User Experience of Mobile Devices: Physical Form, Usability and Coolness

Dimitrios Raptis

Department of Computer Science

Research Centre for Socio-Interactive Design

Aalborg University

December 2014

User Experience of Mobile Devices: Physical Form, Usability and Coolness

Published and Distributed by:

Department of Computer Science, Aalborg University

Selma Lagerlöfs Vej 300, DK-9220 Aalborg Øst, Denmark

ISSN 1601-0590

Submitted to the Faculty of Engineering and Science in
partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at

December 2014

Opponents

Assoc. Prof. Ivan Aaen, Department of Computer Science,
Aalborg University

Dr. Keith Cheverst, School of Computing and
Communications, Lancaster University

Prof. Kasper Hornbæk, Department of Computer Science,
University of Copenhagen

Supervisor

Prof. Jesper Kjeldskov

Co-Supervisor

Assoc. Prof. Mikael B. Skov

User Experience of Mobile Devices: Physical Form, Usability and Coolness

Dimitrios Raptis

Supervisor: Prof. Jesper Kjeldskov

Co-supervisor: Assoc. Prof. Mikael B. Skov

Paper 1: Raptis, D., Papachristos, E., Kjeldskov, J., Skov, M.B., and Avouris, N. Studying the effect of perceived hedonic mobile device quality on user experience evaluations of mobile applications. *Behaviour & Information Technology*, Taylor and Francis (2013).

Paper 2: Raptis, D., Tselios, N., Kjeldskov, J., and Skov, M.B. Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In *Proc. MobileHCI '13*, ACM Press (2013), 127-136.

Paper 3: Raptis, D., Kjeldskov, J., and Skov, M.B. Understanding "Cool" in Human-Computer Interaction Research and Design. In *Proc. OzCHI 2013*, ACM Press (2013), 53-62.

Paper 4: Bruun, A., Raptis, D., Kjeldskov, J., and Skov, M.B. Measuring the Coolness of Interactive Products: The COOL Questionnaire. *Extended version of a CHI 2015 submission.*

ABSTRACT - ENGLISH VERSION

This thesis focuses on mobile devices and it specifically investigates the effect of their physical form on two perceived user experience qualities, usability and coolness. With the term mobile devices, I refer to interactive products that users interact with while being on the move, and with the term physical form I refer to the physical elements that constitute a mobile device as a whole, such as weight, size and materials. The selected research area was addressed through two research questions, one focusing on effects of physical form on usability and the other on effects on coolness.

The thesis consists of four research papers and a statement. The four individual papers provide answers to the two research questions, while the statement acts as a bridge that brings the papers together into a coherent whole. The statement initially discusses what user experience and physical form are, and it continues with related work on usability, coolness and previous studies that investigated their relation to physical form. It then presents a research classification that guided my research work, where the papers are placed in relation to their research purpose: exploration and validation. Explorative studies focus on generating enough data to map out a relatively unknown phenomenon, while validation studies confirm cause-effect relationships that have been identified from previous research. The statement continues with a discussion on the implications of my findings to the broader user experience field and concludes with the results of this thesis.

Four research papers contributed to the results, two for each research question. Two laboratory experiments, one conducted with the purpose of exploration and one with validation focused on the first research question. The explorative study showed that the overall physical form of a mobile device has a significant effect on the perceived usability of an application: the more attractive the physical form, the higher the perceived usability. The other study validated the effect of a particular physical form element on usability and showed that the screen size of a mobile device does not affect perceived usability and effectiveness, but significantly affects efficiency: the larger the screen size the better the efficiency, especially for screen sizes around 4.3 inches.

A literature review and a survey study focused on the second research question. The review paper explored what coolness is and showed that physical form evokes outer and inner coolness: outer coolness is related to the aesthetic attributes of the physical form and inner coolness to perceived personality characteristics. The survey study validated that mobile devices' perceived inner coolness can be measured using a proposed COOL questionnaire (measuring usability, desirability and rebelliousness) and suggested that perceived outer coolness can be measured by existing perceived qualities (attractiveness and classic aesthetics).

ABSTRACT - DANISH VERSION

Denne afhandling fokuserer på mobile enheder og undersøger effekten af fysisk form på brugsoplevelsen og coolness. Mobile enheder dækker over interaktive produkter, som brugerne interagerer med på farten. Med fysisk form menes de fysiske elementer, der udgør en mobil enhed, f.eks. vægt, størrelse og materialer. Dette forskningsområde er belyst gennem to forskningsspørgsmål. Det første spørgsmål fokuserer på effekten af den fysiske form på brugsoplevelsen. Det andet spørgsmål relaterer sig til effekten af fysisk form på coolness.

Afhandlingen består af fire forskningsartikler og en sammenfatning. Sammenfatningen indledes med en diskussion af koncepterne "brugsoplevelse" og "fysisk form". Dernæst gives et overblik over tidligere forskning, som er relevant ift. emnet, herunder forskning der undersøger betydningen af fysisk form på brugervenlighed og coolness. Artiklerne positioneres i et framework relateret til deres forskningsformål: Eksplorative studier og valideringsstudier. Eksplorative studier fokuserer på at skabe tilstrækkelig data, således der kan skabes en indledende forståelse af ukendte fænomener. Valideringsstudier har til formål at bekræfte kausale relationer af koncepter, som er afdækket i tidligere forskning. Sammenfatningen fortsætter ved at diskutere implikationerne af mine forskningsresultater indenfor det bredere forskningsfelt relateret til brugsoplevelser. Afslutningsvist gives en konklusion.

De fire forskningsartikler bidrager med resultaterne, og besvarer de to forskningsspørgsmål. To artikler omhandler laboratoriestudier, hvoraf den ene er eksplorativ af natur og den anden validerende. Disse laboratioriestudier besvarer det første forskningsspørgsmål. Det eksplorative studie viser, at den fysiske form af mobile enheder har en signifikant effekt på opfattelsen af brugervenlighed: Jo mere attraktiv den fysiske form er, des højere vurderes brugervenligheden. Valideringsstudiet undersøger effekten af fysisk form på brugervenlighed. Dette studie viser at skærmstørrelsen på mobile enheder ikke influerer opfattelsen af brugervenlighed i form af gennemførselsrate. Dog har skærmstørrelse en signifikant effekt på gennemførelsestiden, hvor større skærmstørrelse leder til hurtigere opgaveløsning. Dette gør sig især gældende for skærmstørrelser på ca. 4.3 tommer.

En litteraturgennemgang samt en spørgeskemaundersøgelse besvarer det andet forskningsspørgsmål. Litteraturgennemgangen dækker begrebet coolness og udleder, at fysisk form fremkalder ydre og indre coolness. Ydre coolness relaterer sig til æstetiske egenskaber ved den fysiske form og indre coolness relaterer sig til personlighedstræk. Spørgeskemaundersøgelsen validerer, at opfattelsen af indre coolness af mobile enheder kan måles via et udviklet spørgeskema: COOL. Dette måler niveauet af ønskværdighed, brugervenlighed og rebelskhed. Spørgeskemaundersøgelsen viser desuden, at ydre coolness can måles gennem eksisterende kvalitetskoncepter såsom klassisk æstetik og attraktivitet.

ACKNOWLEDGEMENTS

Well, its winter 2004. Not a very cold one, at least not a Danish winter. And I remember thinking: *you should do a Phd!* Here we are ten years after that and this thingy you are reading is my thesis. So, it basically took me as many years as Odysseus did in order to return home from Troy. And maybe with more adventures than he had. But what the hell, the trip is that matters and not the destination!

In this trip, I met, worked and had fun with a lot of people. I want to thank a million of them (exaggerating a bit as always), but the space is limited, so I will just put the first names that pop into my mind. First, I would like to thank my supervisor Jesper Kjelskov and my co-supervisor Mikael B. Skov. In 2011 you gave me the chance to start a PhD in an amazing environment, you pointed me to the correct directions, you gave me the freedom to follow some crazy ideas, you gave me a good spanking when it was required, and you are two of my dearest friends in Denmark. Thank you very very much and you earned forever the nickname: The Bosses! Second, I would like to thank all the people in both the S+I group in Aalborg University and the HCI Group in Greece. Prof. Avouris thank you for introducing me to the incredible HCI world and Jeni thank you for everything ☺. Peter, Ivan, Jeremy, Jan, John, Lise, Jacob, Rikke, Jane, Nis, Γιώργος, Χρήστος, Adrian, Γιάννης, Νίκος, Νίκη, Φιλιώ, Σοφία, Helle, Helle, and Ulla, thank you for your help, support and all the good moments we shared!

Next, I want to thank some very very special people: Serpil my Rock-N-Roll-Queen, Κώστας and Ερατώ the Family-1, Πάνος the Nissan-Lover-With-A-Ford, Ηλίας and Βικού the Family-2, Anders the Twin, Κάχρι the Tall, Ειρήνη the Metal-Queen, Άρης the Oldest-Friend, Ζήνια the Best-Blond-In-The-World, Karsten the Smiling-German, Ubaldo the Spicy-Gun, Λευτέρης the Cold-Shower-Joker, Hernan the Chocolate-Lover, Λάμπρος the Shop-Owner, Ricardo the Looong, Ζωή the Grave-Digger, Roberto the Some-Blonds-Make-Me-Loose-Control, Κατσού the What-You-See-Is-What-You-Get, Manolito the ONE, Νίκος the I-Cannot-Stop-Working, Henrik the Ninja-Assassin and since I am out of nicknames Jurga, Jette, Amanda, Μπλα, Φοιbs, Χαρά, Στέλιος, Βιβή, Νάνσυ, Άλεξ, Μαρία, Λόgo, Ευδοκία, Åsa, Χανιάς, Kitten, Γιάννης, Εύα, Τάσος, Ediz and the rest of family and friends. Thank you all for everything, and if I disappeared on some of you, I am really sorry, but at least something came out, right? I will try not to do that again! And I hope I did not forget someone, as always..

Finally and most importantly, this thesis is dedicated to the best family in the world. Dad, Mam and Sis after so many adventures, I think we did it. The rest of the pages belong to you!

PHD SOUNDTRACK

My research work and this thesis wouldn't have been produced without music. The following list of songs comprises my PhD soundtrack. Hopefully it will be useful for others too :)

1 *When you start:*
Machine Drum - Realization, (Urban Biology, 2002)

2 *When you finish your first paper:*
Trentemøller - Silver Surfer, Ghost Rider Go!!! (Into The Great Wide Yonder, 2010)

3 *When you get rejected and the reviewers have a point:*
Tom Waits - God's Away on Business, (Blood Money, 2002)

4 *When you get rejected and one reviewer is wrooooong:*
Beastie Boys - Sabotage, (Ill Communication, 1994)

5 *When you get accepted:*
The Heavy - How You Like Me Now?, (The House That Dirt Built, 2009)

6 *When you travel for conferences and work:*
Daft Punk - Around the World, (Homework, 1997)

7 *When somebody tells you "maybe you are too old for this":*
Madrugada - Blood Shot Adult Commitment, (Grit, 2002)

8 *When you have to work during holidays:*
Dead Kennedys - Holiday In Cambodia, (Fresh Fruit For Rotting Vegetables, 1980)

9 *When a weekend is about to start and, strangely, you have nooo wooork at aaaall:*
The Cure - Friday I'm In Love, (Wish, 1992)

10 *When you are stuck and don't know what else to do:*
Joy Division - Day Of The Lords, (Unknown Pleasures, 1979)

11 *When you miss home..:*
Goran Bregović - Ederlezi, (Dom Za Vešanje, 1988)

12 *When you have a brilliant idea on how to proceed:*
DJ Shadow - This Time, I'm Gonna Try It My Way, (The Outsider, 2006)

13 *When somebody pisses you off:*
Death In Vegas – Aisha, (The Contino Sessions, 1999)

14 *When you write your thesis:*
William Basinski - Watermusik II, (Watermusic II, 2003)

15 *When you hand in:*
Imam Baildi - Samba Clarina, (Imam Baildi, 2007)

16 *When you prepare for defense:*
Kode 9 and Spaceape - Nine Samurai, (Memories Of The Future, 2006)

17 *When you are done!:*
?

CONTENTS

LIST OF FIGURES AND TABLES

[1] Icons created by Freepik. Source: Flaticon.com, licensed under a Creative Commons Attribution 2.0 Generic License

1 INTRODUCTION

1 INTRODUCTION

This thesis studies the relation between mobile devices and user experience. It specifically investigates the effect of mobile devices' physical form on qualities people perceive when they interact with them.

User experience (UX) is an area of human computer interaction that studies users' interactions with digital artifacts. User experience extends beyond the functionality aspects of users' interactions and studies qualities such as affect, attractiveness, emotions, etc. The importance of user experience lies in its purpose, which is to inform design (Buxton 2010). For this reason, user experience studies have been successfully conducted for a variety of digital artifacts, such as, game consoles (Mandryk et al., 2006), tourist guides (Arhippainen and Tähti, 2003) and ATMs (Tractinsky et al., 2000).

User experience research can be either theoretical or empirical. From its theoretical perspective, it produces the necessary models that explain how users experience digital artifacts (e.g. Forlizzi and Battarbee, 2004). From its empirical perspective, tools are developed that reliably measure the qualities people perceive when they interact with digital artifacts. Both perspectives are evolving, as new models and new perceived qualities are introduced to the field. This has an impact on users' everyday lives because by producing better user experience models and better tools to measure the perceived qualities, we design artifacts that are more meaningful than previous ones (Hallnäs and Redström, 2002).

This thesis contributes to user experience research by focusing on mobile devices and particularly on their physical form. The term mobile device refers to portable digital artifacts that allow users to interact with them while they are on the move. Such devices include mobile phones, tablets, MP3 players and GPS navigation units. The physical form of a mobile device is the sum of different physical elements, such as weight, size, materials used for the casing, overall design, etc. According to Bloch (1995), one of the main factors responsible for evoking users' responses towards a product is its physical form. In user experience research these responses are measured through perceived qualities. Within this research context, this thesis studies the effect of physical form of mobile devices on two user experience qualities: usability and coolness.

In following sections user experience, usability and coolness are introduced, along with a discussion on what defines the physical form of an artifact. The chapter concludes with the two research questions that guide this research work.

1.1 User Experience: Usability and Coolness

ISO 9241-210 (2010) defines user experience as a *"person's perceptions and responses that result from the use or anticipated use of a product, system or service"*. Tullis and Albert (2008) state that user experience research *"is looking at the individual's entire interaction with the thing, as well as, the thoughts, feelings and perceptions that result from that interaction"*. The fact that user experience studies the entire interaction provides a hint on how broad the field is and how many different research directions exist within it. In the following paragraphs I will briefly introduce the different perspectives that fall within user experience research and then focus on the two user experience qualities that constitute the research domain of this thesis, usability and coolness.

A common characteristic of the studies that belong to user experience research is that they adopt a holistic approach in studying users' interactions with products. This was evident from the first studies that argued about the significance of user experience (e.g. Norman et al., 1995), the first studies that provided data about its relevance for HCI (e.g. Tractinsky et al., 2000) and the first user experience research agendas (e.g. Hassenzahl and Tractinsky, 2006). Slowly, many research directions emerged within the field including user experience researchers studying affect (Picard 1997), pleasure (Jordan 1999; Jordan 2000), beauty (Tractinsky et al., 2000), emotions (Scott and Nass, 2002), satisfaction (Lindgaard and Dudek, 2002) and aesthetics (Lavie and Tractinsky, 2004). User experience research directions can be classified as either theoretical or empirical. The theoretical perspective of user experience research deals with understanding what an *experience* is (for more details: Forlizzi and Ford, 2000) and provides models that describe how people experience products (Forlizzi and Battarbee, 2004; Desmet and Hekkert, 2007; Mahlke and Thüring, 2007; Hartmann et al., 2008; Van Schaik et al., 2012). The empirical perspective of user experience research is related to design and evaluation processes and provides tools that measure the parameters that affect users' interactions, investigates the parameters' relations in specific contexts, and develops methods on how to design and evaluate. The main purpose of the theoretical perspective of user experience research is to provide the necessary vocabulary in order to explain users' interactions in a particular context, while the main purpose of the empirical perspective is to improve design. The empirical part of user experience research is the focus of this thesis.

Within the empirical part of user experience research there are studies of users' emotions, physical and psychological responses, behaviors, preferences, etc., that may occur before, during and after using a product. All these can be measured either objectively or subjectively. Objective measures are the ones that are measured directly. For example, someone can measure emotions through skin temperature, pupil dilation and heart rate (e.g. Peter and Herbon, 2006, or Staiano et al., 2012). Subjective qualities are measured indirectly and they can be classified into two categories: established perceived qualities and recently introduced ones. Their differences lie in the existence, or not, of tools that can reliably measure them and studies that show their uniqueness in relation to the rest of perceived qualities. The established category includes perceived qualities such as pleasure, hedonic and pragmatic quality (perceived usability), attractiveness, and flow (for a more detailed list: Vermeeren et al., 2010; Bargas-Avila and Hornbæk, 2011; Bargas-Avila and Hornbæk, 2012). Examples of reliable tools that can measure these qualities are Jordan's pleasure questionnaire (Jordan 2000), the Attrakdiff questionnaire (Hassenzahl 2004) for hedonic and pragmatic quality (perceived usability), and the FSS questionnaire that measures flow (Jackson and Marsh, 1996). Such reliable tools do not exist for the recently introduced qualities, nor are there enough studies that research their relation to the established ones. Examples of recently introduced qualities are novelty (Mugge and Schoormans, 2012) and coolness (Holtzblatt 2011). From these existing qualities this thesis focuses on usability (perceived and objective) and coolness.

Usability is the quality studied first in this thesis and it is defined as *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* (ISO 9241-11, 1998). Usability can be measured either directly (objective usability) or indirectly (perceived usability). Objective usability is measured through a variety of usability metrics and the most common ones being task completion times and rates, and the number of the identified usability problems. Perceived usability is the quality that people perceive in relation to how usable they believe a product is. With the term usable we refer to products that are effective to use, efficient to use, safe to use, easy to learn, easy to remember how to use them, and have good utility (Rogers et al., 2011). Due to its long tradition within the HCI field, perceived usability can be reliably measured through a variety of tools, such as the SUS questionnaire (Brooke 1996).

Coolness is the second studied quality of this thesis. Since it is a recently introduced perceived quality there is no established definition of coolness, and only recently a tool was developed that measures it (Sundar et al., 2014). Other fields, such as marketing, have studied coolness and tried to identify its fundamental characteristics. In HCI it was introduced by Holtzblatt (2011) and Read et al. (2011, 2012) as a quality that describes delightful experiences that bring joy to people's lives. The authors considered perceived coolness as relevant for user experience research having as a starting point that it is a term people use everyday to describe memorable interactions with people, objects and places. A few studies have followed that are also focused on identifying what is coolness and how can it be measured (e.g. Sundar et al., 2014; Farnsworth et al., 2014a; Farnsworth et al., 2014b).

1.2 MOBILE DEVICES: PHYSICAL FORM

From all available digital artifacts, this thesis focuses on mobile devices. Mobile devices are considered an interesting topic for research due to their fundamental characteristics: they have wide-spread acceptance in people's everyday lives and, since they are mobile, they introduce new, interesting contexts of use through applications and services. The large diffusion of mobile devices in people's lives can be realized from their numbers. For example, there are 7 billion active mobile phones around the world (ITU 2014) and according to marketing projections tablets will dominate the entire PC market by 2015 (IDC 2013). Additionally, new contexts of use can be realized from the vast number of mobile applications that are available. For example Google's Play Store had approximately 1.318.869 available Android applications in August 2014 (AppBrain stats 2014).

While using or seeing a mobile device, one of the main sources for evoking user experience perceived qualities is their physical form (Bloch 1995). For products, Bloch (1995) argues that their physical form is specified by designers who choose to blend into a whole, a number of elements such as shape, proportion, materials, color, texture, etc. In architecture F.D.K. Ching argues that physical form denotes *"the formal structure of a work – the manner of arranging and coordinating the elements and parts of a composition so as to produce a coherent image"* (Ching 2007). In the context of this thesis, the physical form of a mobile device is understood as the composition of physical elements that constitute the whole, which users can sense and are specified by designers, such as shape, color, materials, hardware buttons, screen, weight, etc. Examples of different physical forms of mobile devices are presented in Figure 1.

Since designers are responsible for creating the physical form of an object, they were among the first that understood its significance and developed principles on how to design ideal forms. In the beginning of the 20th century the phrase "form (ever) follows function" (Greenough 1947) became the fundamental principle of the Bauhaus movement, which dominated architecture, art and product design, resulting in artifacts characterized by clear lines, simple forms and increased functionality. Industrial designers, such as Raymond Loewy and Henry Dreyfuss, embraced the role of function, but it was soon realized that if the principle was fully applied, it could lead to identical products with similar physical form, or it could even forbid the design of products that did not serve any function. Eventually, a new principle was introduced, which had its roots to the Shaker movement: "Every force evolves a form" (Davenport 1987). Following this principle, designers take into consideration different forces in order to produce the physical form of an artifact, such as aesthetics, culture, ethical considerations, etc.

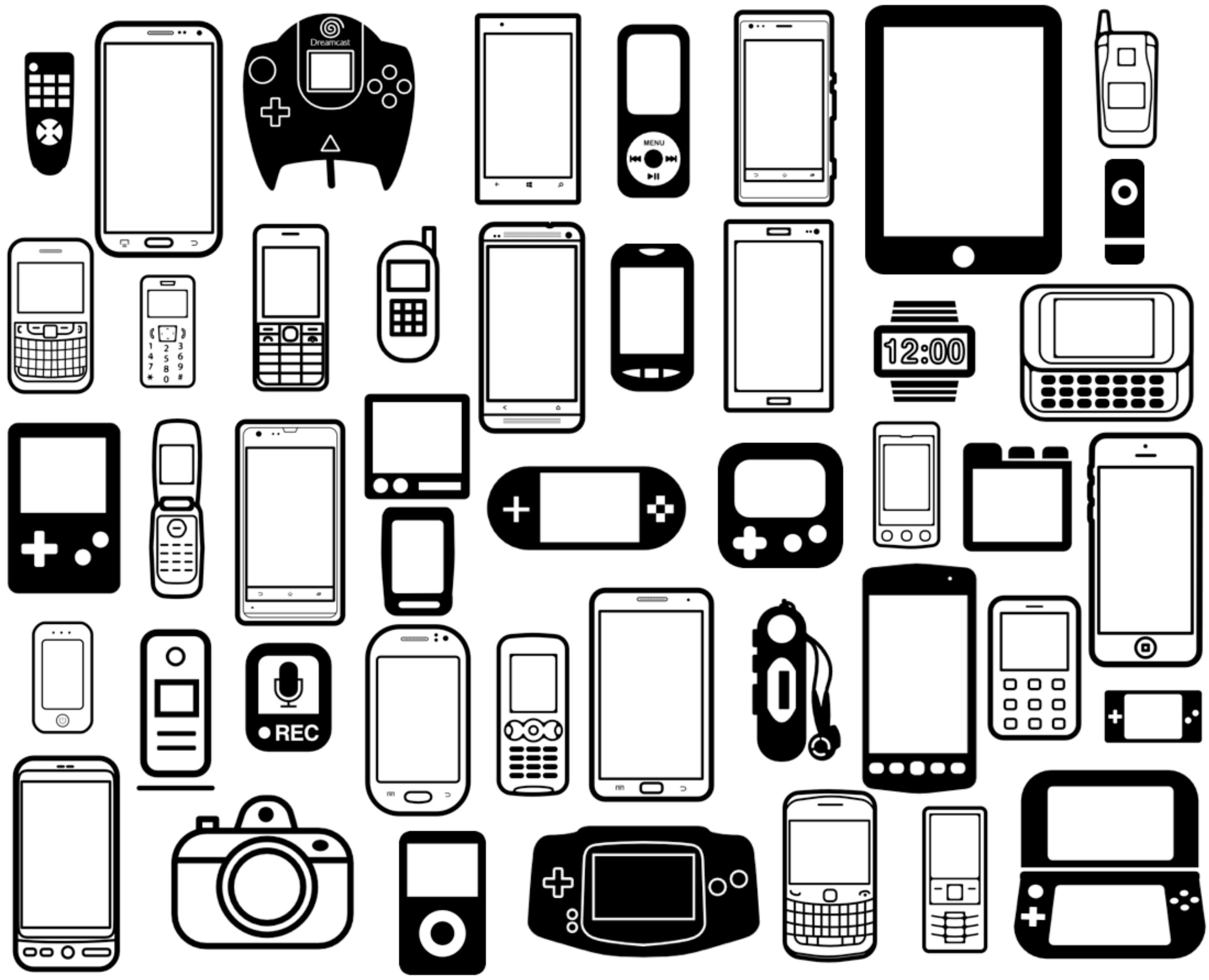


Figure 1. Mobile devices with a variety of physical forms.

If we consider design and evaluation as two sides of the same coin we may observe that the evaluation of digital artifacts within HCI followed a similar path as the one presented above. Initially, function was the dominant force behind evaluations and the effect of physical form on usability was explored. For example, we have findings suggesting that the width and the thickness of a mobile device have a significant effect on controllability and performance (Kim and Jung, 2010). Slowly, and without underestimating the significance of usability, more forces (or qualities) started to be investigated through user experience research, such as pleasure, and aesthetics, and others described in detail in the previous section. For example, there are studies that showed that attractive mobile devices are perceived as more usable (Quinn and Tran, 2010).

In conclusion, there is a strong relation between the physical form of an artifact and user experience research. Since measuring user experience qualities is at the center of the empirical user experience research, this thesis will contribute to this by investigating the effect of the physical form of a mobile device on usability and coolness, through four empirical studies.

1.3 RESEARCH QUESTIONS

From the many digital artifacts that are available, this thesis focuses on mobile devices and specifically investigates the role of their physical form. From the different qualities that are part of user experience empirical research, this thesis focuses particularly on *usability* and *coolness*. The identified research area is approached by two research questions:

RQ1: What are the effects of the *physical form* of a mobile device on *usability*?

RQ2: What are the effects of the *physical form* of a mobile device on *coolness*?

The first two paper contributions of this thesis deal with the first research question. Paper 1 is an exploratory study of the effect of physical form on perceived usability. The second contribution is a validating study that focuses on the effect of a particular physical form element of a mobile device (screen size) to objective and perceived usability. Papers 3 and 4 are related to the second research question. Paper 3 is a literature review that explores the fundamental characteristics of coolness that are evoked by the physical form. Paper 4 is a validating study that begins with the identified characteristics of coolness, and produces a questionnaire that can accurately measure it.

1.4 THE STRUCTURE OF THE THESIS

The next chapter presents related work on usability, coolness and their relation to physical form. Chapter 3 summarizes the paper contributions and briefly presents their findings. Chapter 4 discusses these findings and then this thesis is concluded with Chapter 5 where the results are presented along with research limitations and future research work. In the last chapter the four paper contributions are presented in detail.

2 RELATED WORK

2 RELATED WORK

The previous chapter briefly introduced various user experience research directions, presented the perceived qualities that are measured in the empirical part of user experience research, as well as, defined what is meant by physical form in this thesis. More importantly, it presented the two research questions, which guide this research work. In short, this thesis studies the relations between the physical form of a mobile device, usability and coolness. This chapter will present research work that is related to the two research questions. Therefore previous research on usability and coolness will be discussed, as well as, previous work on their relation to physical form.

2.1 USABILITY

In this section I will briefly introduce usability and usability engineering, present methods that were created to evaluate the usability of products, discuss what usability research is and conclude with how this thesis contributes to user experience research through the first research question.

According to the ISO 9241-11 definition usability is *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"*. Thus, usability describes a set of qualities that a product must have in order to allow its users to be effective, efficient and satisfied. Usability engineering is the process of building usability into a product and creating usable products (Good 1988; Nielsen 1993; Mayhew 1999; Carroll 2000; Rosson and Carroll, 2002). In order to produce usable products Gould and Lewis (1998) suggest four principles: a) focus early and continuously on the users, b) integrate consideration of all aspects of usability, c) test versions with users early and continuously, and d) iterate the design. These principles are applied in practice by usability engineers, which aim to develop products with a high degree of usability. In order to do so they conduct usability evaluations with the purpose of using the findings to redesign.

There are a number of usability evaluation methods and the most common ones are: usability testing, interviews, focus groups, surveys, cognitive walkthrough (Lewis et al., 1990), heuristic evaluation (Nielsen and Molich, 1990), contextual inquiry (Wixon et al., 1990) and the think-aloud protocol (Lewis 1982). A common way of categorizing usability evaluation methods is in relation to the existence, or not, of real users. For example cognitive walkthrough and heuristic evaluation require usability experts to inspect a system, while usability testing requires real users to interact with a system. Among these methods the most dominant one is usability testing and a large number of handbooks have been produced on how to plan, design and perform it (e.g. Dumas and Redish, 1999; Rubin et al., 2008).

Usability evaluation methods have been applied in a variety of environments in order to improve the usability of existing systems. Such cases include safety critical systems (Palangue et al., 1998), clinical information systems (Kushniruk and Patel, 2004), and even nuclear plants (Jung et al., 2002). As technology started being diffused into users' everyday lives, usability became part of user experience research and was studied in new contexts. For mobile devices, the research focus of this thesis, the usability of a variety of mobile applications has been explored in detail. Such cases include mobile guides (Cheverst et al., 2000), museum applications (Kenteris et al., 2009), mobile websites (Buchanan et al., 2001), etc.

While conducting usability evaluations researchers and practitioners use a number of objective and subjective metrics in order to measure usability. Objective metrics are used to measure users' performance while they perform specific tasks with a product, while subjective metrics are measuring user satisfaction and preference/attitude towards a product (Nielsen 2012). Thus subjective metrics are dependent on attitudes and perceptions towards a product, while objective measures are independent of that. Examples of objective usability metrics include: task completion rates, task completion times, total numbers of identified usability problems, user success rate (more details can be found in: Sauro 2010), etc. Although most of these metrics seem rather straightforward to measure there is still a lot of research around them (for example on what a usability problem is, Lavery et al., 1997, and how to classify them Keenan et al., 1999). Apart from the objective performance metrics there are also the subjective metrics that measure users' preferences and attitudes. Such metrics are usually measured by using specific tools. For example there are questionnaires that measure perceived usability, such as SUS (Brooke 1996), USE (Lund 2001) and Attrakdiff (Hassenzahl 2004), or measure perceived workload, such as the NASA TLX (Hart and Staveland, 1988).

Besides applying usability engineering in practice there has also been a lot of research conducted on usability, which has evolved in many directions. For example one direction focuses on usability evaluation methods. There are studies that proposed new usability evaluation methods (e.g. a new evaluation framework for digital libraries, Fuhr et al., 2007), adapted old methods to new technologies and contexts of use (e.g. a new set of heuristics to evaluate computer games, Pinelle et al., 2008), compared different usability evaluation methods in relation to their effectiveness and efficiency (e.g. in healthcare, Jaspers 2009), or demonstrated the importance of good experimental design, by raising concerns for the validity and generality of many studies that produced results that were not supported from their data (Gray and Salzman, 1998). Other studies focused on metrics and developed new ones (e.g. the summated usability metric, Sauro and Kindlund, 2005) or investigated the relationships between objective and subjective measures (e.g. Frøkjær et al., 2000; Hornbæk and Law, 2007). Additionally, many studies focused on practice and identified the required number of users to conduct usability evaluations (e.g. Nielsen and Landauer, 1993; Lewis 1994) and produced guidelines on how to design new systems (e.g. for websites, Nielsen 2001). Finally one of the most influential studies was conducted by Greenberg and Buxton (2008) where they reviewed a number of studies and proposed that usability evaluation can be harmful if applied inappropriately and raised five points for improving the situation. They concluded by suggesting that usability evaluation should be combined with other non-empirical evaluation methods, such as cultural probes, or even abandoned in some cases.

Nowadays, as new contexts of use are introduced through the latest technological developments, new challenges and research directions emerge, both for usability, and other user experience qualities. For example, we do not know how the relationships among user experience qualities change in various contexts of use (Bargas-Avila and Hornbæk, 2011; Bargas-Avila and Hornbæk, 2012). Since user experience is a rather new research field, there are many that argue that usability is an internal part of it (e.g. Kuniavsky 2003; Vermeeren et al., 2010). This is the stance this thesis adopts, as I believe that usability is subsumed by user experience, as suggested by Vermeeren et al. (2010). Thus, this thesis contributes to usability research (and thus to user experience research) through the first research question, by studying the effect of the physical form of a mobile device on objective usability (task completion times and rates) and perceived usability.

2.2 COOLNESS

Coolness is frequently used to describe user approbation of various interactive products. For example, several Apple products are often characterized as being cool. Despite the fact that people often use the term cool to describe objects, activities, persons or places and the term is so much a part of their everyday lives, there is no agreed definition of it.

In an attempt to define coolness other disciplines, such as marketing, or the fashion and music industries, attempted to identify its fundamental characteristics. In relation to objects, three are the most important: 1) coolness is an attitude towards an object, 2) people recognize coolness immediately, and 3) the perception about coolness is shared among the members of a group. The first characteristic of coolness suggests that it is a quality that people assign to objects (Southgate 2003) and it is not something that inheres in the objects themselves (Pountain and Robbins, 2000). Thus, we cannot directly measure the coolness of an artifact, but we can measure the level of coolness that people perceive it has. The process of identifying an artifact as cool, or not, is very fast (Pountain and Robbins, 2000) and it is based on the personal values of an individual, which are created and negotiated inside the groups that an individual belongs to (O'Donnell and Wardlow, 2000).

According to the first identified characteristic, coolness is a quality that people assign to objects. Since human computer interaction deals with the design of interactive products, it seemed relevant to consider coolness as a candidate quality for the field. For example, Sundar et al. (2014) point out that coolness of interactive products became an essential psychological criterion for designers and practitioners when they create new systems, applications, interfaces, or devices. Although the term coolness has been used within HCI for quite some time (for example Ruuska-Kalliokulju et al., 2000), it has only recently appeared in a number of studies that attempted to operationalize it and research it. There are three fundamental studies introducing and researching coolness within human computer interaction. These studies were conducted by Read et al. in 2011 and 2012, and by Holtzblatt in 2011 and all contributed to an initial understanding of coolness both conceptually and methodologically.

In two studies on coolness in HCI, Read et al. (2011, 2012) proposed a conceptual framework of "Being Cool", by "Doing Cool Things" and by "Having Cool Stuff", facilitating the design of cool artefacts and interactive products for teenagers that displayed characteristics such as being rebellious, antisocial, retro, authentic, rich, and innovative. This framework was then used by McCrickard et al. (2012) to identify cool challenges and opportunities while designing for teenagers. In 2011, Holtzblatt discussed the concept of coolness and stressed that cool products bring joy to our lives and contribute to our personal feelings of accomplishment, connection with others, identity, and delightful experiences (Holtzblatt 2011). Methodologically, Holtzblatt presents "The Wheel of Joy" and "The Triangle of Design" as tools for defining the aspects of life and experience that designers should focus on when designing for cool. In a similar manner, De Guzman (2012) and Fitton et al. (2012a, 2012b) present techniques for identifying and analysing coolness through "Cool Card Sorting", or mapping out people's coolness preferences using the "Cool Wall". However, there are also studies that advise caution in introducing coolness to design processes (Cowan et al., 2013).

Recently three studies appeared that extended the theoretical approaches of the previous papers and attempted to decompose and operationalize coolness by conducting experiments. Sundar et al. (2014) presented a detailed literature review, and conducted an experiment that included a variety of digital artifacts in order to identify the parameters that contribute to coolness. The study led to a questionnaire for measuring coolness and they proposed that coolness is decomposed to originality, attractiveness and subcultural appeal. With two recent papers Farnsworth et al. (2014a, 2014b) described the process of experimentally decomposing coolness and suggested that it can be understood by seven constructs: accomplishment, connection, identity, sensation, direct into action, the hassle factor and the learning delta. Finally, Warren and Campbell conducted a study and related perceived coolness, brand preference and the concept of autonomy (Warren and Campbell, 2014).

The above research studies have contributed considerably to an initial understanding of coolness, how to design for it, and how to measure it, but they have also strongly highlighted the need for further research. This thesis contributes to this research area through the second research question, which investigates the effect of the physical form of a mobile device on coolness.

2.3 PHYSICAL FORM

According to Bloch (1995), the physical form of a product evokes cognitive and affective responses from the users, which in user experience research are measured through perceived qualities. Thus, there is a relation between physical form and perceived qualities and there are studies that have identified significant effects. For example, Kim and Sundar (2014) showed that a large screen size could lead to a higher adoption for a smartphone as it promotes both perceived pragmatic and hedonic qualities. Ling et al. (2007) showed that user satisfaction of a mobile phone is mainly affected by attractiveness and size. Han et al. (2004) identified key physical form elements of mobile devices, such as body color, weight and button shape, and provided evidence that they affect satisfaction, luxuriousness, attractiveness and harmoniousness. Their continued work developed a method on how to design the physical form in order to increase user satisfaction and shape users' preferences (Yun et al., 2010). Chuang et al. (2001) moved a step further as they actually redesigned the physical form of a mobile device (through images) to match the preferences of users and they discovered that users prefer mobile phone forms that are "soft" and "compact".

Looking at the relation between usability, coolness and physical form there are a lot of studies available in the related work. Since, according to Bloch (1995) all perceived qualities may be affected by specific elements of the physical form, or can emerge by evaluating the overall form of a product, the identified studies are categorized in: a) studies that focused on specific physical form elements, and b) studies that focused on the overall physical form.

