus exploration	Focus formulation	Postfocus Writing	Postfocus Writing
Work task performer	Group	Group/individual	Group/Individual	Individual	Group
⇩ Sub-tasks ⇨ Stages					
Group work	Forming	Storming	Norming	Performing →	
Information seeking	Seeking/sharing relevant information →		Seeking/sharing pertinent information		
Experiences					
Cognitive	Ambiguity →		Specificity		
Work task knowledge and skills	Low →				High
Search task knowledge and skills	Low ←			→	High
Affective ⇩					
Forming/storming groups	Uncertainty Frustration →				Uncertainty Frustration Stress Disappointment Relief
Norming/performing groups	Confusion Confidence →		Clarity Confidence →		Clarity Confidence Relief

Figure C.2. The Group Member In Context (GMIC)-model (Hyldegård, 2006c, p.348).

C.3 Information processing strategies and search tasks

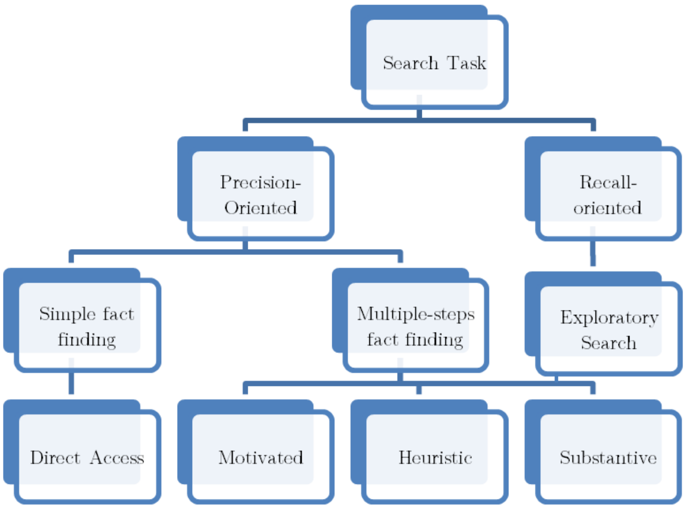


Figure C.3 Possible relationships between types of search tasks and information processing strategies.

C.4 Data granularity

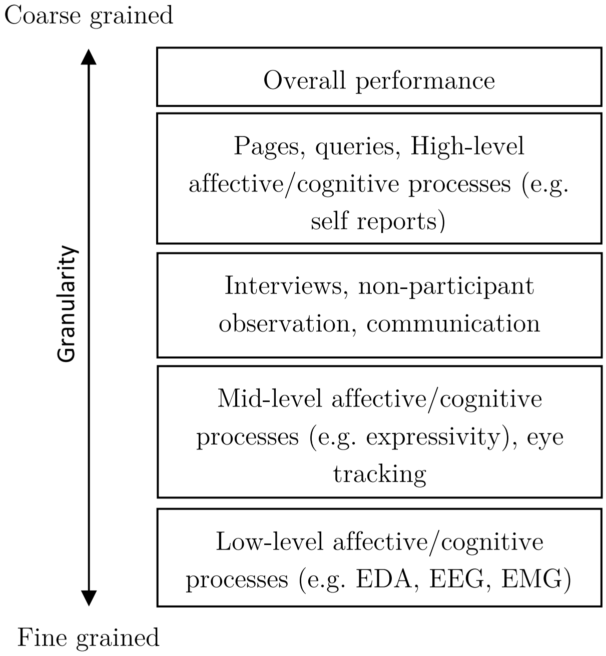
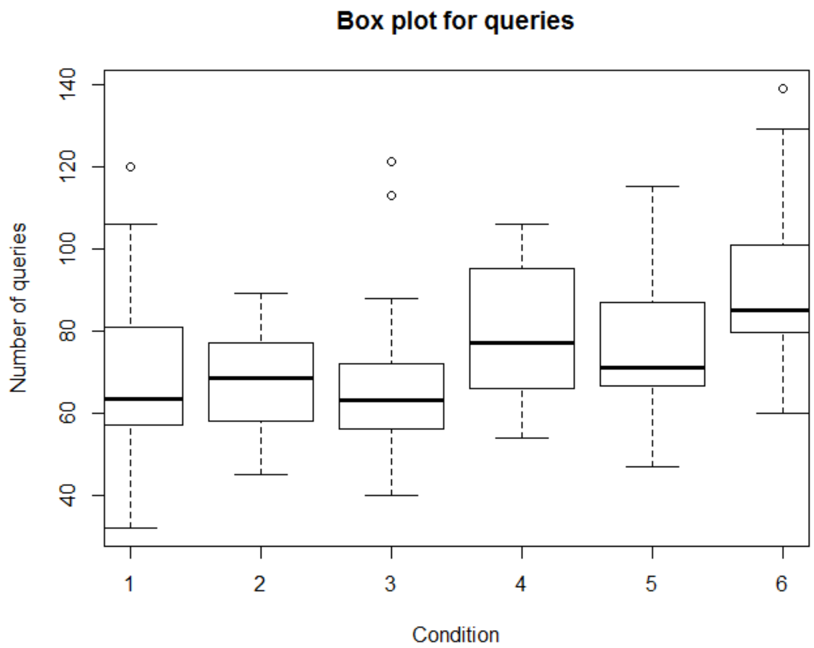


Figure C.4. Granularity levels, data collection, and data types.

Appendix D. Figures and Tables for Data Exploration

D.1 Data exploration



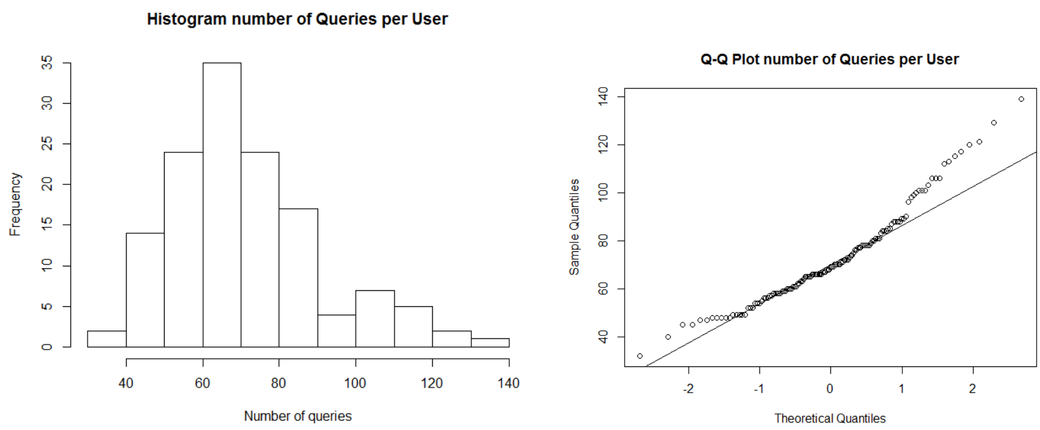


Figure D.1. Example of visual exploration of data through box plots, histograms, and Q-Q plots. Plots in this case correspond to number of queries per users.

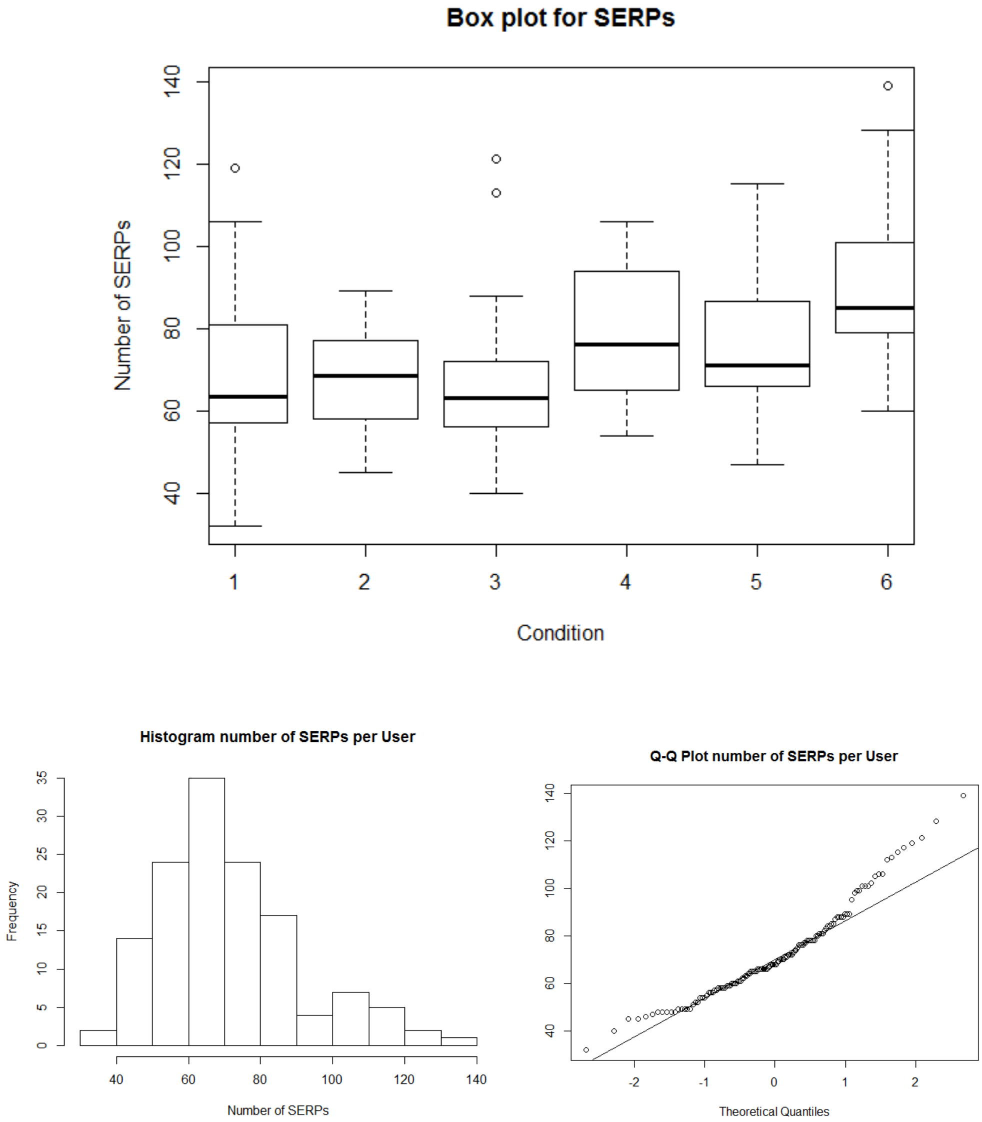


Figure D.2. Example of visual exploration of data through box plots, histograms, and Q-Q plots. Plots in this case correspond to number of SERPs per users.

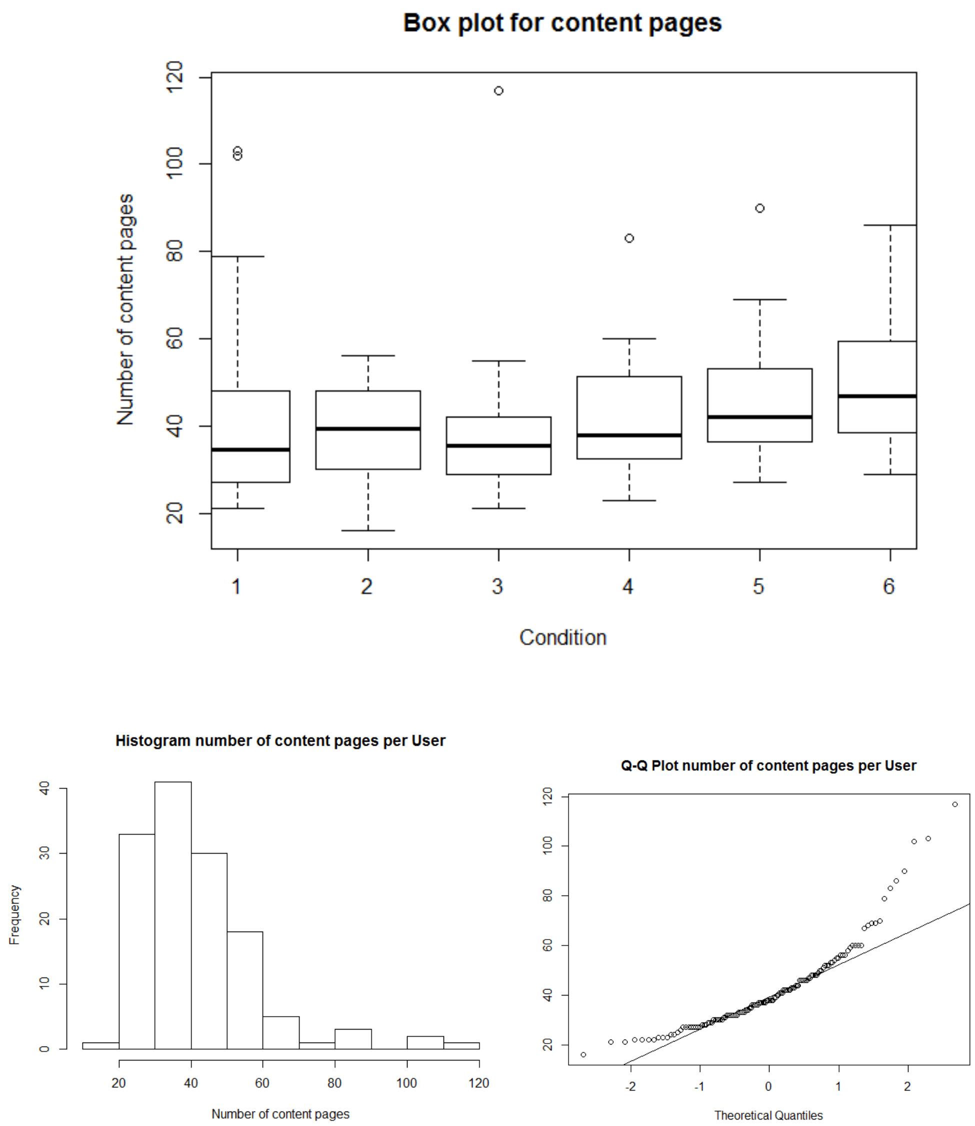


Figure D.3. Example of visual exploration of data through box plots, histograms, and Q-Q plots. Plots in this case correspond to number of content pages (information coverage) per users.

