Master Thesis

Building a web-based experiment to capture and analyze cultural attraction

Renan Oliveira Moreira

2019

Eötvös Loránd University

Faculty of Education and Psychology

Master of Science in Computational and Cognitive Neuroscience

Building a web-based experiment to capture and analyze cultural attraction

Author Supervisor

Renan Oliveira Moreira Dr. Christophe Heintz

Student Head of the Department of Cognitive Science at CEU

2019

Declaration of Originality

I declare herewith, that this thesis presented here is my own original work.

Furthermore, I confirm that:

− this work has been composed by me without assistance;

− I have clearly referenced in accordance with departmental requirements, in both the text and the bibliography or references, all sources (either from a printed source, internet or any other source) used in the work;

− all data and findings in the work have not been falsified or embellished;

− this work has not been previously, or concurrently, used either for other courses or within other exam processes as an exam work;

− this work has not been published.

I appreciate that any false claim in respect of this work will result in disciplinary action in accordance with the university or departmental regulations.

I confirm that I understand that my work may be electronically checked for plagiarism by the use of plagiarism detection software and stored on a third party’s server for eventual future comparison.

Renan Oliveira Moreira February 04th, 2019

Proof of consultation, supervisor’s statement

page should contain a confirmation form which contains the supervisor's consent for thesis submission

ACKNOWLEDGEMENTS

Abstract

Keywords

## Table of Contents

[Abstract I](#_Toc5003381)

[Keywords II](#_Toc5003382)

[Table of Contents III](#_Toc5003383)

[List of figures V](#_Toc5003384)

[List of tables VI](#_Toc5003385)

[Abbreviations VII](#_Toc5003386)

[Part I: Background 1](#_Toc5003387)

[1. Introduction 1](#_Toc5003388)

[1.1 Motivation of the research 2](#_Toc5003389)

[1.2 Research aim and objectives 2](#_Toc5003390)

[1.3 Research question 3](#_Toc5003391)

[1.4 Research approach 3](#_Toc5003392)

[1.5 Thesis structure 4](#_Toc5003393)

[2. Theoretical foundations 6](#_Toc5003394)

[2.1 Culture transmission 6](#_Toc5003395)

[2.2 Cultural Attractor Theory 10](#_Toc5003396)

[Part II: Experiment / Empirical studies 12](#_Toc5003397)

[3. Design and Implementation 12](#_Toc5003398)

[3.1 Architecture 13](#_Toc5003399)

[3.1.1 Angular framework 13](#_Toc5003400)

[3.1.2 Firebase hosting and database structure 14](#_Toc5003401)

[3.2 Scenarios for testing 17](#_Toc5003402)

[3.2.1 Calibration 18](#_Toc5003403)

[3.2.2 Face 18](#_Toc5003404)

[3.2.3 Spiral 20](#_Toc5003405)

[3.2.4 Button 20](#_Toc5003406)

[4. Data collection 21](#_Toc5003407)

[4.1 Calibration phase 21](#_Toc5003408)

[4.2 Testing phase 22](#_Toc5003409)

[4.3 Attractor condition phase 22](#_Toc5003410)

[5. Data analysis 24](#_Toc5003411)

[5.1 Data extraction and cleaning 24](#_Toc5003412)

[5.2 Data calculations 25](#_Toc5003413)

[5.3 Results 25](#_Toc5003414)

[Part III: Conclusion 26](#_Toc5003415)

[6. Discussion and future work 26](#_Toc5003416)

[6.1 Research objectives: Summary of findings 26](#_Toc5003417)

[6.2 Limitations 26](#_Toc5003418)

[6.3 Future research 26](#_Toc5003419)

[References 28](#_Toc5003420)

List of figures

[Figure 1 - Thesis Structure. 4](#_Toc3476709)

[Figure 2 - Data structure of cloud firestore 10](#_Toc3476710)

[Figure 3 - Data structure example from the experiment. Source: Elaborathed by the author 11](#_Toc3476711)

[Figure 4 - Face shaped image used on the scenario 14](#_Toc3476712)