In the first category of studies, while focusing on usability, the most common researched physical form element is screen size. For example, there is a significant volume of research dealing with the effect of screen size on usability for desktop computers (Dillon et al., 1990; Bridgeman et al., 2003; Polys et al., 2005) and mini-notebooks (Lai and Wu, 2014). All these studies showed that usability ratings are affected by the screen size for a variety of tasks (the bigger the better). For the case of mobile devices the first research efforts compared their screens with desktop computers. Such a study was conducted by Findlater and McGrenere (2008) and showed that users are more efficient and more satisfied while using adaptive interfaces on small screens. Another study came from Sweeney and Crestani (2006) where they compared screen sizes between a smartphone, a PDA and a desktop and showed that the screen size is an important factor and it should be taken into consideration while designing how to present information to users. Solely for mobile devices, Jones et al. (2003) reported that internet searching tasks are slower on smaller screens, and Chae and Kim (2004) concluded that both the information structure and the size of a mobile device screen affects users' navigation behavior, but only during complex tasks. Furthermore, large mobile device screens have an impact on mobility (Kim et al., 2011) and controllability (Kim and Jung, 2010). Other physical form elements that were investigated are the size and weight of a mobile device and research has shown that they impact users' preference to hold a device with one hand (Karlson et al., 2008) and their performance in one-handed tasks (Trudeau et al., 2012; Pereira et al., 2013). Finally, users' efficiency is also significantly affected by the position of hardware or virtual keyboard buttons of a mobile device (Kurniawan 2007; Park and Han, 2010; Trudeau et al., 2013; Kim et al., 2013) and by device's thickness (Kim and Jung, 2010).

Additionally, there is a volume of research investigating the effect of overall physical form on usability. For example, there are research findings suggesting that attractive things work better (Norman 2002) and that beauty affects perceived usability (Tractinsky et al., 2000). Specifically for mobile devices, Quinn and Tran (2010) showed that the attractiveness of the physical form has a significant effect on perceived usability and the same was the case for the Sonderegger and Sauer (2010) study, even though they used computer emulated mobile devices. Finally, Sonderegger et al. (2012) showed, through a longitudinal study, that the effect of attractiveness on perceived usability begins to decrease as the exposure time increases.

The relationship between coolness and physical form can be described as relatively unknown and challenging due to the fact that this perceived quality is relatively new in user experience research. As a result, there are not many studies that investigated the relationship between physical form and coolness, both for interactive products in general and specifically for mobile devices. In fact, I have managed to locate only one study that partially investigated the effect of mobile devices' overall physical form on coolness (Sundar et al., 2014). In this study the researchers studied the effect of the physical form of various digital artifacts (some of which were mobile devices) and showed that the physical form evokes three perceived qualities (originality, attractiveness and subcultural appeal). The authors suggest that these three qualities contribute to perceived coolness.

2.4 RESEARCH CLASSIFICATION

As identified in the literature, investigating the effects of products' physical form on user experience qualities is a research area that could be studied more. This thesis contributes to this area of user experience research by posing two research questions. The research questions investigate the effects of the physical form of a mobile device on usability and on coolness. In order to better explain the research approach adopted in this thesis I have produced a classification scheme, which explains the research purposes behind my research work. In Table 1 this classification scheme is presented: the two research questions that focus on usability and coolness are addressed from studies that were conducted with the purpose of *exploration* and *validation*.

		Research Purpose	
		exploration	validation
User Experience	usability		
	coolness		

Table 1. A classification scheme presenting the research approach of this thesis: studies conducted with the purpose of exploration or validation, which investigated the relation between physical form, usability and coolness.

Exploration is the research purpose of studies that are conducted when there is a new research area (Babbie 2007). In such cases, the phenomena under research are at a preliminary stage and there are not enough data to fully explain them (ibid). Exploratory research is usually applied with the purpose to generate enough data that will allow mapping out of the phenomena and pave the way for further research. Consequently, the purpose of exploration is to investigate a phenomenon and identify possible variables that influence it. The term exploration is used in this thesis with a similar meaning as the *understanding* category of research motivations (Basili et al, 1986; Wynekoop and Conger, 1990; Kjeldskov and Graham, 2003), where the research is conducted in order to find meaning behind the studied phenomena. An example of a study with the purpose of exploration could be to identify possible variables behind the increased success of a product.

I use the term *validation* in referring to studies that validate cause-effect relationships that have already been identified from previous research. Thus, while exploratory studies identify parameters that influence a phenomenon, validation studies provide scientific data about their relationships. Validation is part of the ***evaluation*** research purpose (Basili et al, 1986; Wynekoop and Conger, 1990; Kjeldskov and Graham, 2003), where research is conducted with the purpose to evaluate a specific theory, or hypothesis. Furthermore, validation can be considered part of explanatory research (often referred as causal research). Explanatory research is conducted in order to verify cause-effect relationships between variables (Babbie 2007). Thus, the term *validation* is used in this thesis to describe studies that are based on previous research findings and attempt to identify causality, while statistically and experimentally controlling for influential parameters. An example of a study conducted with the purpose of validation could be one that measures the cause/effect relationships between buyers' income, education and social status and the success of a product.

3 CONTRIBUTIONS

3 CONTRIBUTIONS

This chapter presents in short the four individual paper contributions of this thesis. The four papers address the two research questions and explore and validate the effects of the physical form of a mobile device on usability and coolness. The relation between the four paper contributions is presented in Table 2.

User Experience	Research Purpose: exploration	Research Purpose: validation
usability	1	2
coolness	3	4

Table 2. The four paper contributions classified according to their research purpose and the studied user experience quality.

Paper 1: Raptis, D., Papachristos, E., Kjeldskov, J., Skov, M.B., and Avouris, N. Studying the effect of perceived hedonic mobile device quality on user experience evaluations of mobile applications. *Behaviour & Information Technology*, Taylor and Francis (2013).

Paper 2: Raptis, D., Tselios, N., Kjeldskov, J., and Skov, M.B. Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In *Proc. MobileHCI '13*, ACM Press (2013), 127-136.

Paper 3: Raptis, D., Kjeldskov, J., and Skov, M.B. Understanding "Cool" in Human-Computer Interaction Research and Design. In *Proc. OzCHI 2013*, ACM Press (2013), 53-62.

Paper 4: Bruun, A., Raptis, D., Kjeldskov, J., and Skov, M.B. Measuring the Coolness of Interactive Products: The COOL Questionnaire. *Extended version of a CHI 2015submission.*

3.1 PAPER 1: EXPLORING THE EFFECT OF OVERALL PHYSICAL FORM ON USABILITY.

Raptis, D., Papachristos, E., Kjeldskov, J., Skov, M.B., and Avouris, N. Studying the effect of perceived hedonic mobile device quality on user experience evaluations of mobile applications. *Behaviour & Information Technology*, Taylor and Francis (2013).

The first paper reports on a study that was conducted with the purpose of exploring the effects of overall physical form of a mobile device on the perceived usability of a mobile application. Previous studies have identified effects in the past using older mobile phones, and this study aimed at exploring if this was the case with more modern devices.

The paper reports a laboratory experiment that extended previous studies by selecting up-to-date mobile devices, by using a variety of questionnaires to measure perceived usability, and by investigating the effect of physical form on hedonic quality. For the purpose of the experiment two mobile devices were selected that had big differences in their physical form. These differences were in relation to the materials used for the casing, the colors, their design, the position of the hardware buttons and their attractiveness. The 54 participants of the experiment were randomly assigned to one of the devices and were asked to evaluate an unknown mobile application without been provided with any tasks. The two devices offered similar functionality, which was verified through a pilot study that showed that the participants had identical task completion rates and times.

The data from the questionnaires were merged and a factor analysis was conducted, indicating the existence of two factors that were named as new_hedonic and new_pragmatic (perceived usability) qualities. The results of the experiment showed that even though in reality the devices offered the same functionality, their physical form had a significant impact on the perceived usability of the application. Furthermore, specific perceived usability attributes, such as ease of use and ease of learning also had statistically significant differences between the two groups. At the same time there was no difference in hedonic quality, which was an interesting finding since the physical form of the devices was perceived differently in terms of their attractiveness. Two explanations were identified on why only perceived usability was affected by the physical form. The first is that participants considered usability a more important quality, as they were affected by the fact they were inside a usability laboratory. The second is that the used application was a relatively pragmatic one, and this facilitated an effect only on the quality that participants considered as more relevant (perceived usability).

3.2 PAPER 2: VALIDATING THE EFFECT OF SCREEN SIZE ON USABILITY

Raptis, D., Tselios, N., Kjeldskov, J., and Skov, M.B. Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In *Proc. MobileHCI '13*, ACM Press (2013), 127-136.

The second paper reports a study conducted with the purpose of validating the effect of a specific physical form element on the usability of an application. The studied physical form element was the screen size of a mobile device and it was selected due to the relatively large number of studies that had researched its effect on usability in the past, using older mobile devices. Both objective and perceived usability were included in the study.

The study took the form of a laboratory experiment and differentiated from related work by experimentally and statistically controlling for many parameters that could influence the results, and by studying perceived usability as well. Besides screen size, other physical form characteristics were experimentally controlled through the selection process for the three mobile devices that were included in the experiment. The selected mobile devices had the same brand, the same design, the same materials for the casing, the same position of hardware buttons and different screen sizes. Furthermore, *prior experience* with the device and/or the used application and *desire for the device* were statistically controlled, as they are known parameters that can have an effect on usability. The 60 participants of the experiment were randomly assigned to one of the devices and asked to evaluate a specific mobile application, through five tasks of varying difficulty. Data for perceived usability were collected by using the SUS questionnaire, and for objective usability through task completions times and rates. The data were analyzed with a variety of statistical methods.

The findings validated the fact that the screen size also has a significant effect on efficiency for the case of modern mobile devices, but they showed that it does not affect perceived usability, or effectiveness. There are two explanations behind these findings. The first one is that specific physical form elements do not affect perceived usability of an application, but only impact users' performance. The second explanation is that perceived usability did not have any differences between the groups because it was considered a less important quality by the participants, since they interacted with a relatively hedonic application. Both explanations require more studies in order to have conclusive results. Finally, the paper suggests that mobile devices with screen sizes around 4.3 inches could be ideal in terms of performance for common mobile tasks.

3.3 PAPER 3: EXPLORING THE RELATION BETWEEN PHYSICAL FORM AND COOLNESS

Raptis, D., Kjeldskov, J., and Skov, M.B. Understanding "Cool" in Human-Computer Interaction Research and Design. In *Proc. OzCHI 2013*, ACM Press (2013), 53-62.

The third paper reports on a study conducted with the purpose of exploring what coolness is and how it is related to physical form. The study takes advantage of the fact that users assign personality traits to products, and the starting assumption was that if we identify the cool personality traits that are evoked by the physical form, we will understand what coolness is.

The paper reports a literature review and it extended related work by exploring in depth more research domains. The study reviewed books and papers mainly from the domains of marketing, movie and music industry and human computer interaction. In each book or article that was included in the literature review, the characteristics the authors identified as contributing to coolness were collected and coded. Then all of them were analyzed and all duplicates removed until a set of distinct characteristics was produced.

The study proposed that coolness is divided into: a) the perception about the style of a physical form (outer cool), and b) the perceived personality characteristics assigned to the physical form (inner cool). In relation to outer cool the study identified five characteristics in the form of opposites: minimalistic/flamboyant, expensive/cheap, beautiful/ugly, innovative/retro, and illicit/licit. Additionally, the study identified eleven inner cool characteristics: being rebellious and antisocial, embracing authenticity and innovation, seeking exclusivity, pleasure and personal development, being or appearing to be in control, making difficult things appear to be easy, being or appearing to be detached and emotionally neutral, and being strongly tied to a group.

In conclusion, the study provided an explanation on what coolness is, how it is related to the physical form of an artifact, and how it can be decomposed to outer cool and inner cool. The study suggested that the identified outer and inner cool characteristics could be useful for practitioners and designers as they can lead them towards specific directions for design. For example if the physical form of a prototype is perceived as rebellious and authentic, then the final product has a higher chance of being perceived as cool.

3.4 PAPER 4: VALIDATING THE EFFECT OF PHYSICAL FORM ON COOLNESS

Bruun, A., Raptis, D., Kjeldskov, J., and Skov, M.B. Measuring the Coolness of Interactive Products: The COOL Questionnaire. *Extended version of a CHI 2015 submission.*

The fourth paper reports on a study conducted with the purpose of measuring and validating what coolness is in the case of mobile devices. Its purpose was to transform the identified inner cool characteristics that are evoked by the physical form, to measurable factors that could be useful for researchers and practitioners.

The paper reports an online survey study that extended related work by focusing specifically on mobile devices and by including thirteen mobile devices with a large variety in their physical forms. In order to enhance the effect of physical form and minimize the influence of actual usage, the participants did not interact with the actual devices but instead were shown images of them. Furthermore, in order to minimize external parameters that could affect the results all devices were presented to the participants with their screens turned off, all had the same size, and there was no indication of their brand. 2236 participants were asked to answer a number of questions that measured eleven identified perceived inner cool characteristics. None of the participants had any experience with the mobile device they evaluated in the study.

After applying the statistical techniques of exploratory factor analysis (4 times) and confirmatory factor analysis (3 times) the study presented research data supporting that perceived inner coolness can be understood as a three factor structure containing usability, desirability and rebelliousness. The main outcome of the paper is the COOL questionnaire, a reliable tool that can be used to measure the perceived inner coolness of mobile devices, through a set of 16 questions. I believe that it can be a useful tool for practitioners and designers that want to design new physical forms as it can allow them to evaluate their prototypes in relation to coolness. Additionally, the second contribution of the paper is related to user experience qualities in general. The study investigated the relation between the three inner coolness factors and established user experience qualities, such as classic and expressive aesthetics and hedonic and pragmatic quality. It showed that they are statistically different qualities and proposed a PLS model on how they are related for the case of mobile devices.

4 DISCUSSION

4 DISCUSSION

In the previous chapter I presented in short the four paper contributions that constitute this thesis. In this chapter I will discuss the implications of my findings to the broader user experience field. Three main sections constitute this chapter evolving around physical form and usability, physical form and coolness and implications for using the COOL questionnaire in practice.

4.1 PHYSICAL FORM, PERCEIVED USABILITY AND RELEVANCE

Recent research efforts have shown that when people make a judgment about a product they do not identify and rate all the available product qualities, but they perceive instead only the qualities they consider as ***relevant*** to the specific context (Kruglanski and Gigerenzer, 2011; Van Schaik et al., 2012). For example, usability is favored in product choice situations (Diefenbach and Hassenzahl, 2009), but only when the participants are asked about it. In the case of mobile devices, attributes such as cost, beauty, phone features, brand and durability may be more of a priority (Işıklar and Büyüközkan, 2007; Mack and Sharples, 2009; Safa 2013). When information for a quality is not available then people ***infer*** a value for it, based on the available qualities (Kruglanski and Gigerenzer, 2011). For example, when someone sees a product and there is no hands-on experience with it, if he considers usability as relevant he might infer a value about it based on the product's attractiveness. Such processes are triggered by specific rules, which are applied either consciously, or unconsciously (ibid.). Which rule to apply in a specific context is related to the previous experiences of the user and on how effective the rule was in the past (ibid.). Examples of applied rules within user experience literature are: the association of beauty with goodness rule ("what is beautiful is good", Dion et al., 1972), the inference of a value for the usability of a product based on its attractiveness rule ("what is beautiful is usable", Tractinsky et al., 2000), or the application of the reversed rule ("it is usable therefore it is beautiful", Tuch et al., 2012).

The findings of the first paper contribution showed that the overall physical form of a mobile device significantly affects perceived usability, when users interact with a relatively pragmatic application. In the laboratory experiment reported in the second paper contribution perceived usability was not affected by a specific physical form element and a relatively hedonic application was used. These findings suggest that the participants in my experiments might have treated usability as relevant or irrelevant based on the type of the application they interacted with. Thus they could have applied the rule "the application is pragmatic thus usability is relevant/important".

Furthermore, there are research findings in the literature showing that the relationships among the perceived qualities and the applied rules are affected by various parameters, such as: attitude towards brand (Delgado-Ballester and Munuera-Alemán, 2001; Rondeau 2005; De Angeli et al., 2009; Solomon et al., 2010), prior experience (Langton et al., 2007; Quinn and Tran, 2010; McLellan et al., 2012), expertise (Ziefle 2002) and positive or negative reviews (Raita and Oulasvirta, 2010). In the second paper contribution, familiarity was identified as a possible external parameter, as there were data suggesting that participants who interacted with a mobile device that had the same screen size as their own provided better usability ratings.

Based on my findings and the previous discussion I believe that more research on these issues will be beneficial for user experience research as well as human computer interaction in general. Identifying more rules that people apply when they make judgments in their everyday life, or when they are requested to make judgments during evaluations of interactive products, understanding how these rules are related to contexts of use and how they are influenced by external parameters are areas of research that will allow us both to improve our design and evaluation methods, but also to better understand what experience is.

4.2 PHYSICAL FORM, AESTHETICS AND COOLNESS

According to Seva et al. (2011) products evoke responses, which can be classified into three categories: aesthetic impression, semantic interpretation and symbolic association. Aesthetic impression is related to users' responses to a product's appeal, semantic interpretation describes what the product conveys about its function, and symbolic association what the product says about its owner (ibid). These responses occur really fast (Lindgaard et al., 2006; Tractinsky et al., 2006) and within user experience research are measured through qualities that users perceive, consciously, or even unconsciously (Kardes et al., 2004). According to Bloch (1995) these responses lead to behavioral changes. For example, someone might enjoy his new tablet so much that it may lead to an increased usage (behavioral change).

My findings in relation to the second research question of this thesis showed that coolness is decomposed to outer and inner cool, and inner cool has a three-factor structure containing usability, desirability and rebelliousness. By applying the statistical techniques of exploratory and confirmatory factor analysis I have identified that these inner cool factors are different from attractiveness, hedonic quality, classic and expressive aesthetics. This finding shows that coolness should be considered as a new user experience perceived quality, at least for the case of mobile devices. By having as a starting point that outer cool is related to aesthetic impression, usability to semantic association, and desirability and rebelliousness to symbolic association, in the fourth paper contribution I proposed a PLS model (see page 130) that suggested for the case of mobile devices, that outer cool (attractiveness and classic aesthetics), influences inner cool (rebelliousness, usability and desirability) and both of them contribute to an overall judgment on coolness. As with previous studies (e.g. Van Schaik et al., 2012), the issue with a PLS model is that although it accurately quantifies the relationships among constructs, it needs solid theoretical foundations in order to define the direction among the relationships. My model is empirical, as its theoretical basis emerged from a literature review. Therefore even though the directions of the relationships in the model seem plausible (aesthetics is the starting point that shapes all other qualities as it is the case with other models (ibid.)), there is a need for more research as there is always a possibility that the relationships among the qualities are different (for example the relationship between inner and outer cool can be the reverse).

Nevertheless, by treating the proposed PLS model as a first step, I believe that researching in depth if the proposed model is correct and/or if it should be enriched with more qualities, can contribute significantly to a number of user experience research areas. Firstly, we will be able to better understand the relationships between user experience qualities, a challenge already identified in the literature (Bargas-Avila and Hornbæk, 2011; Diefenbach et al., 2014). For example, the model I suggested is different from the one proposed by Van Schaik et al. (2012). Are they applicable in different settings? Do they both ignore some influential qualities? Secondly, we will be able to clarify the relations between hedonic quality and classic and expressive aesthetics, since many studies, including mine, report that they overlap (Diefenbach et al., 2014). Thirdly, we will understand if coolness is an "umbrella" concept that covers many user experience qualities. If this is the case then we will be able to use coolness as a broader evaluation construct that encompasses Attrakdiff's bad/good and ugly/beautiful, and slowly define a specific set of user experience perceived qualities that practitioners could use during user experience evaluations.

4.3 CHALLENGES FOR USING THE COOL QUESTIONNAIRE IN PRACTICE

In the fourth paper contribution I presented the COOL questionnaire, a scale that can reliably measure the three factors that contribute to inner coolness for the case of mobile devices: usability, desirability and rebelliousness. In this section I will discuss five challenges that need to be explored more if the scale is to be used in practice.

While browsing through the literature I identified a study conducted by Hekkert et al. (2003) where they attempted to test the "most advanced yet acceptable" (MAYA) design principle, which was proposed by Loewy (1951). In order to do so they compared designs where the physical form differed in novelty and prototypicality. Novelty is a quality that suggests how novel, different and original a product is in relation to its competition.

Prototypicality refers to the amount an object is representative of a category, or class of objects to which it belongs (Leder et al., 2004). The findings of Hekkert, et al. (2003) suggest that both novelty and prototypicality equally and positively predict preference, but when both are perceived by a user then one dominates, depending on factors like context of use and user characteristics. Since novelty and prototypicality seem to both contribute to my rebelliousness factor, I believe that it would be useful to know how the perception regarding coolness changes in extreme cases. For example, if the physical form of a mobile device is perceived as extremely rebellious (or non-prototypical) is the device cool?

I have identified two more challenges that will have implications for practitioners that consider coolness a design goal: 1) is the scale suitable for other interactive artifacts besides mobile devices?, and 2) how does the scale perform? In relation to the first challenge I believe that the COOL questionnaire is suitable for a variety of artifacts, as was the case with other questionnaires in the past. For example the classic/expressive aesthetics questionnaire (Lavie and Tractinsky, 2004) was designed for websites and today is used for almost any interactive artifact. More studies using the COOL questionnaire with a variety of artifacts will allow us to have conclusive results about its generalizability. Additionally, by conducting more studies we will be able to tackle the second challenge too. At this moment we do not know what it means for an artifact to score, for example, 5 on usability or 3.5 on rebelliousness and thus we do not know if such scores are good, or bad in relation to coolness. In order to understand the behavior of the scale, research approaches that were used for other questionnaire in the past, could be applied (e.g. for the SUS scale, Bangor et al., 2008).

One of the findings of the third paper contribution was that coolness is a value shared among the members of a group. This results in the fact that different groups have very different ideas on what is cool. For example, teenagers that listen to metal music consider different things as cool compared with teenagers who listen to pop music. This translates to the fourth challenge of using the COOL questionnaire in practice, which is related to the participants that practitioners should recruit. If different groups of people have such dramatically differing approaches to coolness then which participants should practitioners involve in their evaluations? There are tools that can allow practitioners to filter the participants (e.g. the Trendsetting Questionnaire, Batinic et al., 2008) but the only way to tackle this challenge is by conducting more studies and research on the degree of differences in relation to coolness for different interactive products. Someone might assume for example that since mobile devices are so integral in the everyday life of different groups of people, then their perception about their coolness will not be so different as it will be with other objects, such as clothes. Of course, only with more studies that compare how different perceptions of coolness among different groups are for a variety of products will we be able to have conclusive results.

Finally, the last challenge for using the COOL questionnaire in practice is related to time. We know for example that while using interactive products for long periods of time, the perception about their usability changes (Sonderegger et al., 2012) and probably this is the case for coolness too. Besides the fact that the perceived coolness of a product might diminish with the time, the added challenge for coolness is that the general perception on what is cool changes rapidly. For example, music bands, clothes, etc., can be cool for short periods of time and then they are considered as uncool. Do interactive products have such rapid changes in their coolness? If yes, what does this mean for design practitioners? Do they need to change often the physical forms of their interactive products?

I believe that tackling these five challenges is of great significance for using the COOL questionnaire in practice, as it will provide practitioners with the necessary knowledge in order to design for coolness. Then, they will have a very useful tool that will allow them to measure how cool their produced physical forms are perceived, will provide a mechanism for redesigning, and will allow them to match the produced physical forms to the targeted groups.

5 CONCLUSIONS

5 CONCLUSIONS

In this thesis I have focused on the effects of mobile devices' physical form on usability and coolness by posing two research questions. In order to answer the two research questions I have conducted four studies that investigated the effect of overall physical form on usability (paper 1), the effect of a specific physical form element on usability (paper 2) and the effect of overall physical form on coolness (paper 3 and paper 4). Papers 1 and 3 report on studies conducted with the purpose of exploration, which refers to studies that explore a domain in order to identify variables that can influence it. Papers 2 and 4 report on studies conducted with the purpose of validation, which refers to studies that verify the relations among the identified variables. The rest of the chapter summarizes the conclusions of my research work and points out limitations and future research directions.

5.1 1ST RESEARCH QUESTION: WHAT ARE THE EFFECTS OF THE PHYSICAL FORM OF A MOBILE DEVICE ON USABILITY?

I have conducted two studies that focused on the first research question and investigated the effect of physical form on usability. Effects on both objective and perceived usability were investigated. The primary findings of my research work for the first research question are:

1. The overall physical form of a mobile device has a significant effect on the perceived usability of an application. The findings suggest that physical forms that are considered as more attractive facilitate higher perceived usability ratings.
2. The screen size of a mobile device (a specific physical form element) does not have an effect on perceived usability and effectiveness, but significantly affects efficiency with an application. Larger screens provide better efficiency during information seeking tasks, and the effect becomes significant for screen sizes around 4.3 inches.

In relation to overall physical form I have identified a significant effect of the physical form of a mobile device on perceived usability when a relatively pragmatic application is used. The findings suggest that in such cases the aesthetic attributes of the physical form affect the perceived usability of the application, when effectiveness and efficiency are at an acceptable level. In other words, devices that are considered as more attractive cause an application to be perceived as more usable. The findings are inline with previous research work that showed that attractiveness has an effect on perceived usability (what is beautiful is usable approach, Tractinsky et al., 2000). Furthermore, the overall physical form of the device does not have any significant effect on the perceived hedonic quality of an application, a quality that was not explored in previous, similar studies.

In relation to specific form elements, I have focused on the screen size of mobile devices, and my findings showed that during information seeking tasks it does not have an effect on perceived usability and effectiveness, but significantly affects efficiency. Larger screens improve the efficiency of the users while they interact with a mobile application, during information seeking tasks. The effect becomes significant for screen sizes around 4.3 inches. Another interesting observation was that perceived usability was not affected, despite the fact that participants were provided with specific tasks (goal mode), where in such cases perceived usability is affected more (Hassenzahl and Ullrich, 2007) in comparison to freely interacting with an application (action mode). The results are inline with previous research work that identified effects of screen size and other physical form elements on objective usability metrics, with the difference that no effect on effectiveness was identified.

In conclusion, the physical form of a mobile device can have an impact on both perceived and objective usability of an application. For perceived usability it seems that the magnitude of the effect is related to the relevance that usability has for people in specific contexts of use. My findings suggest that when users interact with pragmatic applications, usability is considered as important and then the physical form of a mobile device has a significant effect on it, through the attractiveness of the form. Furthermore, my findings showed that specific physical form elements, such as the screen size, can have an effect on objective usability metrics, but more studies are needed with different types of applications, in order to have conclusive results for more physical form elements of modern mobile devices.

5.2 2ND RESEARCH QUESTION: WHAT ARE THE EFFECTS OF THE PHYSICAL FORM OF A MOBILE DEVICE ON COOLNESS?

I have conducted two studies that focused on the second research question and investigated the effect of physical form on coolness. The primary conclusions of my research work for the second research question are:

1. The physical form of a mobile device evokes perceived coolness, which can be decomposed to outer and inner coolness. Outer coolness is related to the aesthetic attributes of the physical form, while inner coolness to the perceived personality characteristics that are assigned to the physical form.
2. Perceived inner coolness of mobile devices can be reliably measured by three factors: usability, desirability and rebelliousness, through the COOL questionnaire. My findings suggest that perceived outer coolness can be measured using existing user experience qualities (attractiveness and classic aesthetics).

My findings show that physical form evokes perceived coolness, which is divided to two parts: outer and inner coolness. When users see a product they perceive the aesthetic attributes of the physical form (outer coolness) and they assign personality characteristics to it (inner coolness), influenced by various parameters, such as previous experience. Outer cool is a combination of five scales of opposites which describe the style of a product: minimalistic/flamboyant, expensive/cheap, beautiful/ugly, innovative/retro, and illicit/licit. Many different combinations of these outer cool characteristics can be perceived as cool from different groups of people, based on the values that the group shares. Inner coolness is related to the personality characteristics that people assign to products they perceive as cool. I have identified eleven inner cool personality characteristics: being rebellious and antisocial, embracing authenticity and innovation, seeking exclusivity, pleasure and personal development, being/appearing to be in control, making difficult things appear to be easy, being/appearing to be detached and emotionally neutral, and being strongly tied to a group.

In relation to perceived inner coolness, I have identified its three-factor structure for the case of mobile devices. Perceived inner coolness can be understood as a combination of rebelliousness, usability and desirability. Based on the aesthetic impression of the physical form of a mobile device (outer cool), users assign values to these three inner cool factors. For the case of mobile devices, inner coolness can be measured reliably using the produced COOL questionnaire. Furthermore, my findings suggest that perceived outer coolness can be measured using existing user experience constructs (attractiveness and classic aesthetics).

In conclusion, the physical form of a mobile device does have an effect on coolness. The perception about the aesthetic characteristics of the physical form (outer cool) shape its perceived inner coolness factors (rebelliousness, usability and desirability) and they both contribute to an overall judgment about the device's coolness. Further research work is needed on how coolness is related to the rest of user experience constructs in different contexts of use and if the COOL questionnaire is applicable for other interactive products.

5.3 PHYSICAL FORM AND USER EXPERIENCE

This thesis contributes to the broader user experience field with a number of secondary findings that go beyond the two research questions. Firstly, this research work showed that existing perceived user experience qualities are not enough in order to measure user experience, for the case of mobile devices. My findings showed that perceived inner coolness acts as an umbrella construct that encompasses one existing user experience quality (usability) and two new ones (desirability and rebelliousness). This finding should urge us to investigate the possibility of new qualities that we currently ignore and users find relevant.

Secondly, I provided a PLS model that proposed specific relationships between outer cool, inner cool and the overall judgment of a device's coolness. My findings suggest that perceived outer cool can be measured by existing user experience qualities (attractiveness and classic aesthetics) and inner cool through the COOL questionnaire. The model suggests that outer coolness influences inner coolness and they both contribute to an overall judgment on a device's coolness. If the model is verified by more studies, it could lead to the specification of a set of perceived qualities and questions that practitioners could use in their evaluations.

Thirdly, I have contributed to the ongoing discussion on the relation between existing user experience qualities. My findings showed that hedonic quality and classic and expressive aesthetics overlap for the case of mobile devices (paper 4). Furthermore, there are also results showing that in evaluations of mobile applications the USE, Attrakdiff and Jordan's Pleasure questionnaires, when combined, actually measure hedonic and pragmatic quality (paper 1). Both findings suggest that practitioners should be careful on the combinations of questionnaires they are using during evaluations, as they might measure the same thing twice, especially when they include mobile devices.

Finally, I have research evidence suggesting that familiarity has an effect on usability ratings. My findings suggest that users that were asked to evaluate a mobile application using a device that has the same size as their own, provide higher perceived usability ratings. If this is the case and my finding is verified with more studies, then familiarity should be added to the list of external parameters (for example prior experience) that influence the evaluation results and should be controlled, either statistically or experimentally.

5.4 LIMITATIONS

The research described in this thesis has some limitations. In my research work I have focused on the effect of mobile devices' physical form on user experience qualities. Due to the imposed time constraints only two qualities were investigated in detail: usability and coolness. Additionally, only one physical form element (screen size) was studied. This is a clear limitation of my research work since including a larger number of perceived qualities and physical form elements would have broadened the research. Nevertheless, going in depth with only two qualities allowed me to have a better understanding as well as richer findings.

Additionally, this thesis is based on laboratory experiments, a literature review and a survey study, and thus, it lacks real world studies. The survey study and the literature review are independent from the setting, but the laboratory experiments have the limitation that they might excluded parameters that could influence the results and exist in the real world. This is a known limitation for any laboratory-based research. Nevertheless, the fact that I conducted laboratory experiments allowed for a better control of parameters that could have an effect on the results, such as brand or prior experience.

Finally, this thesis is about mobile devices, but only mobile phones were used. Similar approaches have been applied in the past for other digital artifacts too, such as websites, without having an impact on generalizability. Nevertheless, only with more studies will there be enough data to know if my findings are generalizable to all types of mobile devices. Furthermore, also in terms of generalizability, it is unknown if the COOL questionnaire is suitable for other types of products, besides mobile devices, such as web applications. The facts that a) the questionnaire was created by following an established approach, and b) a large number of participants were included in its creation process, suggest that it can be suitable for other products too. More studies are needed to confirm this.

5.5 FUTURE RESEARCH

This research work opens up several opportunities for future research. I will briefly present which research directions I consider the most important.

Firstly, one of the limitations of my research is that I only focused on two user experience perceived qualities (usability and coolness) and one physical form element (screen size). Since I have research data showing that the overall physical form and screen size affect these qualities, more user experience perceived qualities should be investigated in detail. Increasing the research domain by focusing on more qualities, by including more types of artifacts and by conducting more real world studies for longer periods of time, will allow us to better understand how physical form affects our experiences.

Secondly, it is important to understand: a) how the relations among the established perceived qualities change in different contexts of use, and b) if they are enough to holistically describe/measure users' experiences. One outcome of my research work is that perceived coolness should be considered as a new user experience quality and I have proposed a model that describes specific relations among specific user experience perceived qualities, for the case of mobile devices. Researching if there are more qualities that we currently ignore and if the proposed model is applicable for other types of interactive products, will allow us to define better user experience models, and this can have an impact on user experience design and evaluation methods.

Thirdly, the COOL questionnaire needs to be used in a variety of evaluations, with a variety of artifacts and while comparing it with other questionnaires. This process will allow us to understand its generalizability and applicability in other domains.

6 REFERENCES

6 REFERENCES

AppBrain Stats (2014) Number of Android Applications. Available online: http://www.appbrain.com/stats/number-of-android-apps

Arhippainen, L., & Tähti, M. (2003) Empirical Evaluation of User Experience in Two Adaptive Mobile Application Prototypes. In *Proc. MUM 2003*, Linköping University E-Press, 27-34.

Babbie, E. (2007) *The Practice of Social Research* (11th edition). Thompson-Wadsworth.

Bangor, A., Kortum, P., & Miller, J. (2008) An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24 (6), 574–594.

Bargas-Avila, J.A., & Hornbæk, K. (2011) Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. In *Proc. CHI 2011*, ACM Press, 2689-2698.

Bargas-Avila, J.A., & Hornbæk, K. (2012) Foci and blind spots in user experience research. *Interactions*, 19 (6), 24-27.

Basili, V.R., Selby, R.W., & Hutchins, D.H. (1986) Experimentation in software engineering. *IEEE Transactions on Software Engineering*, 12 (7), 733-743.

Batinic, B., Wolff, H.G., & Haupt, C.M. (2008) Construction and Factorial Structure of a Short Version of the Trendsetting Questionnaire (TDS-K): A Cross-Validation Using Multigroup Confirmatory Factor Analyses. *European Journal of Psychological Assessment*, 24 (2), 88-94.

Bloch, P.H. (1995) Seeking the Ideal Form: Product Design and Consumer Response. *Journal of Marketing*, 59, 16-29.

Bridgeman, B., Lennon, M.L., & Jackenthal, A. (2003) Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Performance. *Applied Measurement in Education*, 16 (3), 191-205.

Brooke, J. (1996) SUS – A quick and dirty usability scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester, and I.L. McClelland (eds.), *Usability evaluation in industry*, Taylor and Francis, 189-194.

Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., & Pazzani, M. (2001) Improving mobile internet usability. In *Proc. WWW 2001*, ACM Press, 673-680.

Buxton, B. (2010) *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann.

Carroll, J.M. (2000) *Making Use: Scenario-Based Design of Human-Computer Interactions*. MIT Press.

Chae, M., & Kim, J. (2004) Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour and Information Technology*, 23 (3), 165-181.

Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C. (2000) Developing a context-aware electronic tourist guide: some issues and experiences. In *Proc. CHI 2000*, ACM Press, 17-24.

Ching, F.D.K. (2007) *Architecture: Form, Space, and Order*. John Wiley & Sons.

Chuang, M.C., Chang, C.C., & Hsu, S.H. (2001) Perceptual factors underlying user preferences toward product form of mobile phones. *International Journal of Industrial Ergonomics*, 27 (4), 247-258.

Cowan, B.R., Avramides, K., & Beale, R. (2013) Should cool be a design goal? In *Proc. BCS HCI 2013*, BCS L&D Ltd, 1-5.

Davenport, G. (1987) *Every Force Evolves a Form*. North Point Press.

De Angeli, A., Hartmann, J., & Sutcliffe, A. (2009) The effect of brand on the evaluation of websites. In *Proc. INTERACT '09*, Springer-Verlag, 638-651.

De Guzman, E.S. "Cool" Card Sort: Adapting Desirability studies to assess "Cool"-ness (2012) *ACM CHI Workshop on Cool aX Continents, Cultures and Communities*.

Delgado-Ballester, E., & Munuera-Alemán, J.L. (2001) Brand trust in the context of consumer loyalty. *European Journal of Marketing,* 35 (12), 1238-1258.

Desmet, P.M.A., & Hekkert, P. (2007) Framework of product experience. *International Journal of Design*, 1 (1), 57-66.

Diefenbach, S., & Hassenzahl, M. (2009) The Beauty Dilemma: Beauty is valued but discounted in product choice. In *Proc. CHI 2009*, ACM Press, 1419-1426.

Diefenbach, S., Kolb, N., & Hassenzahl, M. (2014) The 'Hedonic' in Human-Computer Interaction – History, Contributions, and Future Research Directions. In *Proc. DIS 2014*, ACM Press, 305-314.

Dillon, A., Richardson, J., & McKnight, C. (1990) The effect of display size and paragraph splitting on reading lengthy text from screen. *Behaviour and Information Technology*, 9 (3), 215-227.

Dion, K., Berscheid, E., & Walster, E. (1972) What is beautiful is good. *Journal of Personality and Social Psychology,* 24 (3), 285-290.

Dumas, J.S., & Redish, J.C. (1999) *A Practical Guide to Usability Testing*. Intellect Ltd.

Farnsworth, C., Holtzblatt, K., Pai, S., Wagg, K., Held, T., Wylen, E., Kennedy, S.L., & Kutty, P (2014a) Measuring Products "Coolness" – Developing a Measurement Instrument. In *Proc. CHI 2014 Extended Abstracts*, ACM Press, 893-896.

Farnsworth, C., Holtzblatt, K., Held, T., & Pai, S. (2014b) Cool in Business: Developing a Data-Based Instrument Measuring "Cool". In *Proc. HCII 2014*, Springer, 232-243.

Findlater, L., & McGrenere, J. (2008) Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. In *Proc. CHI 2008*, ACM Press, 1247-1256.

Fitton, D., Horton, M., Read, J.C., Little, L., & Toth, N. (2012a) Climbing the Cool Wall: Exploring Teenage Preferences of Cool. In *Proc. Ext. Abstracts CHI 2012*, ACM Press, 2093-2098.