D.2 Typing speed

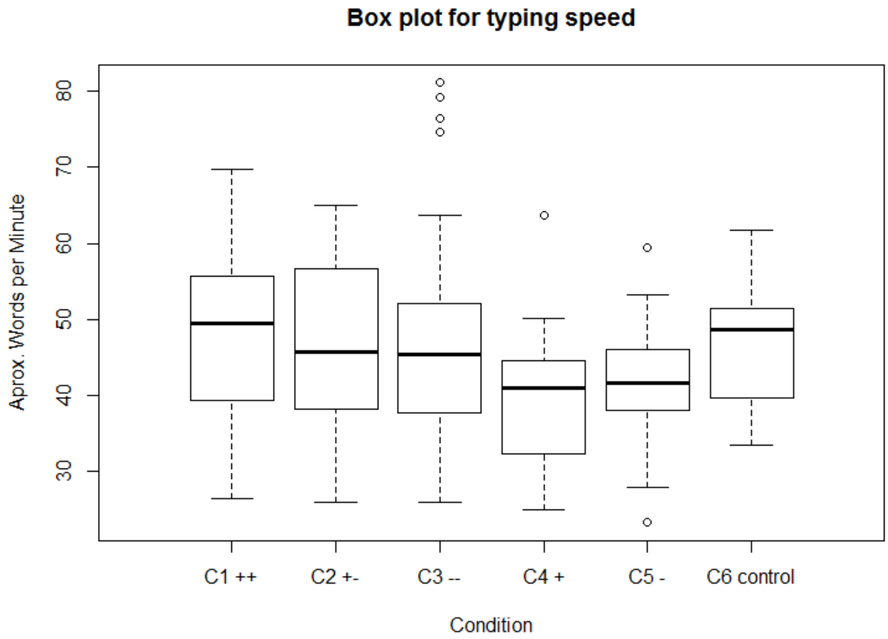


Figure D.4. Box plot for estimated typing speed in each condition. Typing speed was computed based on keystrokes recorded with the Morae software.

Appendix E. Figures and Tables for RQ1

Table E.1. Excerpt of descriptive statistics for number of distinct queries during the stimuli stage.

	# distinct queries				
	Min	**Max**	**M**	**SD**	**Mdn**
Stim^{+}	1	14	5.70	2.92	5.00
Stim^{-}	1	15	5.57	3.60	5.00
Stim^{control}	1	11	6.60	3.17	6.00

Table E.2. Excerpt of descriptive statistics for number of queries including query reuse during the stimuli stage.

	# queries with query reuse				
	Min	**Max**	**M**	**SD**	**Mdn**
Stim^{+}	1	18	6.13	3.47	5.00
Stim^{-}	1	16	5.83	3.88	5.50
Stim^{control}	1	12	6.93	3.73	6.00

Table E.3. Excerpt of descriptive statistics for different measures during the evaluation of prolonged effects, specifically in MT_q2.

MT_q2		Experimental conditions					
		C1++	C2+−	C3−−	C4+	C5−	C6control
Coverage	Min	1	1	1	1	1	0
	Max	10	9	14	6	9	6
	M	4.93	4.53	4.80	2.33	3.67	2.60
	SD	2.37	2.33	3.26	1.54	2.13	1.50
	Mdn	5.00	4.00	4.00	2.00	3.00	2.00
	K-W	H(5)=18.70, *p*<.01					
Useful coverage	Min	0	0	0	0	0	0
	Max	2	1	4	3	2	2
	M	0.93	0.53	1.20	0.47	0.40	0.33
	SD	0.70	0.52	1.15	0.92	0.63	0.62
	Mdn	1.00	1.00	1.00	0.00	0.00	0.00
	K-W	H(5)=14.10, *p*<.05					
# Queries sequence	Min	2	2	2	1	1	1
	Max	13	23	12	5	8	6
	M	5.6	5.73	5.87	1.87	3.13	2.67
	SD	3.46	5.16	3.38	1.46	2.07	1.72
	Mdn	5.00	5.00	5.00	1.00	3.00	3.00
	K-W	H(5)=28.32, *p*<.01					
# Distinct queries	Min	2	2	2	1	1	1
	Max	12	22	12	5	8	5
	M	5.27	5.67	5.80	1.87	2.87	2.53
	SD	2.89	4.92	3.28	1.46	1.92	1.51
	Mdn	5.00	5.00	5.00	1.00	2.00	3.00
	K-W	H(5)=30.46, *p*<.01					
SERPs	Min	2	2	2	1	1	1
	Max	13	23	13	5	9	6
	M	5.60	5.93	6.00	1.87	3.20	2.87
	SD	3.46	5.16	3.51	1.46	2.08	1.85
	Mdn	5.00	5.00	6.00	1.00	3.00	3.00
	K-W	H(5)=27.87, *p*<.01					

(cont.)

MT_q1		Experimental conditions					
		C1^{++}	**C2^{+-}**	**C3^{--}**	**C4^{+}**	**C5^{-}**	**C6^{control}**
# Queries sequence	**Min**	3	3	3	1	1	2
	Max	8	11	8	8	8	12
	M	5.07	6.13	5.20	2.80	3.00	4.00
	SD	1.53	2.23	1.70	1.70	1.73	3.16
	Mdn	5.00	6.00	5.00	2.00	2.00	3.00
	K-W	H(5)=35.60, p<.01					
# Distinct queries	**Min**	3	3	3	1	1	2
	Max	8	11	8	8	8	11
	M	5.07	6.07	5.13	2.73	3.00	3.67
	SD	1.53	2.28	1.64	1.62	1.73	2.89
	Mdn	6.00	6.00	5.00	2.00	2.00	3.00
	K-W	H(5)=37.27, p<.01					
SERPs	**Min**	3	3	3	1	1	2
	Max	8	11	9	8	8	16
	M	5.13	6.13	5.20	2.73	3.07	4.07
	SD	1.51	2.39	1.78	1.62	1.71	4.03
	Mdn	5.00	6.00	5.00	2.00	3.00	3.00
	K-W	H(5)=37.11, p<.01					

Table E.4. Comparison between pairs in collaborative condition with respect to features of chat messages according to LIWC dictionary.

Team	Question						
	MT_q_1	**MT_q_2**	**MT_q_3**	**MT_q_5**	**MT_q_6**	**MT_q_7**	**MT_q_8**
I_4	1>3 2>1,3						
You_6				1>2,3 2>3			
SheHe_7			1>3 2>1,3				
They_8							1>2,3
Ipron_9							1>3 2>3
Article_10						1>2 3>1,2	
AuxVb_12	2>3						
Past_13				1>3 2>1,3			
Adverbs_16							1>2,3 2>3
Conj_18	1>3 2>3						
Quant_20			1>3 2>1,3				
Swear_22	1>2,3		1>2,3				
Social_23				1>2			
Discrep_36		1>3 2>1,3					
Tentat_37					2>1 3>1,2		
Sexual_49					1>2,3		
Relativ_51	2>1,3						
Work_55						1>2,3	
Home_58		2>1,3					
Money_59			2>1,3				

Table E.5. Comparison between group members in collaborative condition with respect to features of chat messages according to LIWC dictionary.

Individual	**Question**							
	MT__q$_1$	**MT__q$_2$**	**MT__q$_3$**	**MT__q$_4$**	**MT__q$_5$**	**MT__q$_6$**	**MT__q$_7$**	**MT__q$_8$**
Funct__1	$1>3$ $2^+>1,2^-,3$ $2^->1,3$				$1>2^+,2^-,3$ $2^+>2^-,3$ $2^->3$			
Pronoun__2	$1>3$ $2^+>1,2^-,3$ 2->1,3							$2^+>2^-,3$ $2^->3$
Ppron__3	$1>3$ $2^+>1,3$ 2->1,2^+,3			$1>2^+,2^-,3$ $2^+>2^-$				
I__4	$1>3$ $2^+>1,3$ $2^->1,2^+,3$							
You__6					$1>2^+,2^-,3$ $2^+>3$ $2^->2^+,3$			
SheHe__7			$1>2^+,3$ $2^->1,2^+,3$ $3>2^+$					
They__8								$1>2^+,2^-,3$
Ipron__9								$1>2^+,2^-,3$ $2^+>3$ $2^->2^+,3$
Article__10							$1>2^+,2^-$ $3>1,2^+,2^-$	
Verbs__11	$1>3$ $2^+>1,2^-$				$1>2^-,3$ $2^+>1,2^-$			
AuxVb__12	$2^->1$ $3>1$				$1>2^+,3$ $2^->1$			
Past__13			$2^+>1,3$ $2^->1,2^+,3$ $3>1$		$1>3$ $2^+>1,2^-,3$ $2^->1,3$			
Adverbs__16								$1>2^+,2^-,3$ $2^+>2^-,3$ $2^->3$
Conj__18	$1>3$ $2^+>1,3$ $2^->1,3$							
Quant__20			$1>3$ $2^+>1,3$ $2^->1,3$					
Swear__22			$1>2^+,2^-,3$					
Social__23					$1>3$ $2^+>3$ $2^->3$			

Table E.6. Comparison between pairs in collaborative condition with respect to features of chat messages according to LIWC dictionary. Specifically psychological processes.

Individual	Question							
	MT_q1	**MT_q2**	**MT_q3**	**MT_q4**	**MT_q5**	**MT_q6**	**MT_q7**	**MT_q8**
CogMech_33					1>2^-,3 2^+>1,2^-			1>2^+,2^-,3 2^+>2^-,3 3>2^-
Cause_35	1>3 2^+>1,2^-,3 2^->1,3							
Discrep_36		1>3 2^+>1,3 2^->1,2^+,3			1>2^-,3 2^+>1,2^-,3 3>2^-			
Tentat_37						2^+>1 2^->1,2^+,3 3>1,2^+		
Percept_42	2^+>1,3 2^->1,3 3>1							1>2^+,2^-,3 2^+>2^-,3 2^->3
Feel_45								1>2^+,2^-,3 2^->2^+,3
Bio_46			1>2^+,2^-,3 2->2^+ 3>2^+,2^-					
Relativ_51	2->2^+							
Space_53	1>3 2^+>1,3 2^->1,3							
Home_58		2^+>1,3 2^->1,3						
AffectWords_69					1>2^-,3 2^+>1,2^-,3 2^->3			

Appendix F. Figures and Tables for RQ2

F.1 Plots for evaluation of short-term effects

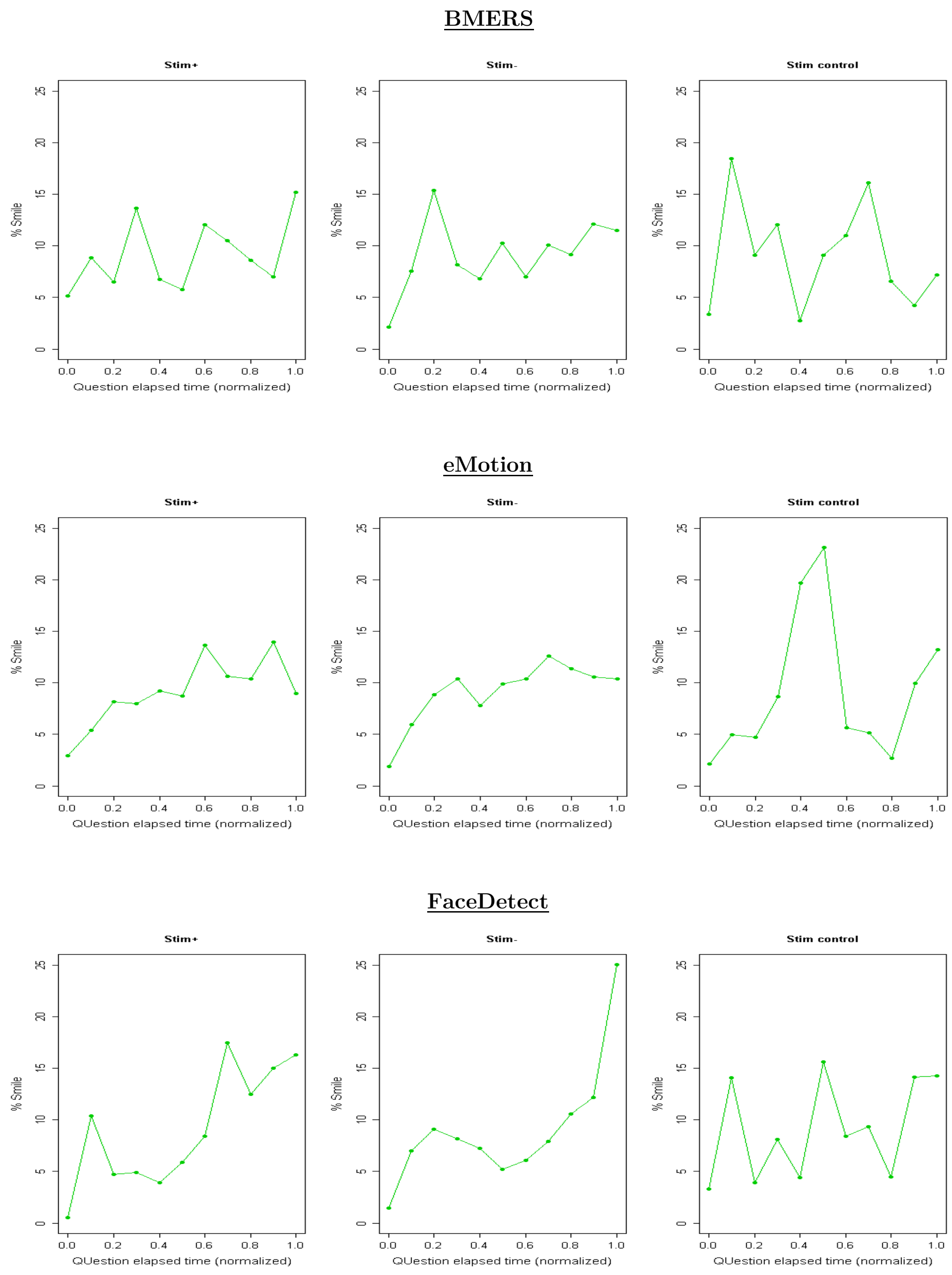


Figure F.1. Smiles recognized in different stages during the development of questions in the evaluation of short-term effects, specifically PreS stage. Results from three different software: *BMERS*, *eMotion*, and *FaceDetect*.