List of tables

[Table 1 - Detailed explanation of data variables 12](#_Toc3278686)

Abbreviations

Pixels px

CAT cultural attractor theory

IT information technology

Npm nodejs package manager

Part I: Background

# 1. Introduction

The word culture can draw multiple interpretations or even meanings for each of us, but some of the most accepted and well-marked definitions can be seen in the Merriam-Webster dictionary as: “the customary beliefs, social forms, and material traits of a racial, religious, or social group” or “the integrated pattern of human knowledge, belief, and behavior that depends upon the capacity for learning and transmitting knowledge to succeeding generations” (Merriam-Webster Dictionary, 2018). These two definitions together offer a great idea of what is culture and how it is vital in our life.

Focusing on the second definition, it offers a valuable content that cannot be overlooked and still up to debate in the scientific community, more specifically when it refers to “...learning and transmitting knowledge to succeeding generations”. This part is extremely important and will be main topic of discussion of the theoretical approach of this thesis. Touching some topics as: How is human culture transmitted? What are the mechanisms that affect the process of transmission? Draw a parallel of culture transmission with genetic transmission, memetics, the cultural evolutionary theory, and more specifically one of those approaches within the cultural evolutionary theory called cultural attraction.

With purpose of exploring deeper the Cultural Attractor Theory (hereafter; CAT), it was created a web-based experiment which allowed multiple users around the world to interact with different hypothesis and scenarios regarding cultural attraction. Those interactions resulted on small amount of data that was analyzed with the goal to find any patterns that could be assumed as a factor of attraction. The whole creation of the experiment and the analysis of the results will be elicited in the empirical part of this thesis.

## Motivation of the research

My Information Technology background studies aligned with my working experience drove me to the construction of the website and the utilization of Machine Learning techniques to analyze data, together with my interests in social sciences, pushed me to develop the research on this topic so interesting called cultural evolution and more specifically cultural attractor theory which is currently discussed in the scientific community and has plenty of flourishing ideas and controversies. One of the controversies for example is considering memes as the unit of human cultural transmission, other example, regarding CAT, is the dilemma found on the article from (Claidière & Sperber, 2007) called “The role of attraction in cultural evolution” which defends a probabilistic view of cultural attraction that it was firstly introduced by (Sperber, 1996), against a deterministic interpretation of cultural attraction proposed by (Henrich & Boyd, 2002) .

## Research aim and objectives

The aim of this research project is to create a web-based experiment that enables testing hypothesis related to CAT, more specifically, a cross selection game that has different scenarios to try out which of those scenarios could be factors of attraction. Besides the construction of the website, it will be collected data for a certain period of time, more precisely 3 months and 10 days, that will be used in the data analysis with the hope of finding any result that could lead us to concluding the presence or existence of factors of attraction on these tested scenarios.

Besides my own data collection and data analysis, the platform of the experiment was constructed with the aim to be made available to other researchers and scientists interested on the topic. The code of the whole website structure is open source and available to be use by everyone, so the only required steps for the utilization would be the modification of the desired tested scenarios and deploying the website. With this approach of adopting open source development, there is the aim of sparking a discussion about the topic and providing a platform where other researchers could test their own hypothesis and leverage on the website.

## Research question

Question the existence of some factors of attraction on possible scenarios.

## Research approach

(Verschuren & Doorewaard, 2010) describes five ways to conduct a research: a survey, an experiment, a case study, a grounded theory approach and desk research. It was adopted the experiment approach because it was desired to compare two different groups of scenarios, one containing possible factors of attraction and other without it, and check the validity of one hypothesis, in our case, the existence of factors of attraction in some specific scenarios.

It was decided to perform the experiment over the internet, in a way that would provide us a larger access to demographic and culturally different participants, this type of experiment is known for several terms as: Internet-based experiment, Web experiment, on(-)line experiment, Web-based experiment, World Wide Web(WWW) experiment, and Internet experiment. (Reips, 2002). The structure of the Web-based experiment followed some of the standards and guidelines suggested on the paper of Reips.