Fitton, D., Read, J.C., Horton, M., Little, L., Toth, N., & Guo, Y. (2012b) Constructing the Cool Wall: A Tool to Explore Teen Meanings of Cool. *PsychNology*, 10 (2), 141-162.

Forlizzi, J., & Ford, S. (2000) The Building Blocks of Experience: An Early Framework for Interaction Designers. In *Proc. DIS 2000*, ACM Press, 419–423.

Forlizzi, J., & Battarbee (2004) Understanding Experience in Interactive Systems. In *Proc. DIS 2004*, ACM Press, 261-268.

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000) Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? In *Proc. CHI 2000*, ACM Press, 345-352.

Fuhr, N., et al. (2007) Evaluation of digital libraries. *Int. Journal on Digital Libraries*, 8 (1), 21-38.

Gray, W.D., & Salzman, M.C. (1998) Damaged Merchandise? A Review of experiments that compare Usability Evaluation methods. *Human Computer Interaction*, 13, 203-261.

Greenberg, S., & Buxton, B. (2008) Usability evaluation considered harmful (some of the time). In *Proc. CHI 2008*, ACM Press, 111-120.

Greenough, H. (1947) *Form and Function: Remarks on Art, Design, and Architecture*. University of California Press.

Good, M.D. (1988) Software usability testing. *Digital Technical Journal*, 6, 125-133.

Gould, J.D., & Lewis, C. (1985) Designing for usability-key principles and what designers think of them. *Communications of the ACM*, 28 (3), 300-311.

Hallnäs L., & Redström J. (2002). From use to presence: On the expressions and aesthetics of everyday computational things. *Transactions on Computer-Human Interaction*, 9 (2), 106-124.

Han, S.H., Kwang, J.K., Myung, H.Y., Hong, S.W., & Kim, J. (2004) Identifying mobile phone design features critical to user satisfaction. *Hum. Fact. and Ergon. in Manufacturing*, 14 (1), 15-29.

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*. North Holland Press.

Hartmann, J., Sutcliffe, A., & De Angeli, A. (2008) Towards a theory of user judgment of aesthetics and user interface quality. *Transactions on Computer-Human Interaction*, 15 (4), article 15.

Hassenzahl, M. (2004) The interplay of beauty, goodness and usability in interactive products. *Human Computer Interaction*, 19 (4), 319-349.

Hassenzahl, M., & Tractinsky, N. (2006) User Experience – A research agenda. *Behaviour & Information Technology*, 25 (2), 91-97.

Hassenzahl, M., & Ullrich, D. (2007) To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, 19 (4), 429–437.

Hekkert, P., Snelders, D., & Wieringen, P.C.W. (2003) Most advanced, yet acceptable: Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology*, 94 (1), 111-124.

Holtzblatt, K. (2011) What Makes Things Cool? Intentional Design for Innovation. *Interactions*, 18 (6), 40-47.

Hornbæk, K., & Law, E.L. (2007) Meta-analysis of correlations among usability measures. In *Proc. CHI 2007*, ACM Press, 617-626.

IDC (2013) IDC Forecasts: Worldwide Tablet Shipments to Surpass Portable PC Shipments in 2013, Total PC Shipments in 2015. Available online: http://www.idc.com/getdoc.jsp?containerId= prUS24129713

Işıklar, G., & Büyüközkan, G. (2007) Using a multi-criteria decision making approach to evaluate mobile phone alternatives. *Computer Standards & Interfaces*, 29 (2), 265-274.

ISO 9421-210 (2010) Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland.

ISO 9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability. International Standardization Organization (ISO). Switzerland.

ITU Report (2014). The World in 2014: ICT Facts and Figures. Available online: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

Jackson, S., & Marsh, H. (1996) Development and validation of a scale to measure optimal experience: the Flow State Scale. *Journal of Sport and Exercise Psychology*, 18, 17-35.

Jaspers, M.W.M. (2009) A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *International Journal of Medical Informatics*, 78 (5), 340-353.

Jones, M., Buchanan, G., & Thimbleby, H. (2003) Improving web search on small screen devices. *Interacting with Computers*, 15 (4), 479-495.

Jordan, P.W. (1999) Pleasure with Products: Human factors for body, mind and soul. In W.S. Green and P.W. Jordan (eds), *Human Factors in product design: Current practise and future trends*, Taylor & Francis, 206-217.

Jordan, P.W. (2000) *Designing Pleasurable Products. An Introduction to the New Human Factors*. Taylor & Francis.

Jung, W.D, Yoon, W.C., & Kim, J.W. (2002) Structured information analysis for human reliability analysis of emergency tasks in nuclear power plants. *Reliability Engineering & System Safety*, 71 (1), 21–32.

Kardes, F.R., Posavac, S.S., & Gronley, M.L. (2004) Consumer inference: A review of processes, bases and judgment inference. *Journal of Consumer Psychology*, 14, 230-256.

Karlson, A.K., Bederson, B.B., & Contreras-Vidal, J.L. (2008) Understanding One-Handed Use of Mobile Devices. In J. Lumsden (ed), *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, IGI Global, 86-101.

Keenan, S.L., Hartson, H.R., Kafura, D.G., & Schulman, R.S. (1999) The Usability Problem taxonomy: A framework for classification and analysis. *Empirical Software Engineering*, 4 (1), 1382-3256.

Kenteris, M., Gavalas, D., & Economou, D. (2009) An innovative mobile electronic tourist guide application. *Personal and Ubiquitous Computing*, 13 (2), 103-118.

Kim, J.H., Aulck, L., Thamsuwan, O., Bartha, M.C., & Johnson, P.W. (2013) The effect of key size of touch screen virtual keyboards on productivity, usability and typing biomechanics. In Proc. *Human Factors and Ergonomics Society Annual Meeting 2013*, 57 (1), 887-891.

Kim, K.J., Sundar, S.S., & Park, E. (2011) The effect of Screen-Size and Communication Modality on Psychology of Mobile Device Users. In *Proc. CHI 2011*, ACM Press, 1207-1212.

Kim, K.J., & Sundar, S.S. (2014) Does screen size matter for smartphones? Utilitarian and hedonic effects of screen size on smartphone adoption. *Cyberpsychology, Behavior and Social Networking*, 17 (7), 466-473.

Kim, T., & Jung, E.S. (2010) Customer-Oriented Design of Mobile Phones for Optimal Controllability. In *Proc. MUE 2010*, IEEE, 1-6.

Kjeldskov, J., & Graham, C. (2003) A review of Mobile HCI research methods. In *Proc. MobileHCI 2003*, Springer, 317-335.

Kruglanski, A.W., & Gigerenzer, G. (2011) Intuitive and deliberate judgments are based on common principles. *Psychological Review*, 118 (1), 97-109.

Kuniavsky, M. (2003). *Observing the User Experience: A Practitioner's Guide to User Research*. Morgan Kaufmann.

Kurniawan, S.H. (2007) Mobile Phone Design for Older Persons. *Interactions*, 14 (4), 24-25.

Kushniruk, A.W., & Patel, V.L. (2004) Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of Biomedical Informatics*, 37 (1), 56–76.

Lai, C.C., & Wu, C.F. (2014) Display and device size effects on the usability of mini-notebooks (netbooks)/ultraportables as small form-factor Mobile PCs. *Applied Ergonomics*, 45, 1106-1115.

Langdon, P.M., Lewis, T., & Clarkson, P.J. (2007) The effects of prior experience on the use of consumer products. *Universal Access in the Information Society*, 6 (2), 179-191.

Lavery, D., Cockton, G. & Atkinson, M. P. (1997) Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, 16 (4), 246-266

Lavie, T., & Tractinsky, N. (2004) Assessing the dimensions of perceived visual aesthetics of web sites. *International Journal of Human Computer Studies*, 60 (3), 269-298.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004) A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95 (4), 489-508.

Lewis, C.H. (1982) Using the "Thinking Aloud" Method In Cognitive Interface Design (Technical report). IBM. RC-9265.

Lewis, C., Polson, P.G., Wharton, C., & Rieman, J. (1990) Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *Proc. CHI 1990*, ACM Press, 235-242.

Lewis, J.R. (1994) Sample sizes for usability studies: additional considerations. *Human Factors*, 36 (2), 368-378.

Lindgaard, G., & Dudek, C. (2002) What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15 (3), 429-452.

Lindgaard, G., Dudek, G., Fernandes, G., & Brown, J. (2006) Attention web designers: you have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, 25 (2), 115-126.

Ling, C., Hwang, W., & Salvendy, G. (2007) A survey of what customers want in a cell phone design. *Behaviour & Information Technology*, 26 (2), 149-163.

Loewy, R. (1951) *Never leave well enough alone*. Johns Hopkins University Press (new edition).

Lund, A. (2001) Measuring Usability with the USE questionnaire. *Usability and User Experience Newsletter of the STC Usability SIG,* 8 (1).

Mack, Z., & Sharples, S. (2009) The importance of usability in product choice: A mobile phone case study. *Ergonomics*, 52 (12), 1514-1528.

Mahlke, S., & Thüring, M. (2007) Studying Antecedents of Emotional Experiences in interactive contexts, In *Proc. CHI 2007*, ACM Press, 915-918.

Mandryk, R.L., Inkpen, K.M., & Calvert, T.W. (2006) Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*, 25 (2), 141 – 158.

Mayhew, D. (1999) *The Usability Engineering Lifecycle: A Practitioner's Handbook*. Morgan Kaufmann.

McCrickard, D.S., Doswell, F., Barksdale, J., & Piggott, D. (2012) Understanding Cool in Computing for African-American Youth. *ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012*.

McLellan, S., Muddimer, A. & Peres, S.C. (2012) The effect of experience on System Usability Scale Ratings. *Journal of Usability Studies*, 7 (2), 56-67.

Mugge, R., & Schoormans, J.P.L. (2012) Product design and apparent usability. The influence of novelty in product appearance. *Applied Ergonomics*, 43, 1081-1088.

Nielsen, J., & Molich, R. (1990) Heuristic evaluation of user interfaces. In *Proc. CHI 1990*, ACM Press, 249-156.

Nielsen, J. (1993) *Usability engineering*. Morgan Kaufmann.

Nielsen, J., & Landauer, T.K. (1993) A mathematical model of the finding of usability problems. In *Proc. of INTERACT 1993 and CHI 1993*, ACM Press, 206-213.

Nielsen, J. (2001) 113 Design Guidelines for Homepage Usability. Available online: http://www.nngroup.com/articles/113-design-guidelines-homepage-usability/

Nielsen, J. (2012) User Satisfaction vs. Performance Metrics. Available online: http://www.nngroup.com/articles/satisfaction-vs-performance-metrics/

Norman, D., Miller, J., & Henderson, A. (1995) What You See, Some of What's in the Future, And How We Go About Doing It: HI at Apple Computer. In *Proc. CHI 1995*, ACM Press, p.155.

Norman, D. (2002) Emotion & Design: Attractive Things Work Better. *Interactions*, 9 (4), 36-42.

O'Donnell, K.A., & Wardlow, D.L. (2000) A Theory on the origins of Coolness. *Advances in Consumer Research*, 27 (1), 13-18.

Palangue, P., Paternó, F., & Wright, P. (1998) Designing user interfaces for safety critical systems. In *Proc. CHI 1998*, ACM Press, p. 200.

Park, Y.S., & Han, S.H. (2010) Touch key design for one-handed thump interaction with a mobile phone: effects of touch key size and touch key location. *International Journal of Industrial Ergonomics*, 40 (1), 68-76.

Pereira, A., Miller, T., Huang, Y.M., Odell, D., & Rempel, D. (2013) Holding a tablet computer with one hand: effect of tablet design features on biomechanics and subjective usability among users with small hands. *Ergonomics*, 56 (9), 1363-1375.

Peter, C., & Herbon, A. (2006) Emotion representation and physiology assignments in digital systems. *Interacting with Computers*, 18 (2), 139-170.

Phau, I., &Lau, K. (2001) Brand personality and consumer self-expression: Single or dual carriageway? *The Journal of Brand Management,* 8 (6), 428-444.

Picard, R.W. (1997) *Affective Computing*. MIT Press.

Pinelle, D., Wong, N., & Stach, T. (2008) Heuristic evaluation for games: usability principles for video game design. In *Proc. CHI 2008*, ACM Press, 1453-1462.

Polys, N.F., Kim, S., & Bowman, D.A. (2005) Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments. In *Proc. VRST 2005*, ACM Press, 46-55.

Pountain, D., & Robbins, D. (2000) *Cool Rules. Anatomy of an Attitude*. Reaction Books.

Quinn, J.M., & Tran, T.Q. (2010). Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In *Proc. CHI 2010*, ACM Press, 353-362.

Raita, E., & Oulasvirta, A. (2010) Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers*, 23 (4), 363-371.

Read, J., et al. (2011) Understanding and Designing Cool Technologies for Teenagers. *Ext. Abstracts CHI '11*, ACM Press, 1567-1572.

Read, C.R., Horton, M., & Fitton, D. (2012) Being Cool-Getting Persona-l. *ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012.*

Rogers, Y., Sharp, H., & Preece, J. (2011) *Interaction Design: beyond human-computer interaction* (3rd Edition). John Wiley & Sons.

Rondeau, D.B. (2005) For mobile applications, branding is experience. *Communications of the ACM*, 48 (7), 61-66.

Rosson, M.B., & Carroll, J.M. (2002). *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. Morgan Kaufmann.

Rubin, J., Chisnell, D., & Spool, J. (2008) *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests (2nd edition)*. Wiley.

Ruuska-Kalliokulju, S., Schneider-Hufschmidt, M., Väänänen-Vainio-Mattila, K., & Von Niman, B. (2000) Shaping the Future of Mobile Devices: Results of the Workshop on Future Mobile Device User Interfaces at CHI 2000, *SiGCHI Bulletin*, p.16.

Safa, M. (2013) Factors affecting consumer buying behavior of mobile phone devices. *Mediterranean Journal of Social Sciences*, 4 (12), 103-112.

Sauro, J., & Kindlund, E. (2005) A method to standardize usability metrics into a single score. In *Proc. CHI 2005*, ACM Press, 401-409.

Sauro, J. (2010) *A Practical Guide to Measuring Usability: 72 Answers to the Most Common Questions about Quantifying the Usability of Websites and Software*. CreateSpace Independent Publishing Platform.

Scott, B., & Nass, C. (2002) Emotion in human-computer interaction. In A. Sears, and J.A. Jacko (eds.) *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, Lawrence Erlbaum, 81-96.

Seva, R.R., Gosiaco, K.G.T., Santos, M.C., & Pangilinan, D.M.L. (2011) Product design enhancement using apparent usability and affective quality. *Applied Ergonomics*, 42 (3), 511-517.

Solomon, M., Bamossy, G., Askegaard, S., & Hogg, M.K. (2010) *Consumer Behaviour: A EuropeanPerspective*. Pearson Education.

Sonderegger, A., & Sauer, J. (2010) The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics*, 41 (3), 403-410.

Sonderegger, A., Zbinden, G., Uebelbacher, A., and Sauer, J. (2012) The influence of product aesthetics and usability over the course of time: a longitudinal field experiment. *Ergonomics*, 55 (7), 713-730.

Southgate, N. (2003) Coolhunting, account planning and the ancient cool of Aristotle. *Marketing Intelligence & Planning*, 21 (7), 453-461.

Staiano, J., Menéndez, M., Battochi, A., De Angeli, A., & Sebe, N. (2012) UX_Mate: From Facial Expressions to UX Evaluation. In *Proc DIS 2012*, ACM Press, 741-750.

Sundar, S.S., Tamul, D.J., & Wu, M. (2014) Capturing "cool": Measures for assessing coolness of technological products. *International Journal of Human-Computer Studies*, 72 (2), 169-180.

Sweeney, S., & Crestani, F. (2006) Effective search results summary size and device screen size: Is there a relationship? *Information Processing and Management*, 42 (4), 1056-1074.

Tractinsky, N., Katz, A.S., & Ikar, D. (2000) What is beautiful is usable. *Interacting with Computers,* 13 (2), 127-145.

Tractinsky, N., Cockhavi, A., Kirschenbaum, M., & Sharfi, T. (2006) Evaluating the consistency of immediate aesthetic perceptions of web pages. *International Journal of Human-Computer Studies,* 64 (11), 1071-1083.

Trudeau, M.B., Udtamadilok, T., Kalson, A.K., & Dennerlein, J.T. (2012) Thump motor performance varies by movement orientation, direction, and device size during single-handed mobile phone use. *Human Factors*, 54 (1), 52-59.

Trudeau, M.B., Catalano,P.J., Jindrich, D.L., & Dennerlein, J.T. (2013) Tablet Keyboard Configuration Affects Performance, Discomfort and Task Difficulty for Thumb Typing in a Two-Handed Grip. *PLoS One*, 8 (6), 1-9.

Tuch, A.N., Roth, S.P., Hornbæk, K., Opwis, K., & Bargas-Avila, J.A. (2012) Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior*, 28 (5), 1596–1607.

Tullis, T., & Albert, B. (2008) *Measuring the User Experience: collecting, analyzing and presenting usability metrics*. Elsevier.

Van Schaik, P., Hassenzahl, M., & Ling, J. (2012) User-Experience from an Inference Perspective. *TOCHI,* 19 (2), Article No. 11.

Vermeeren, A.P.O.S. et al. (2010) User Experience Evaluation Methods: Current State and Development Needs. In *Proc. NordiCHI 2010*, ACM Press, 521-530.

Yun, M.H., Han, S.H., Hong, S.W., & Kim, J. (2010) Incorporating user satisfaction into the look-and-feel of mobile phone design. *Ergonomics*, 46 (13-14), 1423-1440.

Warren, C., & Campbell, M.C. (2014) What makes things cool? How Autonomy Influences Perceived Coolness. *Journal of Consumer Research*, 41 (2), 543-563.

Wixon, D., Holtzblatt, K., & Knox, S. (1990) Contextual Design: An Emergent View of System Design. In *Proc. CHI 1990*, ACM Press, 329-336.

Wynekoop, J.L., & Conger, S.A. (1990) A review of computer aided software engineering research methods. In *Proc. Of the IFIP TC8 WG 8.2 Working Conference on the Information Systems Research Arena of the 90's*.

Ziefle, M. (2002) The influence of user experience and phone complexity on performance, ease of use and learnability of different mobile phones. *Behav. & Inform. Technology*, 21 (5), 303-311.

7 PAPER CONTRIBUTIONS

7.1 PAPER 1: STUDYING THE EFFECT OF PERCEIVED HEDONIC MOBILE DEVICE QUALITY ON USER EXPERIENCE EVALUATIONS OF MOBILE APPLICATIONS.

ABSTRACT

When people interact with digital artifacts they perceive their *pragmatic* and *hedonic qualities*. In the case of interacting with mobile devices and applications users seek utility as they try to satisfy certain needs, but at the same time they have certain feelings and emotions when, for example, they feel attached to their personal phone and/or trust its brand. Due to this strong relation between users and mobile devices a significant problem occurs when researchers want to evaluate the user experience of a mobile application in laboratory settings: the selection of an appropriate mobile device. Towards this end, this paper aims to unveil the effect of perceived *hedonic quality* of a mobile device on the user experience evaluation results of an application. Our results show that the perceived *hedonic quality* of a mobile device significantly affects the perceived *pragmatic* quality of the application, but not the *hedonic* one.

1 INTRODUCTION

The diffusion of mobile technologies has been considerable over the past years. Today, people constantly use their personal mobile devices and take advantage of the hundreds existing applications to accomplish everyday tasks, communicate, play, socialize, etc. During the last decade, the mobile phone has rapidly transformed from a single purpose device, mainly used for making phone calls, to a powerful digital artifact. In addition, since interaction with mobile devices becomes more widespread, researchers' perception regarding the user is evolving. During the 70's the user was understood as a gear in a rational machine, during the 80's as a source of error, during the 90's as a social actor, and now as a consumer (Kuuti, 2001). When such users/consumers want to buy a mobile device they face a vast selection of candidate devices that offer more, or less similar functionality and have differences in their style, design, brand, operating system, etc. Acting as consumers many users don't treat the device they have intentionally selected just as a tool that helps them to deal with various tasks, thus focus only on utility, but as something more (Coates, 2002). For them the mobile device is a *product* that was intentionally chosen among others, it has significant *value* and *meaning* (Hallnäs and Redström, 2002) and it projects elements of their personality (Jordan, 1997; Phau and Lau, 2001). We may also find consumers that move a step further and treat their devices as something so important that needs to be personalized and made more attractive according to their personal taste and style. Personalization can occur at the software level, affecting how usable a product is perceived (Tossell et al., 2012), or at the hardware level and therefore we may find: users that customize their devices in various ways (Moggridge, 2007; Marathe and

Sundar, 2011), special editions of a particular device, or even large fashion industries involved in the production of a device (for example Motorola Razr V3i D&G handset).

At the same time, researchers and practitioners that are involved in developing new mobile applications face the challenge of properly evaluating their prototypes by measuring the users' experience with them (Hassenzal and Tractinsky, 2006). There is a variety of user experience evaluation methods available (Vermeeren et al., 2010), but especially in the case of evaluating mobile applications in laboratory settings researchers and practitioners need to decide about the:

- Reasons behind conducting an evaluation. Thus, they need to decide on how to link the evaluation and the design process and how to use the findings in order to improve their prototype (Hornbæk, 2010).
- Participants: how many participants are going to be involved, what kind of previous experience with the application they should have, what kind of general experiences might affect their perception for the application (e.g. internet usage).
- Evaluation instruments: what needs to be measured (usability, pleasure, user satisfaction, etc.) and how it is going to be measured (direct observations, task completion times and rates, questionnaires, etc.).
- Tasks: select between *goal mode* and *action mode* (Hassenzahl, 2003; van Schaik et al., 2012). In goal mode participants are provided with specific tasks and they try to be as effective and efficient as possible, since the fulfillment of the goal becomes the core of their interaction. On the contrary, in action mode the action itself is to the fore and using the mobile application is at the center of their interaction.
- Mobile devices: which mobile devices are going to facilitate the mobile application during the laboratory evaluations.

If we focus on the decision for the selection of an appropriate evaluation instrument, then we observe that user experience can be measured through a variety of questionnaire constructs such as hedonic and pragmatic qualities, fun, flow, enchantment, etc. Despite the fact that there is a significant amount of research directed towards understanding these constructs, there still are challenges that HCI needs to overcome. The most important one is derived from the fact that we do not know if these constructs are enough to describe user experience because the relations among them have not been studied in detail (Bargas-Avila and Hornbæk, 2011). Thus, the challenge for researchers and practitioners, when they want to make a laboratory evaluation with questionnaires, is to decide which of the user experience constructs they will use in order to evaluate their applications. Furthermore, in relation to the selection of a mobile device for the evaluation, an important challenge is that practitioners and researchers do not know the "complete" digital artefact of their users. If for example they are developing an Android™ application, then they know the software part of the artefact, but they have to deal with a huge variety at the hardware part. The reason for this variety is that there are many Android™ manufacturers that produce devices with different characteristics, for example different screen sizes, weight, colours, materials, etc. These characteristics, which for some researchers may seem of secondary importance in the context of mobile user experience evaluations, are treated as very significant from potential mobile phone consumers (Ling et al., 2007) and can determine the commercial success, or not of a specific device (Kim et al., 2012).

2 RELATED WORK

In this paper we will focus on the cases where the user experience of a mobile application is measured through questionnaires in laboratory settings and the participants first interact with the application and then they are asked to formulate a judgment. There are two different approaches on how this judgment is shaped (van Schaik et al., 2012). The first approach is through *induction* (Kardes et al., 2004) and it implies that the participants are building an overall assessment by carefully considering, analyzing and weighting all the relevant attributes (for example usability, aesthetics, functionality, etc.). The second approach, which is more recent one and supported by a large amount of empirical evidence, proposes that participants are using simple cognitive rules and *infer* a judgment based on the availability of information (Kruglanski and Cigerenzer, 2011). An example from the HCI domain is that participants infer that a product will be usable because it is appealing.

On the other hand, when users interact with a digital artifact their judgment is affected by *integral* and *incidental affect* (Hazzenzahl and Monk, 2010; Hassenzahl et al., 2010; van Schaik et al., 2012). *Integral affect* are the feelings produced and attributed to the artifact and *incidental affect* are the feelings produced by circumstances. These feelings contribute to the formulation of judgments regarding the perceived *hedonic* and *pragmatic quality* of the artifact (Hassenzahl, 2004). *Pragmatic quality* (i.e. perceived usability) refers to a judgment in relation to the possibility the artifact to support participant's "do-goals" and *hedonic quality* is a judgment to the artifact's ability to facilitate pleasure while using it (Hassenzahl et al., 2010). These judgments are also affected by the availability of information regarding the artifact's features. If for example users own an artifact then they know most of the artifact's features and thus *integral affect* is highly influencing their judgments. On the other hand, in the context of evaluating a mobile application in laboratory settings the participants are not familiar with all the application's features as they do not have enough time to fully analyze it, and as a result they *infer* judgments for the *hedonic* and *pragmatic* quality of the application based either on previous experience, or specific artifact features that are easy to perceive (such as the attractiveness of the application).

In the context of mobile evaluations the effect of specific features to user experience have been studied in detail. For example, in relation to perceived *pragmatic quality* of a mobile device we may find studies that compare the performance of various input methods (for example, types of keyboards: Clarkson et al., 2005). On the other hand regarding *hedonic quality* we may find evidence that attractive things work better (Norman, 2004) and studies that specifically investigate the notion of attractiveness in relation to perceived usability (Tractinsky et al., 2000; Chawda et al., 2005). In line to these findings are also recent studies that unveiled the effect of mobile phone's attractiveness on effectiveness and efficiency (Quinn and Tran, 2010; Sauer and Sonderegger, 2010). Attractiveness has also been studied in relation to the first impression and the impression after usability testing, both in the context of websites (Lindgaard et al., 2006; Lindgaard et al., 2011) and mobile computing (Sauer and Sonderegger, 2010), and it showed that the judgment regarding the attractiveness of a digital artifact is mostly shaped by the first impression. Additionally, Ling et al. (2007) examined the impact of specific mobile phone characteristics to user satisfaction, showing that users' general attitude towards a mobile device can be explained by their preference towards specific characteristics, such as physical appearance and body color. Finally, many studies focused on the effect of brand during usability testing (Bolchini et al., 2009; De Angeli et al., 2009).

3 AIM OF THIS STUDY

In this paper we focus on the evaluations of mobile applications in laboratory settings. In these cases the users' experience is influenced both by the mobile device and the mobile application. Furthermore, since in such evaluations participants are asked to focus on the application and formulate a judgment about it, the mobile device they use is treated as the means to conduct the task(s) they have to accomplish. As a result, in laboratory evaluations the perceived *hedonic* and *pragmatic* quality of the device contributes to participant's *incidental affect* and the perceived *hedonic* and *pragmatic* quality of the application to participant's *integral affect*.

In order to see how the HCI community deals with the *incidental affect* created by a mobile device in laboratory evaluations, we reviewed relevant papers in two CHI conferences (2010 and 2011). Relevant were the papers that included the words 'mobile device' or 'smartphone' or 'PDA' and a mobile application was involved. In about 100 papers researchers: a) simply mention the model of the used mobile device without arguing why they chose this device (for example Bidwell et al., 2010), b) state that they used a mobile device of a certain operating system without specifying a model (for example Durrant et al., 2011), c) use various mobile devices and sum up the results (for example Quinn and Tran, 2010), d) state that they selected a mobile device based on specific needs on hardware and/or software (for example Costanza et al., 2010), or e) do not clearly present any details about the device (for example McLoughlin and Ciolfi, 2011). We identified just one case where the researchers tried to minimize the effect of mobile device hedonic features by using custom made mobile device casings to avoid influencing their participants (Chong and Gellersen, 2011).

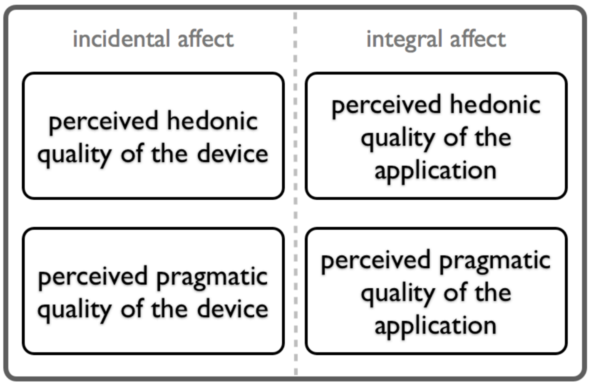


Figure 1. Perceived hedonic and pragmatic quality of a mobile device and a mobile application in the context of user experience laboratory evaluations.

Consequently, in the vast majority of these papers researchers do not present a clear argumentation on why they selected a specific device and they also tend to ignore the possible effect of the perceived *hedonic* and *pragmatic* device quality (incidental affect) on the evaluation results of their mobile applications. We will illustrate why we believe this approach can be misleading through an analogy: Treating the perceived *hedonic* quality of a mobile device as a parameter that has minor effect on user experience is like assuming that drivers of different cars will experience the same road in a similar way, if their cars facilitate the same perceived *pragmatic* quality (same engines, same horsepower etc.), independently from their cars' perceived *hedonic* quality (for example brand, visual design, etc.).

Therefore, in this paper we are interested in this research question: *"Does the incidental affect created by a mobile device's perceived hedonic quality has an effect on the user experience evaluation results of a mobile application in laboratory settings?"*

The rest of the paper reports from an experiment we conducted in order to answer this research question. Through this experiment we discovered that the perceived *hedonic quality* of a mobile device has a significant effect on users' experience with a mobile application. The rest of the paper is organized as follows. First, we present our method and the experimental conditions. Then, we outline our findings and discuss these findings against related research. Finally, we conclude on our work by presenting our suggestions for the reasons that have created these results.

4 METHOD

In order to answer the research question we initially made two pilot studies to test our experimental setup and then we adopted a between-group experimental design (Lazar et al., 2010) and asked two groups of users (each group interacted with one device) to interact with and then evaluate the same application on two different devices. In the following subsections we present our key decisions.

4.1 Device selection

One key decision in our experiment was the selection of devices. In our selection process we had to make sure both that the two devices would offer as similar perceived *pragmatic* qualities as possible and that the perceived *hedonic* qualities would have clear and distinct differences. We argue that our experimental design shares common ground with studies where *framing* on user judgments is under investigation (for example Hartmann et al., 2008). In such studies positively and negatively framed statements/questions of the same *meaning* are presented to participants and then they are asked to make a judgment. In our case we wanted to provide the same *"meaning"* (same application with the same functionality and utility on both devices) framed by different perceived *hedonic qualities* of the devices. In order to achieve a big distance between the perceived *hedonic qualities* we had to select two devices with clear and notable differences on style, visual design, materials used for the casing and age. We chose these device features as the base to frame our study, by considering previous studies that identified which mobile features users consider as important (Ling et al. 2007). Having these features as a starting point we chose to include to the experiment an iPod Touch™, 2nd generation, A1288 model (Device A), which when the study took place it was recently introduced to the Greek market and a Dell X51v™ (Device B) that even then it was considered an old PDA with somehow outmoded design (Figure 2). This specific device selection allowed us to have a big distance on the perceived *hedonic quality* and thus a clear *framing* for the *incidental affect* created by their *hedonic qualities*.

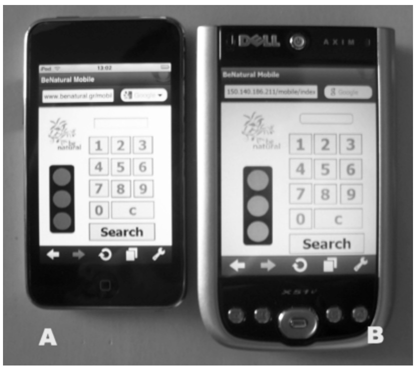


Figure 2. Test application on Devices A and B.

4.2 Application Selection

Any mobile application would have been suitable for our experiment since our purpose was not to evaluate the application per se, but to investigate the effect of perceived *hedonic* mobile device quality on its evaluation results. We chose to use an application called *beNatural* as a test application mainly because it was developed by us and therefore: a) none of the participants had any prior experience with it, and b) we could easily make the necessary software changes. By experimentally controlling for prior experience with the test application we have also eliminated its effect on perceived usability as identified by Quinn and Tran (2010) and McLellan et al. (2012).

In short, *beNatural* (Figure 2) is an application that allows customers to be informed about the environmental impact of a product prior to purchasing it (for example to know if a product is made from recyclable materials). Information about products is contributed to the system's database by the users through a desktop website, following the Wikipedia example. Inquiries about a product are conducted through the mobile application by entering the product's barcode through a numerical keypad (Figure 2). The system's response about a particular product is presented to the users through a traffic light metaphor (Figure 2). Red light indicates that a product should not be bought, yellow light that it could be avoided, green that is suitable to buy it and when all lights are off then there is no information in the database. The responses are personalized according to each user's profile by applying a filtering mechanism, which describing it is out of the scope of this paper.

4.3 Experimental Setup

A series of actions were taken to ensure that both groups were offered similar perceived *pragmatic* qualities for the device and the test application and similar *hedonic quality* for the application. Since we wanted our software to look and feel the same in both devices we installed the Opera Mini™ browser (Figure 2) and we modified the interface of the test application in order to achieve that the size of the application and the size of the on-screen buttons would be as similar as possible in both cases, despite the devices' differences on screen size and resolution. Additionally, we decided that our subjects should interact with the test application in a unified way and thus the group that interacted with Device B was not provided with a stylus and we locked Device A resize capabilities. Consequently, all participants experienced the test application by using their fingers and interacted with a similar interface.

A pilot study was conducted to check if the participants could perceive the big differences in devices' hardware (processor, screen responsiveness, Wi-Fi speed, etc.). Four participants inserted a barcode as fast as they could 15 times, repeatedly on both devices (in total 60 repetitions on each device). For each effort the timer was started when they pressed the first barcode digit and it was stopped when the system responded for a product. A Latin Square design (Edwards, 1951) was adopted in order to avoid first-order and carry-over effects and the barcode they were asked to enter was "1234567890" in order to minimize cognitive load.

All four participants took similar time to accomplish the task in both devices. The average time to insert one barcode and see a response and the average error rate in 60 repetitions were: 10.39sec (SD=0.748) and 1.5errors (SD=0.911) for Device A, and 10.31sec (SD=0.581) and 1.63 errors (SD=0.758) for Device B. The average times were checked for differences between the two devices using T-Tests and there were no statistically significant differences. These minimum differences on insertion-response time and error rate showed that our participants could achieve the same goal at almost the same time, independently from the device they used and that the differences in hardware were not perceived, mainly because the test application was not resource demanding and the task was relatively simple.

In order to have accurate data on the perceived *hedonic quality* differences between the two devices, we asked 15 additional participants to use an online form that contained one image for each device (the devices were depicted from the front and their screens were turned off) and formulate a judgment for the devices' *hedonic* and *pragmatic quality* (AttrakDiff2: Hazzenzahl and Monk, 2010; van Schaik et al., 2012) and *attractiveness* (Quinn and Tran, 2010). The sequence the devices were presented was randomized.

Results showed that *hedonic quality* was perceived on average M=5.16 (SD=0.96) for Device A and M=3.31 (SD=1.45) for Device B, and this difference was significant (t(28) =4.105, p<.001**) according to a two-sample T-test. *Attractiveness* was rated M=3.66 (SD=0.71) for Device A and M=2.43 (SD=0.84) and this difference was also significant (t(28) =-4.303, p<.001**). The difference in the *pragmatic quality* was not significant (t(28)=2.041, p=.052) and it was perceived on average M=4.81(SD=1.39) for Device A and M=3.98(SD=0.73) for Device B.

Consequently, the results from the two pilot studies showed that our experimental setup offered both groups similar *pragmatic quality* and different *hedonic quality* for the device, and similar *pragmatic* and *hedonic qualities* for the application.

4.4 Measures

Two different types of measures were used during the two laboratory evaluation sessions that we conducted next: a demographics questionnaire and an evaluation questionnaire. We created our own version of the demographics questionnaire, but we chose to use existing, established questionnaires for evaluating the participants' experience with the application.

4.4.1 Demographics questionnaire

The demographics questionnaire comprised 10 questions that documented the participants' experience with Devices A and B and general aspects like internet usage, familiarity with mobile devices, etc.

4.4.2 Evaluation questionnaire

The used evaluation questionnaire was created by combining three established questionnaires: USE (Lund, 2001; Tullis and Albert, 2008), AttrakDiff (Hassenzahl, 2004) and "Pleasure while interacting with products" (Jordan, 2000) questionnaires. The produced evaluation instrument contained 58 items in total and it allowed us to obtain a holistic view on how our participants experienced their interaction with the test application.

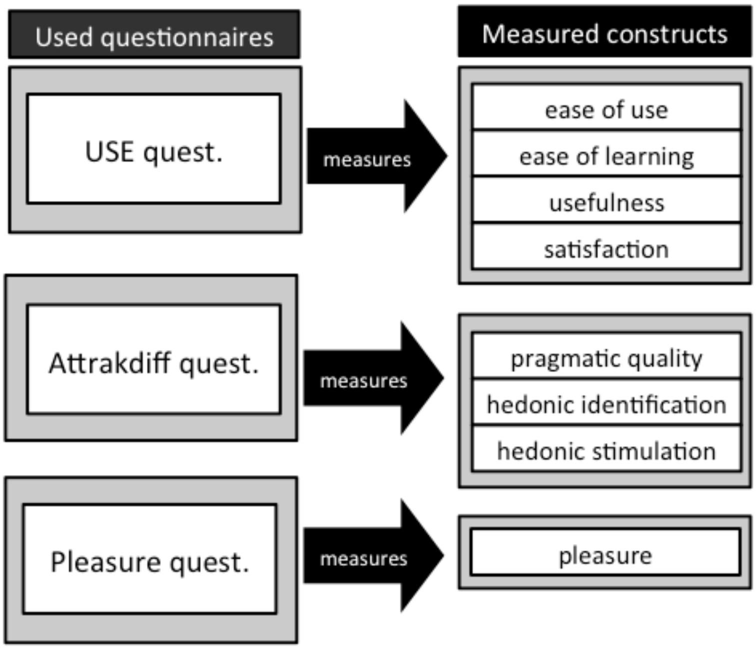


Figure 3. Measured evaluation constructs.