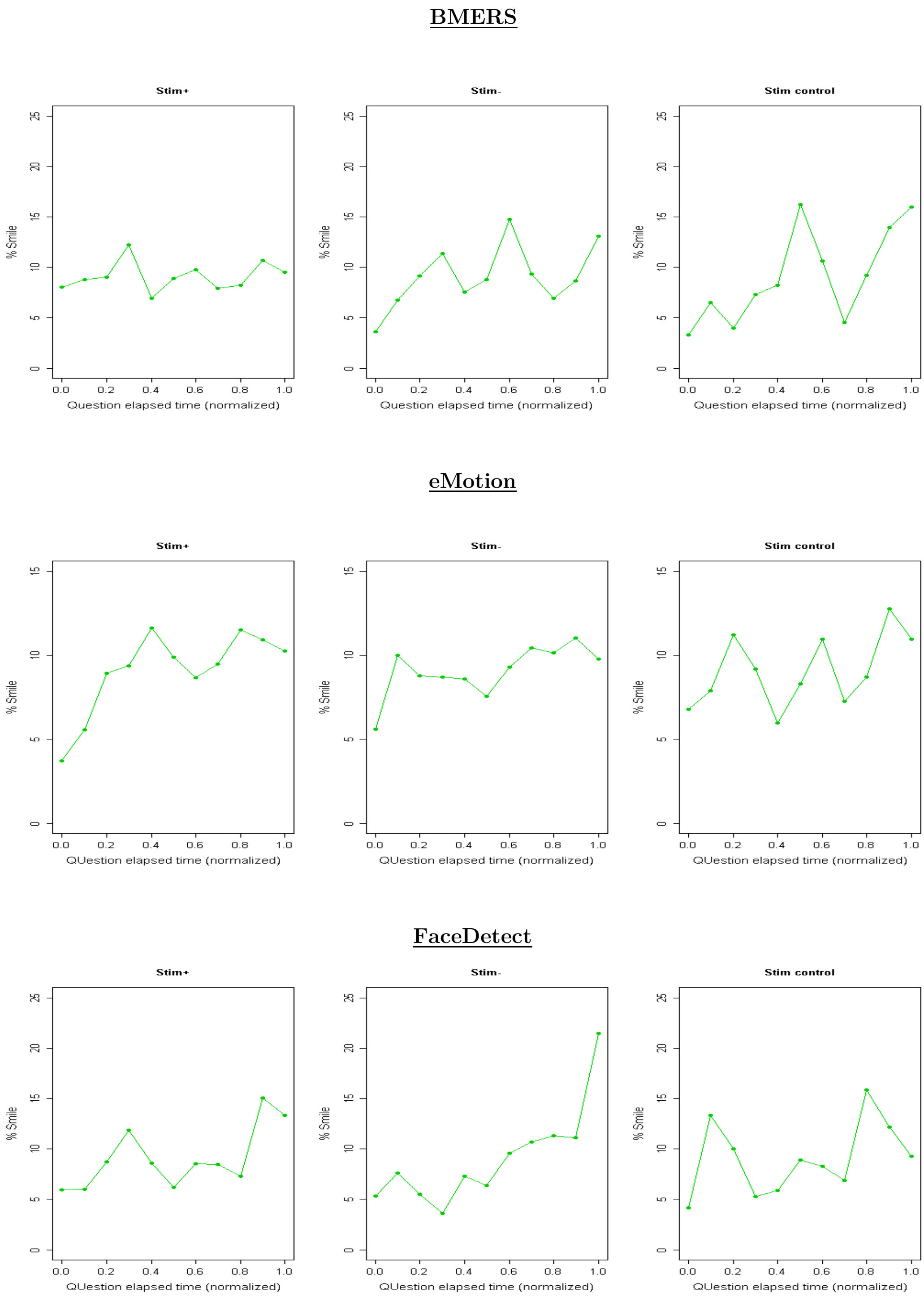


Figure F.2. Smiles recognized in different stages during the development of questions in the evaluation of short-term effects, specifically Stimuli stage. Results from three different software: *BMERS*, *eMotion*, and *FaceDetect*.

Figure F.3. Smiles recognized in different stages during the development of questions in the evaluation of short-term effects, specifically PostS stage. Results from three different software: *BMERS*, *eMotion*, and *FaceDetect*.

F.2 Plots for evaluation of prolonged effects

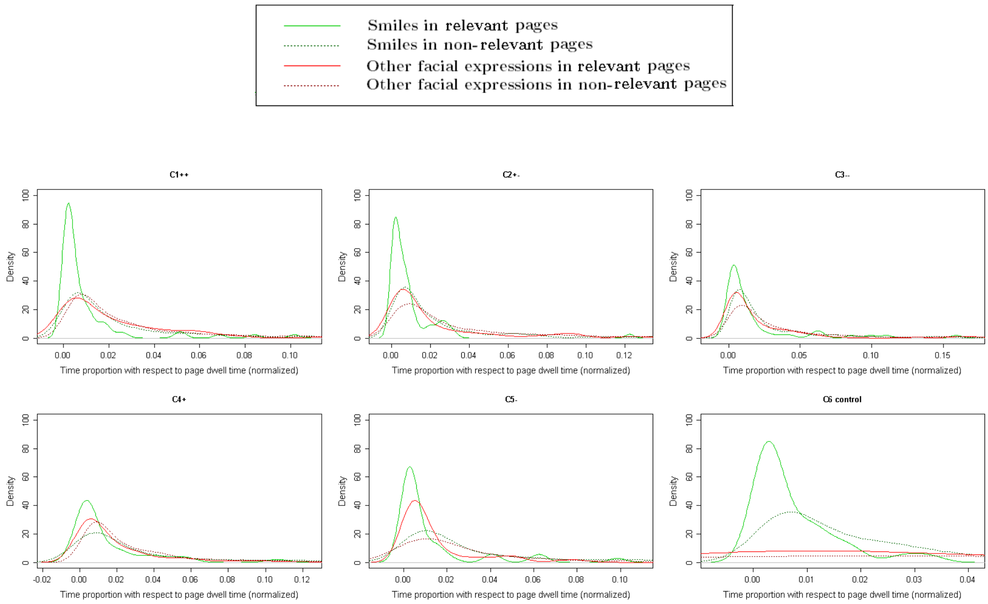


Figure F.4.Density plots for smiles and other facial expressions recognized during the exposure to relevant and non-relevant pages. Results from *BMERS*.

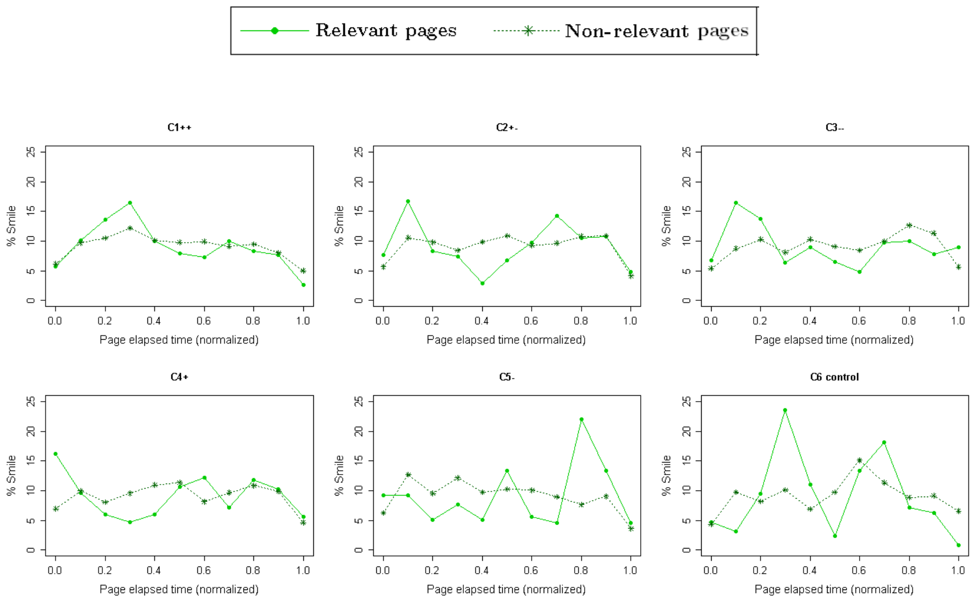


Figure F.5. Smiles recognized during the exposure to relevant and non-relevant pages. Smiles and page elapsed time are normalized. Results from *BMERS*.

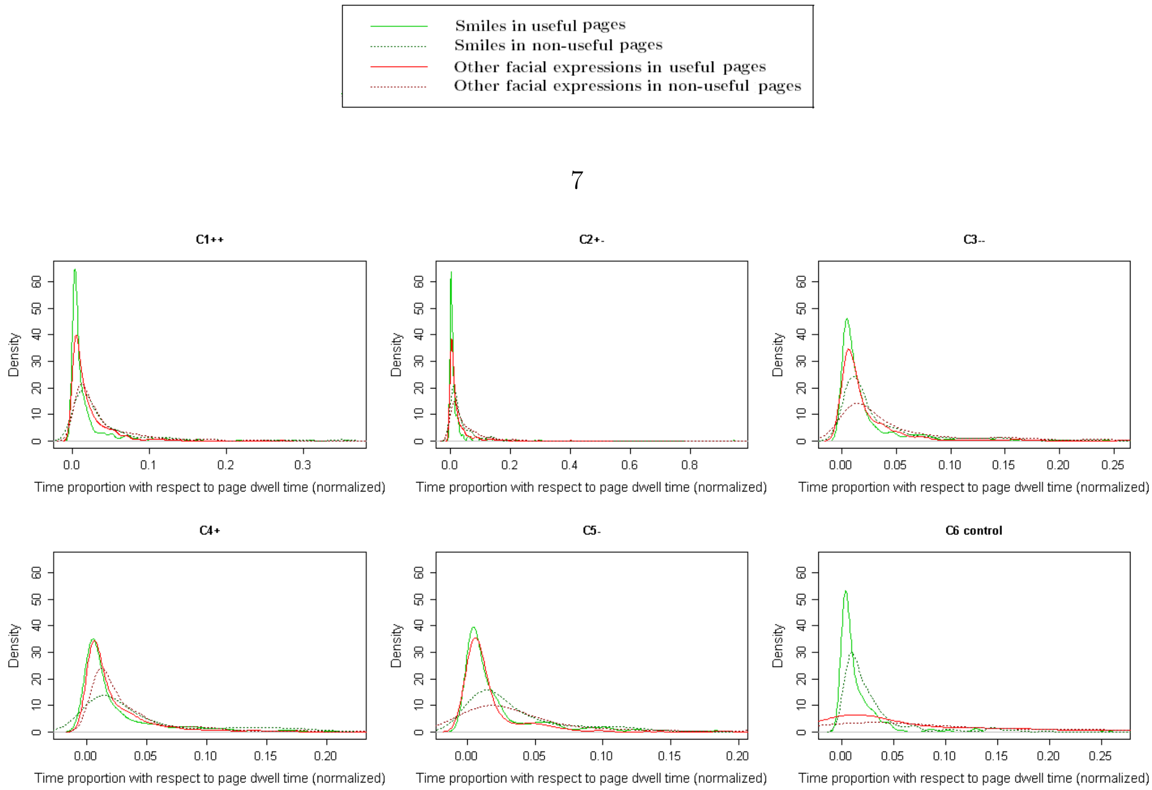


Figure F.6. Density plots for smiles and other facial expressions recognized during the exposure to useful and non-useful pages. Results from *BMERS*.

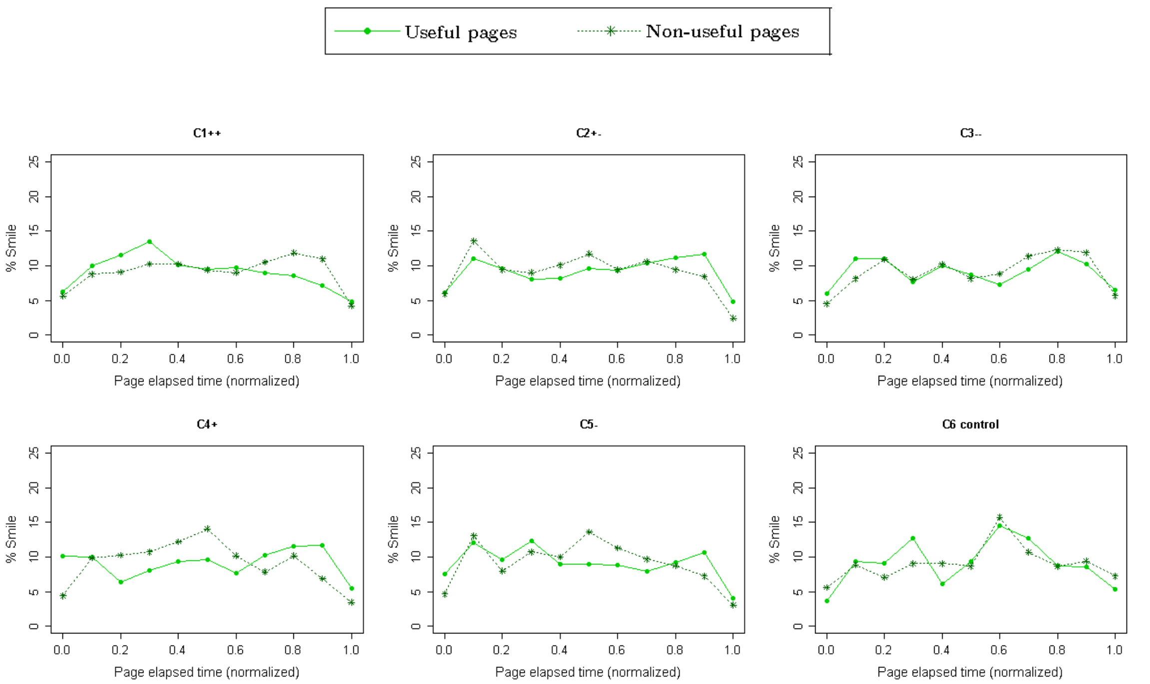


Figure F.7. Smiles recognized during the exposure to useful and non-useful pages. Smiles and page elapsed time are normalized. Results from *BMERS*.