Avoidance of organizational problems, such as scheduling difficulties, as thousands of participants may participate simultaneously.

Self Selection

Experimental Setting – control group.

This master thesis research can be separated in three stages of development: the first one was the exploratory literature studies about CAT and the planning and designing of the experiment. The second stage consisted on the coding and deployment of the website and the third and last one was the data collection, extraction, analysis and conclusions.

First stage: In-depth literature studies were fulfilled to enable the understanding of the research. Those studies together with discussions with Cristopher Heintz made possible the elaboration of the design of the experiment. On this phase, it was also decided which programming language would fit the requirements and would be adopted in the project implementation, Angular 5, details on the decision will be explained on the section [3.1.1 Angular Framework](#_3.1.1_Angular_framework). It was also necessary technical studies regarding web programming, more specifically TypeScript, programming language used by the Angular Framework. This phase on lasted approximately 4 (four) months, from May of 2018 until August of 2018.

Second stage: The implementation, also known as coding of the website, used as the platform for the web-based experiments lasted for approximately 6 (six) months from August of 2018 until January of 2019. This phase also contained the creation of the necessary connections to access the database (Cloud Firestore) and the deployment of the website on the Firebase platform, the details of this will be later explained on the section [3.1.2 Firebase hosting and database structure](#_3.1.2_Firebase_hosting).

Third stage: The last stage of the development of the research is the de facto experiment, when the web-based experiment in online and accessible to the users, allowing the collection of the data. This collection period lasts for 3 months, from X to Y and after that, the data extraction from the database and analyses of the data is performed. In parallel with those tasks, this master thesis which provides the documentation of the research was being written.

## Thesis structure

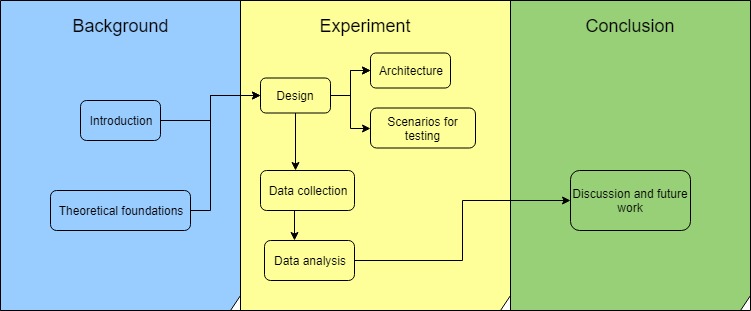
The structure of the thesis follows the instructions from the thesis manual from ELTE PPK, which recommends the separation of the structure into Theoretical and Empirical part as can be seen in the Figure X:

Figure 1 - Thesis Structure. SOURCE: Elaborated by the author.

The Part I, called Background provides a briefly introduction of the research developed, motivation, objectives and approaches utilized. Besides that, it also contains the Theoretical part which provides a base of understanding of the topic that will be researched and explored, eliciting definitions from different authors and researchers of the field and also the current discussions on the academic community.

The second part, called Experiment / Empirical Studies explains detailed all the phases of development of the experiment, the decision reasoning and the procedures utilized, from the design until the implementation and execution and lastly the data analysis.

The last part is the conclusion of the thesis, where it is discussed the results of the experiment, the conclusions reached, possible future works and limitations of the experiment.



# 2. Theoretical foundations

## 2.1 Culture transmission

Based on the definition of (Mesoudi & Whiten, 2008) and (Bisin & Verdier, 2008) culture transmission is the process in which information (knowledge, beliefs, preferences or norms of behavior) are passed from one individual to another one, within and across generations using social learning mechanisms as teaching, imitation or language.

It is possible to separate culture transmission in three different types: vertical, horizontal and oblique. Vertical transmission refers to the passage of traits from the parents to their offspring (top to down), and not necessarily has to be genetic related, for example it could be referent to transfer of knowledge from one parent to their adopted children. Other type is the horizontal transmission, which as the name suggests, in the hierarchy of generations, the transfer is from one side to another, this horizontally transmission consists of individuals on the same generation, as age peers, siblings, etc. The last type of transmission is called oblique transmission and covers the transmission from one individual of a specific generation to another of a younger one, excluding the parent-offspring relation, so for example, a teacher-student relation can be considered oblique transmission (Hershberger, 2014).