Figure 3 depicts the evaluation constructs measured by the evaluation questionnaire. We point out that some of these constructs are either highly correlated, or even measuring similar aspects (for example *easy of use* and *ease of learning* are highly correlated to *pragmatic quality*). Nevertheless, we chose to include all these constructs in order to have multiple data sources.

The following subsections present in detail the user experience constructs that were measured by the evaluation questionnaire.

4.4.2.1 Perceived Usability

We measured perceived usability with USE questionnaire (Lund, 2001; Tullis and Albert, 2008). USE measures perceived usability through *usefulness* (7 items), *ease of use* (4 items), *ease of learning* (3 items) and *satisfaction* (7 items). Participants were provided with statements and they had to formulate their judgment on a 7-point scale (strongly agree-strongly disagree).

4.4.2.2 Perceived Hedonic and Pragmatic quality

Hedonic and *pragmatic qualities* were measured by AttrakDiff (Hassenzahl, 2004). AttrakDiff (Hassenzahl, 2004) is a questionnaire that presents pairs of opposite adjectives, used in this case on a 7-point scale. *Pragmatic quality* (7 items) is related to perceived usability and it indicates if users can fulfill their goals while using a product. *Hedonic quality* is further decomposed to *hedonic identification* (7 items) and *hedonic stimulation* (7 items). *Hedonic identification* measures if users identify with a product and if it can communicate their personal values. *Hedonic stimulation* indicates how challenging and novel a product is perceived by the users. AttrakDiff contains also two items related to the user's overall impression towards an interactive product: *bad/good* and *ugly/beautiful*.

4.4.2.3 Pleasure

Finally, Jordan's *pleasure* questionnaire (Jordan, 2000) indicates to what extend users would characterise their interaction with a product as pleasurable. *Pleasure* was measured by 14 items, on a 7-point scale (strongly agree-strongly disagree).

5 EXPERIMENT

5.1 Participants

54 additional individuals, 30 men and 24 women, aged 22 to 35 (M=25.1) participated in the experiment. From them, 3 were excluded from data analysis as they partially completed the evaluation questionnaire. All of them were undergraduate and postgraduate students from four University of Patras departments in Greece (Departments of Computer Science, Chemistry, Biology, and Educational Sciences and Early Childhood Education). All of them volunteered for the experiment.

5.2 Procedure

Two weeks prior to the experiment all participants were asked to fill in the online demographics questionnaire. From the demographics questionnaire we found out that none of the subjects had any experience with Device B (Dell X51v™) and six of them had limited experience with Device A (iPod Touch™) and/or similar products (for example iPhone™). These six participants were equally distributed to the two groups and all the rest were assigned randomly. 27 participants interacted with Device A and 24 interacted with Device B.

The experiment evolved in three phases, all conducted in the same room in a usability laboratory and each subject participated individually. The initial phase was a five-minute general introduction to the test application, conducted by the same researcher who made sure that all participants heard the same narrative. In short, participants were informed about the basic services offered by the test application, the type of tasks that can be performed on its mobile and the desktop parts and how users can benefit from the test application in general. Since we wanted to focus on the application and not on achieving specific goals we decided to opt for the *action mode* (Hassenzahl, 2003; van Schaik et al., 2012) and we did not provide any tasks to the participants. Therefore, after the introduction each participant interacted freely with the test application for fifteen minutes.

In both groups the application was running in full screen mode to prevent the participants from interacting with the operating system. In order to mimic real world conditions seven products that are common in a supermarket were placed on a shelf next to them. Prior to each session the researcher made sure that five products were always present in the test application's database and two did not exist in order to give participants the possibility to insert a product to the system if they decided to do so. The selection of the two non-existing products was made randomly. All participants inquired for information for at least four products. Help was provided only when asked and only two participants asked for it. The average error rate in inserting barcodes was: 1.16 errors (SD=0.90) for Device A and 1.28 errors (SD=0.94) for Device B, without any significant effects. In comparison to the pilot study the participants made less errors (probably due to the fact that they did not have to be fast) and their average error rate was similar in both groups. None of the participants used the mobile devices' hardware buttons, nor interacted with the operating system.

At the final phase of the experiment participants were instructed to formulate a judgment about their experience with the test application by filling in the evaluation questionnaire.

6 DATA ANALYSIS

As a first step in the data analysis all participants' ratings on the three questionnaires were combined and used as variables in an exploratory factor analysis. The aim was to reconcile for overlapping attributes, considering that the three questionnaires that were used assessed similar constructs (e.g. pragmatic quality and perceived usability), and obtain a reduced number of factors that would hopefully reflect new perceived *hedonic* and *pragmatic* qualities for the application.

After conducting the exploratory factor analysis, the second step of data analysis was to analyse each questionnaire individually, and we tested both groups for statistically significant differences using two sample T-Tests ($p<.05$). T-Tests were used since we had one independent variable with two groups (devices A and B) and many dependent ones (the 8 measured constructs). A possible alternative would have been to conduct MANOVA, but this would mean loss of statistical power since DV's are highly correlated to each other. Another approach would have been to adopt a more conservative alpha (p <.01), but this could lead to type II error. Therefore we decided to perform T-Tests (Saville, 1990) and report the t, p and α values. Furthermore, for each evaluation construct we conducted a two-way analysis of variance (ANOVA) with device as between groups factor and age and gender as within subjects factor. No significant interaction effects were found. Apart from the actual evaluation questionnaire participants were provided with an area for comments and none of them reported anything related to the mobile device.

7 RESULTS

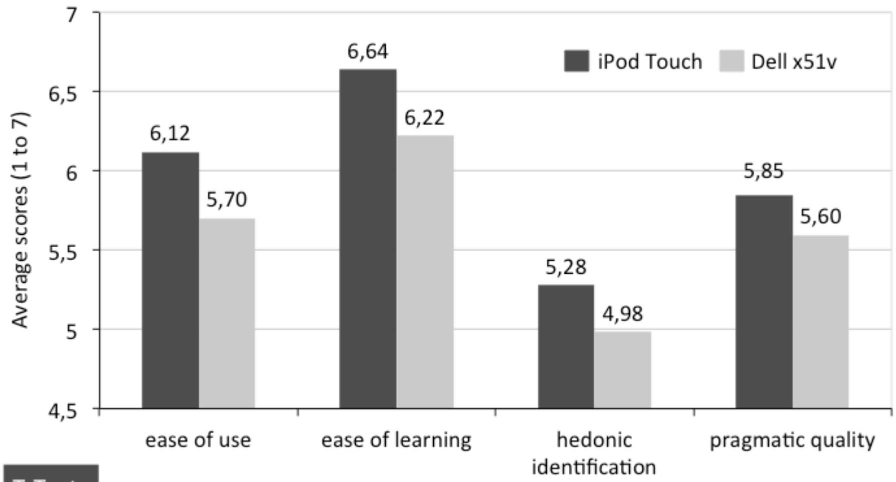
7.1 Exploratory Factor analysis

Our dataset did not allow for an item level response factor analytic procedure as the accumulated number of single items (58) exceeded the number of total observations (51). Therefore the analysis was conducted on the eight evaluation constructs (*ease of use, ease of learning, satisfaction, usefulness, hedonic identification, hedonic stimulation, pragmatic quality* and *pleasure*). The Monte Carlo simulations with data permutations that were conducted on this dataset indicated a two-factor structure as the most appropriate. Examination of factor loadings in the oblique rotated structure revealed that the two emerging factors resembled the pragmatic and hedonic dimensions. *Ease of use* (.951), *satisfaction* (.674), *ease of learning* (.662) and usefulness (.306) contributed on the first factor, while *hedonic identification* (.800), *hedonic stimulation* (.643) and *pleasure* (.609) contributed on the second one. *Pragmatic quality* contributed to both factors (.596 and .735 respectively).

Despite the fact that *pragmatic quality* contributed to both factors, due to its high correlation with perceived usability, we interpreted the two emerging factors as *perceived usability/pragmatic* and *hedonic quality* of the application and we named them as *new_pragmatic* and *new_hedonic* respectively. The factor scores of these qualities were used in a one-way ANOVA analysis with mobile device as a between subject factor. The results of this analysis showed a significant main effect of the mobile device on the *new_pragmatic quality* ($F(1,49) = 6.724$, $p = .013$*). Contrary, a statistical significant main effect of the mobile device could not be shown for the *new_hedonic quality* ($F(1,49) = 1.947$, $p = .169$). On average, iPod Touch™ participants perceived the *new_hedonic quality* of the test application quite high (M=0.17, SD= 1.02), whereas X51v™ participants rated this lower (M=-0.19, SD=0.76). The same was the case for the *new_pragmatic quality* for iPod Touch™ participants (M=0.32, SD=0.88) and X51v™ participants (M=-0.36, SD=0.99).

7.2 Examining each questionnaire individually

Figure 4 depicts the average scores from users' answers on four constructs of particular interest and the results of the two sample T-tests (t, p and α values). All the remaining measured constructs did not have any statistically significant differences. In the following subsections we present our findings for each questionnaire.



T-Tests				
T(49)	2.254	3.060	1.861	1.993
p	.029*	.004*	.069	.052
α	.79	.623	.646	.624

Figure 4. Average scores and results from two sample T-tests for four evaluation constructs with significant, or close to significant differences.

7.2.1 Perceived usability of the mobile application (USE questionnaire)

The perceived usability of the application was measured through *ease of use* (Cronbach α=.79), *ease of learning* (α=.623), *usefulness* (α=.611) and *satisfaction* (α=.731) on a scale from 1 to 7. We found that the differences the two devices had on perceived *hedonic quality* affected significantly some of the usability related constructs of the application. On average, iPod Touch™ participants perceived *ease of use* of the test application quite high (M=6.12, SD= 0.63), whereas X51v™ participants rated this lower (M=5.70, SD=0.70) and this difference, according to a two-sample T-test was significant (t(49) =2.254, p=.029*). We further identified a difference for *ease of learning,* where iPod Touch™ users perceived the test application as very *easy to learn* (M=6.64, SD= 0.43) and this was slightly better than X51v™ users (M=6.22, SD=0.54). This difference was also significant (t(49)=3.060, p=.004*).

7.2.2. Perceived hedonic and pragmatic qualities of the mobile application (AttrakDiff questionnaire)

Through Attrakdiff we measured *pragmatic quality* (Cronbach α=.624), *hedonic stimulation* (α=.754) and *hedonic identification* (α=.646). While our results on application's perceived *hedonic identification* and *pragmatic quality* showed higher values for iPod Touch™ participants, we identified no significant differences between the two groups of users even though we observed a trend both on *hedonic identification* ($t(49)=1.861$, $p=.069$) and *pragmatic quality* ($t(49)=1.993$, $p=.052$). iPod Touch™ users rated *hedonic identification* on average M=5.28 (SD=0.62) and X51v™ users M=4.98 (SD=0.51) and *pragmatic quality* was rated on average M=5.85 (SD=0.48) and M=5.60 (SD=0.40) respectively.

7.3.3. Experienced pleasure with the mobile application (Pleasure questionnaire)

We measured *pleasure* (Cronbach α=.869) using Jordan's questionnaire (Jordan, 2000) and we identified no significant differences between the two groups.

Overall, our results showed that the differences the two devices had on perceived *hedonic quality* affected significantly the perceived *pragmatic quality* of the application, despite the fact the devices offered similar perceived *pragmatic* quality. This result was observed when we examined the three evaluation questionnaires individually, as well as, when we conducted an exploratory factor analysis by combining the questionnaires.

8 DISCUSSION

In relation to previous research work we believe that we moved a step further as we did not focus on studying the effect of the *hedonic* and *pragmatic qualities* of the application (as in Chawda et al., 2005), but we focused on the device. Additionally, we extended the study conducted by Quinn and Tran (2010) by focusing both on *hedonic* and *pragmatic* quality and we eliminated the fact that they used many mobile devices and summed the results. Furthermore, we differentiate from the Sauer and Sonderegger (2010) study, as we chose to maximize the differences on the perceived *hedonic* device quality, by selecting two completely different devices and not two versions of the same phone. Finally, we differentiated from the related papers as we chose to include to our experiment a variety of evaluation questionnaires.

The most important finding of our study is that two user experience evaluations of the same application on two different devices provide significantly different results. In detail, data analysis showed that two groups of users interacting with the same application on devices that facilitate similar perceived *pragmatic* and different perceived *hedonic* qualities have different experiences. We were surprised to find out that the perceived *hedonic quality* of the application was not affected despite the fact that we *framed* the *incidental affect* by choosing two devices with big differences on perceived *hedonic* qualities. On the other hand, we observed a significant effect on the perceived *pragmatic quality* of the application.

We propose that the main reason that the devices influenced only the perceived *pragmatic quality* of the application and not the *hedonic* one is related to the type of the application we used in our experiment. Since the test application was mainly *pragmatic*, as its purpose is to assist users to decide about the environmental impact of a product, then the *incidental affect* that was created by the devices has affected only the user experience constructs that the participants considered as relevant for the application. In other words the *hedonic quality* of the device influenced more the *pragmatic quality* of the application because the participants rated it as more important. We argue that a possible explanation to this situation is the "beauty dilemma", which describes the phenomenon where potential users, even though they consider beauty as important, they discount beauty, visual design, and hedonic aspects, when it comes to the requirements of a new product, in favour of utility and usability (Diefenbach and Hassenzahl, 2009). Therefore we propose that our participants where incidentally influenced by the *hedonic quality* of the device and they transferred this effect only to the *pragmatic quality* of the application, due to "beauty dilemma". Furthermore, the fact that we invited our participants for this experiment to a laboratory might urged them to act more rational (Diefenbach and Hassenzahl, 2009) and thus they unintentionally transferred the perceived *hedonic quality* of the device only to the constructs related to application's perceived usability. Finally, another interesting fact about this finding is that we observed this phenomenon even though we opted for the *action mode* for the evaluation, and in these cases the *pragmatic* constructs do not have as significant impact as in *goal mode* (Hassenzahl and Ullrich, 2007). Therefore it will be interesting to find out if the same phenomenon will be observed when participants are asked to evaluate a leisure application and/or in conditions outside a laboratory. We believe that in this case the *hedonic quality* of the application will be influenced more since the participants will perceive it as more important. Of course, more research data are needed in order to have solid results.

In the following subsections we propose some possible explanations on which of the mobile device features influenced initially the perception regarding the *hedonic quality* of the device which then it influenced the perception of the *pragmatic quality* of the application and we will discuss the implications of our findings on user experience evaluations of mobile applications in laboratory settings, in general.

8.1 The effect of the brand

One of the main factors that shape a consumer's opinion to choose a specific product among a set of possible alternatives is *brand* (Solomon et al., 2010). Furthermore, *brand trust* influences consumers to continue using a specific product, or to buy more items of the same brand (Delgado-Ballester and Munuera-Alemán, 2001). Considering the facts that mobile devices are products and that we chose to include in our experiment two of the most dominant and well-known brands in the area, it is safe to assume that brand might have played a role.

An experience with a brand can be characterized as *indirect*, created by marketing and advertisement, and *direct*, created by the actual interaction with a product (Rondeau, 2005). We believe that our participants had their own preconceptions about the brand of the device they used, even if, as documented by the demographics questionnaire, most of them had only *indirect* experiences with the devices (similar to an opinion that someone has about the brand of a very expensive car, which never drove), and these preconceptions affected the *pragmatic quality* of the application. As shown by Raita and Oulasvirta (2011) the effect of these preconceptions, or expectations towards a mobile device can be rather significant and strongly influence perceived usability. Thus, we believe that our users' attitude towards the brand might have imposed a halo effect (Nisbett and Wilson, 1997) that significantly influenced their judgment on the *pragmatic quality* of the application. Due to this halo effect users transferred positive traits associated with the brand of the Device A to the mobile application in a similar way that most humans consider beautiful as good (Dion et al., 1972). Thus, they transferred the fact that Apple™ is associated with easy to use and learn products to their judgment of the mobile application (*ease of use* and *ease of learning* had statistically significant differences, Figure 4). Our argument is also further enhanced by a previous study, in the context of laptops, which affirmed that brand attachment affects the users' perception regarding perceived usefulness, beauty and pleasure (Tzou and Lu, 2009).

8.2 The effect of attractiveness

Based on studies that show that *attractiveness* can possibly affect perceived usability (Tractinsky et al., 2000; Chawda et al., 2005; Quinn and Tran, 2010; Sauer and Sonderegger, 2010) we argue that the differences the two devices had in relation to attractiveness (pilot study results) influenced our participants, and for this reason the perceived *pragmatic quality* of the test application was significantly higher for the iPod Touch™ users. Thus the *attractiveness* of the device contributed to the perceived *hedonic quality* of the device and this perception was unintentionally transferred to the *pragmatic quality* of the application because it was considered as more important.

8.3 Implications to mobile user experience evaluations in laboratory settings

The results of our experiment showed that users interacting with the same application on two devices with similar perceived *pragmatic* and different *hedonic* qualities had different experiences. Their experiences varied significantly on how they perceived the *pragmatic quality* of the application. Further studies are necessary in order to validate this finding and investigate if this result can be generalized and it is not circumstantial. Nevertheless, this experiment offers a strong indication that the *incidental affect* created by a mobile device during user experience laboratory evaluations, does affect the evaluation results of a mobile application. As a result, this finding may lead to a significant change on the way researchers set up and conduct their user experience laboratory evaluations. Since the device can significantly affect the evaluation results of an application, the selection of an appropriate device becomes a decision as important as selecting the participants, deciding about the evaluation tools, etc.

A classification that categorizes mobile devices' characteristics has emerged from our effort to select two suitable devices for our experiment and it can act as a starting point for researchers and guide them on how to select an appropriate device for their user experience evaluations. We argue that each mobile device owns some *interaction*, *technical* and *physical* characteristics that shape the perceived *pragmatic quality* of the device, and some *intangible* and *physical* characteristics that influence the perceived *hedonic quality* of the device (Fig. 5).

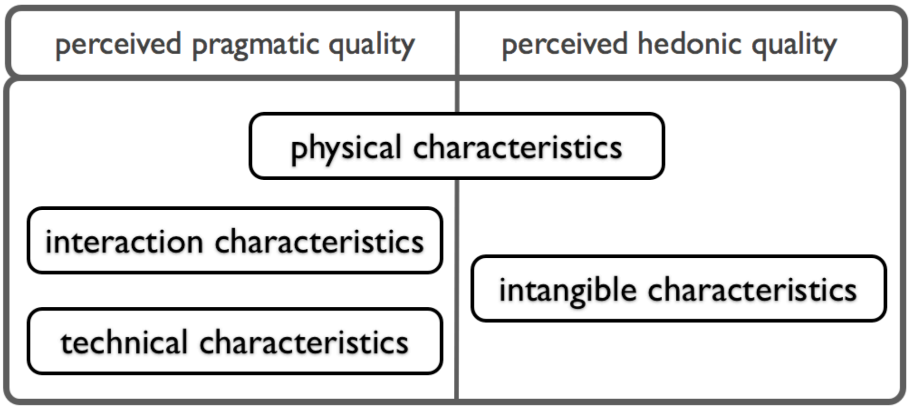


Figure 5. A proposed classification for the mobile device characteristics that influence the perceived hedonic and pragmatic quality of a mobile device.

Technical characteristics are related to the hardware components of the device and can have an effect on the quality of the experience. For example, when users are experiencing delays in a network connection, or in rendering mechanisms they can become frustrated. *Interaction* characteristics are related to the metaphors and the affordances (De Souza, 2005) facilitated by the device. Using a stylus or fingers, the size and the position of software and hardware buttons can be classified as *interaction* characteristics. *Physical* characteristics are the most difficult to categorize since they can influence both *pragmatic* and *hedonic* quality. For example, the weight of a device can belong to both categories, but the used materials, the visual design, etc., are more related to *hedonic* quality. A detailed list of mobile devices' *physical* characteristics is presented by Han et al. (2004) who provided details on various hardware features and associated them with user satisfaction and by a more recent study that ranks these features in relation to how important users think they are (Ling et al., 2007). Finally, *intangible* characteristics are related to the added value and *intrinsic* meaning (McCarthy and Wright, 2004) of an experience. *Intrinsic* meaning refers to the value of an experience that is enjoyed for its own sake (i.e. appreciating a painting). Some of the *intangible* characteristics are brand, trust and attractiveness.

An example that can assist in better realizing this classification is the activity of driving a car. While driving users are perceiving the *pragmatic* quality of their car as they feel the power of the engine (*technical characteristic*), hold the leather steering wheel (*physical characteristic*), or operate the air-condition (*interaction characteristic*). At the same time, they also perceive its *hedonic* quality as they admire the beautiful design of the car's interior (*physical characteristic*), or feel younger and powerful inside their expensive convertible (*intangible characteristic*).

In our case this classification was helpful to provide the same perceived *pragmatic* and different *hedonic* quality for the mobile devices. We made sure that the differences in *technical* characteristics were not perceived, we offered similar *interaction* characteristics and ensured that the *pragmatic physical* characteristics were as similar as possible. Consequently, we believe that this classification can act as a starting point and guide researchers to select the appropriate device(s) when they evaluate the user experience of a mobile application in laboratory settings.

9 CONCLUSIONS

This paper shows that there are worth noting differences on users' judgments for the same application when they interact with devices that have different perceived *hedonic* and similar perceived *pragmatic* qualities, in the context of laboratory user experience evaluations. In our experiment the *incidental affect* that was shaped by the perceived *hedonic* quality of a mobile device, significantly affected the perceived *pragmatic* quality of the mobile application. Further studies are needed with different devices and applications in order to test if this phenomenon is circumstantial due to our experimental setup, or if there is a general *device effect* that significantly affects users' interaction with any mobile application.

Nevertheless, the results presented in this paper are worth reporting since they provide evidence that mobile devices play a more significant role than expected in shaping the evaluation results of a mobile application in laboratory settings. To use the car and road analogy, our results show that different cars do influence the opinion of a driver regarding the road he is driving, even if they facilitate the same *pragmatic* quality. The most important consequence of this finding is that yet another problem is added to every laboratory evaluation of a mobile application that is intended for a variety of devices: the selection of the device. We propose that researchers and practitioners should include to their mobile application evaluations as many candidate devices as possible since the impact of every device can not only significantly affect the users' judgments for a particular application, but is also more related, at least in this case, to the perceived *pragmatic* quality of the application. If it is not possible to include many devices then we propose that our classification of mobile devices characteristics can act as a starting point in order to select the appropriate one(s). Ideally, if researchers want to eliminate the effect of the device then they can, either select a neutral device in terms of design and brand, or provide their subjects with devices with custom made design (as Chong and Gellersen, 2011) to influence less the evaluation results.

Furthermore, the most important finding of this study, besides the fact that the perceived device *hedonic* quality affected significantly applications' perceived *pragmatic quality*, is that it did not have an effect on its perceived *hedonic* quality. A possible explanation for this phenomenon is the fact that we used an application that its function was primarily perceived as *pragmatic*. We propose that it is really important for user experience research to investigate if the same phenomenon will appear when we use *hedonic* applications (for example leisure applications) since we believe that in these cases the perceived *hedonic quality* of an application will be affected more. Additionally, we need to further investigate the differences on users' judgments between *action* and *goal* mode (Hassenzahl and Ullrich, 2007) for these cases. We consider such research directions as important for HCI since we will gain a deeper understanding on the underlying mechanisms that our participants use in order to evaluate an application, we will understand how our experimental setups affect their judgments and this knowledge can contribute significantly on the user experience evaluation methods we apply.

As a future direction we intend to continue experimenting with different combinations of hardware and software in order to check the generalizability of our findings. Additionally, we plan to move one step further and check for possible overlaps and correlations between specific mobile device characteristics (such as device weight and screen size) and user experience constructs.

ACKNOWLEDGEMENTS

We would like to thank all the participants and Rondiri, F., Rontogiannis, N., and Stoica, A. for their effort to develop the test application. The research behind this paper was partly financed by the Danish Research Councils (grant number 09-065143).

REFERENCES

Bargas-Avila, J.A., and Hornbæk, K. Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 2689-2698.

Bidwell, N.J., Reitmaier, T., Marsden, G., and Hansen, S., 2010. Designing with Mobile Digital Storytelling in Rural Africa. In *Proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10)*, 10-15 April 2010 Atlanta. New York: ACM Press, 1593-1602.

Bolchini, D., Garzotto, F., and Sorce, F., 2009. Does Branding need web usability? A Value-Oriented Empirical Study. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II (INTERACT '09)*. Berlin: Springer-Verlag, 652-665.

Chawda, B., Craft, B., Cairns, P., Rüger, S., and Heech, D., 2005. Do attractive things work better? An exploration of search tool visualizations. In *Proceedings of 19th BCS Conference on Human Computer Interaction: Volume 2 (British HCI '05)*, 5-9 September Edinburgh. BCS Press, 46-51.

Chong, M.K. and Gellersen, H, 2011. How Users Associate Wireless Devices. *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 1909-1918.

Clarkson, E., Clawson, J., Lyons, K., and Starner, T., 2005. An empirical study of typing rates on mini-QWERTY keyboards. In *CHI '05 extended abstracts on Human factors in computing systems (CHI EA '05)*, 2-7 April Portland. New York: ACM Press, 1288-1291.

Coates, D., 2002. *Watches tell more than time: Product design, information and the quest for elegance.* New York: McGraw-Hill.

Costanza, E., Panchard, J., Zufferey, G., Nembrini, J., Freudiger, J., Huang, J., and Hubaux, J.P., 2010. SensorTune: a Mobile Auditory Interface for DIY Wireless Sensor Networks. In *Proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10)*, 10-15 April 2010 Atlanta. New York: ACM Press, 2317-2326.

De Angeli, A., Hartmann, J., and Sutcliffe, A., 2009. The effect of brand on the evaluation of websites. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II (INTERACT '09)*, 24-28 August Uppsala. Berlin: Springer-Verlag, 638 – 651.

De Souza, C.S., 2005. *The semiotic engineering of Human-Computer Interaction*. Cambridge, Massachusetts: MIT Press.

Delgado-Ballester, E. and Munuera-Alemán, J.L., 2001. Brand trust in the context of consumer loyalty. *European Journal of Marketing*, 35 (12), 1238 – 1258.

Diefenbach S. and Hassenzahl, M., 2009. The "Beauty Dilemma": Beauty is valued but discounted in product choice. In *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, 4-9 April Boston. New York: ACM Press, 1419-1426.

Dion, K., Berscheid, E. and Walster, E., 1972. What is beautiful is good. *Journal of Personality and Social Psychology,* 24 (3), 285-290.

Durrant, A., Rowland, D., Kirk, D.S., Benford, S., Fischer, J. and McAuley, D., 2011. Automics: Souvenir Generating Photoware for Theme Parks. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 1767-1776.

Edwards, A.L., 1951. Balanced Latin-Square Designs in Psychological Research. *The American Journal of Psychology*, 64 (4), 598-603.

Hallnäs L. and Redström J., 2002. From use to presence: On the expressions and aesthetics of everyday computational things. *Trans. on Computer-Human Interaction*, 9 (2), 106 – 124.

Han, S.H., Kim, K.K., Yun, M.H., Hong, S.W. and Kim, J., 2004. Identifying mobile phone design features critical to user satisfaction. *Human Factors and Ergonomics*, 14 (1), 15-29.

Hartmann, J., De Angeli, A. and Sutcliffe, A., 2008. Framing the User Experience: Information Biases on Website Quality Judgement. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (CHI '08), 5-10 April Florence. ACM Press, 855-864.

Hassenzahl, M., 2003. The Thing and I: Understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A.F. Monk, and P.C. Wright (eds), *Funology: From Usability to Enjoyment*. Dordrecht: Kluwer Academic Press, 31-42.

Hassenzahl, M., 2004. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19 (4), 319–349.

Hassenzahl, M., and Tractinsky, N., 2006. User Experience – a Research Agenda. *Behaviour & Information Technology*, 25 (2), 91-97.

Hassenzahl, M., and Ullrich, D., 2007. To do or not to do. Difference in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, 19 (4), 429-437.

Hassenzahl, M., and Monk, A., 2010. The inference of Perceived Usability from Beauty. *Human-Computer Interaction*, 25 (3), 235-260.

Hassenzahl, M., Diefenbach, S., and Göritz, A., 2010. Needs, Affect and Interactive Products – Facets of User Experience. *Interacting with Computers*, 22 (5), 353-362.

Hornbæk, K., 2010. Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology*, 29 (1), 97-111.

Jordan, P.W., 1997. Products as personalities. In M.A. Hanson, ed. *Contemporary Ergonomics 1997*. London: Taylor & Francis, 449-453.

Jordan, P.W., 2000. *Designing Pleasurable Products. An Introduction to the New Human Factors*. London: Taylor & Francis.

Kardes, F.R., Posavac, S.S., and Gronley, M.L., 2004. Consumer inference: A review of processes, bases and judgment inference. *Journal of Consumer Psychology*, 14, 230-256.

Kim, K., Proctor, R.W. and Salvendy, G., 2012. The relation between usability and product success in cell phones. *Behaviour & Information Technology*, 31 (10), 969-982.

Kruglanski, A.W., and Gigerenzer, G., 2011. Intuitive and deliberate judgments are based on common principles. *Psychological Review*, 118 (1), 97-109.

Kuutti, K., 2001. Hunting for the lost user: From sources of errors to active actors and beyond. In *Proceedings of Cultural Usability Seminar,* 24 April Helsinki. Available from: http://mlab.uiah.fi/culturalusability/papers/Kuutti_paper.html [Assessed 19 July 2012]

Lazar, J., Feng, J.H., and Hochheiser, H., 2010. *Research methods in human-computer interaction*. West Sussex: Wiley & Sons.

Lindgaard, G., Fernandes, G., Dudek, C. and Brown, J. Attention web designers: You have 50 millisec to make a good first impression!. *Behaviour & Information Technology,* 25 (2), 115-126.

Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P., 2011. An Exploration of Relations Between Visual Appeal, Trustworthiness and Perceived Usability of Homepages. ACM Transactions on Computer-Human Interaction,18 (1), Article No. 1.

Ling, C., Hwang, W. and Salventy, G., 2007. A survey of what customers want in cell phone design. *Behaviour & Information Technology*, 26 (2), 149-163.

Lund, A., 2001. Measuring Usability with the USE questionnaire. *Usability and User Experience Newsletter of the STC Usability SIG,* 8 (1).

Marathe, S. and Sundar, S.S., 2011. What Drives Customization? Control Or Identity? In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 781-790.

McCarthy, J. and Wright, P., 2004. *Technology as Experience*. Cambridge, Massachusetts: MIT Press.

McLellan, S., Muddimer, A. and Peres S.C., 2012. The effect of experience on System Usability Scale Ratings. *Journal of Usability Studies*, 7 (2), 56-67.

McLoughlin, M. and Ciolfi, L., 2011. Design Interventions for Open-Air Museums: Applying and Extending the Principles of "Assembly". In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 553-556.

Moggridge, B., 2007. Interaction Design (IDEO). *Innovation Forum*, Potsdam, Germany. Available from: http://interface.fh-potsdam.de/innoforum/english /09_videos.php [Accessed 19 July 2012]

Nisbett, R.E. and Wilson, T.D., 1997. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology,* 35 (4), 250–256.

Norman, D.A., 2004. *Emotional design: Why we love (or hate) everyday things*. NY: Basic Books.

Phau, I. and Lau, K., 2001. Brand personality and consumer self-expression: Single or dual carriageway? *The Journal of Brand Management*, 8 (6), 428-444.

Quinn, J.M. and Tran, T.Q., 2010. Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In *Proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10)*, 10-15 April 2010 Atlanta. New York: ACM Press, 353-362.

Raita, E. and Oulasvirta, A., 2011. Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers*, 23 (4), 363-371.

Rondeau, D.B., 2005. For mobile applications, branding is experience. *Communications of the ACM*, 48 (7), 61-66.

Sauer, J. and Sonderegger, A., 2010. The influence of product aesthetics and user state in usability testing. *Behaviour & Information Technology*, 30 (6), 787-796.

Saville, D.J., 1990. Multiple Comparison Procedures: The Practical Solution. *The American Statistician*, 44, 174-180.

Solomon, M., Bamossy, G., Askegaard, S. and Hogg, M.K., 2010. *Consumer Behaviour: A European Perspective*. Essex: Pearson Education.

Tossell, C.C., Kortum, P., Shepard, C., Rahmati, A. and Zhong, L., 2012. An empirical analysis of smartphone personalisation: measurement and user variability. *Behaviour & Information Technology*, 31 (10), 995-1010.

Tractinsky, N., Katz, A.S. and Ikar, D., 2000. What is beautiful is usable. *Interacting with Computers,* 13 (2), 127-145.

Tullis, T. and Albert, B., 2008. *Measuring the User Experience: collecting, analysing and presenting usability metrics*. Burlington: Morgan Kaufmann.

Tzou, R.C. and Lu, H.P., 2009. Exploring the emotional, aesthetic, and ergonomic facets of innovative product on fashion technology acceptance model. *Behaviour & Information Technology*, 28 (4), 311-322.

Van Schaik, P., Hassenzahl, M., and Ling, J., 2012. User experience from an Inference Perspective. *ACM Transactions on Human-Computer Interaction*, 19 (2), 1-25.

Vermeeren, A.P.O.S., Law, E.L. Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K., 2010. User experience evaluation methods: Current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NordiCHI '10)*, 16-20 October 2010, Reykjavik, Iceland. New York: ACM Press, 521-530.

7.2 PAPER 2: DOES SIZE MATTER? INVESTIGATING THE IMPACT OF MOBILE PHONE SCREEN SIZE ON USERS' PERCEIVED USABILITY, EFFECTIVENESS AND EFFICIENCY.

ABSTRACT

Given the wide adoption of smartphones, an interesting debate is taking place regarding their optimal screen size and specifically whether possible portability issues counterbalance the obvious benefits of a larger screen. Moreover, the lack of scientific evidence about the concrete impact of mobile phones' screen size on usability raises questions both to practitioners and researchers. In this paper, we investigate the impact of a mobile phone's screen size on users' effectiveness, efficiency and perceived usability as measured using System Usability Scale (SUS). An experiment was conducted with 60 participants, which interacted with the same information seeking application on three different devices of the same brand that differed on their screen size. A significant effect of screen size on efficiency was derived, leading to an important finding that users who interact with larger than 4.3in screens are more efficient during information seeking tasks.

1 INTRODUCTION

Over the recent years mobile phones have been transformed from simple devices used just for calls and text messaging to powerful personal computing artifacts. Nowadays, activities such as web browsing, document processing, social networking, media reproduction and gaming constitute the typical usages of a mobile phone. Moreover, traditional mobile input paradigms, such as using a reduced set of keys, or a stylus are fading out in favor of touch interaction. Consequently, in relation to interaction, a modern mobile phone is characterized by its screen, a few physical buttons and a casing.

While users interact with their personal mobile phone and use mobile applications they are affected by various parameters such as the attractiveness of the device, the attractiveness of the applications, the brand of the device, their prior experience, etc. Even though there is a significant research effort directed on discovering the effect that all these parameters have on user experience, surprisingly there is a little volume of research in investigating if users are affected by specific mobile device characteristics (for example screen size, position and size of the physical buttons, materials used for the casing, etc.). Such a research goal is of significant importance for the community given the wide range of mobile phones' characteristics and the fact that we know very little on how they may shape and/or alter the actual user experience with a mobile application.

This paper is an initial step in the process of understanding mobile device characteristics and we chose to focus on the effect of screen size. Towards this end an experiment was designed and conducted to measure the effect of screen size on usability metrics (effectiveness and efficiency) and perceived usability, while interacting with a test application on 3 devices with representative screen sizes. We believe that our results will help designers and practitioners in better understanding how screen size affects their subjects, when they design and evaluate mobile applications.

The rest of the paper is organized as follows: First an overview of the related work is provided, followed by the detailed description of the experimental design. Subsequently, details about the participants, their characteristics, the used materials and the experimental setting are also presented. Then the obtained data are discussed and analyzed with respect to the research goals. Finally, the results and their implications are discussed in comparison with other related studies.

2 RELATED WORK

The focus of this research is to investigate the impact of a mobile device screen size on usability. We studied in our experiment both perceived usability, as it is obtained by evaluation questionnaires, and specific usability metrics (effectiveness, efficiency) by applying the ISO 9241 definition of usability [3].

2.1 Parameters influencing usability

Perceived usability has been a subject of research for many years. Through this research effort many parameters affecting it have been identified (Figure 1).

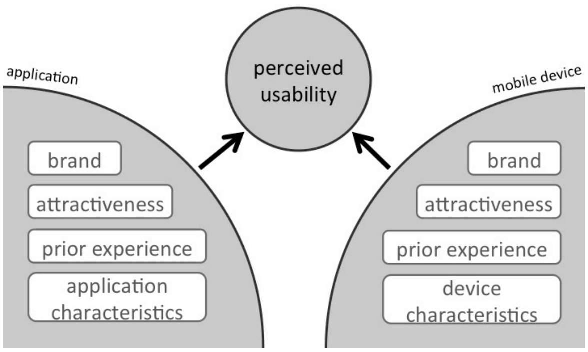


Figure 1: Parameters that affect perceived usability while interacting with a mobile device and an application.

When users interact with digital artifacts attractiveness plays an important role. Thus, we may find evidence that attractive things work better [23] and studies that specifically investigated the notion of attractiveness in relation to perceived usability [8, 31]. In the context of mobile computing Quinn and Tran [25] showed that attractiveness, effectiveness, and efficiency have an independent influence on usability ratings, with attractiveness having the largest impact. They argued that it is necessary to test the simultaneous influence of these variables (attractiveness, effectiveness, and efficiency) on perceived usability and they concluded that "Clearly, attractiveness is one factor that must be considered when interpreting participant-rated usability" [25, p. 361]. Therefore, it is possible that an attractive phone could be rated high in perceived usability regardless of (low or high) effectiveness and efficiency [25]. Furthermore, there are studies illustrating that brand has also a strong effect on perceived usability [9].