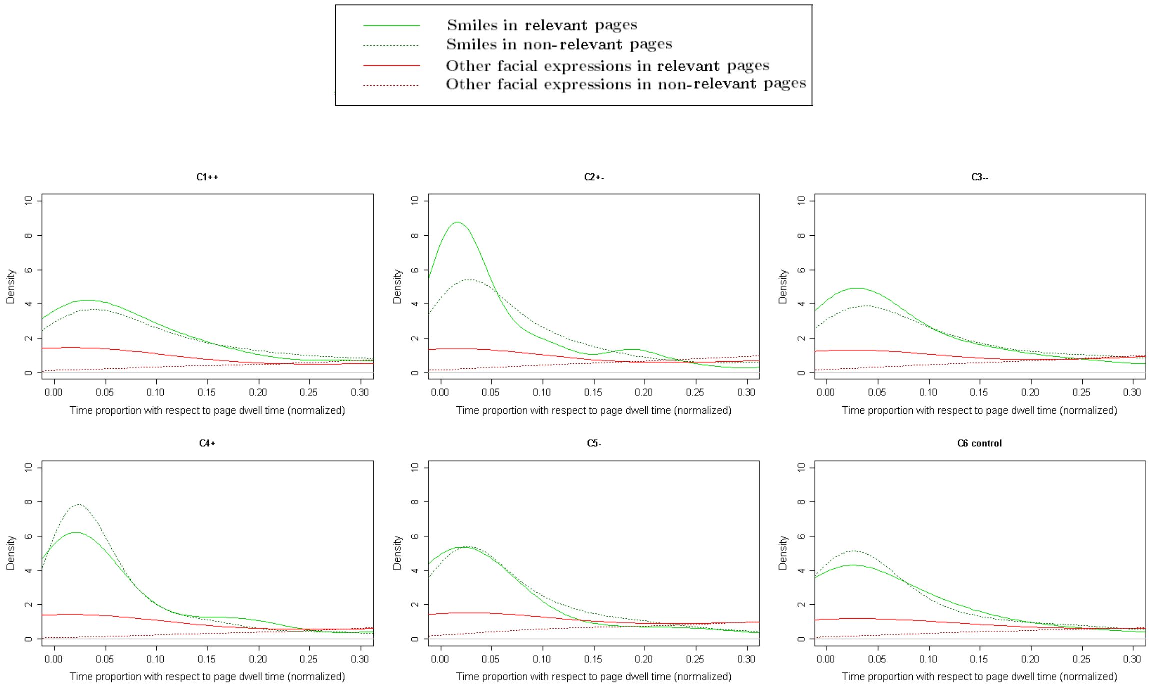


Figure F.8. Density plots for smiles and other facial expressions recognized during the exposure to relevant and non-relevant pages. Results from *eMotion*.

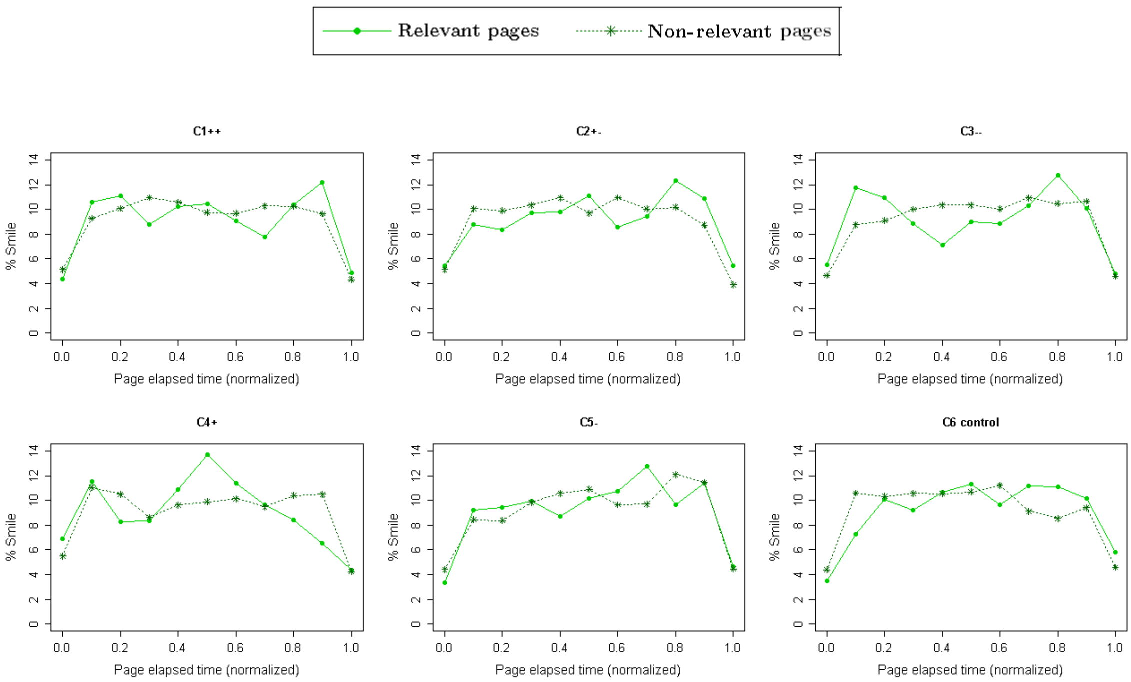


Figure F.9. Smiles recognized during the exposure to relevant and non-relevant pages. Smiles and page elapsed time are normalized. Results from *eMotion*.

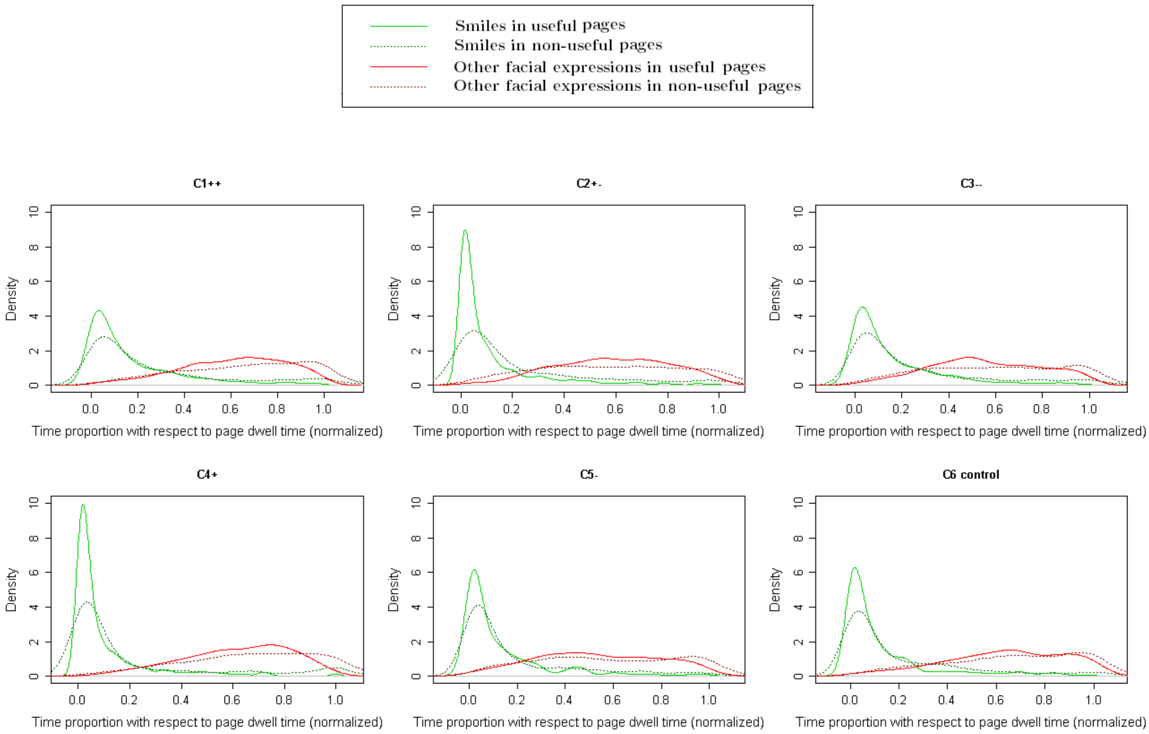


Figure F.10. Density plots for smiles and other facial expressions recognized during the exposure to useful and non-useful pages. Results from *eMotion*.

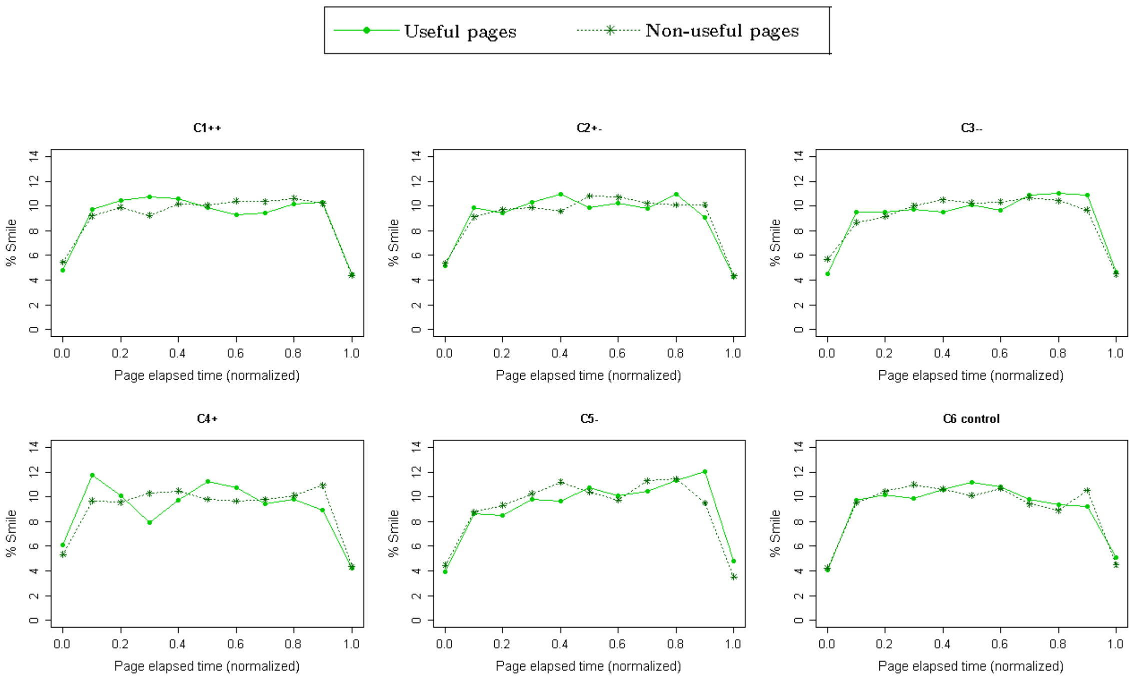


Figure F.11. Smiles recognized during the exposure to useful and non-useful pages. Smiles and page elapsed time are normalized. Results from *eMotion*.

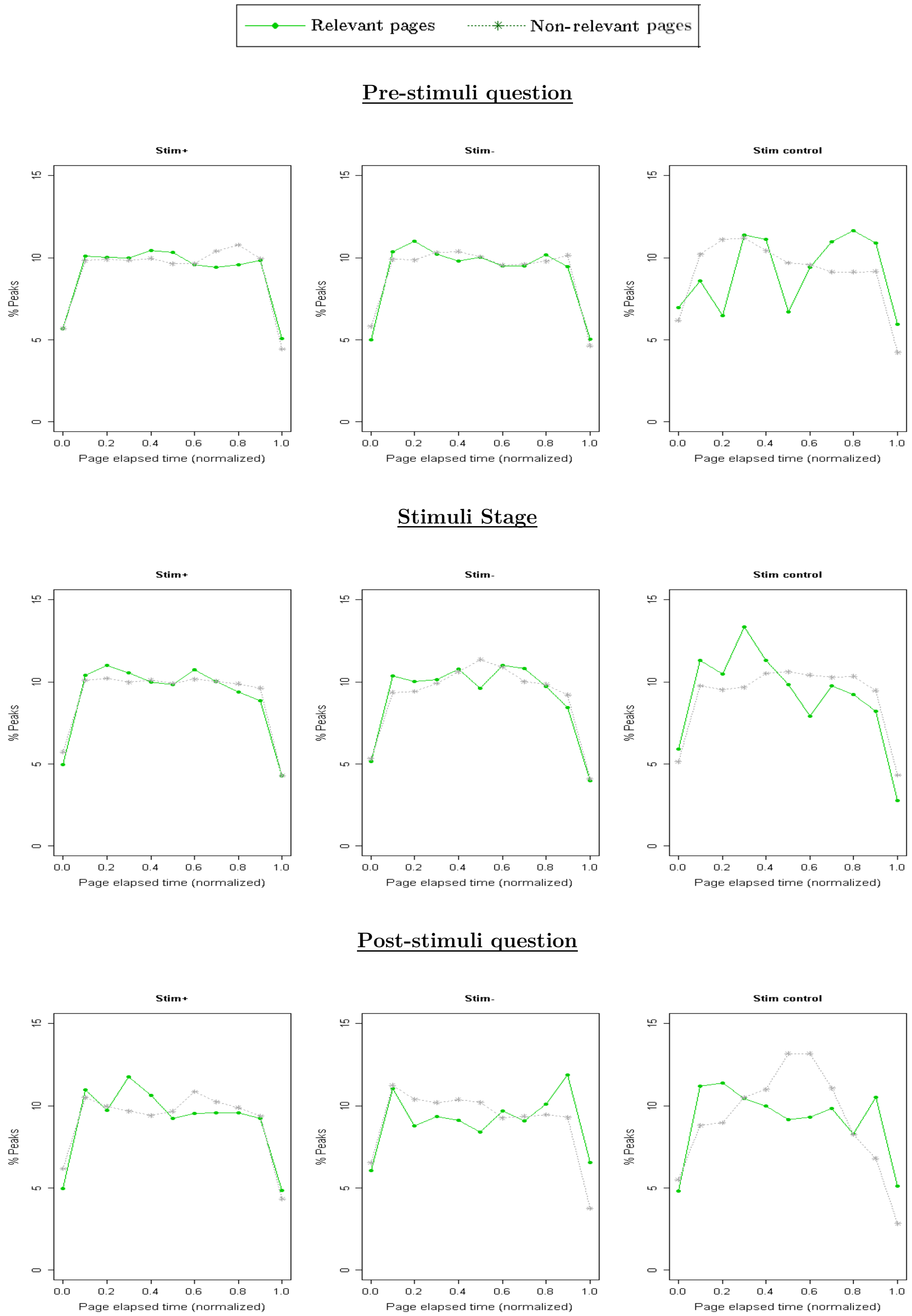
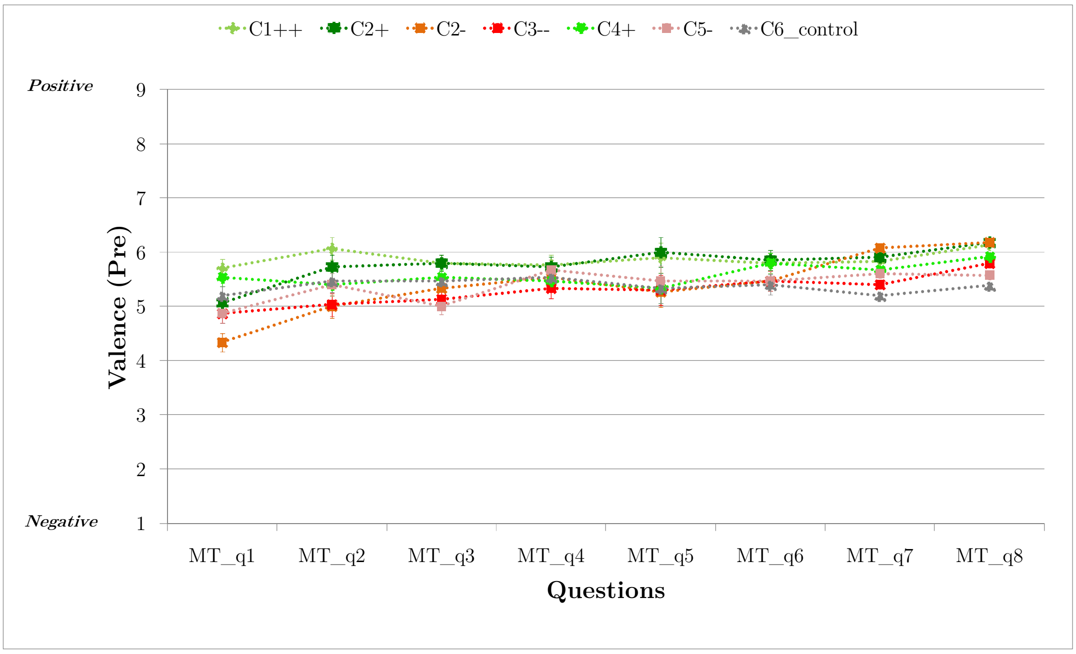