Culture transmission can be compared to genetic transmission, on the sense that both forms acquire traits or information from other individual (genetic traits/information versus cultural traits/information) and are intertwined in the process of evolution of our species, with the differences of the latter being more studied and with plenty of experiments conducted by population geneticists. While the former still has plenty of open discussions and far less experimental research (Mesoudi & Whiten, 2008).

Just like in genetics, the approach to explain the cultural transmission principles and questions can also be viewed in an evolutionary approach. This means that the cultural transmission theories called cultural evolution follow the principles in accordance with the Darwinian ideas of *variation*, *differential fitness* and *inheritance*. Accordingly with (Mesoudi & Whiten, 2008) and (Mesoudi, Whiten, & Laland, 2004) cultural traits vary across and within individuals and groups. Regarding differential fitness, some of the cultural traits will not be preserved and copied due to competition for expression, attention or memory space and some ideas are more attractive than others. And just like genetics, cultural traits are also inherited, but using social learning.

First studies and etc

Difference replication , and not replication. Memes

4 Questions

Content based, model based transmission.

Chain transmission

Studies of cultural transmission will be most valuable if they are pursued within a framework of cultural evolution. This body of theory contends that human culture evolves according to basic Darwinian principles, in important respects similar to those by which biological species evolve (Campbell 1974; Cavalli-Sforza & Feldman 1981; Boyd & Richerson 1985; Plotkin 1994; Mesoudi et al. 2004; Richerson & Boyd 2005; Mesoudi et al. 2006b). These Darwinian principles are variation, differential fitness and inheritance, and just as Darwin (1859/1968) showed these basic principles to characterize the evolution of biological organisms, they can also be observed in human culture (Mesoudi et al. 2004): (i) cultural traits (beliefs, attitudes, skills, knowledge, etc.) vary across and within individuals and groups; (ii) not all cultural traits are equally likely to be preserved and copied due to competition for expression, attention or memory space, some ideas are more memorable or attractive than others, and some models are more likely to be copied; and (iii) cultural traits are inherited or transmitted from model(s) to learner(s) via social learning.

As indicated in point (iii), cultural transmission is a fundamental component of cultural evolution. Without transmission there can be no evolution, and the form that this transmission takes can significantly influence the evolutionary dynamics of culture. As such, the cultural evolution literature already contains definitions, classifications and rigorous mathematical analyses of many aspects of cultural transmission.

Really good paper Alex Mesoudi1 The multiple roles of cultural transmission experiments in understanding human cultural evolution

----

There are problems associated with any effort to trace the pedigree of cultural evolutionary theories back to Darwin himself. One of the reasons for this is that cultural evolutionary theories often define themselves in opposition to those which claim that genetic inheritance is the only significant inheritance mechanism. Clearly one cannot cast Darwin as a cultural evolutionist in this manner, for he had no notion of genetic inheritance to oppose. Having said this, Darwin did believe that what was learned in one generation could be inherited in later generations. But far from distinguishing cultural inheritance from organic inheritance, Darwin thought that all inheritance should be explained by the transmission of ‘gemmules’.

Fathers: Herbert Spencer and Charles Darwin – cultural evolution

Work of Lumsden and Wilson (1981), Cavalli-Sforza and Feldman (1981), and Boyd and Richerson (1985). All of these authors have attempted, in one way or another, to produce formal models that can integrate the effects of cultural inheritance into more standard biological models of evolution.

Dan Sperber (1996), Richerson and Boyd 2005

Difference between genetics vs culture: These sorts of cultural evolutionary models do not assume that cultural inheritance works in the same way as genetic inheritance. Yet they remain recognizably evolutionary in style, primarily because they seek to explain the changes in trait frequencies in a population over time.