Sauro [27] examined the influence of prior experience with a website on the users' rating. Using a large dataset with 62 websites he found that experienced users rated the websites as 11% more usable [27]. Suzuki et al. [29] examined the impact of task completion time on perceived mobile phone usability. They report that a negative correlation between time-on-task and perceived usability becomes significant with as little as an hour's time doing tasks with an unfamiliar phone. However, the correlation was not significant for novice users, when initially inspecting a new phone model.

Furthermore, Raita and Oulasvirta [26] devised an experiment where 36 subjects read a positive, or a negative product review for a novel mobile device (while a control group read nothing) before a usability test. Their results demonstrated a strong amplifying effect of the positive expectation on the post-experiment questionnaire ratings.

2.2 The effect of screen size on usability

Even though there is more than enough evidence that the previously mentioned parameters have a strong effect on perceived usability, there is a little research towards identifying possible effects of specific mobile phone characteristics on it. As a result, there are not many studies dealing with possible effects of weight, screen size, button size and other phone characteristics on usability, especially for the new modern phones where the interaction is touch enabled and screen size plays a crucial role.

Regarding screen size most of the previous research deals with desktop environments (for example [5, 24]), or compares between desktops and small screens (for example [10, 12, 30]). In the context of mobile computing Jones et al. [14] report that Internet searching tasks are slower in smaller screens. Maniar et al. [21] examined the effect of non-touch mobile phones' screen size on video based learning. Using phones with 3.78, 2.28 and 1.76in screen, they found that the smallest screen significantly deteriorated the students' learning effectiveness. However, they did not find any significant differences between the phones with the larger displays. In addition, non-touch mobile phone screen effect on efficiency was studied in relation to users' navigation activities [7]. They concluded that both information structure and screen size significantly affect the navigation behavior and the participants' satisfaction, but only when the task complexity increases.

Kim et al. [16] used three mobile devices (3.5, 5.7 and 9.7in) and found out that the largest screen led to higher participants' enjoyment, while the smaller screen-size elicited greater perceived mobility. However, the effect of screen size on enjoyment was found to be significant only between the 5.7 and 9.7 inches. They also state that participants who used the 3.5in device reported that they were more likely to use a similar mobile device in the near future, than participants with the 5.7in device. In addition, participants reported that the 3.5in and 9.7in device were more useful than the 5.7in device.

3 EXPERIMENT

Motivated by related work we decided that the main goal of this experiment was to measure the effect of a mobile phone's *screen size* on users' perceived usability, effectiveness and efficiency while our subjects interacted with a mobile application. Our research hypotheses:

- H_1: The mobile phone's *screen size* will have an effect on participants' perceived usability (a larger screen will increase the perceived usability ratings).
- H_2: The mobile phone's *screen size* will have an effect on participants' task completion times (a larger screen will increase efficiency).
- H_3: The mobile phone's *screen size* will have an effect on participants' task completion rates (a larger screen will increase effectiveness).

In order to investigate these hypotheses we adopted a between-groups experimental design [19] and asked three groups of users to interact with the same application on three different devices and then formulate their view about it (each group interacted with one device).

Figure 2: The mobile phones used in the experiment, running IMDB. From left to right SamsungTM: a) Galaxy Ace, b) Galaxy Note, and c) Galaxy SII.

Model	Size (mm)	Weight	Scr. Size	Resolution
Galaxy Ace	112.4x59.9x11.5	113g	3.5in	320*480
GalaxyNote	145.8x66.1x8.5	178g	5.3in	800*1280
Galaxy SII	125.3x66.1x8.5	116g	4.3in	480*800

Table 2: **Mobile phones' characteristics.**

3.1 Device selection

One key decision in this experiment was the selection of devices. Since there is substantial research evidence that brand has a significant effect on perceived usability [9], we chose three Samsung™ devices. The reason for choosing this brand was that it offered us the possibility to select phones with different screen sizes and extremely similar visual design (Figure 2), thus allowing us to control for attractiveness. Therefore, the fact that Samsung™ offered phones that had different screen sizes, but the same materials used for the casing, both in front and the back, same physical buttons, almost the same appearance, same colors and similar age, constituted Samsung™ as an ideal candidate. For these reasons, we have selected a Samsung™ Galaxy Ace, a Galaxy Note and a Galaxy SII (Figure 2) and their screen size varied from 3.5 to 5.3 inches in order to cover a variety of common sizes. Table 1 presents in detail the mobile phones' characteristics.

3.2 Application selection

Since we wanted to investigate the effect of *screen size* on users' perceived usability, effectiveness and efficiency we chose to avoid leisure applications, such as mobile games, where pleasure, fun and flow might be more influential than usability. Furthermore, we chose to focus on information retrieving applications because in these applications usability metrics, such as effectiveness and efficiency, are frequently used as evaluation criteria. For these reasons we decided to use IMDB (Internet Movie DataBase, version 2.3.1) as test application.

3.3 Experimental setup

A series of actions were performed in order to ensure that the devices would offer a similar experience. We installed to all three devices the same Android version (2.3), the same launcher ("Go Launcher"), the same theme ("Android 4" theme) and the same keyboard ("GO keyboard"). Screen orientation was locked and since we wanted all the participants to interact with the test application using their fingers, we removed the Galaxy Note's stylus. In order to ensure that the technical characteristics of the devices would not affect the participants we underclocked the CPUs of all the phones (using "No-frills CPU Control" application) to match the slower device (Galaxy Ace). Thus, all the devices operated at 800Mhz.

By using these specific devices and by performing these actions we made sure that the three groups would interact with the same operating system and the same application on devices of the same brand that had similar hardware performance and an identical physical design. In other words, we have experimentally controlled for the parameters depicted in Fig 1 (for example brand and attractiveness), we created identical hardware performances by downgrading the CPUs and conducting a performance test, and we provided a similar IMDB interface, thus controlling as much as possible for the different screen characteristics (Fig 2).

3.4 Participants

60 university students, 48 male, 12 female, aged 19-30 (M=23.48, SD=2.34) participated in the experiment. All of them were undergraduate and postgraduate students, attending Aalborg University in Denmark. 49 of them were Computer Science students, where the rest from various university departments. All of them volunteered for the experiment and a voucher was awarded to one participant at the end of the procedure, after conducting a draw.

We had specifically designed our experiment in order to perform ANOVA to analyze the gathered data and the reason for opting for 60 participants was that 20 (per group) is the minimum number of participants to safely conduct it [28]. Additionally, the fact that we chose participants that belonged to same age group eliminated the possible effect of participants' *age* on perceived usability as identified by [2].

When the students expressed online their interest to participate to the experiment they provided us with details about their personal mobile phone. We found out that 8 participants had previous experience with Samsung™ phones and among them, 4 had experience with the Galaxy SII. These participants were distributed among the three groups. The rest were assigned randomly.

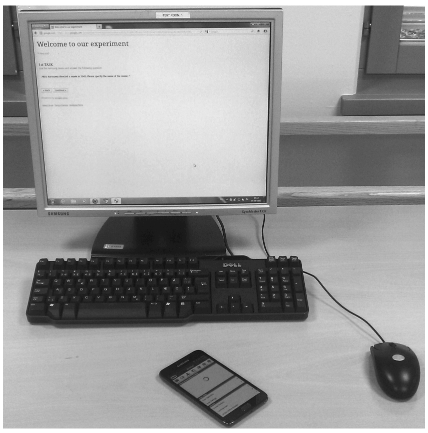


Figure 3: The experimental setup at the usability laboratory.

3.5 Procedure

The experiment evolved in two phases, both conducted at the same room in a usability laboratory and each subject participated individually. None of the participants knew the research scope of the experiment. Prior entering the usability lab a researcher informed each participant that their purpose was to accomplish five tasks while using a test application and then fill-in an online evaluation questionnaire. Then the researcher handed over a randomly selected mobile device with the test application already running in full screen. Each participant entered the usability laboratory alone and after a short introductory text, the first out of five tasks was presented through a desktop computer (Figure 3). The five information seeking tasks the participants had to accomplish, as well as their characteristics, were:

1. "Akira Kurosawa directed a movie in 1943. Please specify the name of the movie." (*easy, heavy scrolling*),
2. "Charlie Chaplin was the writer and the director of the movie "The Kid". Please specify when was the birth date of the actor that played the role of the Man." (*easy, light scrolling*),
3. "What is the title of the ninth episode of the second season of the documentary series "Through the Wormhole"?" (*medium, medium scrolling)*,
4. "When did the user "D-V" wrote his review of the movie "Kin Dza Dza"?" (*difficult, light scrolling*),
5. "Please specify the director of the movie that is on the position 186 at IMDB's Top 250." (*medium, heavy scrolling*).

The aforementioned information seeking tasks were selected by applying three criteria. First, they should have a varying level of difficulty for locating the answer (easy, medium and difficult). Second, the amount of required scrolling to find the requested information had also to vary (light, medium and heavy scrolling) since we wanted to check if scrolling had an effect on effectiveness and efficiency. Third, we selected these tasks in order to minimize the possibility our participants to be familiar with the requested movies/information.

The researchers were located at the laboratory's observation room and each participant's session was monitored and videotaped. Additionally, the "droid VNC" application was installed in all three mobile devices. This application streamed (over Wi-Fi) the interface of the mobile device to a desktop computer that was located in the observation room and in combination with the rest of the usability laboratory's equipment, it allowed us to track task completion times and rates without disturbing the participants.

3.6 Measures

Three different measures were used in this experiment: a demographics questionnaire, an attractiveness questionnaire and an evaluation questionnaire.

3.6.1 Demographics questionnaire

The demographics questionnaire comprised 19 questions that documented the participants' experience with mobile devices, their previous experience with the test application and general aspects like favorite mobile brands, favorite screen size, their own personal device and their will to buy the used in the experiment device.

3.6.2 Attractiveness questionnaire

In order to measure the attractiveness of the mobile phones we adopted a slightly modified version of an attractiveness questionnaire [25] that was provided to the participants on a 7-point scale. We used a modified version with only five questions since Quinn and Tran [25] do report in their paper only the five questions out of the seven they used in their study.

3.6.3 Evaluation questionnaire

In order to evaluate the test application and collect data about perceived usability we used the System Usability Scale questionnaire (SUS, [6]). The reasons for choosing SUS were the facts that it is free, very simple and short (10 items on a 5-point scale) and more importantly that it has been found remarkably robust on various studies (for example [1, 2, 4, 20]). The SUS questionnaire was used with the modification proposed by [11] (replacing the word "cumbersome" with "awkward") and the word "system" was replaced by the word "IMDB".

4 DATA ANALYSIS AND RESULTS

4.1 Extracted variables

The collected data were organized and analyzed using Excel 2010 and SPSS v19.0. The following subsections present the variables that we extracted from the questionnaires and used for the data analysis.

4.1.1 SUS Score

The average SUS scores and standard deviations in the three groups were: 83.12 (SD=11.21) for Galaxy Ace, 88.12 (SD=7.02) for Galaxy Note and 82.12 (SD=9.08) for Galaxy SII (Table 2). In all cases the perceived usability of the test application was rated close to excellent (as proposed by [1]), since the SUS score was close to 85.58.

4.1.2 Attractiveness

We measured the *attractiveness* of the mobile devices through Quinn and Tran's [25] attractiveness questionnaire (α=0.90). Since they report only five out of seven questions in their paper we conducted factor analysis in order to measure the effectiveness of the five-question version. We have observed one dominant factor and it was effectively measuring *attractiveness* (α=0.813). On average our participants rated the *attractiveness* of: a) Galaxy ACE 4.43 (SD=0.474), b) Galaxy Note 4.94 (SD=0.440), and c) Galaxy SII 4.51 (SD=0.475).

4.1.3 Prior experience with the application

The participants were divided into two groups: the ones *without experience* (16 participants) and the ones *with prior experience* with the test application (44 participants).

4.1.4 Prior experience with Android

Since only eight participants had experience with a Samsung device, we collected the participants' experience with mobile phones in general, based on the device they owned. We have identified three categories: *Android* users (27 participants), *iOS* users (12 participants) and *Other* users (21 participants).

4.1.5 Will to buy the used device

Through the demographics questionnaire we measured the will of the participants to buy the mobile device we had provided them during the experiment, through two questions: "I have a strong desire in buying this Samsung phone", "This Samsung phone will definitely be my next phone" on a 1 to 7, strongly disagree-strongly agree scale. Factor analysis identified only one factor and these two items could accurately measure *will to buy* (α=0.904). *Will to buy* was measured on average 1.95 (SD= 1.19) for the Galaxy Ace, 2.8 (SD=1.61) for the Galaxy Note and 2.42 (SD=1.76) for the Galaxy SII. The participants were divided in three categories: *low will* (37 participants), *medium will* (17 participants) and *high will* (6 participants).

4.1.6 Favorite screen size

We asked the participants to select their favorite screen size from a set of predefined choices. The participants were divided in four categories: a) *small screens* (3.0"-3.7", 14 participants), b) *medium screens* (4.0"-4.5", 20 participants), c) *large screens* (more than 4.7", 7 participants), and d) the ones that stated they *did not care* about the screen size (19 participants).

4.1.7 Owned screen size

Since we were informed about the personal device each of our participants owned, we divided them to the same categories as *favorite screen size*. Thus, the participants were categorized as: a) *small screen users* (3.0"-3.7", 16 participants), b) *medium screen users* (4.0"-4.5", 32 participants), and c) *large screens users* (more than 4.7", 12 participants).

4.1.8 Favorite brand

Through the demographics questionnaire we asked the participants to select from a list their two favorite mobile manufacturers. Then we divided them in two categories: a) the ones who stated that Samsung™ is one of their favorite mobile device manufacturers (31 participants), and b) the ones who stated other brands (29 participants).

4.1.9 Adjective rating

The *adjective rating* was proposed by [1] as a qualitative replacement for the SUS questionnaire. This variable is measured through one question: "Overall I would rate the user-friendliness of IMDB as:" on a 1 to 7, worst imaginable-best imaginable scale. Overall IMDB was rated on average 5.55 (SD=0.70).

		Task completion rates (number of successful participants)						SUS Score	
	Screen (inches)	**T1**	**T2**	**T3**	**T4**	**T5**	**Overall success (%)**	**Mean**	**SD**
Galaxy Ace	3.5	19	13	20	18	15	85	83.12	11.21
Galaxy Note	5.3	20	15	18	19	13	85	88.12	7.02
Galaxy SII	4.3	20	16	19	18	16	89	82.12	9.08
Overall		**19.7**	**14.7**	**19.0**	**18.3**	**14.7**	**86.3**	**84.5**	**9.47**

Table 2: Descriptive statistics for participants' task success rates and SUS score in relation to the used mobile device.

4.2 The effect of attractiveness

The first step in data analysis was to check if the mobile phones were rated differently in relation to *attractiveness*, despite the fact that they were intentionally selected to have extremely similar designs. In order to test how the three groups perceived the attractiveness of the devices we conducted one-way between subjects ANOVA having one independent variable (device) and one dependent variable (attractiveness). No significant effects were observed: $F(2,57)=1.532$, $p=.225$. The attractiveness of the devices was treated in a similar way among the three groups, a result that verified that our selection of three devices with extremely similar design (Figure 2) was effective. However, attractiveness has been included to our data analysis in order to control for possible interaction effects.

4.3 The effect of screen size on perceived usability (SUS)

The next step was to investigate the effect of *screen size* on the SUS score and test our first hypothesis (H_1). The variables that were included to this part of data analysis were the *screen size* (independent variable), the SUS score (dependent variable) and the demographic variables of *attractiveness, favorite brand, prior experience with the application, prior experience with*

Android, will to buy the used device, favorite screen size and *owned screen size*. We did not include age in our analysis since all our participants belonged to the same age group. A General Lineal Model was applied to analyze the collected data. The first approach was to treat all the demographic variables as covariates (a variable that can be observed along the dependent variable, but is not possible to be experimentally controlled) and perform ANCOVA. Since ANCOVA loses power when many covariates are included, we reduced their number by performing principle component analysis (regression method). Two components were extracted, reducing the number of covariates from six to two. The first component was *prior experience* and it was extracted from *prior experience with Android* and *prior experience with the application*. The second component was extracted from the rest of the demographic variables (*attractiveness, will to buy the used device, favorite screen size* and *favorite brand)* and it was named as *desire for the device*. After checking that homogeneity of variance, homogeneity of regression and normality assumptions were not violated, we have performed ANCOVA with *prior experience* and *desire for the device* as covariates. Results showed that there was no main effect of *screen size* ($F(2,55)=.978$, $p=.383$), but there was a significant effect of *prior experience* ($F(1,55)=5.890$, $p=.019$*) and *desire for the device* ($F(1,55)=4.910$, $p=.031$*), at the $p<.05$ level. Furthermore, from the answers the 44 participants *with prior experience* with the test application had provided in four questions ("how often do you use IMDB: in general, to find data about actors, to find data about movies, to rate movies", 1-5, rarely-often scale) we observed that they had diverse experiences. Therefore we further divided them in two categories: the ones with *low experience* (23 participants) and the ones with *high experience* (21 participants). We repeated the same ANCOVA analysis and had similar results as before for *screen size* ($F(2,55)=.855$, $p=.431$), *prior experience* ($F(1,55)=4.391$, $p=.041$*) and *desire for the device* ($F(1,55)=4.635$, $p=.036$*).

Subsequently, we investigated the possibility the participants' SUS scores to be influenced by the mobile device they own and/or their favorite screen size. Initially, we performed a Chi-Square test between *favorite screen size* and *owned screen size* in order to check if the participants' selection for their favorite screen size was biased by the device they owned. Our results showed that a statistical significant association between *owned screen size* and *favorite screen size* ($\chi^2(6, N=60)=23.154$, $p=.001$**). Then we have performed a two-way ANOVA between *favorite screen size* and *screen size* on SUS score, having always in mind that through this approach we risked to exclude influential interactions from the analysis. Significant effects were observed ($F(6, 48) = 3.663$, $p = .004$**). Data analysis showed that *favorite screen size* had an effect ($F(3, 48) = 5.887$, $p = .002$**) and the same was the case with *screen size* ($F(2, 48) = 6.525$, $p = .003$**).

one way ANOVA	*screen size* on *attractiveness*	F(2,57)=1.532	p=.225
ANCOVA	*screen size* on SUS	F(2,55)=.978	p=.383
	covariate 1: (*prior experience*)	F(1,55)=5.890	p=.019*
	covariate 2:(*desire for the device*)	F(1,55)=4.910	p=.031*
two way ANOVA	*screen size* * *favorite screen size* on SUS	F(6,48)=3.663	p=.004**
	screen size on SUS	F(3,48)=5.887	p=.002**
	favorite screen size on SUS	F(2,48)=6.525	p=.003**
3x5 mixed design ANOVA	*screen size* on task completion times	F(2.37,57)=10.3	p<.001**
one way ANOVA	*screen size* on task completion times (per task)	1) F(2,57)=2.855	p=.066
		2) F(2,57)=0.583	p=.562
		3) F(2,57)=6.016	p=.004**
		4) F(2,57)=0.237	p=.789
		5) F(2,57)=4.559	p=.015*
Chi-Square test	*screen size* on task completion rates (per task)	1) χ^2(2,N=60)=8.17	p=.98
		2) χ^2(2,N=60)=3.83	p=.85
		3) χ^2(2,N=60)=5.94	p=.94
		4) χ^2(2,N=60)=8.03	p=.98
		5) χ^2(2,N=60)=3.83	p=.85

Table 3: Data analysis results

4.4. The effect of screen size on efficiency and effectiveness

Efficiency was measured through task completion times (Table 4). A 3x5 mixed design ANOVA with one within-subjects factor (five levels: Tasks 1-5, Table 4) and one between-subjects factor (*screen size)* was adopted. Since the Mauchly's test for sphericity was violated a Greenhouse-Geisser correction was applied. A significant main effect of *screen size* on participants' efficiency was unveiled (F(2.37,57)=10.3, p<.001**), thus confirming H_2. Pairwise comparisons between the participants' task completion times unveiled significant differences between the users who used Galaxy Ace (3.5in) and the ones who used Galaxy Note (5.3in), p=.007**. No significant differences were unveiled between Galaxy SII (4.3in) and Galaxy Ace (p=.107), nor between Galaxy SII and Galaxy Note (p=.253). This finding suggests that a significant gain on efficiency possibly occurs when the screen size increases more than 4.3in.

Additionally, the effect of *screen size* on *efficiency* was checked for each of the tasks by applying one-way between-subjects ANOVAs (Table 3). From these results it was evident that there was a significant effect of *screen size* on *efficiency* for tasks 3 and 5. This finding suggests that not all tasks benefit from a larger display, but only the tasks that are not easy and require a significant amount of scrolling. Effectiveness was measured through task completion rates (Table 2). Success rates for all 5 tasks were 85%, 85%, 89% for the Galaxy Ace, Note and SII respectively. H_3 was rejected as there was not observed any significant effect of *screen size* on *effectiveness* after conducting Chi Square tests (Table 3).

4.5 Correlations across usability metrics

To further understand the effect of a mobile phone screen size we examined the correlations between perceived usability, total task completion times (using z scores) and total task completion rates. No significant correlation was found between SUS ratings and total task completion rates (Pearson's r=0.248, p=.056). The same was the case between total task completion rates and total task completion times (Pearson's r=-0.151, p=.25). On the contrary, a significant modest correlation was found for total task completion time and SUS ratings (Pearson's r=-0.357, p=.005**). The latter, agrees with the findings reported by [25] and [29]. In specific, Quinn and Tran [25] reported a significant correlation, which was very close to our findings (r=-0.33), while Suzuki et al. [29] reported a significant negative correlation for novice users (-.263) and expert users (-.563). However, contrary to our findings Quinn and Tran [25] have also identified significant correlations between total task completion rates and SUS scores. Further investigation for each phone unveiled a significant correlation between total task completion rates and SUS only for the Galaxy Ace (r=.501 p=.029*) and not for Galaxy SII (r=0.080, p=.737) and Galaxy Note (r=0.386, p=.093).

Moreover, the correlation between the participants' *adjective rating* and their SUS score was examined. The correlation was very high (Pearson's r=0.679, p<.001), but not as high as the one reported by Bangor et al. ([1], 0.806). Nevertheless, we can also confirm that "an associated adjective rating scale is a legitimate complement to the SUS statements and overall SUS score" [1, pp. 588].

There were no significant differences between male and female participants' SUS scores (male: M=83.95, SD=9.72, female: M= 86.45, SD=8.49, two tailed T-test, t=0.81, p=.42), which is a contradictory finding to that provided by [17] who reported that males assigned higher SUS scores than females. However, since the number of our female participants was rather low (12), more data are needed to make conclusive claims about the impact of gender on mobile application's SUS evaluation.

Finally, we have identified a significant difference between the participants' SUS rating *with prior experience* (44 participants) and *without experience* (16 participants) to the test application (with experience: M=86.3, SD=7.69, without experience: M=79.4, SD=12.1, two tailed T-test, t=2.141, p=.04*). The participants *with prior experience (prior exposure)* rated the application on average 8.7% higher, a result which is in line with [25], [27] and [22] who found 9.1%, 11% and 15-16% difference, respectively. Table 5 summarizes the results.

		Task completion Time (sec)											
	Screen	T1		T2		T3		T4		T5		Total	
	(inches)	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Galaxy Ace	3.5	90.0	43.7	77.6	88.7	39.7	19.7	93.0	90.2	61.4	39.2	361.7	188.1
Galaxy Note	5.3	50.6	77.7	45.8	40.3	24.3	12.5	71.3	53.7	36.8	20.8	228.9	134.1
Galaxy SII	4.3	79.0	30.1	58.7	41.6	31.2	7.0	71.4	73.3	43.4	12.7	283.7	120.4
Overall		**73.2**	**55.9**	**60.7**	**61.6**	**31.8**	**15.2**	**78.6**	**73.4**	**47.2**	**28.2**	**291.5**	**157.7**

Table 4: Descriptive statistics for participants' task completion time in relation to the used device.

SUS and total task completion times	r=-0.357	p=.005**
total task completion rates and total task completion times	r=-0.151	p=.25
SUS and total task completion rates	r=0.248	p=.056
total task completion rates and SUS (per device)	Galaxy Ace: r=.501	p=.029*
	Galaxy SII: r=0.080	p=.737
	Galaxy Note: r=0.386	p=.093
adjective rating and SUS	r=0.679	p<.001**
gender on SUS	t=0.81	p=.42
prior exposure on SUS	t=2.141	p=.04*

Table 5: Correlations across usability metrics.

5 DISCUSSION

ANCOVA analysis showed that the *screen size* of a device did not have a significant effect on SUS score, but *prior experience* and *desire for the device* did have a significant effect. Consequently, our first research hypothesis (H_1) was rejected. Furthermore, our results confirm previous studies, which provide evidence that *prior experience* affects perceived usability (such as [18, 22, 27]) and that hedonic qualities [13], in this case *desire for the device*, do also have a significant effect. Additionally, data analysis suggests there is a significant effect of mobile phones' screen size on efficiency in information seeking tasks, as we have observed that participants who used larger screens were more efficient (H_2 was confirmed). However, the magnitude of the effect is related to the nature of the task. Tasks that are not easy and require a significant amount of low level interactions (such as scrolling) seem to greatly benefit by the additional screen area. This result is inline with previous studies [7, 14] with the difference that we have used modern touch enabled mobile devices. Despite the large differences on task completion times, the task completion rates were almost equivalent. This finding suggests that differences on screen size did not impact effectiveness (H_3 was rejected) possibly due to our participants' reported high mobile usage. Regarding the correlations across usability metrics, task completion times were significantly correlated with the participants' SUS ratings. Additionally, we have identified a correlation between task completion rates and SUS only for the smallest device.

Three important findings have emerged from data analysis and results. The first finding is related to everyday mobile device usage. Through mobile devices users interact with hundreds of applications for a variety of purposes. If a mobile phone user has as purpose to be mainly engaged to information seeking activities (such as internet browsing) then our study suggests that she will be more efficient when she interacts with a device that is larger than 4.3in. Of course, more research is needed with devices that have larger screens in order to measure if and at what size this positive effect on efficiency stops. Furthermore, if the reasons behind using a mobile device are related more to leisure activities, such as watching videos or playing games, then a large screen will also be beneficial, as it will lead to higher enjoyment [16]. On the other hand though, extremely large mobile phone screens (such as 5.7in) might also have the opposite effect as they decrease the level of perceived mobility [16], have an effect on portability as they reduce battery life and decrease also the ability to use the phone with one hand [15]. For these reasons, a screen size around 4.3in seems to be satisfactory for both purposes. Perhaps, this could also explain why many mobile device manufacturers are currently producing and highly advertising devices with screens around 4.3in (for example Apple™ iPhone 5 – 4.0in, Samsung™ Galaxy SIII - 4.8in, HTC™ One - 4.7in, Nokia™ Lumia 920 – 4.5in).

This first finding is also important for designers. If users of information retrieving applications are more efficient while using devices with larger than 4.3in screens, then perhaps designers should make specific design decisions for the smaller screens in order to increase their efficiency too. For example in smaller than 4.3in screens they could avoid using scrolling by providing an alternative information architecture, or enhanced search mechanisms. Furthermore, in situations where designers are developing applications that are tailored for specific devices/contexts we recommend to them to select devices with at least 4.3in when they design complex applications that contain a large amount of information. For instance, if a museum is considering to provide its visitors with a mobile device and a guide application in order to interact with the exhibits while being inside the space, then if the designers adopt a device with at least 4.3in screen they will significantly increase visitors' efficiency.

As a secondary finding, we argue that our results can also be beneficial for researchers and especially practitioners in the context of mobile usability evaluations. If they choose to assess their applications/prototypes by measuring perceived usability through SUS then the screen size of the device they selected for the evaluation will not have an effect on the SUS ratings, for information seeking tasks and at least for the range of screen sizes that we used in our experiment. In addition, since *adjective rating* [1] was confirmed as an efficient supplement of SUS, then it can be used when quick data are needed about perceived usability. On the other hand, if they choose to assess the usability of their information seeking applications through traditional usability metrics, such as effectiveness and efficiency, then devices with screens larger than 4.3in will increase users' efficiency. For the latter cases we recommend to practitioners to use a variety of screen sizes during their evaluations in order to have more holistic results, or if this is not possible, to use one that has a screen size around 4.3in. Nevertheless, in all cases they should take into consideration the parameters that can influence usability (Figure 1).

Finally, if we focus on the two-way ANOVA results (with its limitations since we excluded many variables from the analysis) we have a third, preliminary finding that *favorite screen size* or *owned screen size* has a strong effect on perceived usability, since our participants rated IMDB significantly differently when they used a device that had the same screen size as the one they owned, possibly due to familiarity. Of course, due to our experimental setup, which was not designed to study this phenomenon, this result needs to be further studied in detail, but nevertheless we have preliminary evidence that *owned screen size* might play a crucial role in our interpretation of mobile perceived usability. If more research is conducted on this issue and our result is verified, then there might be a need for an extension of the Quinn and Tran study [25], which demonstrated that attractiveness affects perceived usability. Due to the fact that they used a variety of mobile devices, if these devices had a variety of screen sizes then *owned screen size* might have also affected perceived usability, besides attractiveness.

6 CONCLUSION

In this experiment we have tried either to control for, or take into consideration the parameters that might affect usability (Figure 1) in order to measure the effect of *screen size* on perceived usability, effectiveness and efficiency. We have experimentally controlled for brand, attractiveness and application and device characteristics and statistically controlled for prior experience and desire for the device. We had three research hypotheses: *screen size* would have an effect on perceived usability (H_1), on efficiency (H_2) and effectiveness (H_3). H_1 and H_3 were rejected. On the contrary, there was a significant effect of *screen size* on efficiency (H_2 was confirmed).

Three important findings have emerged. The first one is related to everyday mobile device use. Mobile users that interact with a device with the purpose of mainly performing information seeking tasks, such as Internet browsing, will be more efficient if they use a device with a screen larger than 4.3in. The same is the case for users that want to mainly use their device to play games, or watch media, as the larger screen size will lead to higher enjoyment, but with the counter argument that very large screens decrease the portability of the device and reduce the ability to use the phone with one hand. Therefore a mobile device with a screen size around 4.3in seems to be beneficial for both cases. Our second finding is related to usability evaluations of information retrieving applications. According to our results, researchers that measure perceived usability through SUS will not observe any differences if they evaluate an application on devices with different screen sizes. On the contrary, researchers that assess usability through usability metrics, such as effectiveness and efficiency will observe that larger screens will lead to higher efficiency.

Our findings do show that screen size matters. It matters for the typical everyday mobile users, since it can have an effect on their efficiency in many of their everyday information seeking activities and it matters also for practitioners and researchers as it can influence their design decisions, as well as, the way they conduct mobile usability evaluations.

As a future work we plan to continue at the same direction and study the effect of *screen size* on a more expanded set of tasks, such as map navigation, in order to cover more of the common tasks that people do with their devices. We would like also to use devices with larger screens in order to measure if and when the identified positive effect on efficiency stops. Additionally, we plan to investigate in detail our third preliminary finding that the participants of a mobile usability evaluation might be influenced on their SUS scores when they are asked to evaluate an application using a mobile device that has the same screen size as the one they own. Finally, another direction could be the study of the *screen size* effect over time, as the familiarity with a device and an application increases.

ACKNOWLEDGEMENTS

We would like to thank all the participants in this experiment and Kasper Hornbæk for his valuable comments. The research behind this paper was partly financed by the Danish Research Councils (grant number 09-065143).

REFERENCES

Bangor, A., Kortum, P., and Miller, J. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24, 6 (2008), 574–594.

Bangor, A., Kortum, P., and Miller, J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, 4, 3 (2009), 114-123.

Bevan, N. International Standards for Usability should be more widely used. *Journal of Usability Studies*, 4, 3 (2009), 106-113.

Borsci, S., Federici, S., and Lauriola, M. On the dimensionality of the System Usability Scale: A test of alternative measurement models. *Cognitive processing*, 10, 3 (2009), 193-197.

Bridgeman, B., Lennon, M.L., and Jackenthal, A. Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Performance. *Applied Measurement in Education*, 16, 3 (2003), 191-205.

Brooke, J. SUS – A quick and dirty usability scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester, and I.L. McClelland (eds.), *Usability evaluation in industry*. Taylor and Francis (1996), 189-194.

Chae, M., and Kim, J. Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour and Information Technolog*y, 23, 3 (2004), 165-181.

Chawda, B., Craft, B., Cairns, P., Rüger, S., and Heech, D. Do attractive things work better? An exploration of search tool visualizations. In *Proc. HCI 2005, 2* (2005), 46-51.

De Angeli, A., Hartmann, J. and Sutcliffe, A. The effect of brand on the evaluation of websites. In *Proc. Interact 2009*, Part II (2009), 638 – 651.

Findlater, L., and McGrenere, J. Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. In *Proc. CHI 2008*, ACM Press (2008), 1247-1256.

Finstad, K. The System Usability Scale and Non-Native English Speakers. *Journal of Usability Studies*, 4, 1 (2006), 185-188.

Grabea, M.E., Lombard, M., Reich, R.D., Bracken, C.C., and Ditton, T.B. The role of screen size in viewer experiences of media content. *Visual Communication Quarterly*, 6, 2 (1999), 4-9.

Hassenzahl, M. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19, 4 (2004), 319–349.

Jones, M., Buchanan, G., and Thimbleby, H. Improving web search on small screen devices. *Interacting with Computers* 15, 4 (2003), 479-495.

Karlson, A.K., Bederson, B.B., and Contreras-Vidal, J.L. Understanding One-Handed Use of Mobile Devices. *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, (2002), 86-101.

Kim, K.J., Sundar, S.S., and Park, E. The effect of Screen-Size and Communication Modality on Psychology of Mobile Device Users. In *Proc. CHI 2011*, ACM Press (2011), 1207-1212.

Kobsa, A., Sonawalla, R., Tsudik, G., Uzun, E., and Wang, Y. Serial hook-ups: A comparative usability study of secure device pairing methods. In *Proc. SOUPS '09: 5th Symposium on Usable Privacy and Security*, ACM Press (2009), Article No10.

Langdon, P.M., Lewis, T., and Clarkson, P.J. The effects of prior experience on the use of consumer products. *Universal Access in the Information Society*, 6, 2 (2007), 179-191.

Lazar, J., Feng, J.H., and Hochheiser, H. *Research methods in human-computer interaction*. Wiley & Sons, West Sussex, UK, 2010.

Lewis, J.R., and Sauro, J. The Factor Structure of the System Usability Scale. In *Proc. HCII 2009*, Springer-Verlag (2009), 94-103.

Maniar, N., Bennett, E., Hand, S., and Allan, G. The effect of mobile phone screen size on video based learning. *Journal of Software*, 3, 4 (2008), 51-61.

McLellan, S., Muddimer, A., and Peres S.C. The effect of experience on System Usability Scale Ratings. *Journal of Usability Studies*, 7, 2 (2012), 56-67.

Norman, D.A. *Emotional design: Why we love (or hate) everyday things*. Basic Books, NY, 2004.

Polys, N.F., Kim, S., and Bowman, D.A. Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments. In *Proc. VRST 2005*, ACM Press (2005), 46-55.

Quinn, J.M., and Tran, T.Q. Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In *Proc. CHI 2010*, ACM Press (2010), 353-362.

Raita, E., and Oulasvirta, A. Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers*, 23, 4 (2010), 363-371.

Sauro, J. Does Prior Experience Affect Perceptions Of Usability? (2011) Available online: http://www.measuringusability.com/blog/prior-exposure.php

Simmons, J.P., Nelson, L.D., and Simonsohn, U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22, 11 (2011), 1359 -1366.

Suzuki, S., Bellotti, V., Yee, N., John, B.E., Nakao, Y., Asahi, T., and Fukuzumi, S. Variation in importance of time-on-task with familiarity with mobile phone models. In *Proc. CHI 2011*, ACM Press (2011), 2551-2554.

Sweeney, S., and Crestani, F. Effective search results summary size and device screen size: Is there a relationship? *Information Processing and Management*, 42, 4 (2006), 1056-1074.

Tractinsky, N., Katz, A.S., and Ikar, D. What is beautiful is usable. *Interacting with Computers*, 13, 2 (2000), 127-145.

7.3 PAPER 3: UNDERSTANDING "COOL" IN HUMAN-COMPUTER INTERACTION RESEARCH AND DESIGN.

ABSTRACT

Recently a discussion has been initiated on what is cool and how HCI can use the concept in practice and design for it. This paper aims to provide a better understanding on cool as a concept from a theoretical and a practical perspective. From the theoretical perspective, we selected the HCI papers that focus on cool and we present their core findings. Then we performed a literature review on the concept of cool and we have identified its fundamental characteristics, through cool personalities and cool styles. From a practical perspective, we have studied how other domains have managed to successfully produce cool objects and we provide four suggestions on how to design cool digital artifacts. Finally, in this paper we also identify possible research directions in relation to cool, which if we manage to address we can increase our understanding on what is user experience and this can lead to the creation of better digital artifacts. Overall, this paper is a contribution towards researching and designing for cool, a research topic, which we believe it will initiate fruitful discussions in the HCI field.

1 INTRODUCTION

What is cool? Why we are referring to cool movies, cool people, or cool objects almost everyday? Are there some specific attributes that we perceive on an artifact and we characterize it as cool? If yes, can we define them? Why is cool relevant to HCI?

HCI is studying our interactions with digital technologies and researches our experiences with them. User experience (UX, Hassenzahl and Tractinsky, 2006) is one of the main research streams of HCI and it encompasses the idea that our interactions with digital artifacts transcend *effectiveness* and *efficiency*. Although the usability literature also moved towards this direction with *user satisfaction* as a flag concept, UX studies our experiences from a broader perspective by encompassing concepts such as affect, emotion, hedonic and pragmatic qualities, fun, flow, enchantment, etc. Despite the fact that there is a significant amount of research directed towards understanding these concepts, there still are challenges that HCI needs to overcome. The most important ones are derived from the fact that we do not know if these concepts are enough to describe UX because the relations among them have not been studied in detail (Bargas-Avila and Hornbæk, 2011; Bargas-Avila and Hornbæk, 2012).