Figure F.12. Aggregated EDA peaks during the exposure to relevant and non-relevant pages in different stages of the evaluation of short-term effects. Smiles and page elapsed time are normalized.



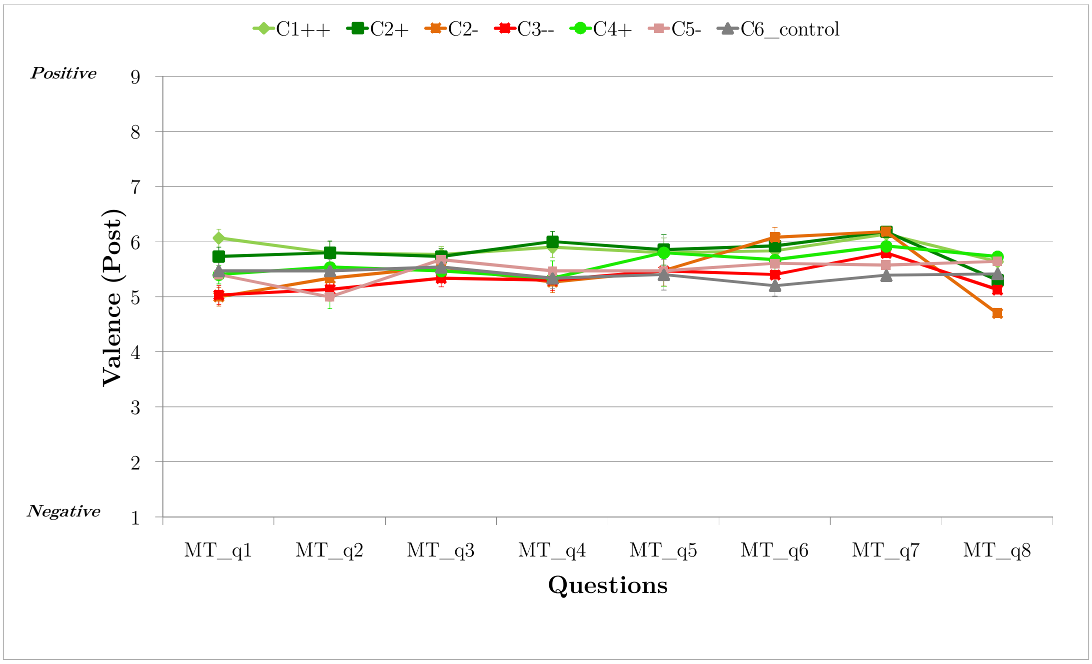
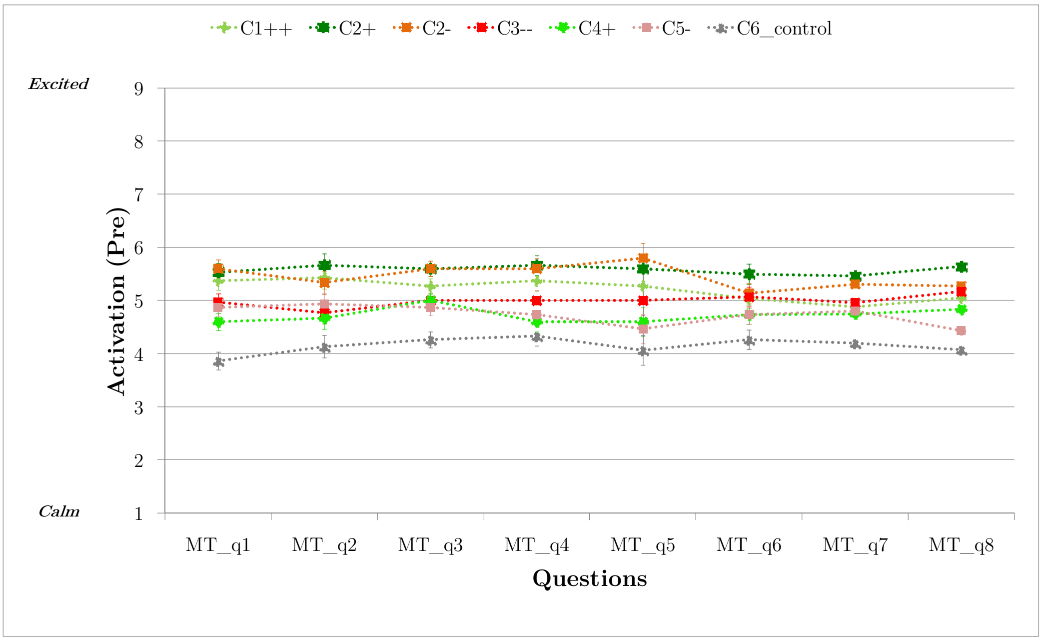


Figure F.13. Distribution of participants per question in terms of feeling-based conditions. (PreS) Pre-stimuli question, (Stim) Stimuli, and (PostS) Post-stimuli question.



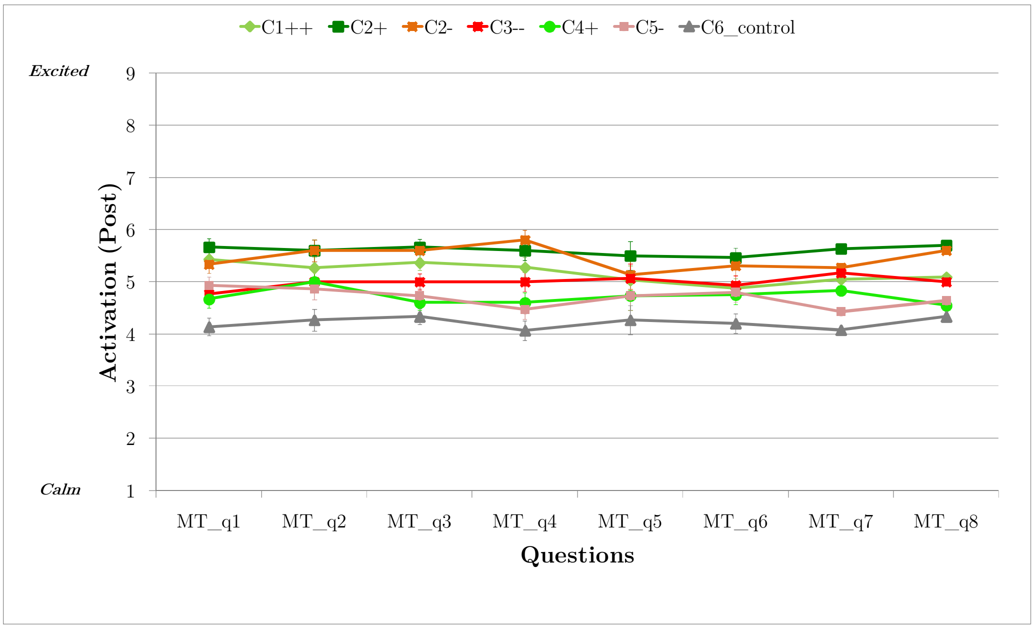
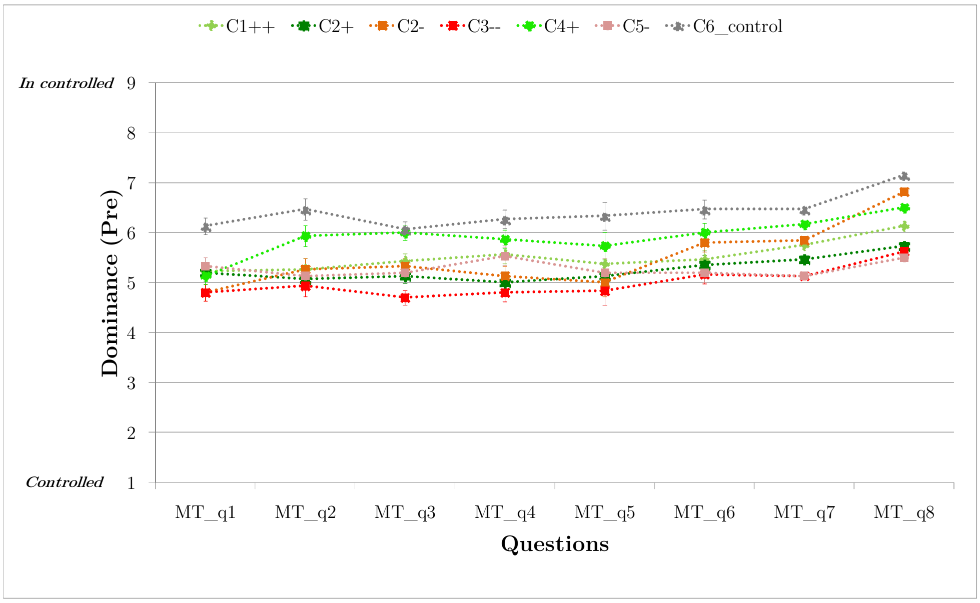


Figure F.14. Distribution of participants per question in terms of feeling-based conditions. (PreS) Pre-stimuli question, (Stim) Stimuli, and (PostS) Post-stimuli question.



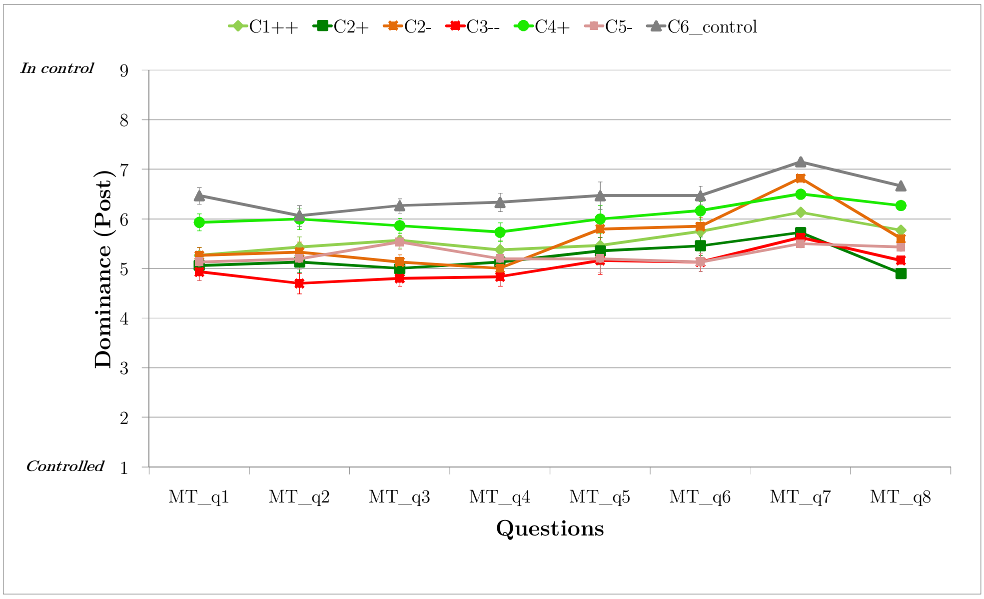


Figure F.15. Distribution of participants per question in terms of feeling-based conditions. (PreS) Pre-stimuli question, (Stim) Stimuli, and (PostS) Post-stimuli question.