No one can deny that cultural inheritance is an important factor in explaining how our species has changed over time. Cultural inheritance is not merely a process that acts in parallel to genetic evolution, it is intertwined with genetic evolution.

Importance of culture: Cultural changes bring about alterations to the environment, which in turn affect both how genes act in development, and what selection pressures act on genes.

Should I Talk about memes? *Cultural units are not replicators*: Attractor vs Replicator. Attractor - culturally shared patterns of thought, which enable representations to spread through a population without literal copying.

… most cultural items are ‘re-produced’ in the sense that they are produced again and again—with, of course, a causal link between all these productions—but are not reproduced in the sense of being copied from one another…Hence they are not memes, even when they are close ‘copies’ of one another (in a loose sense of ‘copy’, of course). (Sperber 2000, 164–65)

Cultural evolutionary Theories: We want to know what makes some ideas fitter than others. in the cultural realm we will need to look at local psychological dispositions to explain why some ideas are more likely to spread than others.

5 theories that can be put together

<https://plato.stanford.edu/entries/evolution-cultural/#ExpRolCulEvoThe>

Five Misunderstandings about Cultural Evolution (2002) Joseph Henrich, Robert Boyd and Peter J. Richerson:

Second, Sperber (1996), Atran (2001) and Boyer (1998) emphasize that unlike genes, ideas are not transmitted intact from one brain to another. Instead, the mental representations in one brain generate observable behavior, a “public representation” in Sperber’s terminology

Mental representations will be replicated from one brain to another only if most people induce a unique mental representation from a given public representation. Moreover, inferential processes often systematically transform mental representations, so that unlike genetic transmission, the cultural transmission is highly biased toward particular representations. Following Sperber (1996), we call the representations favored by processes of psychological inference (including storage and retrieval) ‘cognitive attractors.’

cognitive attractors will rapidly concentrate the cultural variation in a population. Instead of a continuum of cultural variants, most people will hold a representation near an attractor. If there is only one attractor, it will dominate. However, if, as seems likely in most cases, attactors are many, other selective forces will then act to increase the frequency of people holding one attractor and decrease others. The weak selective forces (‘weak’ relative to the strength of the attractors) will actually determine the final distribution of representations in the population.

In the formalization, individuals acquire their mental representations by observing the behavior of others. Two cognitive mechanisms affect this learning process. First, inferential transformation captures the manner in which cognitive processes of acquisition, storage and retieve alter mental representations in ways to favor some representations over others—cognitive attactors. Because the two extreme represetations, “Moon as person“ and “moon as rock“ are easier to think, they act as cognitive attractors in our example. Individuals who observe behaviors that result from intermediate representations tend to infer mental representations closer to one of the two attractors. The second process, selective attention, captures the tendency for individuals to pay particular attention to some individuals more than others. For example, it could be in a modernizing environment, where the representations favored by science are prestigious, people who hold the “moon as rock“ representation are more succesful than those who hold the alternative, and thus they attract more attention (and are more likely to be learned from). Finally we assume the effects of inferential transformation are much stronger than the effects of selective attention.

Henrich and Boyd 2002:

First, cultural transmission processes are

usually incomplete and imperfect, so, unlike genetic systems, accurate

replication rarely occurs. Replication is the exception, rather than the

rule. Second, unlike DNA replication, inferential processes “transform”

these representations during their transmission and reconstruction. This

suggests that mutation-like processes are much more important than selection-

like processes in shaping cultural variation. Third, unlike genes, cultural

representations are rarely discrete units, suggesting that the idea of

a ‘replicator’ (or meme) makes little sense for most types of cultural

representations.2

## 2.2 Cultural Attractor Theory

One of the models to explain the process of cultural evolution is Cultural Attraction Theory (CAT) which differs from other evolutionary approaches as it develops the idea of constructive convergence in cultural transmission [2]. Constructive convergence refers to processes of cultural transmission that cause systematic transformation rather than faithful replication of cultural items. However, the transformations are biased and favour the production of some cultural items. These transformation biases can be explained by the existence of factors of attractions, which stabilize the distribution of cultural items at a macro level (whole populations and across generations) [3].