From a practical perspective all this research effort towards understanding user experience is being applied in order to produce better digital artifacts and HCI has managed to successfully provide the practitioners with valuable knowledge on how to do so. At the same time, an interesting fact is that many of the people that practitioners are designing for are using cool to describe their extraordinary interactions with the world: "My car is cool", "Her boyfriend is cool", "Skydiving is cool", etc. As a result, cool is a concept that is deeply rooted to the everyday life of our subjects and therefore we believe it will be beneficial both for researchers and practitioners to understand *why* they perceive an object, a person, or an activity as cool.

Consequently, we see a direct relation between HCI and the concept of cool both on a theoretical and on a practical level. Therefore, we propose that by researching on cool, HCI will be able to establish a deeper understanding on our users' experiences with digital artifacts, produce appropriate tools to guide practitioners to design for cool and thus significantly contribute in the creation of better digital artifacts. As a result, in this paper our research aim is to operationalize cool, by understanding what it is and by proposing suggestions on how to use it in practice.

In the rest of the paper we initially discuss how cool has been used in HCI up to now and then we present the results of a literature study on cool where we have identified its basic characteristics. Then we discuss the practical implications of the concept of cool and we present some suggestions on how our findings can be applied from practitioners in order to increase the *coolness* of their designs. Finally, we conclude our paper by suggesting future research directions.

2 COOL IN HCI

We have searched in the literature in order to locate studies that dealt with the concept of cool, both in HCI and IT in general. The vast majority of the papers that use cool they don't refer to it as a concept, but they use it to attract more attention. Towards this end we may find papers that use the word cool as name for a programming language (Chandra et al., 1994).

Recently though, the HCI community touched upon cool as a concept and is tentatively discussing its implications. Holtzblatt et al. (2010) provided us with one of the first attempts to introduce cool to HCI and she presented a list of challenges for cool. We have summarized those challenges into:

- *What?* In order to design for cool we need to know what is it and understand its fundamental characteristics.
- *Which?* Which products are perceived as cool and how can we identify them?
- *How?* How can we design for cool?

Most of the related work that deals with cool as a concept contributes to *what* is cool. One of the first attempts focused on teenagers and presented us with the hierarchy of Cool: "Being Cool", by "Doing Cool Things" and by "Having Cool Stuff" [Read et al. (2011) and Read et al. (2012)]. According to the authors this hierarchy creates a design space where HCI can contribute by producing designs that can be perceived as cool. Furthermore, they argue that cool is defined by six characteristics: rebellious, antisocial, retro, authentic, rich and innovative [Read et al., 2011] and the significance of these cool categories was explored by [McCrickard et al. (2012)]. A different approach regarding what is cool comes from Holtzblatt (2011) where she focuses on cool experiences that bring *joy* in our lives and contribute to our personal motivations for life (accomplish, connection, identity and sensation). Furthermore, she presents an interesting discussion on *how* to design for joy and *how* to put it into action. Finally, there are two papers that present us with techniques on how to identify cool, like the "Cool Card Sort" (de Guzman, 2012) and the "Cool Wall", for collecting insights into cool preferences among teenagers (Fitton et al., 2012a; Fitton et al., 2012b).

All these papers constitute the initial attempts to introduce the cool concept to HCI. Since the volume of research on cool is rather low at this moment, most of the identified challenges are partially addressed. Through this paper we aim to contribute towards overcoming these challenges and in the following sections we present our own findings on *what* is cool as they emerged from a literature review, and we provide practical suggestions both on identifying *which* artifacts are perceived as cool and on *how* to design cool artifacts.

3 WHAT IS COOL?

We have managed to identify three different perspectives on who introduced the concept of cool for the first time. Southgate (2003) argues that Aristotle was the first one that introduced *coolness* in his book "*Nicomachean Ethics*". Pountain and Robbins, (2000) and Gioia (2009) argue that *coolness* was defined in Renaissance Italy by Bardessare Castiglione in his "*Book of the Courtier*" (1516) where he proposed *sprezzatura* as a way of life (suggesting that we have to cultivate an appearance which allows us to be perceived as doing extremely difficult actions, effortless). Finally, some suggest that *coolness* was initially performed by African warriors as a way to appear detached in the face of danger (Pountain and Robbins, 2000; Thompson, 1973; MacAdams, 2001).

Independently from its origin, modern cool (cool as we know it today) appeared among black slaves in USA and it was slowly adopted by the majority of black individuals as a "strategic style" that allows black people to tip society's imbalanced scales in their favor (Majors and Billson, 1993) and to potentially transcend oppressive conditions and express themselves as men (Majors, 2001). Contemporary cool (cool as a characteristic of counterculture) became the means that small groups of black Jazz musicians resisted against the dominant white culture by encompassing a different way of dressing and behaving. This rebellious approach both against the dominant white culture and the old-fashioned way of performing Jazz, urged a number of Jazz musicians like Charlie Parker, Dizzy Gillespie, Thelonious Monk, Billie Holiday, etc., to contribute to a new, cooler style of Jazz: Bebop. The same essential elements that created Bebop appeared also in writing, with Beat authors like Jack Kerouac, painting with Jackson Pollock and later with Andy Warhol, and theatre with the creation of the "Living Theatre" from Judith Malina and Julian Beck (MacAdams, 2001).

Slowly cool transformed from a characteristic of people that belonged to counterculture (contemporary cool) into a significant attribute of mainstream culture (modern cool) and it still affects mass society through music, movies, books and products in general. As a result, this characteristic of a small rebellious community, it is now an attitude that is shared by most young people, and to certain extent by the parents as well (Poynor, 2000). At this moment we have to stress that cool was not a purely American phenomenon, as we observe similar transformations in other places too, for example in UK (Levy, 2002).

In our everyday life we use cool to describe people, objects and activities. When we assign the word cool, for example to an object, we basically *perceive* some of its characteristics as cool at a specific moment and inside a social context (Figure 1). If the context changes then the same characteristics can be perceived as uncool. As a result there is interplay between an individual and an object, inside a social context.

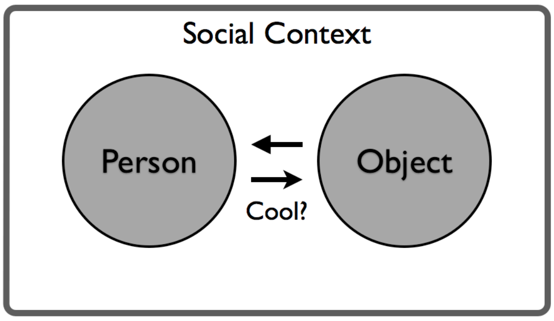


Figure 1: Person-object interplay inside a social context.

The first fundamental issue with this interplay is to understand if cool describes a set of characteristics that belong to the object:

> *"Cool is a quality of people, not of objects. Objects can only said to be cool as much as cool people use them."* (Southgate, 2003)

> *"Cool is not something that inheres in artifacts themselves, but rather in people's attitude to them."* (Pountain and Robbins, 2000)

From these quotes it is evident that there are not specific characteristics in an object that constitute it as cool. Pountain and Robbins (2000) and Southgate (2003) argue that cool is a property of the individual (it is a personal attitude) and inside a certain social context some of the object's characteristics are *perceived* as cool. If hypothetically the color black is cool when we observe it on a very expensive car that does not mean that all black objects will be perceived as cool too. Furthermore, Pountain and Robbins (2000) and Southgate (2003) provide us reasoning on how this cool perception is created: if cool people decide that an object is cool then the majority of the rest of the people will eventually perceive it as cool too.

> *"Cool is a phenomenon that we can recognize when we see it."* (Pountain and Robbins, 2000)

Another fundamental characteristic of this interplay is that we instantly know if an object, a person, or an activity is cool the moment we see it (Pountain and Robbins, 2000; Gladwell, 1997). As a result, we don't need to think which are the characteristics of an object we perceive as cool; we instantly decide on that. The perception of cool is immediate.

> *"Cool is an expression of a community."* (Thompson, 1973)

> *"Coolness is a set of shared meanings (e.g. language, self-presentation, artistic expression, values, attitudes) within a peer group which signify group affiliation."* (O'Donnell and Wardlow, 2000)

> *"Cool is a militant act, a way of staying below the radar screen of the dominant culture without loosing the respect of one's peers."* (MacAdams, 2001)

Finally, the third characteristic of this interplay is related to social context, or the set of shared meanings (O'Donnell and Wardlow, 2000). The social context that influences the perception of cool is decided, defined and reshaped inside a group of peers. Individuals are cool only if they are accepted and bestowed by their peers (Connor, 2003; MacAdams, 2001; Gioia, 2009) and the same stands for the objects that the group accepts as cool. The groups identify themselves both by adopting a unique manifestation on what is cool (O'Donnell and Wardlow, 2000) and also by their strong rejection of what is not cool (Bird and Tapp, 2008). Consequently, the constant changes that happen inside a group in relation to what is cool or not act as a mechanism that ties a group together and demonstrates its distinctiveness from the rest of the groups and the mainstream culture (Saxton, 2005). Furthermore, this process of defining what is cool is dynamically constructed and constantly negotiated inside the group (Rodkin et al., 2006).

From our literature study it was evident that cool can be decomposed into (Figure 2): *Cool personalities* and *cool styles* (or *inner cool* and *outer coo*l respectively, according to Nancarrow et al., 2001). Cool personalities possess some specific characteristics that make them cool and they are expressed by adopting a cool style. Although MacAdams (2001) argues that cool personality and cool style are inseparable we will use this distinction, for the purpose of this paper, as a way to decompose cool and we will provide details on our findings regarding the cool personality and the cool style in the following sections.

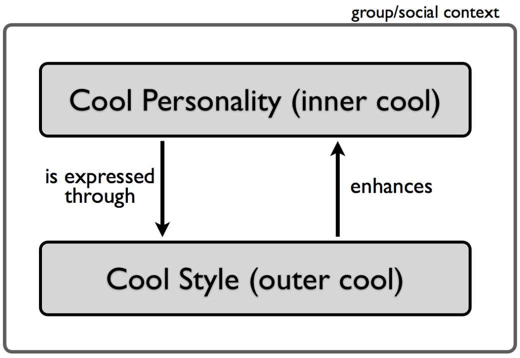


Figure 2: Cool personality and cool style.

4 THE COOL PERSONALITY

Although the perception on what constitutes a cool personal style may change rapidly, even during short periods of time, there are specific attributes that characterize a cool personality, which remain constant and independent from the style they currently adopt. In the following paragraphs we will present these characteristics.

> *"Cool is a permanent state of private rebellion."* (Pountain and Robbins, 2000)

Pountain and Robbins (2000) argue that a cool personality is characterized by a *permanent state* (permanent because when someone enters this state it never leaves) of *private rebellion* (private because cool is not transformed to a mass, group resistance to any kind of authority and/or the conformity imposed by the mainstream culture, but it instead remains a personal "battle" characterized by detachment and irony). The authors move also one step further and they propose that the cool personality is defined by three personality traits: narcissism, ironic detachment and hedonism (Pountain and Robbins, 2000). In fact during the 50's cool individuals embraced an apolitical stance in favor to personal development and pleasure seeking that can be observed even today (Pountain and Robbins, 2000; MacAdams, 2001). Furthermore, this pleasure seeking behavior can be so profound that it may reach the level of self-destructiveness and even flirting with death and it is often related to "live for today" philosophy (Pountain and Robbins, 2000). Additionally, this strong focus on personal development and pleasure inside a group of peers is also expressed by the individual's resistance to anything related to the norms imposed by society, or any value that contradicts the values of the group. Thus, cool personalities express a high level of antisocial behavior.

"Coolness means poise under pressure and the ability to maintain detachment, even during tense encounters." (Majors and Billson, 1993)

At the core of the cool personality lies the ability to appear emotionally neutral, disengaged and nonchalant (Stearns, 1994). The cool personalities are detached and maintain this detachment and the sense of being-in-control in every situation they might encounter. To do so they adopt a calm, ironic pose that enhances the perception of having things under control (Pountain and Robbins, 2000; MacAdams, 2001). Through this way, cool acts as a mask, or as an emotional mantle (MacAdams, 2001; Stearns, 1994). No matter how difficult an unexpected situation might be, or how dangerous an action might be one should maintain his/her cool and his/her emotional neutrality and posture without showing any feelings, such as anger, or fear, and it does not matter if s/he does the right, or wrong thing as long as it is done the right way (MacAdams, 2001). Furthermore, besides unexpected situations, cool is strongly related to the way expected/normal activities are performed. Activities, that are usually difficult to perform, must appear as trivial and easy. The more easy they appear the more perfect they look and the cool personalities appear in control, as "knowing what they are doing". Instances of such attitude can be found in sports, when for example Michael Jordan got his "Air" nickname, because he was cool enough to be perceived as walking on air while playing basketball (Pountain and Robbins, 2000; Majors, 2001). Furthermore, the cool personality expresses this behavior not only in difficult/dangerous situations, but also general in life (Moore, 2004; Nancarrow et al., 2001).

"Authenticity is the truest hallmark of cool behavior." (Southgate, 2003)

Cool is strongly related to authenticity. Cool personalities resist and oppose to anything that they consider as fake, or copied. They perceive music, books, cars, etc., as cool as long as they are authentic, out of the ordinary and part of counterculture. But even these cool objects, or activities become uncool as soon as they enter the mainstream culture, since they become unauthentic and corrupted (Pountain and Robbins, 2000). Taking this attitude into consideration there is a strong relation between cool and innovation, since cool personalities are excited with novel things and constantly seek things that challenge the mainstream (Pountain and Robbins, 2000).

"Cool itself is intrinsically judgmental and exclusive." (Pountain and Robbins, 2000)

The cool personalities are judgmental towards the rest of the world, or the "sheep" (Nancarrow et al., 2001), which do not know what is cool at the moment and follow what the mainstream culture dictates. They are also highly confident that they know what is cool (Charles, 2002) and they find satisfaction through exclusive objects or activities that the "sheep" do not know: for example by paying a visit to an unknown bar.

In summary, cool is a personal attitude toward objects, people and activities from cool personalities which are characterized as rebellious against the dominant culture and any form of authority (ranging from parents to governments), antisocial, always trying to appear in control in every situation, making difficult activities to become/appear as easy, embracing authenticity and exclusivity and strongly tied to a group of peers, while investing in personal development and pleasure by projecting an emotional neutrality. Table 1 summarizes the fundamental characteristics of cool personalities.

cool personality
rebellious
antisocial
embraces authenticity/innovation
seeks exclusivity
appears/is in control
makes things become/appear easy
seeks pleasure/personal development
detached/emotionally neutral
strongly tied to a group

Table 1: Cool personality characteristics.

5 THE COOL STYLE

Through our review on cool it was evident that cool is not a set of characteristics that belong to an object. In fact cool is an attitude that belongs to people and is demonstrated through style. According to Gioia (2009) a cool style is characterized by somebody's attire, accessories, language and pose, which are cool only if they are accepted by one's peers. One's attire, language and pose are elements of one's personality, but accessories are more related to his style (even though this style is defined by the cool personality). Cool and style are inseparable (MacAdams, 2001) and thus in order to design cool digital artifacts it is important to design artifacts that concurrently match a cool personality and fit a cool, personal style. The difficulty in this approach is that cool style is always evolving and changing, influenced both by the cool personality and the social context. Therefore we argue that in order to define a cool style we must describe the extreme forms that it can take. We have applied this approach to our literature study and a set of cool style characteristics have emerged. In the following paragraphs we will present our findings and we will demonstrate them through examples.

5.1 Cool style is: minimalistic.......flamboyant

Cool style can be instantiated both in a minimalistic, or a flamboyant form (or elegant/gigantic according to Pountain and Robbins, 2000) depending on the social context and the perception of cool that a group of peers has. A typical example can be found in clothing. While on one hand some people adopt a style that focuses on elegance and minimalistic-simplistic elements (for example black tie suits), on the other hand we might see groups of people that prefer flamboyant clothing with flashy details that immediately attract attention (for example hip-hop buggy pants and gigantic gold chains, or “rock” leather jackets covered with steel nails). Thus, at the same time two groups of people (or even the same group if we observe how differently Bebop musicians were dressed) are using extremely different accessories that they perceive as cool: flashy, flamboyant ones, or minimalistic and elegant.

5.2 Cool style is: expensive.......cheap

The cool personality is related to authenticity, exclusivity and rebelliousness. In order to express these characteristics though a cool style, individuals usually acquire specific objects that can be either expensive, or cheap. A typical example of an expensive cool style can be found in cars: Ferrari’s, Porsche’s and other expensive cars are perceived by most people as cool, since they are exquisite and very difficult for common people to acquire. On the other hand, authentic and exquisite objects can also be cheap, as for example an old VW Beetle, which was transformed from an “ugly” German car to a symbol of counterculture (Imseng, 2011). Another example of this duality can be also found in shoes, where for example, it is currently considered as cool to wear expensive clothes with sneakers, such as All Star shoes, which is a cheap, but nevertheless a cool item.

5.3 Cool style is: beautiful.......ugly

Cool style can be beautiful, shining and adorable and cool can be ugly and dark. Both style instantiations can appeal even to the same people in different contexts. A typical example can be found in music. While many teenagers enjoy hearing melodic, pop songs from artists like Adele, at the same time we currently observe the dominance of another, more dark and more “ugly” style of music, which is also perceived as cool: dubstep. One of the dubstep definitions provided by Urban online Dictionary is: “The purest form of musical crack cocaine; most effective when ingested intra-subwooferly. Found almost universally at dance parties attended by cool people everywhere”. For many, dubstep is ugly, violent and not even music, but for the cool personalities it is the ultimate vehicle for personal expression, in a similar way as Rave, or Grunge music was during the ‘90s. Towards this end, a known example of a whole generation intentionally adopting ugly stuff as a way to rebel against the dominant culture can be found during the ‘90s among the Grunge generation. Grunge style adopted the ugliest clothes there could be found on every shop in order to project its revolt against marketing (Rushkoff, 2008). This relation also to “ugliness” is also demonstrated by the usage of the word “sick” to describe something excellent and wonderful, which entered the mainstream slang at about the same time as grunge music, or the use of the word "wicked" with the same connotation previously (Pountain and Robbins, 2000).

5.4 Cool style is: innovative…….retro

Cool adores innovation and new exciting stuff and anything that is moving towards the future (MacAdams, 2001). A cool style can adopt the latest technological innovations in order to stand out from the masses and the rest of the people who do not know what is cool yet. A typical example of considering new technologies as cool can be found in the adoption of Apple products, especially on the early days where the company was unknown to the majority of people. At the same time we observe that some groups of individuals are opting for retro objects that clearly signify that they come from another age (Nancarrow and Nancarrow, 2007). Thus, for example, we can observe an increase in usage of old cell phones (Kirman, 2012), which are primitive in comparison to the modern ones, but nevertheless their retro characteristics are perceived as cool. The same tendency towards retro can be found also in clothing with a booming market that deals with vintage products, or even in computer gaming where there are specific communities that focus on playing, maintaining and expanding old computer games, despite the fact there are modern versions of them (for example the Championship Manager 2001/2002 online community).

5.5 Cool style is: illicit…….licit

Cool personalities seek pleasure and in this quest they often reach the level of self-destructiveness because pleasure can sometimes found to things/activities that are not legal and thus they often flirt with criminality and danger (Pountain and Robbins, 2000; Gioia, 2009; MacAdams, 2001). For example, it is a fact that many Bebob Jazz musicians were excessively using heroin (MacAdams, 2001). Nevertheless, this interplay between licit and illicit is inherent in a cool style, with one requirement: flirting with criminal behavior is cool, as long as the individual is in control of the situation; then it is not cool anymore. For example there are research data that children characterize as cool their peers that are at the same time popular and aggressive (tough) (Rodkin et al., 2006). Those individuals according to Horton et al. (2012) are perceived as cool as long as they lie in zone of non-harmful rebelliousness and they become uncool when they reach the level of criminal rebelliousness. A typical example of this adoration on illicit objects can be also found in music video clips where "gangsta-rap" is enhanced by showing off guns. On the other hand, licit objects and activities can of course be considered as cool, as long as they match the style of the individual.

Table 2 summarizes the different characteristics that a cool style may have.

cool style
minimalistic.....flamboyant
expensive.....cheap
beautiful.....ugly
innovative.....retro
illicit.....licit

Table 2: Cool style characteristics.

6 COOL IMPLICATIONS

Up to this point we focused on *what* is cool and through a literature study we identified that cool is a property of people and not of objects and we also presented the fundamental characteristics of cool personalities and cool styles. In the following subsections we will apply a more pragmatic approach and discuss how practitioners can take advantage of the concept in order to design for cool.

Let's imagine a scenario where a group of designers and developers wants to create an application, which will be perceived as cool. The first thing they need to do is to understand *what* is cool and which of its characteristics can be included to their design. Since designers often use existing artifacts as inspiration, then in order to find out *how* to instantiate these characteristics it is important to have an objective way to understand *which* of the existing artifacts are perceived as cool.

6.1 Which artifacts are perceived as cool?

In order to answer *which* of the existing digital artifacts are perceived as cool practitioners and researchers need to understand how cool is created in general.

The first step in the creation of cool comes from cool personalities, which are the Innovators (Taylor, 2009) and they are about 2-3% of the population (The Merchants of cool, 2001). They either envision something new and innovative (Frank, 1997), or they personalize existing objects and make them cool (MacAdams, 2001). Thus, innovators are the ones that create and identify cool (Gladwell, 1997; MacAdams, 2001). The objects the Innovators use and the activities they do, are adopted by the Trendsetters (the ones who mimic Innovators), then the process continues with the Early Adopters, and finally objects and activities become part of the mainstream culture (Taylor, 2009; The Merchants of Cool, 2001). Cool is adopted by one group after the other and in this specific order (Rogers, 2003). As soon as these objects reach the mainstream they are immediately perceived as uncool by the Innovators (Gladwell, 1997). At this moment the rest of the people still perceive them as cool, but most of them will eventually perceive them as uncool in the end. Very few of these objects though reach the level of ***universally cool*** (Schiller, 2012), or ***classic*** (Nancarrow et al., 2001) and even though they are absorbed by the masses, they are still perceived as cool by the majority of people. Such examples are Ray-Ban sunglasses, or Absolut Vodka (The Merchants of Cool, 2001). Figure 3 depicts the different transformations of an object in relation to cool.

From Figure 3 it is evident that in order to identify ***which*** objects will be perceived as cool by the mainstream, we need to identify ***which*** objects Innovators perceive as cool at the moment. If we consider we know from the literature that people select specific products because they project elements of their personality (Jordan, 1997; Phau and Lau, 2001), then we propose that our identified cool personality characteristics can be used as the starting point for locating which of the digital artifacts are perceived as cool by the Innovators. Thus, if Innovators perceive a digital artifact as rebellious, exclusive, authentic and projecting some or all of the characteristics of the cool personality, then the rest of the people will also perceive it as cool after some time. The challenge for practitioners is that they need to be careful on ***who*** they ask to assess the coolness of an artifact as different types of people will give different responses depending if they are Innovators, or not. A tool for locating Innovators is the Trendsetting Questionnaire (TDS-K) from Batinic et al. (2008).

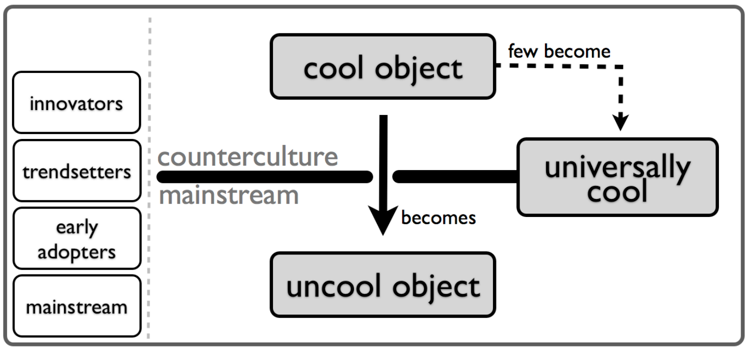


Figure 3: Cool transformations.

Towards this end, HCI should focus on providing the practitioners with the necessary instruments to objectively measure the coolness of existing artifacts. If we consider that cool is a personal attitude and that there are many available techniques in measuring attitudes (Oppenheim, 2001), we argue that HCI should aim in producing an instrument to measure ***perceived coolness***. We believe that our identified cool personality characteristics can contribute to the creation of this instrument and until it is produced they can also guide practitioners into defining possible cool artifacts. For example we believe that such candidate cool artifact is Ekkomaten (Basballe, 2012), which is a digital, historical guide for the Danish city of Aarhus that seems to violate many established design guidelines: it is bulky, unconventional, some can even say ugly. Yet, interacting with it resulted in exciting experiences, especially for all of us who became familiar with it during NordiCHI 2012 conference. We can assume that Ekkomaten was perceived as cool as it was rebellious in comparison to typical digital guides, authentic, exclusive, it allowed its users to be in control and its style was flamboyant, retro and somehow ugly. Of course more research data are needed for concrete results, but nevertheless by using the cool personality and style characteristics practitioners can have an indication on which artifacts are perceived as cool.

6.2 How can we design for cool?

The third challenge is *how* to approach the cool concept as a practitioner and use it in order to produce artifacts that are perceived as cool. In relation to cool we believe that practitioners:

1. Can create *inceptive cool experiences* by designing innovative artifacts that do not exist before, or by envisioning new uses for an artifact that nobody had thought before.
2. Can transform existing artifacts by *cooling them up* (make them perceived as cool).

In the following subsections we present our suggestions on these two types of design activities.

6.2.1 Creating inceptive cool experiences

Creating inceptive cool experiences is the most challenging task for practitioners, because it is a question that has broader implications that only designing for cool. We are basically asking how to create *disruptive technologies* (Christensen, 1997), which permanently change the way we experience the world. For example, how can we produce digital artifacts that create the same cool feeling as holding an iPad for the first time? If we try to evaluate the iPad in relation to cool we will observe that when it was released: it was authentic, rebellious, exclusive, it allowed its users to be in control and to appear as making difficult tasks as easy and it gave pleasure to its owner. Additionally, it was instantiated through a minimalistic, expensive, beautiful and innovative style. Of course some might say that our cool personality and cool style characteristics are easier to apply in order to evaluate something that already exists, than using them to design something new. We agree that they do not constitute a design method for cool, but we believe that they create a design space that cool can emerge. At the same time, we suggest that practitioners should take advantage of a core characteristic of the cool personality: it's tight relation to a group of peers. We propose that in order to produce inceptive cool experiences with digital artifacts practitioners need to study in detail the group of peers they are designing for and produce artifacts that have *value* and *meaning* (Hallnäs and Redström, 2002) for that specific group. In order to do this, it is important not only to be engaged to fieldwork but to do *better* fieldwork by increasing the depth and the duration of their field studies (Kjeldskov and Paay, 2012).

6.2.2 Cooling existing artifacts up

Let's imagine one of our typical users who owns a modern mobile phone and a tablet. We know from user experience literature that this user did not select these devices only because they are efficient and effective. He selected the specific devices because they offer to him something more: he perceives them as beautiful, desirable, cool. Furthermore, he usually applies the same approach throughout his everyday life and thus he does not buy the first pair of shoes he sees, but he selects the one pair that matches his style. As a result, we believe that creating *inceptive cool experiences* is equally important as *cooling* existing digital artifacts *up* and the challenge for practitioners and researchers is to transform the hundreds of the available digital artifacts that do not offer the added value that users enjoy in their personal life (for example the pleasure of having an iPad, or a cool pair of shoes) into more cool and desirable artifacts. Examples of artifacts that need to be cooled up are the banking systems that often have anachronistic interfaces that each day thousands of employees interact with.

In order to understand the possible ways of cooling existing digital artifacts up, we believe that it is useful to become familiar with *how* other domains have managed to create objects that are perceived as cool. Since you cannot sell cool per se, but you can sell the "idea of cool" (MacAdams, 2001) as a label for fashionable hipness (Gioia, 2009), big brands and advertisement companies constantly pump into the mainstream what the Innovators perceive as cool, in order to enhance consumerism. For business, this quest of seeking for cool never ends (Gladwell, 1997; The Merchants of Cool, 2001) and in fact they are so successful that they manage to create objects that directly match a cool personality and fit inside a cool style. For example, during the 60's American business successfully offered the public authenticity, individuality, difference and rebellion (Frank, 1997). A typical example of such targeted consumerism are the products that take advantage of the "cool rebel" Ernesto Che Guevara: from cheap T-shirts to expensive Louis Vuitton handbags. Frank (1997) describes this process in detail and the same phenomenon is elegantly presented in the BBC4 documentary "The century of the Self" (Curtis, 2002).

Thus, the interesting question for practitioners is: if business managed to successfully *cool things up,* can they do the same? We propose four distinct approaches that practitioners can follow in order to design for cool: 1) focus on the Innovators, 2) focus on the mainstream, 3) focus on universally cool items, and 4) allow cool to emerge. In the following subsections we describe these approaches in detail.

6.2.2.1 Focus on the Innovators

We believe that the first approach that practitioners can adopt in order to produce artifacts that are perceived as cool is to do exactly what the business does and perform *coolhunting*. In order to grasp what Innovators perceive as cool businesses perform coolhunting, aiming to identify a subcultural phenomenon before it reaches the mainstream culture, in order to to be the ones that will take advantage of it (Frank, 1997; The Merchants of Cool, 2001). Basically, they observe what the Innovators are doing/wearing/listening at the moment and they incorporate these to their designs, as they know that the mainstream will eventually follow the Innovators. As a result, coolhunting is more about observing and incorporating and less about inventing (O'Donnell and Wardlow, 2000). The challenges though for successfully performing coolhunting are: a) only Innovators can identify/do cool (Gladwell, 1997; MacAdams, 2001) and thus practitioners need to include Innovators into their design teams, b) coolhunting is expensive (Bird and Tapp, 2008). Nevertheless, we believe that practitioners can invest more resources on coolhunting since: a) it can fit really well with the development circle of digital artifacts since it provides the necessary time to built an artifact, which will be perceived as cool by the mainstream when it will be released, and b) it is proven from other domains that it is a really successful technique in order to design for cool.

6.2.2.2 Focus on the mainstream

A second approach for practitioners on how to design perceived cool digital artifacts is not focus on Innovators through coolhunting but to observe objects, activities and people that are at this moment considered as cool by the mainstream. By excessively studying what the mainstream considers as cool at this moment practitioners can have a variety of inspirational ideas on which style their designs should adopt in order to be cooled up. For example, there is a large volume of literature (e.g. Evans, 2007) that presents to us cool cars, hotels, webpages, etc. We argue that during design sessions practitioners can identify specific design characteristics from those perceived cool objects and try to incorporate them to their designs. The disadvantage of this approach, though, is that it requires a fast development circle since the perception of what is cool can change rapidly. As a result this approach could be used in order to make minor adjustments, for example in the interface of an application.

6.2.2.3 Focus on universally cool items

Our third suggestion on how to design for cool is to compare our designs with *universally cool*, or *classic* objects. These objects share the common characteristic that they are perceived as cool by the vast majority of our subjects, independently from where they belong (mainstream or counterculture) and if they are Innovators or not. Thus, these objects somehow stand beyond the influences of social context and are perceived as cool almost universally. We propose that by asking their subjects to compare their designs with these universally cool items (an approach which can be really successful according to Oppenheim, 2001), practitioners will be able both to evaluate their designs in relation to cool, and to identify possibilities for new designs. In order to identify these universally cool items practitioners can apply the previously described approaches for finding *which* products are perceived as cool, but they should focus not only to Innovators, but to the majority of people.

6.2.2.4 Allow cool to emerge

Finally, our fourth suggestion is somehow the reversed from the previous three. Instead of trying to produce digital artifacts that will be perceived as cool, practitioners can provide the freedom to their users to personalize them. Thus, instead of trying to create a "cool jacket", they can allow the users to "attach pins to the jackets" they offer to them, and thus make them as cool as they perceive. For example, we know that many users personalize their mobile phones (Moggridge, 2007) in order to match their style. By allowing this freedom to exist in all levels of their digital artifacts (for example the interface of the applications) practitioners will not produce cool per se, but they will allow cool to be created by users.

7 CONCLUSIONS

In this paper we systematically decomposed the concept of cool through a literature review in order to contribute to the three challenges we have identified from related work: *what* is cool, *which* objects are perceived as cool and *how* we can design for cool. We approached cool from a cool personality and a cool style perspective and we have presented their fundamental characteristics. In summary, cool personalities (the Innovators) are antisocial, rebellious, seeking pleasure, personal development, innovation, authenticity and exclusivity, by appearing as being-in-control and as performing difficult tasks as easy, without showing emotions, while being strongly tied to a group. These personalities are expressed through a cool style, which can be minimalistic/flamboyant, expensive/cheap, beautiful/ ugly, innovative/retro and illicit/licit depending on the perception that their group of peers has in relation to cool. Furthermore, we have presented our suggestions on how these characteristics can be applied to practice in order to produce digital artifacts, which will be perceived as cool by focusing on Innovators, focusing on the mainstream, focusing on universally cool items and by allowing cool to emerge.

Overall, we believe that this paper contributes on having a better understanding of cool as a concept and we will conclude it with some possible research directions that we believe are interesting for HCI. Since cool has already been introduced to HCI then one of the first directions for research should be the investigation of cool's relation to user experience constructs (Bargas-Avila and Hornbæk, 2011; Bargas-Avila and Hornbæk, 2012). Is there any relation between cool and usability, cool and hedonic and pragmatic values, or cool and the rest of user experience constructs? For example, we know that beautiful objects are perceived as usable (Tractinsky et al., 2000), but are cool objects perceived as usable too? We strongly believe that such a research effort will be beneficial not only in understanding cool, but also in better understanding user experience.

Additionally, we believe that it is important to study in detail the effect of first impression in relation to cool. Since most of the literature agrees that we instantly decide if an object is cool and if we consider that HCI has research findings on the significance of first impression on user experience (for example Lindgaard et al., 2006, provided valuable data on the effect of first impression in relation to visual appeal), we propose that it is worth investigating the impact of first impression in relation to cool. By researching more on the subject we argue that we will be able to gain a deeper understanding on the parameters that affect user experience and we will be able to assist practitioners in taking into consideration these parameters during the development processes.

Finally, we argue that it is important to investigate the fact that cool is strongly influenced by social context and as a result the perception of cool is so diverse in different groups, places, cultures and generations. We believe that by studying these different contexts we will be able to have a deeper understanding on the parameters that affect the perception of cool and we will provide better suggestions for design to practitioners.

Overall, we argue that cool is a valuable construct for HCI as it will provide us with new insights both on how to design better artifacts and on what is user experience. Since cool is making its initial steps into the HCI field we believe that a lot of research needs to be done, and we proposed some research and design directions that HCI could take. We suggest that our proposed cool personality and style characteristics can be the starting point towards these directions. As a future work we plan to continue researching on cool and start addressing the research and practical challenges we pointed out.

ACKNOWLEDGMENTS

The research behind this paper was partly financed by the Danish Research Councils (grant number 09-065143).

REFERENCES

Bargas-Avila, J.A., and Hornbæk, K. Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. In Proc. CHI 2011, ACM Press (2011), 2689-2698.

Bargas-Avila, J.A., and Hornbæk, K. Foci and Blind Spots in User Experience Research. Interactions, 19, 6 (2012), 24-27.

Basballe, D.A., Breinbjerg, M., and Fritsch, J. Ekkomaten: an auditory interface to the 18th century city of Aarhus. In Proc. NordiCHI 2012, ACM Press (2012), 742-745.

Batinic, B., Wolff, H.G., and Haupt, C.M. Construction and Factorial Structure of a Short Version of the Trendsetting Questionnaire (TDS-K): A Cross-Validation Using Multigroup Confirmatory Factor Analyses. European Journal of Psychological Assessment, 24, 2 (2008), 88-94.

Bird, S. and Tapp, A. Social Marketing and the meaning of Cool. Social Marketing Quarterly, 14, 1 (2008), 18-29.

Championship Manager 2001/2002 online community: http://www.champman0102.co.uk/

Chandra, R., Gupta, A., and Hennessy, J.L. COOL: An object-based language for parallel programming. Computer, 27, 8 (1994), 13-26.

Charles, G. Cool for Cash. Marketing Week, 25, 33 (2002), 24.

Connor, M.K. What is cool?: Understanding Black Manhood in America. Agate Publishing (2003).

Christensen, C.M. The innovator's dilemma: when new technologies cause great firms to fail. Harvard Business School Press (1997), Boston, Massachusetts.

Curtis, A. The century of the Self. BBC 4 documentary (2002).

De Guzman, E.S. “Cool” Card Sort: Adapting Desirability studies to assess “Cool”-ness. ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012.

Evans, D. Coolhunting. A guide to high design and innovation. Southbank Publishing (2007), London.

Fitton, D., Horton, M., Read, J.C., Little, L., and Toth, N. Climbing the Cool Wall: Exploring Teenage Preferences of Cool. Ext. Abstracts CHI’12, ACM Press (2012), 2093-2098.

Fitton, D., Read, J.C., Horton, M., Little, L., Toth, N., and Guo, Y. Constructing the Cool Wall: A Tool to Explore Teen Meanings of Cool. PsychNology, 10, 2 (2012), 141-162.

Frank, T. The Conquest of Cool. Business Culture, Countercurture, and the Rise of Hip Consumerism. The University of Chicago Press (1997), Chicago.

Gioia, T. The birth and death of the Cool. Speck Press (2009), Golden Colorado.

Gladwell, M. The Coolhunt, New Yorker, March 17, 1997. Retrieved from: http://www.gladwell.com/1997/ 1997_03_17_a_cool.htm

Hallnäs L., and Redström J. From use to presence: On the expressions and aesthetics of everyday computational things. Transactions on Computer-Human Interaction, 9, 2 (2002), 106-124.

Hassenzahl, M. and Tractinsky, N. User Experience – a Research Agenda. Behaviour & Information Technology, 25, 2 (2006), 91-97.

Holtzblatt, K., Rondeau, D.B., and Holtzblatt, L. Understanding “Cool”. Ext. abstracts CHI ‘10, ACM Press (2010), 3159-3162.

Holtzblatt, K. What Makes Things Cool? Intentional Design for Innovation. Interactions, 18, 6 (2011), 40-47.

Horton, M., Read, J.C., Fitton, D., Toth, N., and Little, L. Too Cool at School – Understanding Cool Teenagers. PsychNology Journal, 10, 2 (2012), 73-91.