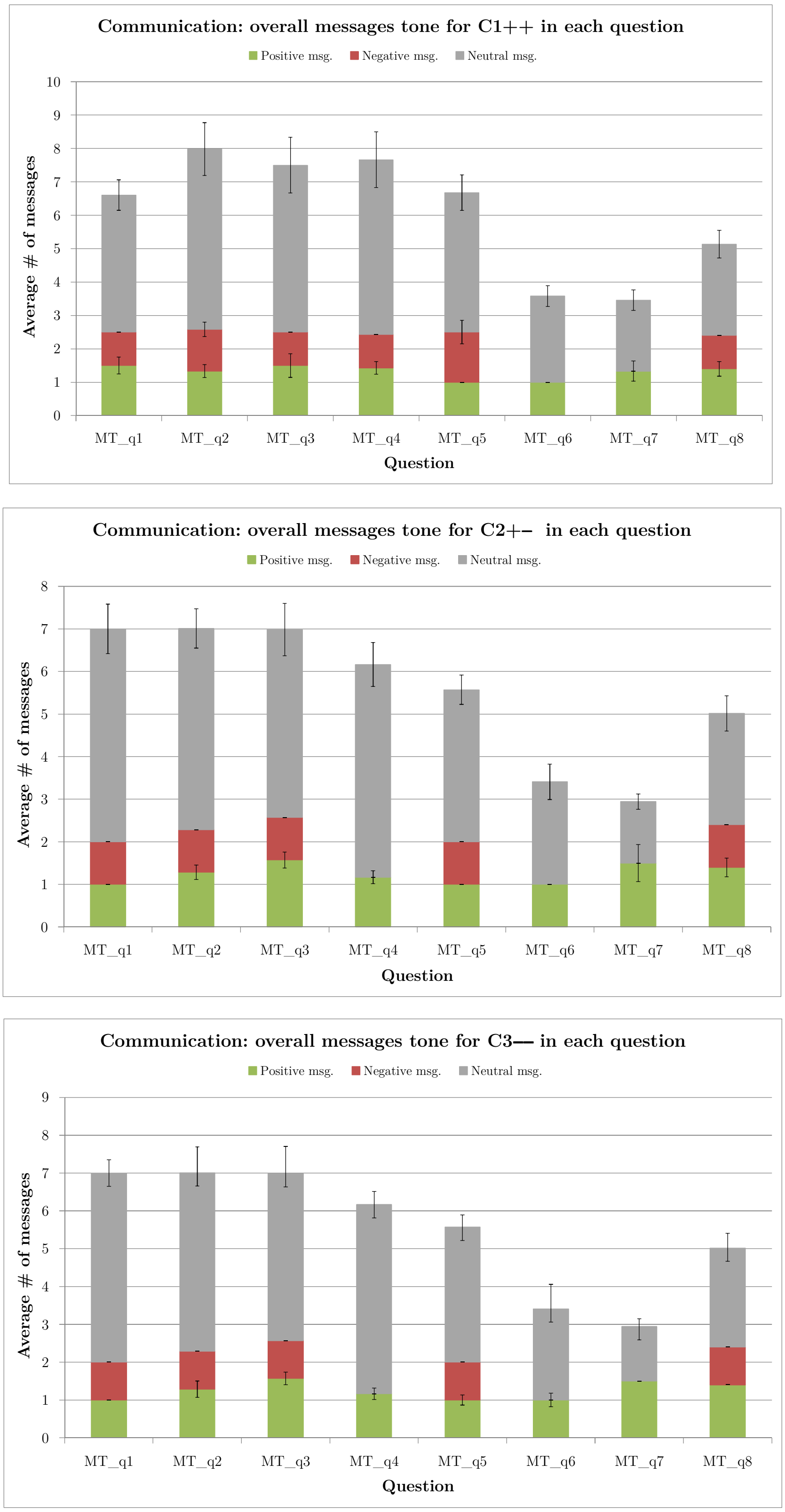
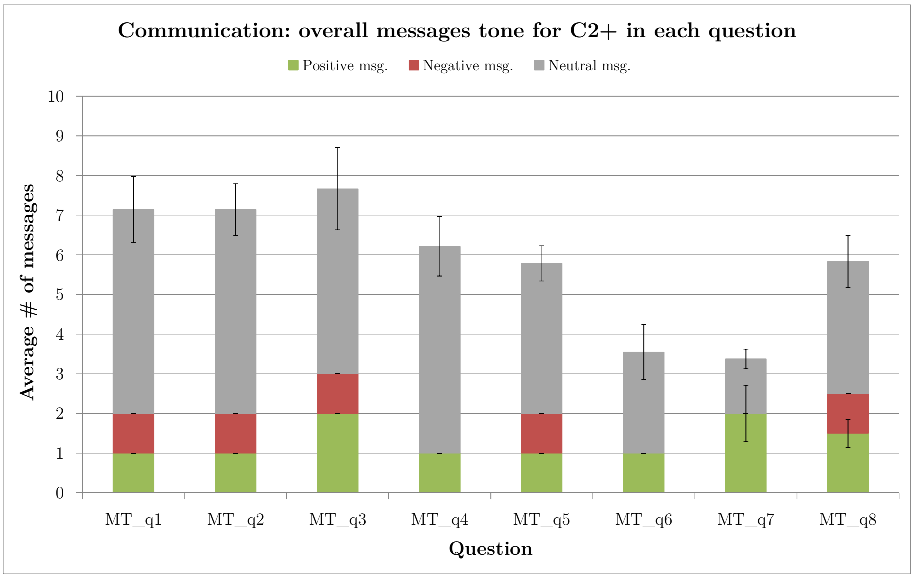


Figure F.16. Communication tone in each question of the evaluation of prolonged effects. Results for pairs.



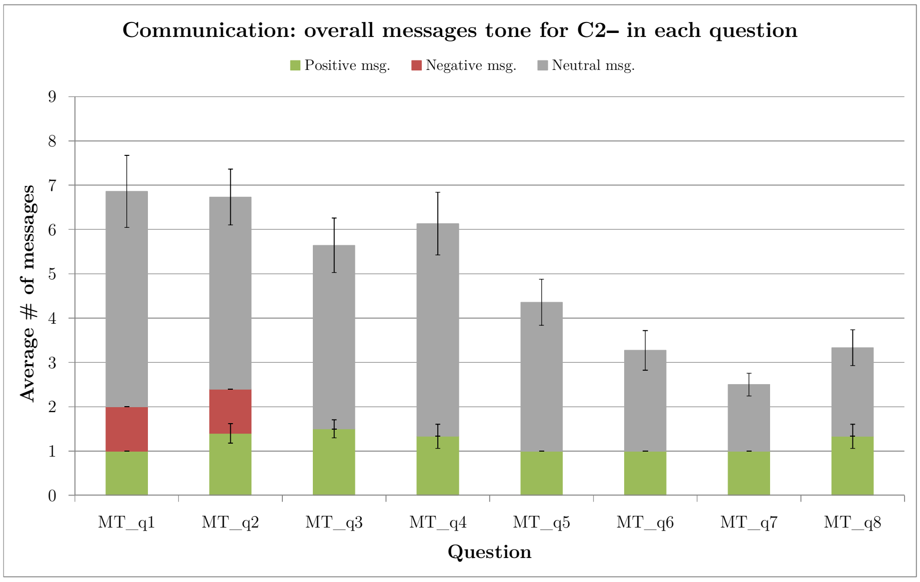


Figure F.17. Communication tone in each question of the evaluation of prolonged effects. Results at the individual level in condition $C2^{+-}$.

Appendix G. Figures and Tables for RQ4

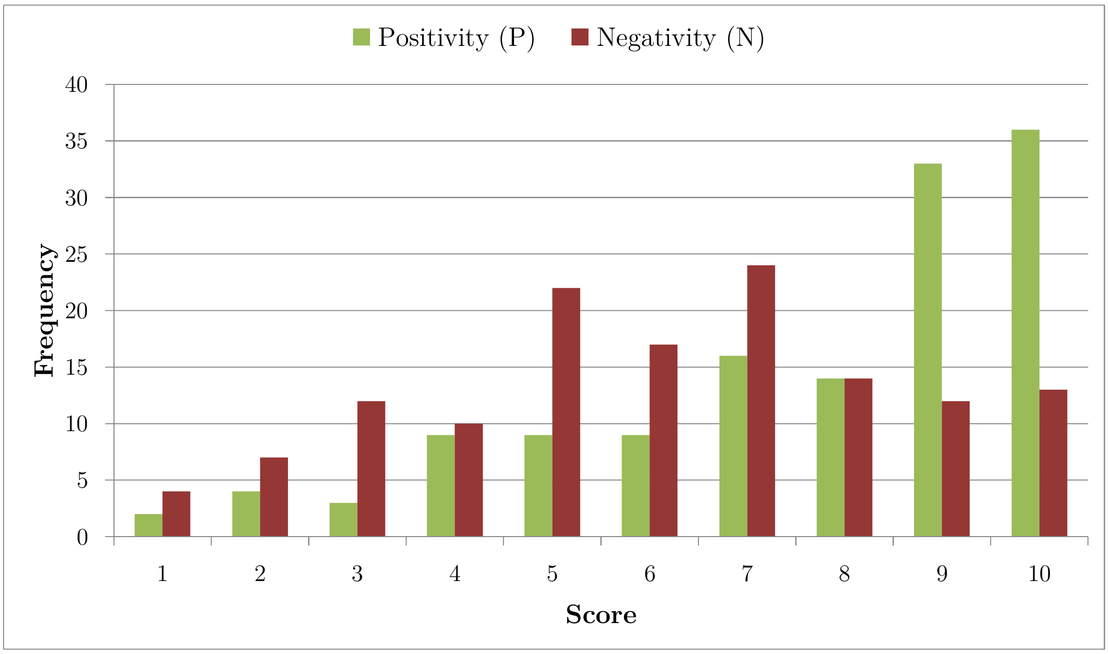


Figure G.1. Histogram of positivity and negativity scores reported through the positivity self test (Fredrickson et al., 2003) in regard to affective states experienced by participants 24 hours before their sessions.

Appendix H. Coding book for communication

The following list of messages corresponds to the communication logs of pairs of users performing an information search task. This communication data will be quantitatively and qualitatively analyzed. For the particular case of quantitative analyses, data will be coded following the coding scheme explained below. This coding scheme is multidimensional and involves 13 categories that will help to characterize each message.

H.1 Coding Procedure

In order to keep consistency in the criteria applied when coding each message, coding must be performed by COLUMN.

The coding procedure must be performed INDIVIDUALLY. Agreement scores will be computed afterwards to check reliability.

In order to help you to contextualize each message and assign codes to them, the questions that participants had to respond while communicating are provided in the last column (QUESTION TEXT).

H.2 Coding Scheme

As indicated above, this coding scheme consists of 13 specific categories or dimensions:

1. The first category is the most general one, which establishes the type of message. The types of messages that can be found in the communication data are: Task Coordination (TC), Task Social (TS), Task Content (TN),

Non-task related (NT), and Non-codable (NC). Below each category is explained and examples are provided:

- **TC**: All types of statements regarding coordination, which involve decision making about how the task should be performed, such as instructions explicitly indicating what to do next.
    - e.g.: “We should start writing now”, “Can we search the second one?”, “How do you want to do this?”, “I will work on the reactions and you work on the consequences”, “Shall we respond?”, “Send me the information”, “I don’t see what you found, where did you find it?”, "Wait!!!", "Should we go with this one?", "do you agree with this?", "Megalodon Shark?"
- **TS**: All types of statements that concern group functioning, effort, or attitude as well as opinions in regards to information obtained or information sources.
    - e.g.: “This task is really hard”, “we did good”, “Wow, so many animals were killed during the spill”, “I think my answer is the best”, “I’m confused about this”, "Hello", "Hi"
- **TN**: All types of statements that are related to the content of task, which include information assessment, topic discussion, layout, structure, and work revision.

    - e.g. “I found something about consequences”, “Ok, I found stuff on the impact on economy life, people and animals”, “Check my snippet”, "I am not finding anything", "did you find anything?", "Megalodon Shark?", (this latter messages is also coded as TC).
- **NT**: All types of statements that are not related to the development of the task.
    - e.g. “What are you going to do after this?”, “Did you watch the movie yesterday”, “I am hungry”.
- **NC**: All types of statements that do not belong to any category specified. Messages that cannot be contextualized based on the previous or next message, must be coded under this category.
    - e.g. “mmmm”, “Ok”, “Let’s see”.
    - Note the examples above are considered as NC assuming that they cannot be contextualized. See the following note with a complete example about context.

Important note about CONTEXT: Messages must be evaluated in context of the conversation in which they appear. One criterion that can be applied for this is by looking at the time column to verify if two messages are closely related in time. Another important element to evaluate context is the column

Question, each question defines a new context of conversation. For example:

Q1 t0: User 1 – "Should we respond?" -> TC

Q1 t1: User 2 – "Ok" -> TC

In this case both messages must be coded as TC. On the other hand, if the two messages appear as follow:

Q1 t0: User 1 – "Should we respond?" -> TC

Q2 t1: User 2 – "Ok" -> NC

Then only the first one must be coded as TC and since the second message cannot be clearly coded, the NC code is assigned.

There are some cases of misspelling or messages that are sent in parts. For example:

Q1 t0: User 1 – "Ye" -> TC

Q1 t1: User 1 – "s" -> Blank

In such cases DO NOT code the second message (leave it in blank).

2. While most messages are defined by one and only one type of message, some messages may include aspects of other types of messages. Therefore, the second category indicates if messages contain aspects from other types, namely: TC, TS, TN, NT. (e.g. "Well, go with that. Good Job!" this message is coded first as TC since it indicates a coordination about

how to proceed. Moreover includes an statement about group functioning "Good job!", thus it is coded in the second category as TS.

3. The third coding category, namely HELP, indicates whether or not (1 or 0 respectively) a message is intended to request help, support, or confirmation from a collaborator (e.g. "so his wife tried to contact abraham lincoln?", "should I just say serpent? or the full name given on wiki", "right?", etc.).
4. The fourth coding category is SNIP, which_indicates if messages refer explicitly (E) or implicitly (I) to information sharing events via the snip tool of the system. Explicit (E) information sharing via SNIP occurs when users send messages pointing to specific snippets in the common repository (e.g. "did u see my snip?" or "check if my snippet it's correct"). In order to code messages as Implicit (I), it is necessary to evaluate if in previous messages participants exchange or not information about sources, page content, specific snippets, etc. (e.g. "jesus ur good" in this case the user expressing this message is aware of the progress of his/her teammate without explicit information sharing or messages as see in previous messages in the context of a the particular question (context). This inferred by looking at previous messages, where there are not explicit references to information sharing; another example is: "yea im on a page about lincoln too", in this case the user commented on a shared resource without previous messages pointing to it, thus it is inferred that the author of this message checked on the snippets and sources shared by his/her collaborator)

5. The fifth category corresponds to CONTROL, which indicates if a team member takes the control of team coordination or actions within the team by providing instructions or typing imperative statements. (e.g. "I will provide the answer", "Go ahead", "respond the question", "search about X", etc.)
6. The sixth category, namely AGREEMENT, indicates if messages convey agreement (A) or disagreement (D), specially during decision making processes. Typically answers to participant's questions or statements about strategies are followed by statements of agreement (e.g. "if your confident ill go with you") or disagreement (e.g. "No, that is not right. We should continue searching")
7. The seventh category, STRATEGY, indicates whether or not (1 or 0 respectively) messages correspond to the definition of strategies to approach the task (e.g. "let's search independently and then compare our findings", "let's search about X and then Y", etc.). Typically strategy messages are linked to TC messages, but not every TC message is about strategy.
8. The eight category is search process, which indicates whether or not (1 or 0 respectively) messages reflect aspects of information search. Typically such messages report what was searched, what was found, what queries are being used, or look for evaluating information sources (e.g. "I googled about X", "I got the answer, its Y", "I got it from Z", "is Z a reliable source?", etc.).
9. The ninth category is UNCERTAINTY, which indicates whether or not (1 or 0 respectively) users' messages express feelings/thoughts/opinions

related to problems finding information. Messages that can be associated to feelings of confusion or frustration with regard to the information that needs to be found are typical examples of UNCERTAINTY (e.g. “I don't find anything”, “I am lost, we don't have anything”, “I'm getting tired of this”, “I can't find it”). Messages about uncertainty must be coded as NEGATIVE (N) under the coding category 13 (see below for more details)

10. The tenth category is SUCCESS, which unlike the previous one indicates whether or not (1 or 0 respectively) messages express feeling of success or victory when completing the task. (e.g. “I got the answer!”, “Hey, the answer is X”, “Go for it! You got it”). Messages about success must be coded as POSITIVE (P) under the coding category 13 (see below for more details)
11. The eleventh category is AWARENESS, which indicates whether or not (1 or 0 respectively) messages are related to the exchange of status such as reporting current page, remaining time, or checking if they are in the same question (e.g. “we have 25 seconds”, “I am in Wikipedia”, “Where are you now?”, etc.)
12. The twelfth category, QUESTION/ANSWERS (Q and A respectively), indicates whether a message is a question or request (e.g. “What are you searching?”, “how do you feel?”, and “please send me the information”) or answer (e.g. “I am looking facts about X”, “sure”, and “I am tired” as respective answers for the previous questions).
13. The final category corresponds to TONE. This one indicates whether a message is positive (P), negative (N), or neutral ( ). Typically messages

are considered as positive if they involves jokes, positive expressions such as words (e.g. “Great!”, “We are doing good”, “you are really good on this”, “We did it!”) or paralinguistic cues such as emoticons or abbreviations (e.g. “LOL”, “ :) “ , “ :P “, “hahahaa!”, “this is good shit”). On the other hand, negative messages are those denoting frustration, confusion, or irritation (e.g. “I am not getting anything”, “this is really bad”, “we won’t make it”, “this does not look good”, “oohh, that is so sad”, “I feel terrible about that”, “Wrong”), or paralinguistic cues such as emoticons or abbreviations (e.g. “ : S “, “:’(“)

H.3 Special Notes About Dataset

The dataset consists of four columns with data about context, time, message, and user.

- The context is defined by the number of question, which is located in the first column of the dataset. Each question lasted five or less minutes. This value can be used to determine the context of a conversation. Every time that this number changes, it means that a new question started. In some cases, messages may still refer to the previous question.
- The second column corresponds to time, this data can be used as a reference to evaluate if two or more messages are related to the same topic of discussion. If messages in a sequence are too apart from each other, it may be possible that messages belong to different discussions, even if they are within the context of the same question. Note that there is no standard threshold to determine whether or not

messages apart are about the same or different topics. This decision will depend on the criteria of the coder.

- The third column is message, which contains the actual message that is being coded.
- Finally, the fourth column corresponds to user it is an indicator of the author of the message. This way the coder can keep track of the communication flow between both team members.

H.4 Notes

- For binary categories, DO NOT put "0" if the message does not fit in such categories.
- For tone, if a message is neutral/objective DO NOT put any code.

Appendix I. Preliminary Studies

I.1 Reference study

A reference study used in the preparation of this dissertation was conducted by Shah (2010c) with 42 pairs of users. The experimentation was conducted using Coagmento (Shah, 2010b; González-Ibáñez & Shah, 2011), a CIS system and an experimental study design consisting of two exploratory search tasks carried out by pairs of users under three awareness conditions: contextual awareness, personal peripheral awareness, and support history. Each experimental condition relates to different levels of social presence as discussed in section 3.1.3.

A key design decision from this study is that participants were required to sign up in pairs with someone with whom they had previous experience collaborating. This decision was made in order to create a more realistic collaborative environment in which group members could feel confident working with each other. This aspect relates to common ground, which was also discussed in section 3.1.3.

From the data collected in this study, Shah and González-Ibáñez (2010) explored the applicability of Kuhlthau's (1991) ISP model to CIS. While the authors indicated that this model constitutes a reasonable reference to investigate CIS, they also reported limitations when attempting to map CIS to the ISP stages. Another interesting aspect of this study was the evaluation of communication messages to determine the affective tone of communication as well as perceived relevance of information. The authors adopted a valence-based approach to classify messages as positive, negative, or neutral. With regard to perceived relevance, the authors focused on discussion messages that were produced around the information collected during the collaborative search process.

Based on findings from this evaluation, González-Ibáñez and Shah (2010) further investigated how affective processes are exchanged during the information search process and how the affective tone of communication relates to team performance. The authors were able to distinguish three performance clusters that were linked to the ratio between positive and negative messages exchanged during the collaboration process. This approach was inspired by studies in positive psychology conducted by Losada and Heaphy (2004) and Fredrickson and Losada (2005) that claim that specific proportions of positivity and negativity correlates to higher or lower levels of performance. Additional details about positivity ratio and its relation to performance can be found in section 3.1.2.2.

While this study provided insights about affective processes in CIS and their possible relationships to performance, the experiment was not designed to investigate affective processes, as neither affective manipulations nor observations with specialized instruments were implemented. Moreover, the granularity of the data analyzed to determine affective processes (i.e. communication logs) was not precise enough to assist researchers with the validation and interpretation of findings.

I.2 Time-space study

The reference study presented above provided interesting findings about CIS and affective processes and at the same time raised several questions. First of all, the study explored CIS under the assumption that searching information in collaboration with others lead to better results than those achieved individually. Second, although the study focused on the impact of different awareness conditions, these were constrained by text-based implementations, meaning that the impact of other communication channels was not addressed, bringing up questions regarding media richness and social presence. Some questions raised in this study are: (1) What are specific differences of CIS with respect to individual information seeking?, (2) To what extent, if any, is CIS better than information seeking performed by single users?, (3) How does the richness of communication channels correlate to search performance of teams?, (4) To what extent, if any, is the richness of communication channels linked to socio-affective factors?, (5) How does the location of group members influence affective processes?, (6) How does the location of group members affect search performance?, (7) Does working in a CIS task at the same time differ from doing so at different moments?, (8) If so, how does searching information collaboratively at different times influence affective processes and performance?, and (9) How does information synthesis performed as a result of CIS differ from that performed by individual users?

I.2.1 Recruitment procedure and participants

In order to address these and other questions, a large laboratory study involving 160 participants in 80 pairs and 10 individual participants was conducted. For the case of collaborative pairs, participants were required to sign up in pairs with someone with whom they had previous experience collaborating, thus following the recommendations derived from the reference study.

Recruitment calls were sent through different listservs. The signup process was performed through an online system that collected basic information from participants (i.e. first name, last name, email address, and brief description of prior collaboration experience for pairs), and a time slot of their preference that was automatically locked after the registration was completed. Following this, the system automatically sent an email to participants reminding them of the date, time, and details about the study. Finally, the system informed the researcher of the new registration.

Participants in this study were Rutgers students (both undergraduate and graduate) from different disciplines. Students were encouraged to participate in the study through artificial/external motivations consisting of $10 compensation for each participant. In some cases students were also allowed to choose between the $10 or class credit if they were enrolled in specific courses. Moreover, in order to motivate participants to take the task seriously, the experiment was framed as a competition, where the best three performing pairs or solo users had the possibility to win cash rewards. Specifically, $50 were offered for first place, $25 for second place, and $15 for third place. Rewards were given to each participant, so if a pair ranked with the highest performance, each group member received $50.

I.2.2 Experimental conditions

The 10 individuals were used to create a baseline ($C0_{Single}$) to perform comparisons with collaborative pairs, while each collaborative pair was randomly assigned to the experimental conditions until there were 10 pairs in each condition. These conditions varied by communication channels (independent variable) and collaborative support, both implemented in different collaboration modes across time and space.

The organization of these conditions is depicted in Figure J.1 and the description of each condition is as follows:

- $C1_{F2F\_same}$: f2f communication and collaborative text editor while working co-located at the same computer.
- $C2_{F2F\_diff}$: f2f communication and collaborative text editor while working co-located with different computers.
- $C3_{Text}$: text-chat communication and collaborative text editor while working remotely located.
- $C4_{Audio}$: audio-chat communication and collaborative text editor while working remotely located.
- $C5_{Video}$: video and audio chat communication and collaborative text editor

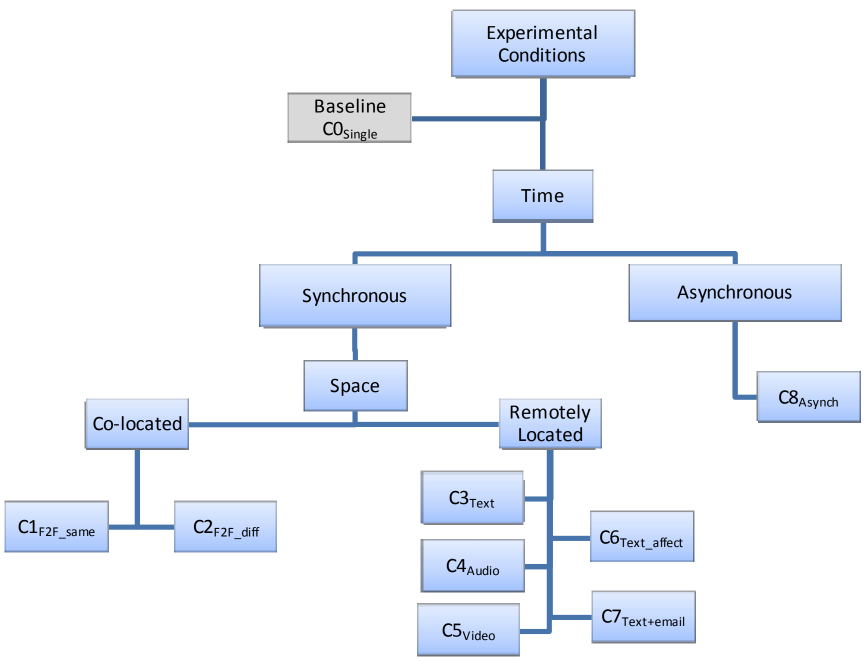


Figure J.1. Experimental conditions of the time-space study.

while working remotely located.

- $C6_{Text\_affect}$: emotionally enriched text-chat communication and collaborative text editor while working remotely located.
- $C7_{Text+email}$: text chat communication and email-based text editor while working remotely located
- $C8_{Asynch}$: asynchronous or sequential collaboration with f2f communication enabled at the beginning and in the middle of the experimental session. Text chat enabled for the first group member to leave messages to his/her partner. Collaborative text editor enabled.

A representation of the different experimental setups is illustrated in Figure J.2. Note that in the conditions where participants worked at remote locations, they were placed in different rooms so that communication was restricted to the channels enabled in the particular condition.

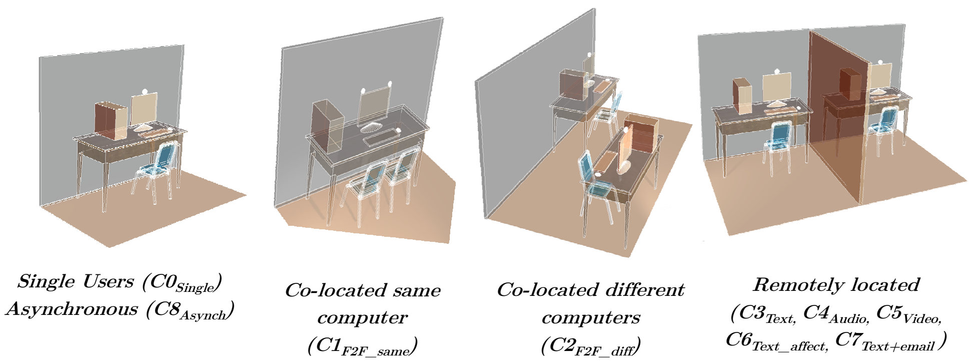


Figure J.2. Experimental setups.

I.2.3 Task

Both collaborative pairs and individual participants were required to perform an exploratory search task, regardless of the experimental condition to which they were assigned. The search task consisted of gathering relevant information about the causes, effects, and consequences of the British Petroleum oil spill, which originated in the Gulf of Mexico in the first half of 2010. Participants were also asked to write a short report addressing specific aspects such as a description of how the event took place, reactions from the company and different governments, and attempts to fix the problem.

The study was conducted the second half of 2010, thus the topic was recent and popular in the media. Pilot runs were conducted before starting the actual study, which indicated a vast number of sources on this topic available online.

I.2.4 Experimental system

A customized version of Coagmento (Shah, 2010b; González-Ibáñez & Shah, 2011) provided support for information search, collaboration, communication, and information synthesis. This version of Coagmento included a renewed chat, an information rating system, an integrated collaborative text editor, and a shared repository containing snippets, search history, and bookmarks. The features that were not used in the study were removed in order to control potential intervening variables. The implementation of additional communication channels, as required by the experimental conditions, was achieved by integrating third-party software such as Skype, Gmail, and *eMotion* (Sebe et al., 2007). These software were activated and positioned in specific areas of the screen depending upon the experimental condition. A snapshot of the customized version of Coagmento is presented in Figure J.3.

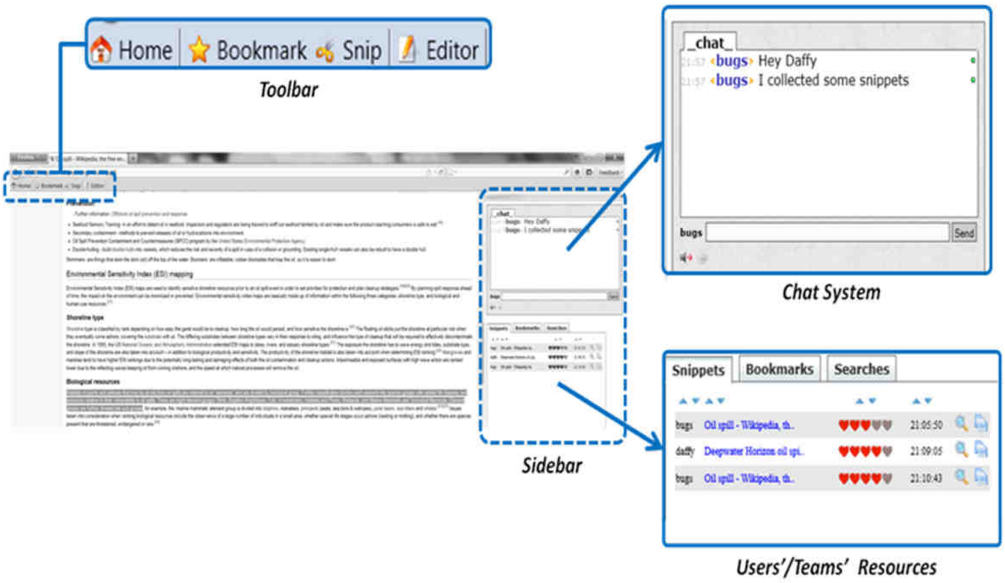


Figure J.3. A snapshot of the experimental system with parts of it shown in detail (Shah & González-Ibáñez, 2011, p.916).