Imseng, D. Think Small. The story of the world’s greatest ad. Full Stop Press (2011), Uster, Switzerland.

Jordan, P.W. Products as personalities. In (ed.) M.A. Hanson, Contemporary Ergonomics. Taylor & Francis (1997), 449-453.

Kirman, B. Hipsters, Trendies and Rebels: If Fun is Cool, is Game Design Cool Design? ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012.

Kjeldskov J., and Paay J. A longitudinal review of Mobile HCI research Methods. In Proc. MobileHCI 2012, ACM Press (2012), 69-78.

Levy, S. Ready, Steady, Go. Swinging London and the invention of Cool. Fourth Estate (2002).

Lindgaard, G., Fernandes, G., Dudek, C., and Brown, J. Attention web designers: you have 50 milliseconds to make a good first impression! Behaviour and Information Technology, 25, 2 (2006), 115-126.

MacAdams, L. Birth of the Cool. Beat, Bebop, and the American Avant-Garde. The Free Press (2001), New York.

Majors, R., and Billson, J.M. Cool Pose: The Dilemmas of Black Manhood in America. Touchstone (1993), New York.

Majors, R. Cool Pose: Black Masculinity and Sports. In (eds) S.M. Whitehead and F. Barrett, The Masculinities Reader, Polity Press (2001), Oxford.

McCrickard, D.S., Doswell, F., Barksdale, J., and Piggott, D. Understanding Cool in Computing for African-American Youth. ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012.

Moggridge, B. Interaction Design (IDEO). Innovation Forum 2007, Potsdam, Germany. Available from: http://interface.fh-potsdam.de/innoforum/english/ 09_videos.php [Accessed 31 May 2013]

Moore, R.L. We’re Cool, mom and dad are swell: Basic slang and generational shifts in values, American Speech, 79, 1 (2004), 59-86.

Nancarrow, C., Nancarrow, P., and Page, J. An analysis of the concept of cool and its marketing implications. Journal of Consumer Behaviour, 1, 4 (2001), 311-322.

Nancarrow, C., and Nancarrow, P. Hunting for cool tribes. In (eds) Cova, B., Kozinets, R.V., and Shankar, A., Consumer Tribes, Butterworth-Heinemann (2007), 129-143.

O'Donnell, K.A., and Wardlow, D.L. A Theory on the origins of Coolness. Advances in Consumer Research, 27 (2000), 13-18.

Oppenheim, A.N. Questionnaire Design, Interviewing and Attitude Measurement. Continuum (2001), London.

Phau, I. and Lau, K. Brand personality and consumer self-expression: Single or dual carriageway? The Journal of Brand Management, 8, 6 (2001), 428-444.

Pountain, D., and Robbins, D. Cool Rules. Anatomy of an Attitude. Reaction Books (2000), London.

Poynor, R. Cool. Optic Nerve, 54, 3 (2000), 64-65.

Read, J.C., et al. Understanding and Designing Cool Technologies for Teenagers. Ext. Abstracts CHI '11, ACM Press (2011), 1567-1572.

Read, J.C., Horton, M., and Fitton, D. Being Cool-Getting Persona-l. ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012.

Rodkin, P.C., Farmer, T.W., Pearl, R., and Van Acker, R. They're Cool: Social Status and Peer Group Supports for Aggressive Boys and Girls. Social Development, 15, 2 (2006), 175–204.

Rogers, E.M. Diffusion of Innovation, 5th ed. Free Press (2003), New-York.

Rushkoff, D. The Pursuit of Cool. (2008) Available online: http://www.rushkoff.com/articles-individual/ 2008/10/28/the-pursuit-of-cool.html

Saxton, G. Collections of Cool. Young Consumers, 6, 2 (2005), 18.

Schiller, J. Who is the coolest of them all. ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012.

Southgate, N. Coolhunting, account planning and the ancient cool of Aristotle. Marketing Intelligence & Planning, 21, 7 (2003), 453-461.

Stearns, P.N. American Cool: Constructing a Twentieth-century Emotional Style. New York University Press (1994), New York.

Taylor, M. The book of Cool. Running Press Book Publishers (2009), Philadelphia.

The Merchants of Cool. Frontline Documentary (2001). Available online: http://www.pbs.org/wgbh/pages/ frontline/shows/cool/view/

Thompson, R.F. An aesthetic of Cool. African Arts, 7, 1 (1973), 41-91.

Tractinsky, N., Katz, A.S. and Ikar, D. What is beautiful is usable. Interacting with Computers*,* 13*,* 2 (2000), 127-145.

7.4 PAPER 4: MEASURING THE COOLNESS OF INTERACTIVE PRODUCTS: THE COOL QUESTIONNAIRE

ABSTRACT

Coolness has recently started to be explored as a design goal for interactive products. While research has also started to explore coolness in HCI, we still need to further operationalize the concept and explore how we can measure it. Our primary contribution in this paper is the COOL Questionnaire, which can be used to measure the coolness of interactive products. Through related literature we identified that coolness is decomposed to outer cool (the style of a product) and inner cool (the personality characteristics assigned to it). We identified eleven inner cool characteristics, which were used to perform exploratory and confirmatory factor analyses on datasets produced by 2236 participants that assessed 16 mobile devices. We identify three factors that can measure the inner coolness of interactive products: desirability, rebelliousness, and usability. Our second contribution is the validation of the relationships between inner cool, outer cool on the overall perception of product coolness. We conclude, by discussing our work against related work on coolness and human computer interaction.

1 INTRODUCTION

Designing for coolness has become increasingly important over the last years for practitioners and researchers working within design of interactive products. As such, coolness is now being integrated into interactive design processes and thus become a part of interaction design and HCI. Coolness is frequently used to describe user approbation of various interactive products. For example, several Apple products are often characterized as being cool or perceived as being cool. As pointed out by Sundar et al. (2014) coolness in interactive products has become an essential psychological criterion for designers and developers when creating new systems, applications, interfaces, or devices.

While coolness as a goal of the design process of interactive products is somewhat new, designers have for many years designed for other product qualities, like usability (e.g. effectiveness and efficiency), which is a central property of most interactive products, and more recently designers have focused on user experience of such products, e.g. pleasure (Jordan 1997). While usability and user experience are still extremely relevant for many (and perhaps most) industrial design projects, we are currently witnessing a growing interest in interactive products that people perceive as being cool. This implies that our scope as interaction designers needs to be broadened to include this additional dimension, and for interaction design researchers it presents an opportunity to explore new challenges for the process of design. For example, interaction designers need to be able to measure the coolness of their design ideas.

For many years, we have conducted research on how to design interactive products, and as a result we have developed theories, guidelines, methods, and principles on usability and user experience providing design practitioners with an array of tools and constructs for achieving usability and user experience. Recent research studies have started to investigate coolness as a concept within interaction design, e.g. Fitton et al. (2012a, 2012b), Holtzblatt et al. (2010), Holtzblatt (2011), Read et al. (2011, 2012), and McCrickard et al. (2012). As an example, studies outline techniques for identifying and analysing coolness through card sorting Fitton et al. (2012a, 2012b). However, we still need deeper understandings of the role of coolness in interactive products design and especially we need specific tools or techniques that can be applied in design projects.

Our main contribution in this paper is the design and construction of The COOL Questionnaire, a questionnaire for assessing the inner coolness of interactive products. Informed by previous literature on coolness, we first derive at a series of key characteristics that describe the essence of inner coolness. Based on this, our questionnaire was developed using statistical techniques on exploratory factor analysis and confirmatory factor analysis (Section 5). Furthermore, our second contribution is that we validated a theoretical model on how the perception of product coolness is shaped (Section 6). The result – The COOL Questionnaire – is presented, and the conception of its question items is explained. We conclude by discussing our work in relation to that of others, its relation to user experience evaluation, its generalizability, as well as avenues for further research.

2 RELATED WORK

Within human-computer interaction and interaction design, a number of research studies have investigated the notion of "cool" and it's meaning in relation to interactive product design. This research has contributed to our understanding of the term cool both conceptually and methodologically and also different ways of working with it in design.

In an early study on cool in HCI, Read et al. (2011, 2012) developed a conceptual framework of "Being Cool", by "Doing Cool Things" and by "Having Cool Stuff" facilitating the design of cool artifacts and interactive products for teenagers displaying characteristics such as being rebellious, antisocial, retro, authentic, rich, and innovative (Read et al., 2011). The framework has later been adopted in other research studies, e.g. McCrickard et al. (2012) that investigated challenges and opportunities in designing cool user interfaces for another group of young people.

In 2011, Holtzblatt (2011) discussed the concept of cool and stressed that cool products bring joy to our lives and contribute to our personal feelings of accomplishment, connection with others, identity, and delightful experiences. Methodologically Holtzblatt presents "The

Wheel of Joy" and "The Triangle of Design" as tools for defining the aspects of life and experience that designers should focus on when designing for cool. In a similar manner, de Guzman (2012) and Fitton et al. (2012a, 2012b) present techniques for identifying and analyzing coolness through "Cool Card Sorting", or mapping out people's coolness preferences using the "Cool Wall".

In a recent study, Sundar et al. (2014) presented a detailed literature review leading to a questionnaire for measuring cool. In this research, which is remarkably similar to what we have been doing ourselves, the concept of cool is mapped out and discussed in light of the broad general literature on the topic, leading to the formulation of a series of potential questionnaire items. These are then developed into a questionnaire through a process of exploratory factor analysis and confirmatory factor analysis. Testing the questionnaire concludes that the coolness of a product can be conceptualized as a matter of supporting originality, attractiveness and subcultural appeal.

The above research studies have contributed considerably to our initial understanding of "cool", how to design for it, and how to measure it. However, this research has also strongly highlighted the need for further research. Although early conceptualizations have been put forward, "cool" is still a very difficult concept to operationalize in interaction design. It is difficult to grasp and difficult to measure. As a result, it is very difficult to design for. In response to this challenge, our efforts have, similar to those of Sundar et al. (2014), focused on breaking down the concept of cool into smaller entities, and then using these systematically as building blocks for a questionnaire that can holistically measure the coolness of an interactive product. Furthermore, we move a step further by validating a theoretical model on how people shape the perception of product coolness.

3 UNDERSTANDING COOL

Our understanding of cool is based on an extensive review of literature from interaction design, marketing, and the music/movie industry that we have conducted (Raptis et al., 2013). From this literature we have identified three overall notable principles of coolness: 1) Coolness consists of *inner* and *outer cool*, 2) Coolness is recognized immediately, and 3) Coolness is grounded in people's communities.

3.1 Coolness consists of Inner and Outer Cool

When using the word cool in our everyday life, it is most often to describe some quality that we see in people and in objects, like when we say "James Dean is cool" or "my new phone is cool". This use of the word is quite broad, and in fact we are often using the same word to describe slightly different aspects of people or objects that are "cool".

Coolness can be divided into inner cool and outer cool (Nancarrow et al., 2001). Inner cool is about someone's or something's personality or character. According to Jordan (1997) this can apply to both people and objects because people often assign personality traits to physical things in order to help us understand and relate to them (Janler and Stolterman, 1997; Jordan 1997). Outer cool is about how something presents itself through a certain style in physical appearance. For people this is typically a matter of somebody's clothes, accessories, language and pose (Gioia 2009). For objects it is a matter of similar aesthetic qualities in its design, for example, materials, lines, colors etc. For example, if someone says that James Dean is cool in the movie "Rebel Without a Cause" this does not only refer to the way he looks (outer cool) but also to his perceived personality in that movie (inner cool). Also, when a physical object is characterized as cool, e.g. an iPhone, this does not only refer to its physical appearance (outer cool) but also to its perceived character as a design object (inner cool).

Obviously, inner and outer cool are interrelated to one another where outer cool is typically expressed through cool style, and inner cool conversely enhances the coolness perception of a person's or an object's personality and character (as illustrated in Figure 1).

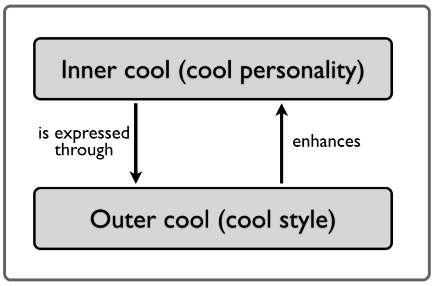


Figure 1. Inner cool and outer cool

Inner cool represents a number of characteristics that are perceived by others as being cool. Selecting broadly from the literature on cool (Raptis et al., 2013) these are characteristics like being rebellious and antisocial, embracing authenticity and innovation, seeking exclusivity, being, or appearing to be, in control, making difficult things appear to be easy, being, or appearing to be, detached and emotionally neutral, and being strongly tied to a group (Fig. 2).

being rebellious
being antisocial
embracing authenticity
embracing innovation
seeking exclusivity
seeking pleasure
seeking personal development
being/appearing in control
making hard things appear easy
being detached or emotionally neutral
being strongly tied to a group

Figure 2. Characteristics of inner coolness

Whereas the characteristics of inner cool are rather high-level, and therefore relatively stable, characteristics of outer coolness are often very specific related to the latest fashion and therefore also fairly ephemeral. Outer cool also vary considerably across cultural background, social groups, etc. Abstracting a bit, however, the literature on cool style and appearance states that expressions of outer coolness often belong to one end of a particular continuum, opposing something else that is then seen as uncool by that particular culture or social group (Figure 3). As an example, one group of teenagers might find cheap retro clothes cool, while another group of teenagers may find that coolness is expressed by the exact opposite.

minimalistic … flamboyant
expensive … cheap
beautiful … ugly
innovative … retro
illicit … licit

Figure 3. Opposing characteristics of outer coolness

3.2 Coolness is recognized immediately

The literature generally agrees that coolness is something we recognize immediately based on what is observed in the moment (e.g. Pountain and Robbins, 2000; Frank 1997) and thus, coolness is not something we need to think about a lot. This is the case for both people and objects, and it is based on an immediate assessment of both their inner and outer coolness. As an example, take a look at the man smoking a cigarette at an outside café in Figure 4. Is he cool? Regardless of one's response to this question, the opinion is immediate, and it is based in a subjective assessment of the inner and outer coolness we perceive. For the authors of this paper he is instantly perceived as not cool. But it is most likely that other people will disagree in an instance – especially if they belong to a similar culture or social group. The same goes for objects. As an example, take a look at the car in Figure 5. Is this a cool car? Again, some people might say "yes", and others might say "no" - but regardless of opinion, it is arrived at immediately.

Figure 4. Cool guy?

Figure 5. Cool car?

3.3 Coolness is grounded in people's communities

While ones perception of coolness in people and objects is subjective, and therefore often different from person to person, it is not simple a matter of individual preference. Instead the "rules" about what is cool and what is not cool are deeply grounded in the communities we belong to, e.g. cultural, social, political, and sexual groups (e.g. O'Donnell and Wardlow, 2000). It is through participation in these different communities that our individual perception of coolness is shaped. Young adults who like to skateboard belong to a very different community in society than middle-aged bank employees do. Consequently they have very different ideas about what is cool. One group might like to listen to Eminem while the other might like Johnny Cash. But these communities have not formed because of shared taste in music. The taste in music is formed through participation in the community. As another example, returning to the cigarette smoking man in Figure 4, this person might well be perceived as cool by a particular community of people who sees him as seeking exclusivity through his flamboyant and expensive style. Other communities probably disagree. The same goes for the car in Figure 5, which might well be perceived in some communities as being rebellious because it is cheap and ugly. In other communities it might just be considered junk. As a consequence we cannot universally conclude about the coolness of the man, or the car, without also contemplating the communities they are parts of, and are considered in.

4 MEASURING COOL

Based on our reading of literature on cool and related work within HCI and interaction design, we have set out to produce a tool for measuring the perceived coolness of an interactive product, such as an interactive device or an application. In doing so, one might suggest to simply ask the people to "*please rate the coolness of this design*". However, by asking such question, we will not gain much insight about *why* a particular product scores the way it does. Thereby we learn very little as designers about what we have done well or not so well, and we are poorly suited for improving our design or repeating our success. If, on the other hand, we are able to measure some more specific aspects of what makes a design cool or not, then it is more likely that we will be able to create something equally, or more cool. But what exactly should we try to measure?

Going back to the discussion about the difference between inner and outer cool, then measuring only on something's style (outer cool), for example if it is considered minimalistic or flamboyant, does not give us a very good idea about its coolness without also investigating if minimalistic and flamboyant is considered cool or uncool by the people or community asked. The measures for outer coolness are simply too dependent on individual taste to work absolute measures for cool. Looking at something's inner coolness, however, the characteristics provided by the literature are much more general and detached from properties that are subject to specific preference. They are generally agreed upon characteristics of something with a cool character or personality, independent of its specific style or appearance. Informed by this, our COOL Questionnaire is designed to measure the perceived inner coolness of a product through a series of questions that correlate with the characteristics listed in Figure 2.

4.1 The Cool Questionnaire

The primary contribution of this paper is our proposal for measuring inner cool through a validated questionnaire, COOL Questionnaire, with 17 items to be measured on a 7-point Likert scale (Figure 6). The questionnaire can specifically be used to measure an interactive product's perceived inner coolness, and can be applied to an empirical study of coolness in the form presented below.

In order to produce the listed questionnaire items, we went through a process of careful application of statistical methods (described in detail in the Section 5). This process took us from the 11 characteristics identified from theory in the literature (see Figure 2) to three factors suitable for measuring the perceived inner coolness of an interactive product.

	strongly disagree						strongly agree
This device is cool							
This device can make me better							
This device is meant for people like me							
This device can make me happy							
This device can make me look good							
This device can make me look in control of things							
This device totally connects with me							
This device is unconventional							
This device moves against the current							
This device is different							
This device is outside the ordinary							
This device is rebellious							
This device is easy to use							
This device is easy to operate							
This device is easy to learn							
This device is simple to use							
This device is effortless to use							

Figure 6. The COOL Questionnaire

The three factors of perceived inner cool measured by the COOL Questionnaire are: *desirability, rebelliousness* and *usability*. From our statistical studies, we can see that these three factors are positively correlated. In practice this means that if a product is perceived as cool, then all of these factors are likely to be rated higher by comparison to ratings for a product that is not perceived as cool.

The COOL Questionnaire measures desirability through six items related to how a specific interactive product appeals to personal desire, for example *"This device can make me happy",* or *"This device can make me look good".* Rebelliousness is measured through five statements regarding the conventionality of the device, for example *"This device is outside the ordinary"* or *"This device is unconventional".* Usability is measured through five items related to ease of use, for example *"This device is easy to operate"*, or *"This device is effortless to use".*

4.2 Deciding on coolness by inferring inner cool from outer cool

The second contribution of this paper is that we validate an inference model on how people shape their perception of product coolness. As mentioned before cool can be decomposed to inner cool and outer cool. However, due to the very limited number of empirical studies on cool, the relation between these is unclear. Through another statistical procedure we propose a model of inference describing the relationship between inner cool and outer cool (see Figure 7 for an illustration). Methodological details are presented in Section 6.

The empirically validated model and theory suggests that in cases where we observe an artifact for the first time we immediately make a perception of its coolness (Pountain and Robbins, 2000). This perception is initially based on the outer cool characteristics, i.e. its style. Based on this we shape a perception on its inner cool characteristics. As an example, when we observe a mobile device for the first time we infer inner cool characteristics (desirability, rebelliousness and usability) and we do this on the basis of its appearance, i.e. its attractiveness and aesthetic appeal. The combination of inner cool and outer cool shapes our perception on whether we find a device cool or not. Thus, inner cool becomes an intermediate factor (a mediator) for perceived coolness.

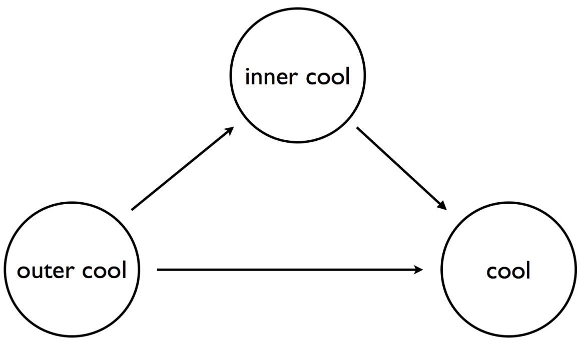


Figure 7. Outer cool - inner cool inference model

5 DEVELOPING THE COOL QUESTIONNAIRE

Creating questionnaires require meticulous statistical exploration and confirmation in order to ensure robustness, for example to be certain that the proposed question items measure particular factors reliably. In order to create the COOL Questionnaire we applied two statistical procedures by following the example of Lavie and Tractinsky (2004): Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA).

EFA is based on an iterative process where questionnaire items (individual questions) are removed from an initial pool of items, based on the degree by which they contribute to measuring a particular factor. In our case these factors are the inner cool characteristics shown in Figure 2. The extent of item contribution is measured through factor loadings where a high loading essentially means that the item correlates highly with the given factor. If an item has a low loading (in our case below a cutoff level of 0.6) on all factors, then that question should be removed. This is an iterative process where, after removing a number of items, another EFA should be conducted on a new dataset until each of the remaining items have a sufficiently high level of correlation with a single factor. In relation to this it is important to note that reliability is also about having items that correlate highly with *only one* factor and not with multiple, as one particular item should not measure multiple constructs. This is partly what lies behind the concept of "multicollinearity" which should be low (for details: Lavie and Tractinsky, 2004).

CFA is, as the name implies, of confirmatory nature. Where the EFA is about exploring (removing and sometimes adding items depending on the emerging factors) the CFA method is applied to make a final validation of these factors and the items that measure these (Schreiber et al., 2006). Through CFA we apply the notion of loadings in the same way as in the EFA described above. In CFA there are item loadings on factors, but also covariance between factors denoting how variances between any two pairs of factors are correlated. The goodness of a factor model is determined by a range of fit-indices, which collectively indicate whether or not the factor structure is appropriate and reliable. More specifically, we can say that when getting acceptable levels on the fit indices we can expect to observe similar patterns of loadings and covariance on a different data sample (assuming this has the same demographic characteristics and based on evaluation of similar devices).

For validity it is crucial that the CFA is not conducted on the same datasets as the EFA, otherwise you end up confirming a factor model on the same data on which it was derived, and thereby run the risk of inflated model-fit measures. In CFA several measures are applied to determine model-fit. For details on this we refer to Schreiber et al. (2006).

5.1 Experimental setup

In order to measure the perceived inner coolness of an interactive product we produced the COOL Questionnaire based on 11 characteristics derived from literature. The process of doing this consisted of three phases: 1) creating candidate questions, 2) EFA studies and 3) CFA studies.

In the first phase in the process we created candidate items for each of the 11 characteristics. Two of the authors created 15 items per characteristic (165 items in total). This was done by taking inspiration from other questionnaires in the literature and from our own understanding of each characteristic. For example, for *authenticity*, some of the produced candidate question items were: *"This device is one of a kind"*, *"This device is original"*, and *"This device mimics other devices"*. Afterwards, six colleague senior researchers evaluated the appropriateness of the 165 initial candidate questions. Through this process the questions per characteristic were reduced from 15 to 13 (143 in total).

In the second phase we conducted iteratively four EFA studies on different datasets in order to discover the underlying factor structure of inner coolness as well as the items that contributed most to each factor. In the fourth EFA study we also included the Attrakdiff2 questionnaire (Van Schaik et al., 2012), an aesthetics questionnaire (Lavie and Tractinsky, 2004) and an attractiveness questionnaire (Quinn and Tran, 2010). This was done in order to test for external validity, i.e. to assess whether or not the COOL questionnaire measured the same constructs proposed by other user experience questionnaires.

Finally, in the third phase, we conducted three CFA studies to confirm the factor structure suggested in phase two. All studies were conducted in September and October 2013.

In all studies we used the same experimental setup. The participants were asked to assess a mobile device through a webpage, which on the left side showed the device (facing forward) and on the right side listed the items being measured, along with the question *"This device is cool"*. All questions were presented in random order, and were all rated on a 7-point Likert scale from "strongly disagree" to "strongly agree".

5.1.1 Products

The products of coolness assessment for the development of our questionnaire were a number of mobile devices. This was a deliberate choice because of the general trend that coolness has become an important factor for people when deciding what product to buy, and not to buy. In total, our participants made assessments of 16 different mobile devices, with an average of 98.6 participants per device in the EFA studies, and 152.3 participants per device in CFA studies. Since we wanted to measure inner coolness of the devices, and not be influenced by bias in their outer appearance, we had to experimentally control for such outer parameters. To achieve this, we presented the participants with mobile devices of the same colour, without any visible indication of brand, and with their screens turned off. Additionally, in order to increase the generalizability of the questionnaire, we used a large number of mobile devices in all EFA and CFA studies.

5.1.1 Participants

Across the EFA and CFA studies we had a total of 2236 participants (916 females), with a mean age of 28.36 (SD=7.58). 1322 of them participated in the EFA studies and 914 in the CFA studies. Details about the number of participants per study can be found in Table 1 and Table 4. All studies were based on participants from the United States, and all participants were recruited through Amazon Mechanical Turk (MTurk). Several studies have shown MTurk participants to be reliable for such studies, including studies within human-computer interaction (Boujarwah et al., 2012; Gupta et al., 2012; Heer and Bostock, 2010; Heimerl et al., 2012; Gottlieb et al., 2012) as well as in other fields such as political science (Berinsky 2012). Using MTurk allowed us: a) to reach the large sample sizes needed for our statistical analyses without compromising the quality of data, b) to include people from different communities with varying perceptions of what is cool, and c) to reach natively English speaking participants, thereby reducing language barriers. The latter is mentioned as a specific challenge experienced in Sundar et al. (2014).

All participants who reported prior experience with the mobile devices assessed were removed from the datasets. The reason for removing those participants is that research has shown, that prior experience with an IT product affects evaluation results (Langton et al., 2007; Sauro, 2011) and that we wanted to measure perceived coolness at first sight, and not after some time of actual usage. Participants assessed only one mobile device each, and participated only in one study. For this they were paid an incentive of $0.35US.

5.2 Key EFA and CFA decisions

A number of decisions were made in order to conduct EFA and CFA. For EFA the most important decisions were in relation to sample size, the method and the criteria for factor extraction, and the factor rotation method. Regarding sample size we tested its suitability through the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO). By following recommendations from literature (e.g. Bulmer 1979; Field 2009) we have used Principal Axes Factoring as extraction method, and since we did not know beforehand if the extracted factors would be orthogonal or not, we opted for the oblique rotation method. The number of the extracted factors was determined through a Scree test and through parallel analysis (using Monte Carlo PCA, Watkins). In all four EFA studies we removed items by applying two criteria: low communalities ($<.5$) and low factor loadings ($<.6$). Finally, in order to conduct EFA, homogeneity of variances should be secured. In order to test this assumption we used Bartlett's Test of Sphericity, which had to be significant.

The CFA was based on Structural Equation Modelling (SEM) with Maximum Likelihood Estimation. When conducting SEM it is necessary to conduct a pre analysis to examine whether SEM assumptions are met in the data sample (Schreiber et al., 2006). This is then followed by a post analysis in which model verification is conducted. In the pre analysis we examined our CFA samples in terms of missing data, normality, linearity and multicollinearity as suggested by Schreiber et al. (2006). In this regard we had no missing data. In terms of normality, SEM leans on the assumption of normally distributed data. In our case all CFA datasets had univariate normality. Normality is assessed by measures of Skewness and Kurtosis. Skewness is an expression of the extent to which the distribution leans towards left or right relative to the normal. Kurtosis determines the level of peaknedness of the distribution. A perfect normal distribution has skewness and Kurtosis values of 0. In our SEM datasets, skewness values lie between -.6 and .6 and Kurtosis values lie between -1.3 and 1.3. According to Bulmer (1979) these are within acceptable boundaries, i.e. individual variables are assumed to be normally distributed. Multivariate normality is another assumption on which SEM is based. This denotes a situation where all variables are normally distributed with respect to each other. We did not obtain multivariate normality in our data (Madia's PK > 3). Non-normality can inflate $\chi 2$, hereby increasing the likelihood of making a Type I error that is to reject a model, which should actually have been accepted. However, as suggested by Kline (2012) SEM is still accurate despite non-normality and, given that the effect of non-normality is pessimistic rather than optimistic, we chose to proceed with post analysis.

Linearity between latent and manifest variables is also an assumption made in SEM analysis. Based on the strong factor loadings ($>.67$) identified during all of our EFA studies we assume linearity in our CFA datasets. Furthermore, SEM is based on the assumption of absence of multicollinearity, that is, latent factors do not correlate to an extent to which they can be said to measure the same construct. Based on the relatively low correlations between latent variables ($<.5$) we assume absence of multicollinearity in our CFA datasets.

Finally, we applied SPSS as the statistical software with the AMOS package v. 22 for model construction. This software provides a broad range of indices, which we applied to determine the degree of model fit, e.g. Ratio of $\chi 2$ to df, Normed Fit Index (NFI), Comparative Fit Index (CFI), Goodness-of-Fit Index (GFI) and several others. Table 5 provides overview of all applied fit-indices and we refer to Schreiber et al. (2006) for more details on these.

5.3 Exploratory Factor Analysis (EFA) results

1322 participants were used in all four EFA studies and they evaluated a total of 13 mobile devices through the experimental setup we have described before. In all EFA studies Bartlett's Test of Sphericity was significant (<.001) for all the models we have produced. Table 1 presents an overview of the 4 EFA studies.

	n	Devices	i
EFA1	310	LG Optimus LS II and a concept Blackberry	143
EFA2	195	Caterpillar B15, Apple iPhone 5, and Huawei Ascend P6	96
EFA3	537	HTC One Mini, Sony Experia Z Ultra, Nokia 8210, Motorola RAZR, and a concept phone	61
EFA4	280	Apple iPhone 5c, Nokia Lumia 625, and Huawei U8300	26+23
Total	1.322	A total of 13 mobile devices, mean=98.6 participants per device, SD=33.5	

Table 1. The four EFA studies. n=number of participants, i=number of items used as input

The purpose for conducting EFA1 was twofold: a) we wanted to reduce the total number of items, and b) we wanted to check the appropriateness of each item. A total of 7 models were produced in EFA1 with three, four and five factor structures. Items were removed only if they did not belong to any of the factors in any of the seven models. In all seven models KMO was >.940 indicating that we had the appropriate sample size to produce them. Another interesting observation of EFA1 was that all the negative items were clustered all together in one factor. Since there are studies that pinpoint the issues of including negative phrased questions in questionnaires (for example Sauro and Lewis, 2011), we decided to change them to positive. Additionally, a number of items were rephrased, as they were characterized as confusing by the participants. After EFA1 the total number of items was reduced from 143 to 96, which were used as input for EFA2.

In EFA2 our data suggested a three or possibly four factor structure. Again a total of 5 models were produced and the lowest KMO was .957. The process of removing items was the same as EFA1 and we kept all the items that belonged to at least one of the factors in at least one of the models. After EFA2 the number of items was reduced from 96 to 61. In EFA3 a three-factor structure for perceived inner coolness was confirmed through Scree Tests and Parallel Analysis (Monte Carlo PCA, Watkins). Our best model had a KMO of .954 and the number of items was reduced from 61 to 26.

	Factor:	A	B	C	D	E
	Eigenvalue:	*9.112*	*4.476*	*2.458*	*1.033*	*0.841*
	Cronbach α:	*.952*	*.928*	*.914*	*.864*	*.737*
Desirability	This device can make me better	**.861**	.160	.012	-.136	.016
	This device is meant for people like me	**.835**	-.091	-.003	.040	.051
	This device can make me happy	**.831**	-.036	.020	.080	-.008
	This device can make me look good	**.787**	.019	-.001	.093	.016
	This device can make me look in control of things	**.771**	.072	.065	.029	.062
	This device totally connects with me	**.770**	-.060	.007	.187	.004
	This device can please me*	**.731**	-.026	.048	.076	.109
Rebelliousness	This device is unconventional	-.040	**.874**	-.016	.002	-.024
	This device moves against the current	-.042	**.836**	.007	.035	-.002
	This device is different	-.063	**.825**	.018	.055	-.022
	This device is outside the ordinary	-.056	**.806**	-.056	.049	.172
	If it was a person this device would be a rebel *	.202	**.789**	-.057	-.049	-.042
	This device is rebellious	.121	**.736**	.145	.052	-.160
Usability	This device is easy to use	-.05	.003	**.928**	.053	-.002
	This device is easy to operate	.025	.052	**.850**	-.033	-.009
	This device is easy to learn	-.109	.049	**.820**	.068	.051
	This device is simple to use	.102	-.095	**.770**	-.067	-.020
	This device is effortless to use	.103	-.019	**.675**	-.016	.110
New attractiveness	I find this device: boring/interesting [1]	.138	-.025	.043	**.898**	-.081
	I find this device: plain/eye-catching [1]	-.036	.124	.011	**.825**	.030
	I judge the device to be: dull/captivating [2]	.168	.059	-.042	**.604**	.188
New classic aesthetics	This device has clean design [3]	.048	-.017	.068	-.025	**.747**
	This device has clear design[3]	.092	-.014	.077	.071	**.617**
	Sum of Squared Loadings (Total variance explained): 77.91%					

Table 2. Item loadings per factor in EFA4. A= Desirability, B=Rebelliousness, C=Usability, D=New Attractiveness and E=New Classic Aesthetics. * removed in CFA1, [1] originates from attractiveness questionnaire, [2] originates from Attrakdiff2, [3] originates from aesthetics questionnaire.

In order to check the external validity of the COOL questionnaire we included three established user experience questionnaires in EFA4: Attrakdiff2 (Van Schaik et al., 2012) that measures *hedonic* (4 items) and *pragmatic quality* (4 items), an aesthetics questionnaire (Lavie and Tractinsky, 2004) that measures *classic* (5 items) and *expressive aesthetics* (5 items), and an attractiveness questionnaire (Quinn and Tran, 2010) that measured the *attractiveness* (5 items) of mobile devices. By applying the cut-off criteria we ended up with five factors: three from our own dataset and two from the external questionnaires. Based on the question items going into each of our three factors (Table 2) we chose to name them *desirability, rebelliousness* and *usability*. Items from the external questionnaires: A) mainly clustered around the two factors of attractiveness and classic aesthetics and B) they did not contribute to our factors related to cool. The remaining external items were merged into two factors: *new attractiveness* and *new classic aesthetics.* Thus, a total of 18 question items from the established questionnaires were discarded due to loadings being below the cut-off level of 0.6. Cumulatively this five-factor model explained 77.91 % of the variance and Bartlett's Test of Sphericity was significant (<.001), while KMO was .921. Thus, we found homogeneity in variances and the sample size is deemed adequate. We ended up having 18 items for perceived inner coolness and in EFA4 we removed a further 8 out of 26 items.

Furthermore, Table 3 presents the correlations between the 5 factors. This shows that the factors of new attractiveness and new classic aesthetics correlate with the three perceived inner coolness factors. However, loadings range from .412 to .605, which shows that none of our perceived inner coolness factors (desirability, usability and rebelliousness) have a 1-1 correlation to measures of attractiveness or aesthetics, i.e. these factors are not measuring the same construct. This finding, along with the fact that most items from established questionnaires were below the cut-off level, provides initial evidence for the external validity of COOL Questionnaire.

	desirability	**rebelliousness**	**usability**	**new attractiveness**
desirability	1.000			
rebelliousness	.223	1.000		
usability	.412	.033	1.000	
new attractiveness	.605	.422	.195	1.000
new classic aesthetics	.511	-.077	.487	.345

Table 3. Factor correlation matrix for EFA4

In terms of the three emerged factors on inner coolness, desirability accounted for 39.62% of the variance and consisted of seven question items. We found this factor to be influenced by items initially proposed to belong to the inner cool characteristics (Figure 2) of *seeking personal development, strongly tied to a group, seeking pleasure* and *being/appearing in control*. Rebelliousness was made up of six items and accounted for 19.46% of the variance. All six came from those initially proposed to belong to the inner characteristic of *being rebellious*. Usability accounted for 10.646% of the variance and consisted of five items from the characteristics of *being/appearing in control* and *making hard things appear easy*.

Note here that we went from the initial 11 inner cool characteristics to the 3 factors mentioned above. Thus, through the EFA studies we discarded numerous (125) of the original items, and even entire characteristics such as *seeking exclusivity, embracing innovation, embracing authenticity, being antisocial* and *being detached* as their related question items did not provide sufficient contributions to any of the emerging factors. A detailed list of the items that were removed on each EFA study can be found in Appendix 1.

5.4 Confirmatory Factor Analysis (CFA) results

In order to prove stability for a single sample dataset Schreiber et al. (2006) argues that there is a general consensus on applying sample sizes of at least 10 participants per estimated parameter. Another approach is to prove stability by conducting multiple CFA's on different datasets. In our study we did both as detailed below.

914 participants were used over three CFA studies and they evaluated a total of 6 mobile devices (Table 4). None of the participants had taken part in the EFA studies.

Study	n	Factor model	Devices	i
CFA1	355	Three factors: Desirability, Rebelliousness, Usability	Alcatel One Touch 282, ASUS Padfone 2 and Samsung Galaxy S4 Mini	18
CFA2	559	Five factors: Desirability, Rebelliousness, Usability, New Attractiveness, New Classic Aesthetics	Apple iPhone 5c, Nokia Lumia 625, and Huawei U8300	16 + 5
CFA3	914	Five factors (merged datasets): Desirability, Rebelliousness, Usability, New Attractiveness, New Classic Aesthetics	All of the above	16 + 5
A total of 6 devices, mean=152.3 participants per device, SD=37.2				

Table 4. The three CFA studies. n=number of participants, i=number of question items used as input.