I.2.5 Session workflow and data collection

Experimental sessions lasted approximately one hour. During this period, participants had to complete a consent form, fill in questionnaires before and after completing the task, watch a tutorial to get familiar with the system, perform the information search task with the support of the resources specified in the corresponding experimental conditions, and participate in a brief interview at the end of the session.

In order to measure affective processes, participants self-reported how they felt before and after completing the search task. This self-assessment was reported through the PANAS questionnaire, which was integrated in the system as part of the session workflow. In addition to self reports, participants' faces were video recorded during the entire search session in order to perform offline automatic facial expression analyses with specialized software.

Browsing activity was logged by Coagmento. This included visited pages, search result pages (SERPs), bookmarks, snippets, chat messages, relevance judgments, reports, and timestamps for later analyses. Other types of data were recorded with third-party software, which included Camtasia for desktop activity, Pamela call recorder to save voice and video conversation whthin the different conditions, and emails from Gmail for $C7_{Text+email}$.

I.2.6 Focused research

With the data collected in this large user study, specific investigations were conducted in order to address some of the research questions derived from the reference study. This section provides a brief overview of published and non-published findings derived from this study.

- ***Synergy***

The first study conducted with this dataset focused on investigating the synergic effect in CIS; in other words, to determine whether or not working collaboratively during exploratory search task lead the team to achieve something greater than the sum of its individual parts (e.g. more information coverage and more information diversity).

This study was presented at SIGIR (Shah & González-Ibáñez, 2011). Here the authors compared the pairs from three collaborative conditions (i.e. $C1_{F2F\_same}$, $C2_{F2F\_diff}$, and $C3_{Text}$), participants in the baseline condition ($C0_{Single}$), and an artificial condition created by generating all possible combinations of single users in $C0_{Single}$. The latter condition was specifically used to test the synergic effect.

The researchers focused on search performance (using some of the measures described in section 3.2.2.1) and cognitive load, which was measured using a simplified version of the NASA-TLX (Hart & Staveland, 1988). The authors found that collaborative pairs in conditions $C2_{F2F\_diff}$, and $C3_{Text}$ outperformed individual participants with respect to different aspects such as recall, F-measure, coverage, relevant coverage, unique coverage, useful pages, and likelihood of discovery. They also found that pairs $C1_{F2F\_same}$ showed similar levels of performance to those found in individuals,

In terms of synergic effect, the authors found that real pairs in $C3_{Tex}$ outperformed artificial pairs in aspects such as unique coverage, useful coverage, likelihood of discovery, and diversity. Pairs in $C2_{F2F\_diff}$, on the other hand, only outperformed artificial pairs in terms of useful coverage. The authors also discovered that $C3_{Tex}$ outperformed $C2_{F2F\_diff}$ in terms of diversity.

Finally, the authors found that the reported cognitive load of participants working collaboratively was not significantly higher or lower than that reported by solo users.

These results, though limited by the experimental approach, suggested potential benefits of CIS. However, questions were raised for future investigation. For example, it was not clear why working remotely with text-chat communication lead participants to formulate more diverse queries than those formulated by pairs working co-located with different computers. In order to properly address this question, an in-process evaluation was necessary.

- ***Space***

In a follow-up study the researchers investigated the possible effects of different space conditions and variations in communication channels within CIS. Some of the results from this study were published in González-Ibáñez, Shah, and Haseki (2013). In order to approach this problem, the researchers focused on three collaborative conditions: $C2_{F2F\_diff}$, $C3_{Text}$, and $C4_{Audio}$. In order to better understand the differences among these conditions, the researchers took a closer look at the collaboration process and products derived from it.

To evaluate the collaborative process, the authors investigated communication processes within each team. This was conducted using the communication evaluation procedure described in the evaluation framework introduced previously in this chapter (section 3.2.2.2). As a result of this process, the authors identified that communication processes of pairs in $C1_{F2F\_same}$ and $C2_{F2F\_diff}$ were less task-oriented than the communication processes carried out by pairs in $C3_{Text}$, and $C4_{Audio}$.

In terms of communication volume and effort, it was found (not surprisingly) that the conditions with audio-based communication enabled (i.e. $C2_{F2F\_diff}$ and $C4_{Audio}$) exchanged significantly more messages than pairs in $C3_{Text}$. It was found, however, that those in $C3_{Text}$ were able to reach significantly higher communication balance than within the other collaborative conditions.

In regards to performance measures, the authors confirmed that remotely located conditions showed higher search diversity than that achieved by participants working at the same location. It is believed that this situation is a side effect of sharing the same space with collaborators due to implicit or explicit influences of group members on their peers' thoughts.

An evaluation of cognitive load revealed that participants working in condition $C4_{Audio}$ conducted searches with significantly lower cognitive load than that reported by participants in $C1_{F2F\_same}$.

Additional components in this study include information synthesis and affective load. In regard to information synthesis, the authors applied automatic readability measures and also manual grading procedures. However, none of the evaluations reported significant differences across conditions. With respect to affective load, a significant difference was found between $C2_{F2F\_diff}$ and $C4_{Audio}$, in which the affective load was lower for participants in the latter condition.

Unpublished analyses showed that the performance and communication processes of pairs in $C5_{Video}$, $C6_{Text\_affect}$, $C7_{Text+email}$ was poorer among different aspects than those found in pairs of $C3_{Text}$ and $C4_{Audio}$. Interviews with participants also revealed negative aspects of the above mentioned conditions. For example, participants in $C5_{Video}$ indicated that having a video stream of their partners enabled while working on the task was distracting. They felt that this feature was unnecessary and that only audio should be enabled for tasks like the one they had to perform. Other participants reported that it would be ideal for them to have the ability to enable and disable channels when needed. Similarly, participants in $C6_{Text\_affect}$ indicated that having affective support (which was implemented with a continuous signal that represents the mood of each participant and that was derived from online facial expression analysis) was attractive at first, but then they did not really pay attention to that component of the system. They only focused on text chat and opted to express affective tone through paralinguistic cues.

Finally, $C7_{Text+email}$, while similar to $C6_{Text\_affect}$, differed from the latter and also from the other conditions in that the construction of the final report was performed using Gmail. This was done in order to simulate the document creation process in which group members work independently and exchange partial increments of the document as email attachments. Participants in $C7_{Text+email}$ described that the writing process exchanging different emails was confusing and inefficient. This contrasted with the opinions expressed by participants in all the other conditions, who highlighted the collaborative editor as one of the most interesting and useful features provided by Coagmento.

- ***Time***

Later, a new study was conducted with the aim to investigate effects of synchrony (i.e. synchronous and asynchronous) in CIS. In order to accomplish this research objective, $C8_{Asynch}$ was added to the set of experimental conditions used in the space study presented above. In this condition, pairs of users worked sequentially and the search session was split into two blocks (25 minutes each) so that only one group member could work on the task at a given time.

Each pair was given five minutes at the beginning of the session to communicate. For a given pair, participants were left alone in the laboratory during this time so that they could have a face-to-face discussion about the task, define strategies, and decide who was going to take the first block. After this short meeting, only one participant remained in the laboratory, whereas the other waited outside. Throughout the following 25 minutes, the participant who remained in the room proceeded according to the plan defined in the initial meeting. While performing the task, this participant could leave messages as well as information resources for his/her partner so that he/she could catch up later in his/her block. The participant waiting outside, on the other hand, was not allowed to work on the task nor establish any sort of communication with his/her partner. Once the first block was completed, participants were allowed to reunite for 5 more minutes in order to talk about what was done and what remained. After this meeting, the participants switched positions (i.e. the one who started left the laboratory, whereas the other remained in it). Using the next 25 minutes, the participant that remained in the lab completed the task.

As expected, analyses revealed that communication in $C8_{Asynch}$ was task oriented showing high levels of balance in terms of communication volume and effort. Due to time constraints, communication was found to be very concise with most of the messages oriented to discussing the content and attaining task coordination.

In regard to productivity, it was found that pairs in $C8_{Asynch}$ were able to discover significantly less useful information than those in $C4_{Audio}$, however diversity in $C8_{Asynch}$ was significantly higher than that achieved by pairs in synchronous conditions.

With regard to information synthesis, no significant differences were identified. Likewise, comparisons in terms of affective load did not show significant differences. However, it was found that the cognitive load of participants in $C8_{Asynch}$ was significantly lower than that of participants in $C2_{F2F\_diff}$.

Results from this study were presented at CHI and ASIS&T. Details of this research are provided in González-Ibáñez, Shah, and Haseki (2012a; 2012b).

- ***Affective processes***

The evaluation of affective processes in this study was limited to expressive components (i.e. facial expressions) and self-reported affects through PANAS. With regard to emotion expressiveness, the original goal was to evaluate the expression of basic emotions in all the conditions; however, during the analyses it was found that the conditions in which the participants could talk constituted an important source of noise for the facial expression recognition systems, due to voluntary activity in the facial muscles as a result of speech, which as explained in the previous studies, was abundant. For this reason, only four conditions (i.e. $C0_{Single}$, $C3_{Text}$, $C6_{Text\_affect}$, $C7_{Text+email}$) and partially $C8_{Asynch}$ met the conditions necessary to properly conduct automatic facial expression analyses.

From this group, only two conditions were selected, namely, $C0_{Single}$ and $C3_{Text}$. The video data captured from each participant in these conditions was pre-processed in order to clean, segment, and synchronize the data with other sources of information such as browsing and communication logs. After this pre-processing stage, the videos were analyzed with three different software solutions using the method described in section 2.3.1.

Based on the empirical evaluations with the video data as well as the literature about the systems used to automatically classify facial expressions, it was decided that only smiles would be investigated. There were two major reasons for this decision: (1) high accuracy of the systems in detecting smiles, and (2) relevance to investigating expressed happiness as an aspect of positive psychology.

The results presented at ASIS&T (González-Ibáñez, Shah, & Córdova-Rubio, 2011), indicated that participants working in collaboration with others ($C3_{Text}$) smiled significantly more often than individual users ($C0_{Single}$); this in spite of the limitations of a text-based channel to carry affective information as described in Table 3.1. A minute-by-minute analysis suggested that smiling in $C3_{Text}$ would be related to communication episodes, whereas smiling in $C0_{Single}$ would be linked to specific information-related events such as finding a page to be relevant for the task being solved.

As part of the evaluation of facial expressions, random search sessions were manually inspected to validate results obtained from the automatic facial expression analyses software. During this observation process, it was possible to identify a variety of facial expressions that did not map to basic emotions identified by the system. Some of the expressions as well as gestures were related to attention, confusion, boredom, and fatigue.

When relating facial expressions to the information search, it was possible to observe different kinds of affective reactions when people were exposed to information that later was annotated by themselves as relevant. For example, in one occasion a participant expressed disgust while reading the content of a webpage. Immediately after that reaction, the participant marked that page as relevant. In this example, the emotional reaction was derived from the content of the page, in particular images of dying birds due to the oil spill. While disgust typically has a negative connotation, in this scenario the result was positive since relevant information was found.

As described in the above studies, one of the aspects measured was affective load, which represents an adaptation of Nahl's (2005; 2009) formulation using a combination of responses to specific questions within PANAS and NASA TLX. Results based on this measure revealed that participants in some conditions did better than others in managing affective load produced as a result of information search and collaboration.

A comparison of the first six collaborative conditions with regard to the overall affective experience was conducted by contrasting participants' responses to the PANAS questionnaire, which was responded at the begginning and at the end of the sessions. Results from this analysis showed increasing positive affects[54] at the end of the sessions for participants in $C4_{Audio}$. In addition, variations in terms of negative affects were rather small. These two aspects resulted in a higher positivity ratio.

The same analysis also showed that participants in $C2_{F2F\_diff}$ reported a prominent decay in the experience of positive affects and a noticeable increment of negative ones. While none of the results were statistically significant at $p<.05$, they provide some insights about how different communication channels could contribute to have a more pleasant or unpleasant experience in CIS.

---

[54] Affects is used in this context under PANAS terminology.

Appendix J. Pilot study summary

Between April 23rd and May 4th a pilot study was conducted in order to evaluate design decisions, experimental protocol, task, analyses techniques, and the system that will be used in a research aiming to investigate the implications of positive and negative affects in the information search process of individuals and teams. The pilot study involved 12 participants randomly assigned to five experimental conditions. Sessions in the study lasted one-hour, in which the participants filled in questionnaires, received affective stimuli, and performed a common information search task. In each session, facial expressions, eye tracking data, electrodermal activity, desktop activity, users' actions, and search as well as communication logs were collected. In addition, as sessions were conducted, observations about participants' behaviors, system problems, and research protocol were made. Results from this pilot study led to introduce changes in the experimental design, define a rigorous research protocol, and perform some adjustment to the experimental system.

As a result of this pilot study, the following actions and changes were performed:

1. Improvement of the experimental design for successfully measuring the direct effects of affective stimuli in the information search process of both individuals and teams.
2. Implementation of a two stages experiment design, this in order to measure the implications of affective stimuli in the long run.
3. Preliminary hypotheses were formulated.
4. Difficulty level of questions was fixed in two for the search task designed for the main study.

5. A new version of Coagmento named Coagmento Collaboratory was developed.
6. Coagmento Collaboratory was tested and improved to better support the requirements of the main study.
7. Stimuli was successfully evaluated and improved based on the results of the pilot study.
8. Several steps were incorporated in the research protocol: this includes procedures regarding sensors calibration, instructions, and researcher responsibilities, to name a few.
9. New layout of the system was implemented to provide visual feedback to the researcher and participants about the eye tracker status while collecting data; this in order to avoid the loose of data.
10. Web cameras were evaluated and changes were implemented.
11. Questionnaires were improved.

A detailed explanation of the pilot study, analyses, results, and decisions made for the main study are available for consultation.