The first CFA study was based on the three-factor model of perceived inner cool as emerged from EFA4 and we assessed the model-fit of the factors desirability, rebelliousness and usability. This model consisted of 35 distinct parameters to be estimated, hence we needed at least 350 participants to ensure stability. We based the dataset on 355 participants, where each evaluated a single device. We applied a total of 3 different devices across all participants. In the first run all model fit indices suggested a close fit with the exception of PCLOSE. This was below the 0.05 threshold. PLCOSE represents a p-value indicating whether the null hypothesis of RMSEA = 0.05 (indicating a close fitting model) is significantly different from the observed value. Thus, a PLCOSE value larger than 0.05 indicates no significant difference from this null hypothesis, i.e. in that case model fit would be close. To increase model fit we went through two iterations of reviewing modification indices in order to determine which items to remove, i.e. removing items causing high modification indices reduces the noise on model fit. In each iteration we removed one item and recalculated model fit indices. We removed the following two items: "*This device can please me*" and "*If it was a person this device would be a rebel*". Thus, after two iterations and removing two items we reached the final model consisting of the three factors of desirability, rebelliousness and usability as well as 16 underlying items. Table 5 shows the respective item loadings and model-fit indices obtained in the first CFA where indices are within acceptable thresholds.

In CFA2 we evaluated a five-factor model including the three factors related to perceived inner cool (with the 16 underlying items) and the two factors of new attractiveness (3 items) and new classic aesthetics (2 items). This model consisted of 52 distinct parameters to be estimated, which indicates that at least 520 participants were needed. CFA2 was based on a different dataset than CFA1 and we had 559 participants. In total they evaluated three devices. At the first run we found all model-fit indices to be within acceptable ranges (Table 5).

In the third and final CFA study we merged the datasets from CFA1 and CFA2 and re-evaluated the five-factor model confirmed in CFA2. Thus, CFA3 is based on 914 participants evaluating a total of 6 different devices. Model-fit indices were all within acceptable ranges after the first run. All item loadings on factors are significant. We further validated the five-factor model by examining the matrix of standardized residuals. Residuals denote the deviations between observed values and the estimated parameters, i.e. residuals express the level of observed differences between all pairs of variables (items and factors) and their estimated loadings. A model with a good fit will have the residuals centered around zero. We examined the matrix of standardized residuals provided in AMOS for excessively high values and found none larger than ±2, which indicates a good model fit (Schreiber et al., 2006).

Based on the model-fit indices, sample sizes and CFA replications we found a three-factor model of the perception of perceived inner cool and a five-factor model also including new attractiveness and new classic aesthetics to be supported by our data. Thus, these models can be considered reliable in the prediction of data.

Table 6 presents the correlations correlation matrix between the five factors obtained in CFA3. These correlations are comparable to those found in the final EFA study (see Table 3). It again shows the factors of new attractiveness and new classic aesthetics to be correlating with the three factors of perceived inner coolness. None of our inner coolness factors (desirability, usability and rebelliousness) have a 1-1 correlation to measures of new attractiveness or new classic aesthetics. However, we found the correlation between desirability and new attractiveness to be .79. This indicates that those concepts are closely related, yet measuring separate dimensions.

		CFA1	CFA2	CFA3
Desirability	This device can make me better	.79	.79	.85
	This device is meant for people like me	.86	.84	.87
	This device can make me happy	.87	.86	.86
	This device can make me look good	.86	.86	.86
	This device can make me look in control of things	.77	.80	.82
	This device totally connects with me	.87	.87	.90
Rebelliousness	This device is unconventional	.77	.72	.74
	This device moves against the current	.62	.77	.84
	This device is different	.79	.85	.76
	This device is outside the ordinary	.74	.84	.79
	This device is rebellious	.73	.73	.86
Usability	This device is easy to use	.90	.90	.90
	This device is easy to operate	.89	.88	.88
	This device is easy to learn	.88	.85	.84
	This device is simple to use	.90	.88	.85
	This device is effortless to use	.71	.75	.75
New attractiveness	I find this device: boring/interesting	-	.94	.88
	I find this device: plain/eye-catching	-	.86	.94
	I judge the device to be: dull/captivating	-	.90	.86
New classic aesthetics	This device has clean design	-	.84	.82
	This device has clear design	-	.80	.79
Model-fit indices				
Ratio of $\chi 2$ to df (CMIN/df, acceptance threshold ≤ 3)		2.2	2.4	2.8
Normed Fit Index (NFI, acceptance threshold ≥ .95)		.94	.95	.97
Incremental Fit Index (IFI, acceptance threshold ≥ .95)		.97	.97	.98
Tucker-Lewis Index (TLI, acceptance threshold ≥ .95)		.96	.97	.97
Comparative Fit Index (CFI, acceptance threshold ≥ .95)		.97	.97	.98
Goodness-of-Fit Index (GFI, acceptance threshold ≤ .95)		.93	.93	.95
Adjusted Goodness-of-Fit Index (AGFI, acceptance threshold ≤ .95)		.90	.91	.94
Root Mean Square Error of Approx. (RMSEA, accept. threshold ≤ .06)		.06	.05	.04
p of close fit (PCLOSE, acceptance threshold > .05)		.06	.40	.97

Table 5. Item loadings per factor and model-fit indices for the three CFA studies. All are within acceptable thresholds indicating good model-fit.

	desirability	rebelliousness	usability	new attractiveness
desirability	1.000			
rebelliousness	.199	1.000		
usability	.341	-.051	1.000	
new attractiveness	.794	.417	.144	1.000
new classic aesthetics	.579	-.131	.534	.438

Table 6. Factor correlation matrix for CFA3

To further test for consistency and validity of our factors we measured Composite Reliability (CR), Cronbach α, and Average Variance Extracted (AVE). For all five factors (usability, desirability, rebelliousness and attractiveness, classic aesthetics), all values, except for rebelliousness in CFA1, were within acceptable ranges: AVE>.5 (suggested by Fornell and Larcker, 1981), CR>.7 (suggested by Hair et al., 1998), and Cronbach α.>.8 (suggested by Furr and Bacharach, 2013). We then assessed discriminant validity using the Fornell and Larcker (1981) technique. For all pairs of factors, both AVEs were larger than the shared variance (square of their correlation). More details can be found in Table 7.

	CFA1 (n=355)			**CFA2** (n=559)			**CFA3** (n=914)		
	CR	**Cron α**	**AVE**	**CR**	**Cron α**	**AVE**	**CR**	**Cron α**	**AVE**
desirability	.934	.933	.702	.934	.934	.704	.937	.934	.711
rebelliousness	.790	.780	.443	.912	.912	.677	.897	.889	.636
usability	.933	.929	.738	.926	.924	.714	.927	.929	.719
new attractiveness	-	-	-	.910	.940	.772	.922	.924	.799
new classic aesthetics	-	-	-	.823	.820	.699	.789	.801	.651

Table 7. Results for factors' Composite Reliability, Cronbach α and Average Variance Extracted for the three CFA studies.

Based on the above, we can conclude that the three-factor model for perceived inner coolness is stable and different from the rest of perceived qualities for the case of mobile devices.

6 VERIFYING THE INNER COOL-OUTER COOL INFERENCE MODEL

Since there are very few empirical studies on measuring cool, we do not know how inner cool and outer cool relates to the overall perception of coolness. Does outer cool influence inner cool during our first impression with a mobile device? Do both significantly influence our perception on the coolness of the device? In order to answer these questions we have applied Partial Least Squares (PLS) analysis (Vinzi et al., 2010). PLS is an ideal technique to use when having models with high complexity and when a theoretical understanding of a domain needs to be tested (Jöreskog and Wold, 1982; Falk and Miller, 1992). The PLS technique could be viewed as complementary to CFA, as it essentially allows a researcher to test the appropriateness of a theoretical inference model by assessing the strengths of relationships between factors.

PLS simultaneously tests the measurement model (sometimes referred to as outer model – the relationships between questionnaire items) and the structural model of inference (or inner model - the paths between factors). In order to conduct PLS we merged the data from EFA4 and CFA2 and we ended up with 839 participants and five factors: desirability (6 items), rebelliousness (5 items), usability (5 items), new attractiveness (3 items) and new classic aesthetics (2 items). Data was analyzed using the SmartPLS software (Ringle et al., 2006). There are two requirements for having robust results in PLS in relation to sample size (Henseler et al., 2009): a) we need ten times more participants than the factor with the most indicators (question items), and b) we need ten times the largest number of paths that are directed at a particular factor. Our sample size satisfied both criteria.

The first part of the process was to assess the measurement model (the questionnaire items) by looking at: Individual item reliability, internal consistency and discriminant validity. Individual item reliability can be checked by examining the loadings of the items on the construct they intend to measure. As a rule of thumb, items with a loading higher than 0.7 should be accepted (Duarte and Raposo, 2010). In our case all items had a loading ranging from 0.82 to 0.945 (more details in Appendix 2). Internal consistency and discriminant validity was examined the same way as in CFA studies through the Composite Reliability Index (CRI), Average Variance Extracted (AVE) and Cronbach α. All values where within acceptable ranges.

The second part of the process was to produce and evaluate the structural model of inference. According to theory, the proposed factors of desirability, usability and rebelliousness contribute to inner cool while new attractiveness and new classic aesthetics contribute to outer cool. Thus, inner cool and outer cool act as second-level formative factors. PLS produces the standardized regression coefficients (path estimates) between factors and we assessed the significance of path coefficients through bootstrap analysis (with N=5000, as proposed by Henseler et al., 2009). Since we had two second-level formative factors we have used the hierarchical components approach, which is the most popular when estimating higher order factors with PLS (Chin et al., 2003; Tenenhaus et al., 2005; Wilson 2007). In this process the second order factors are measured by all the variables from all first order factors, e.g. inner cool is measured by the factors of desirability, rebelliousness and usability etc. This process can be conducted with SmartPLS (Ringle et al., 2006) in two steps.

In the first step, we produced the standardized regression coefficients between desirability, usability and rebelliousness to inner cool, and followed by producing the coefficients between new attractiveness and new classic aesthetics to outer cool. The path coefficients between desirability and inner cool was β=0.727, t= 12.497, p<.001***, while the paths for usability and rebelliousness on inner cool were β=0.375, t=7.607, p<.001*** and β=0.189, t=2.261, p<.05* respectively. Additionally, the path coefficient between new attractiveness and outer cool was β=0.773, t=17.763***, p<.001. The coefficient between new classic aesthetics and outer cool was β=0.395, t=11.292, p<.001***. More details can be found in Appendix 2. Consequently, all the links between first-level formative factors and second-level ones were statistically significant.

The second step in the process was to analyze coefficients between the inner cool and outer cool factors on the overall impression about perceived coolness of a mobile device. Overall impression of cool was measured by the control question "This device is cool". Figure 8 presents our inference model.

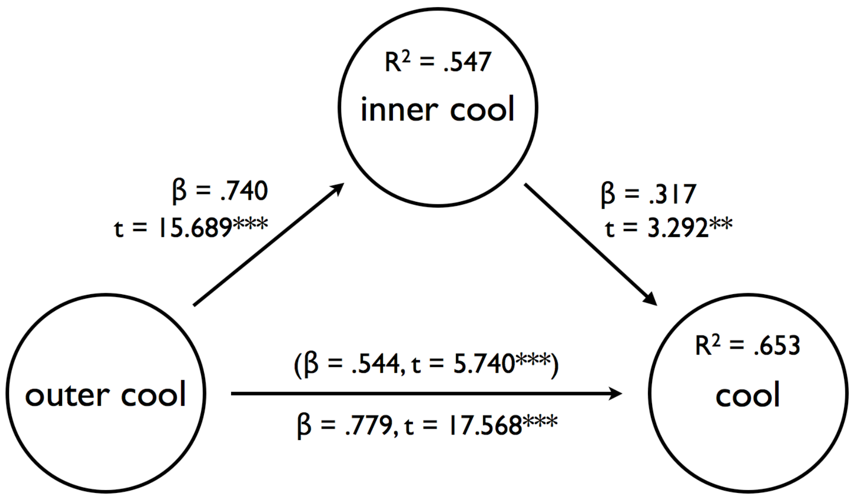


Figure 8. PLS structural model for the second-order formative constructs of inner and outer cool and the overall impression about the coolness of a mobile device. (Values in parentheses indicate indirect effects). β stands for standardized regression coefficients, t for T-statistic and R^2 for percentage of explained variance. ** p<.01, *** p<.001.

PLS is primarily used for predicting the goodness of a theoretical model by checking the strength of each path coefficient and the percentage of explained variance (R^2). According to Chin (1988), R^2 values of 0.67, 0.33, and 0.19 are characterized as substantial, moderate and weak respectively. In our case, we had two moderate R^2 values. The path between outer cool and inner cool was β=0.740, t= 15.689, p<.001***, while the path between outer cool and cool was β=0.317, t= 3.292, p<.01**. Finally the path between outer cool and cool was β=0.779, t= 17.5688, p<.001*** without inner cool and β=0.544, t= 5.740, p<.001*** with the presence of inner cool (indirect effect). Thus we found all path coefficients to be significant.

Finally, in order to test the significance of the mediation effect of inner cool (whether inner cool can be excluded from the model) we used a Sobel test (Sobel 1982). The Sobel test value (3.11) was statistically significant ($p<.05$), which means that inner cool partially mediates outer cool in determining the overall impression of cool and cannot be ignored.

In essence, the path going directly from outer cool to cool means that measurements of attractiveness and aesthetics can work as an approximation of the overall perception of cool ($\beta=.779$). However, given that the effect of inner cool is significant, inferring overall coolness based only on outer cool would be inaccurate, i.e. inner cool must also be measured. Thus, this confirms that when observing a mobile device for the first time, the perception of coolness is initially based on its style (outer cool). We then shape our perception on its inner cool and both factors contribute significantly on the overall perception of coolness.

7 DISCUSSION

Our aim with this paper was to identify the factors that contribute to coolness and produce a tool that can reliably measure them. Based on this, we designed and developed the COOL Questionnaire using the statistical techniques of exploratory factor analysis and confirmatory factor analysis. Our data suggest that the perception of inner coolness in an interactive product is determined by:

1. *Desirability*. The degree to which individuals believe that a product supports personal development, pleasure, and its ability to let its owner appear being in control.
2. *Rebelliousness*. The degree to which individuals believe that a product stands out from other devices.
3. *Usability*. The degree to which individuals believe that a product is usable.

Essentially, this means that a user would perceive a product as being cool if it is desirable, rebellious, and usable. This perception is affected by its impression regarding the product's outer coolness, which in this case was measured by five questions from established, external questionnaires that were clustered around two factors named as new attractiveness and new classic aesthetics.

7.1 Coolness and User Experience Research

The data from our study showed that desirability, rebelliousness and usability reliably measure the perceived inner coolness of a mobile device. In comparison to other questionnaires that measure cool (Sundar et al., 2014), the novelty in the creation process of our questionnaire lies in the fact that we included external questionnaires (Lavie and Tractinsky, 2004; Quinn and Tran, 2010; Van Schaik et al., 2012). During data analysis we observed that items from the external factors of hedonic quality and expressive aesthetics where clustered around attractiveness and classic aesthetics, while pragmatic quality had low loadings on our usability factor and its items were removed. Furthermore, we checked for discriminant validity and internal consistency and the scores were within acceptable ranges.

This paper contributes to user experience research in two ways. Firstly, we showed that coolness should be considered as a new dimension of user experience, at least in case of mobile devices. This is based on the fact that other established questionnaires measured different constructs than ours, i.e. established questionnaires contributed to measure attractiveness and classic aesthetics and not constructs related to inner coolness.

Secondly, we contributed to the on-going discussion regarding the relationships between established user experience qualities in different contexts of use (Bargas-Avila and Hornbæk, 2014). Our data showed that, for the case of mobile devices that perceived hedonic quality, expressive aesthetics and attractiveness overlap as well as that coolness is decomposed of outer and inner cool. Desirability, usability and rebelliousness measure inner cool and two factors emerging from external questionnaires measure outer cool. We argue that further studies should be conducted with different interactive products based on perceived as well as actual usage. In this respect it would be relevant to compare our COOL questionnaire with other established user experience qualities and re-evaluate whether or not the relationships are the same as the ones described in this paper. This could lead to a specific set of qualities and questionnaire items that practitioners could use in their evaluations.

7.2 Coolness and Interaction Design

In relation to the applicability of coolness within interaction design, we showed that coolness can be considered a valid design goal and the COOL questionnaire can be useful in assessing the perceived inner coolness of mobile devices. Our work extends previous work on coolness within interaction design, e.g. Fitton et al. (2000a, 2000b) and Read et al. (2011, 2012), as we provide a validated questionnaire that measures perceived inner coolness. Holtzblatt (2011) raised the question of how designers can identify aspects of coolness for specific product categories and how this can be incorporated into the product requirements. We developed our questionnaire to facilitate assessment of the inner coolness of interactive products – i.e. mobile phones in our present study.

Sundar et al. (2014) recently published a similar study on coolness where they promote the following characteristics of cool: *subculture, attractiveness* and *originality*. We believe that their underlying items of the originality factor are similar to those included in the rebelliousness factor from our study. Furthermore, we initially posed all the questions in the subculture factor identified in Sundar et al. (2014). However, our EFA studies indicated that these questions did not load sufficiently high onto any factors, which is why we removed them.

Similarly, items that Sundar et al. (2014) included in their attractiveness factor are comparable to several items that we either discarded during our EFA studies, e.g. *"This device is sexy"*, *"This device has style"*, *"This device is classy"* and *"This device is innovative"*, or belonged to the attractiveness factor of Quinn and Tran questionnaire (2010). More details can be found in Appendix 1. Through our EFA studies we also found five questions to be loading on the factor of usability, which was confirmed through the CFA studies. In comparison, Sundar et al. (2014) initially had a factor denoted as utility, which bears close resemblance to the factor of usability in our study. However, Sundar et al. (2014) eventually chose to remove utility. Here we differentiate again with the Sundar et al. (2014) study as our data show that usability is an important factor of our model that significantly contributes to inner cool and therefore we did not remove it from our questionnaire.

So, why do we see these differences between studies? The first explanation is related to the different theoretical backgrounds. Through an extended literature review (Raptis et al., 2013) we decomposed cool into outer cool and inner cool and, as demonstrated by our PLS model, we believe attractiveness is a contributing factor to outer cool. On the other hand Sundar et al. (2014) treat their three factors of subculture, attractiveness and originality as directly measuring the overall coolness of an artifact. We believe that this approach it is statistically correct, but needs to be extended in order to explain why some artifacts, even though they are perceived as ugly, are considered as cool (for example the old VW Beetle, Imseng 2011). Through our theoretical understanding, our COOL questionnaire and the proposed PLS model we believe that future researchers will be able to understand why these cases occur.

7.3 Other types of interactive products

In both CFA studies we had an average of 152 participants assessing each mobile device. Therefore, we argue that our model is generalizable in terms of the perception of inner coolness of mobile devices. On the other hand, this points potentially to a limitation of generalizability *beyond* mobile devices, which is relevant to discuss.

In the study by Sundar et al. (2014) they assessed 18 different devices and software applications. This is commendable in terms of attempting to test the questionnaires generalizability, and something to aspire to in future research. However, we do caution that the number of participants needed to make such study across different products typically need to be considerably higher than the 16-36 participants per device reported in (Sundar et al., 2014). For example Lavie and Tractinsky (2004) used more than 75 participants per website assessed to produce their questionnaire. According to Velicer and Fava (1998) one needs at least 10 participants per questionnaire item, per type of object assessed. For that reason exactly, we deliberately decided *not* to include a variety of different devices and applications in our study at this time, and followed the example of Lavie and Tractinsky (2004). The advantage of this is that we have a very solid questionnaire within our focus. The disadvantage is an unknown potential limitation of generalizability that should be addressed in the future. That said, its our hypothesis that due to our experimental setup and large numbers of participants, the COOL Questionnaire is able to measure perceived inner coolness of other interactive products too, as has been the case with Lavie and Tractinsky's (2004) and the Attrakdiff2 (Hassenzahl et al., 2003; Van Schaik et al., 2012) questionnaires.

7.4 How does Cool change over time?

Another question that we find important to discuss further is the relation between perceived coolness and actual use of interactive devices and systems in practice. As discussed earlier, coolness is something we recognize immediately from our first impression of objects and people. But as pointed out by Holtzblatt (2011), is our perception of coolness in an interactive product constant, or does it change over time? Perhaps most likely the former assumption, but what makes us perceive something cool or not cool in the long run? And what characterizes cases where coolness of a device or application has increased or decreased after extensive use, or use over a period of time? If perceived coolness does indeed change over time, then interaction designers would obviously benefit from understanding this process better, and therefore it would be of interest for interaction design researchers to investigate.

8 CONCLUSIONS

We have addressed the concept of cool in interaction design by researching literature on the concept outside the field of HCI, and how it has previously been explored within HCI. Based on the insight gained from this, we have constructed and validated a questionnaire for measuring the inner coolness of interactive products – we call the questionnaire the COOL Questionnaire, which is the key contribution of this paper. The questionnaire was developed from an offset in eleven inner cool characteristics derived from the literature on coolness. Through a process of iteratively applying the statistical methods of exploratory and confirmatory factor analyses (EFA and CFA), we arrived at three factors that measure the perceived coolness of an interactive product namely *desirability, rebelliousness*, and *usability*, as well as a questionnaire design with 17 specific question items that measure these three factors of inner cool. Additionally, we have also contributed to a better understanding both on coolness and on user experience in general through our PLS study in the context of mobile devices. Our results show that our identified constructs and a number of items from established UX constructs contribute to *inner cool* and *outer cool* and both shape our perception of whether a product is cool or not.

Our questionnaire is meant as a practical tool that can be used to measure the perceived inner coolness of an interactive product during a design process. We believe that this could be useful for interaction design practitioners when designing for cool in industrial projects. Furthermore, we think it will be useful for interaction design researchers as a practical tool to understand the coolness dimension of user experience.

9 FUTURE WORK

The study of coolness in HCI and interaction design has just recently started, and current studies seem to open a multitude of avenues for further research. This is also stressed by Holtzblatt et al. (2010) and Holtzblatt (2011) who list several challenges for research on cool e.g. UX design process integration.

Our work points specifically at several avenues for further research. Firstly, one could investigate the applicability of the COOL Questionnaire to other types of interactive products than the ones used in our study. Secondly, our study was based on initial perceptions based on images of mobile phones. As pointed out by Holtzblatt et al. (2010) and Holtzblatt (2011), it would also be relevant and to investigate coolness through actual usage and over time where prediction becomes very relevant and essential for designers. Thirdly, our findings suggest that the COOL questionnaire measures inner cool while attractiveness and expressive aesthetics contribute to outer cool. With more studies we will be able to know if the proposed PLS model is valid and generalizable. Fourthly, since cool is considered to be community specific, it would be interesting to study if there are any design characteristics, which are considered universally cool across communities. Finally, we believe that it would be relevant to investigate the use of empirical tools like the COOL Questionnaire in design processes for conceiving cool interactive design and products.

ACKNOWLEDGEMENTS

We would like to thank all MTurk participants and Dr. Veronika Hinkle for her valuable help. The study was partly financed by the Danish Research Council (grant number 09-065143).

REFERENCES

Bargas-Avila, J.A., and Hornbæk, K. Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. In *Proc. CHI 2011*, ACM Press (2011), 2689-2698.

Berinsky, A.J., Huber, G.A., and Lenz, G.S. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk, *Political Analysis,* 20 (2012), 351–368.

Boujarwah, F., Abowd, G., and Arriaga, R. Socially Computed Scripts to Support Social Problem Solving Skills. In *Proc. CHI 2012*, ACM (2012), 1987-1996.

Bulmer M.G. *Principles of Statistics*. Courier Dover Publications, 1979.

de Guzman, E.S. “Cool” Card Sort: Adapting Desirability studies to assess “Cool”-ness. *ACM CHI Workshop on Cool aX Continents, Cultures and Communities 2012*.

Duarte, A.O., and Raposo, M.L.B. A PLS Model to Study Brand Preference: An Application to the Mobile Phone Market. In V. Esposito Vinzi et al. (eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields*. Springer (2010), Berlin.

Falk, R.F., and Miller, N.B. *A primer for soft modelling*. The University of Akron Press, Ohio, 1992.

Field, A. *Discovering Statistics using SPSS*. SAGE Publications, 2009.

Fitton, D., Horton, M., Read, J.C., Little, L., and Toth, N. Climbing the Cool Wall: Exploring Teenage Preferences of Cool. *Ext. Abstracts CHI’12*, ACM Press (2012a), 2093-2098.

Fitton, D., Read, J.C., Horton, M., Little, L., Toth, N., and Guo, Y. Constructing the Cool Wall: A Tool to Explore Teen Meanings of Cool. *PsychNology*, 10, 2 (2012b), 141-162.

Fornell, C., and Larcker, D. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18 (1981), 39-50.

Frank, T. *The Conquest of Cool. Business Culture, Countercurture, and the Rise of Hip Consumerism*. The University of Chicago Press, Chicago, 1997.

Furr R.M., and Bacharach, V.R. *Psychometrics: An Introduction*. SAGE Publications (2013).

Chin, W.W. The partial least squares approach to structural equation modelling. In G.A., Marcoulides (ed.), *Modern Methods for Business Research*. Lawrence Erlbaum Associates (1998), 295–336.

Chin, W.W., Marcolin, B.L., and Newsted, P.N. A partial least squares latent variable modeling approach for measuring interaction effects: Results from a monte carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research,* 14, 2 (2003), 189–217.

Gioia, T. *The birth and death of the Cool*. Speck Press, Golden Colorado, 2009.

Gottlieb, L., Choi, J., Kelm, P., Sikora, T. and Friedland, G. Pushing the limits of mechanical turk: qualifying the crowd for video geo-location. *In Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia (CrowdMM)*. ACM Press (2012), 23-28.

Gupta, A., Thies, W., Cutrell, E. and Balakrishnan, R. mClerk: enabling mobile crowdsourcing in developing regions. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM (2012), 1843-1852.

Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. *Multivariate Data Analysis*. Prentice Hall International (1998), London.

Hassenzahl, M., Burmester, M., and Koller, F. Attrakdiff: Ein Frageboten zur Messung wahrgenommener hedonischer und pragmatischer Qualität. *Mensch & Computer 2003. Interaction in Bewegung*, (2003), 187-196.

Heer, J., and Bostock, M. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proc. CHI 2010*, ACM Press (2010), 203-212.

Heimerl, K., Gawalt, B., Chen, K., Parikh, T. and Hartmann, B. CommunitySourcing: engaging local crowds to perform expert work via physical kiosks. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM Press (2012).

Henseler, J., Ringle, C., and Rayner, R. The use of partial least squares modelling in international marketing. In T. Cavusgil, R.Sinkovicks, and P. Ghauri (eds), *New Challenges in International Marketing*, Emerald (2009), 277-319.

Holtzblatt, K., Rondeau, D.B., and Holtzblatt, L. Understanding "Cool". *Ext. abstracts CHI '10*, ACM Press (2010), 3159-3162.

Holtzblatt, K. What makes things cool? Intentional design for innovation. *Interactions,* 18,6(2011), 40-47.

Janlert, L.E., and Stolterman, E. The character of things. *Design Studies*, 18, 3 (1997), 297–314.

Jöreskog, K.G., and Wold, H. The ML and PLS techniques for modelling with latent variables: Historical and comparative aspects. In H. Wold and K.G. Jöreskog (eds), *Systems under indirect observation: Causality, structure, prediction*, Vol 1 (1982), 263-270.

Jordan, P.W. Products as personalities. In (ed.) M.A. Hanson, *Contemporary Ergonomics*. Taylor & Francis (1997), 449-453.

Imseng, D. *Think Small. The story of the world's greatest ad*. Full Stop Press (2011), Uster, Switzerland.

Kline, R.B. *Principles and practice of structural equation modeling* (3rd ed.). Guilford (2012), NY.

Langdon, P.M., Lewis, T., and Clarkson, P.J. The effects of prior experience on the use of consumer products. *Universal Access in the Information Society*, 6, 2 (2007), 179-191.

Lavie, T., and Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Human-Computer Studies*, 60 (2004), 269-298.

McCrickard, D.S., Doswell, F., Barksdale, J., and Piggott, D. Understanding Cool in Computing for African-American Youth. *CHI Workshop on Cool aX Continents, Cultures and Communities 2012*.

Nancarrow, C., Nancarrow, P., and Page, J. An analysis of the concept of cool and its marketing implications. *Journal of Consumer Behaviour*, 1, 4 (2001), 311-322.

O'Donnell, K.A., and Wardlow, D.L. A Theory on the origins of Coolness. *Advances in Consumer Research*, 27 (2000), 13-18.

Pountain, D., and Robbins, D. *Cool Rules. Anatomy of an Attitude*. Reaction Books, London, 2000.

Quinn, J.M., and Tran, T.Q. Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In *Proc. CHI 2010*, ACM Press (2010), 353-362.

Raptis, D., Kjeldskov, J., and Skov, M.B. (2013) Understanding "Cool" in Human-Computer Interaction Research and Design. In *Proc. OzCHI 2013,* ACM Press (2013), 53-62.

Read, C.R., et al. Understanding and Designing Cool Technologies for Teenagers. *Ext. Abstracts CHI '11*, ACM Press (2011), 1567-1572.

Read, C.R., Horton, M., and Fitton, D. Being Cool-Getting Persona-l. *CHI Workshop on Cool aX Continents, Cultures and Communities 2012*.

Ringle, C.M., Wende, S., and Will, A. SmartPLS 2.0. University of Hamburg (2006), Hamburg.

Sauro, J. Does Prior Experience Affect Perceptions Of Usability? (2011) Available online: http://www.measuringusability.com/blog/prior-exposure.php

Sauro, J., and Lewis, J.R. When designing usability questionnaires, does it hurt to be positive? In Proc. CHI 2011, ACM Press (2011), 2215-2224.

Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A. and King, J. Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *Journal of Educational Research,* 99, 6 (2006), 323-338.

Sobel, M. E. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, 13 (1982), 290-312.

Sundar, S.S., Tamul, D.J., and Wu, M. Capturing "cool": Measures for assessing coolness of technological products. *Int. J. Human-Computer Studies*, 72 (2014), 169-180.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M., and Lauro, C. PLS Path Modeling. *Computational Statistics & Data Analysis*. 48, 1 (2005), 159-205.

Van Schaik, P., Hassenzahl, M., and Ling, J. User-Experience from an Inference Perspective. *TOCHI,* 19, 2 (2012), Article No. 11.

Velicer, W.F., and Fava, J.L. Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods,* 3 (1998)*,* 231-251.

Vinzi, V.E, Chin, W.W., Herseler, J. and Wang, H. *Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields*. Springer (2010).

Watkins, M.W. *Monte Carlo PCA for Parallel Analysis*. Online: http://edpsychassociates.com/Watkins3.html

Wilson, B. *Using PLS to Investigate Interaction Effects Between Higher Order Brand Constructs*. In V. Esposito Vinzi et al. (eds), Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields. Springer (2010), Berlin.

APPENDIX 1

The items used in the EFA and CFA studies without the ones that were removed after EFA1. They are not included because after EFA1 many items were rephrased. On each item it is depicted the EFA or CFA study that it was removed or added. For example, the first item "*I would love to have this device*" was removed in EFA4. The items in bold were not removed in any of EFA, or CFA studies and thus belong to the final version of the questionnaire.

I would love to have this device (-EFA4)	This device is normal (-EFA2)	**This device is outside the ordinary**
If I buy this device, my friends will follow (-EFA3)	This device is a step forward (-EFA2)	This device is pleasurable (-EFA3)
If it was a person this device would be a rebel (-CFA1)	This device is against the rules (-EFA3)	This device is popular (-EFA2)
It will be so nice to show this device to my friends(-EFA3)	This device brings new things (-EFA2)	This device is progressive (-EFA2)
My friends will love this device (-EFA2)	This device is all fun (-EFA2)	**This device is rebellious**
This device is ahead of the competition (-EFA3)	This device can empower me (-EFA3)	This device is relaxed (-EFA2)
This device can boost my performance (-EFA2)	This device is authentic (-EFA3)	This device is revolutionary (-EFA3)
This device can create new opportunities for me (-EFA2)	This device is calm (-EFA2)	This device is risky (-EFA3)
This device is at the center of attention (-EFA2)	This device is classy (-EFA3)	This device is satisfactory (-EFA2)
This device can give me authority (-EFA2)	This device is common (-EFA3)	This device is sexy (-EFA2)
This device can have a positive impact on my image (-EFA3)	This device is conservative (-EFA2)	**This device is simple to use**
This device can help me improve (-EFA3)	This device is desirable (-EFA4)	This device is special (-EFA2)
This device can help me master new skills (-EFA2)	This device is detached (-EFA2)	This device is state of the art (-EFA4)
This device can make difficult activities look easy (-EFA2)	**This device is different**	This device is stereotypical (-EFA3)
This device is suitable for people the age of my parents (-EFA2)	This device has quality (-EFA3)	**This device can make me better**
This device is straightforward (+EFA3 -EFA4)	**This device is easy to learn**	This device is trivial (-EFA2)
This device can make me look good	**This device is easy to use**	**This device is unconventional**
This device can make me look in control of things	This device is easy-going (-EFA2)	This device is underground (-EFA2)
This device can make me look more powerful (-EFA3)	This device is eccentric (-EFA2)	This device is unique (-EFA3)
This device can make me more productive (-EFA3)	This device is effective (-EFA3)	This device can please me (-CFA1)
This device can make me proud (-EFA3)	This device is efficient (-EFA3)	**This device moves against the current**
This device can make me unique (-EFA4)	This device is exciting (-EFA4)	This device resists authority (-EFA2)
This device is worth recommending (-EFA4)	This device is exclusive (-EFA3)	This device sets new standards (-EFA2)
This device can save me a lot of time (-EFA3)	This device is fresh (-EFA2)	This device starts a new era (-EFA3)
This device disobeys known norms (-EFA3)	This device is innovative (-EFA3)	This device stimulates me (-EFA3)
This device disrupts social order (-EFA2)	This device is magical (-EFA2)	**This device totally connects with me**
This device will influence future devices (-EFA2)	This device is mainstream (-EFA3)	This device does not care (-EFA2)
This device easily convinces me (-EFA2)	This device is meant for few (-EFA3)	This device has style (-EFA4)
This device fits well with my personal values (-EFA3)	This device has character (-EFA2)	**This device is easy to operate** (+EFA3)
This device is meant for people like me	This device is obedient (-EFA2)	**This device is effortless to use** (+EFA3)
This device has my name on it (-EFA3)	This device is obvious (-EFA2)	**This device can make me happy**
This device is easily recognizable (-EFA2)	This device is one of a kind (-EFA3)	This device is useful (+EFA2 -EFA3)
Very few individuals have this device (-EFA3)	This device is opposing society(-EFA2)	This device is normal (+EFA2 -EFA3)
This device imitates other devices (-EFA2)	This device is original (-EFA2)	

APPENDIX 2

PLS measurement and structural models for the first-order formative constructs of inner cool and outer cool. Desirability, usability and rebelliousness contribute to inner cool and new attractiveness and new classic aesthetics to outer cool. All paths are significant.

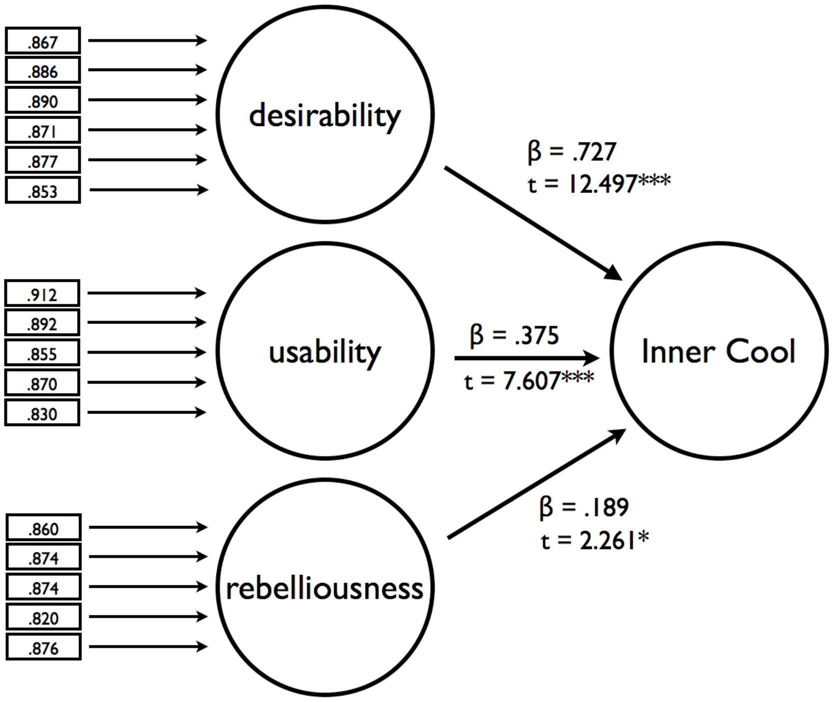


Figure 9. PLS measurement and structural model for the first-order formative construct of inner cool, with factor loadings per item, standardized regression coefficients (β) and T-statistics (t). * p<.05, *** p<.001

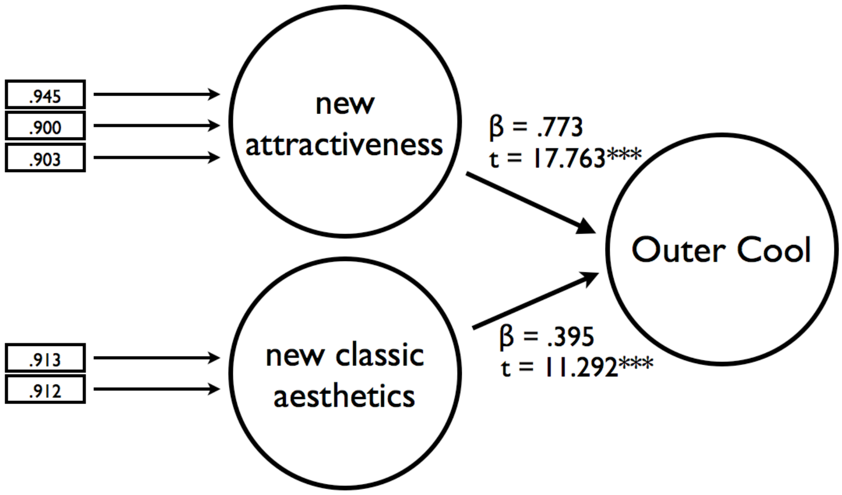


Figure 10. PLS measurement and structural model for the first-order formative construct of outer cool, with factor loadings per item, standardized regression coefficients (β) and T-statistics (t). *** p<.001.