[2] Heintz, C. (2017) Cultural Attraction Theory. International Encyclopedia of Anthropology, Wiley Online Library.

[3] Claidière, N., Sperber, D. (2007). The role of attraction in cultural evolution. Journal of Cognition and Culture 7 (2007)

89-111.

Attraction can have Cognitive disposition, but also psychological or environmental. And it can change over time depending on the factor…but slowly

Attraction vs selection

2.3.1 Determistic

2.3.2 Probabilistic

Part II: Experiment / Empirical studies

# 3. Design and Implementation

The design and implementation of this project is a result of a collaborative work and ideas between the supervisor of the project and head of the Cognitive Science department of the Central Eastern University, Dr. Christophe Heintz and me, student graduating at the Master of Science program of Computational and Cognitive Neuroscience at the Eötvös Loránd University.

Due to the deeper and wider knowledge of Cultural Attraction Theory and the longer history on the field of Cognitive Science, the design of the experiment is a creation from Dr. Heintz ideas. The implementation of the project, which contains development and deployment of the website are result of my expertise on the field of Information Technology, due to my bachelor’s degree in computer science and work experience on the aforementioned area.

The design of the experiment consists of a cross location game where the user sees a diagonal cross and is asked to answer the position of it. At the beginning of the experiment, the cross will be demonstrated in a white background screen for one second, then it will happen a fill-out screen for half-second in a black background screen, with the purpose of changing the visual fixation point of the user, and then automatically after the half-second, it appears another white background screen where is expected the participant to answer the position of the diagonal cross.

The idea behind this experiment, is that, during the cognitive process of visualization of the cross the participant/player/user acquires information (the position of the cross), and then it transmits this information forward (the feedback position of the cross), but this process of perceiving/replying is believed to be imperfect and susceptible to external factors and transformations. The key concept and the goal of this experiment is to visualize if there is any possible scenario or factors, that would evidence the transformation of knowledge, in an individual level. Meaning that it would not happen an exact replication of the information, but a factor that plays a role on attracting these transformations to common points.

In the case of the designed experiment, the input and output are directed from/to the experimenter, because the goal of study is to analyze possible factors of attraction that are in play in the individual cognitive process of receiving information and forwarding it. But for a more elaborated research, it could be used a transmission chain to analyze these transformations in a larger scale, where the output of the user would be an input of another user.

On the following chapters, it will be detailed the architecture of the website, including the technical aspects of it, and the conditions tested, the reasoning behind each of them, and how they are related with the theory of factors of attraction. The code of the website won’t be presented nor explained on this paper because it’s not the aim of this project approaching the nuances of computer programming, but rather test the Cultural Attractor Theory using a web platform which enables high availability and collection of data. That said, the code has comments in itself with the goal to facilitate the understanding of anyone interested on it and it is available on the personal GitHub repository of the author of this master’s thesis. That can be accessed at: https://github.com/RenanOm92/factorsAttractionFirebase.

## 3.1 Architecture

Talk about the screens (home screen, instructions, start screen, fill-out, feedback screen, results screen)

### 3.1.1 Angular framework

The project was developed using the framework Angular, which provides the possibility of creating applications compatible with cross platforms (web, mobile web, desktop)[insert reference <https://angular.io/>], fulfilling one of the aims of the design of the experiment which was having a website accessible by every type of user, at any moment, from anywhere. Generating this way, a spreader sample of whom would interact with the website, deeply connected with our goal that is to examine culture transmission, meaning that the bigger the reach of the experiment, the more accurate would be the conclusions draw by it.

Angular is also an open source project led by Google and the community, which provides a great and stable environment for development, but at the same time still being updated with new features and being one of the most popular Frameworks for web development [insert reference <https://hackernoon.com/is-the-angular-decline-a-myth-e4cf563b72d6>]. Another important advantage of Angular it’s the two-way data binding feature, which enables the Model and the View being synchronized, allowing them to communicate between each other and any changes on the data would affect the visualization of it, used on the project for the calculation of the random position of the cross visualized in the start of each round.

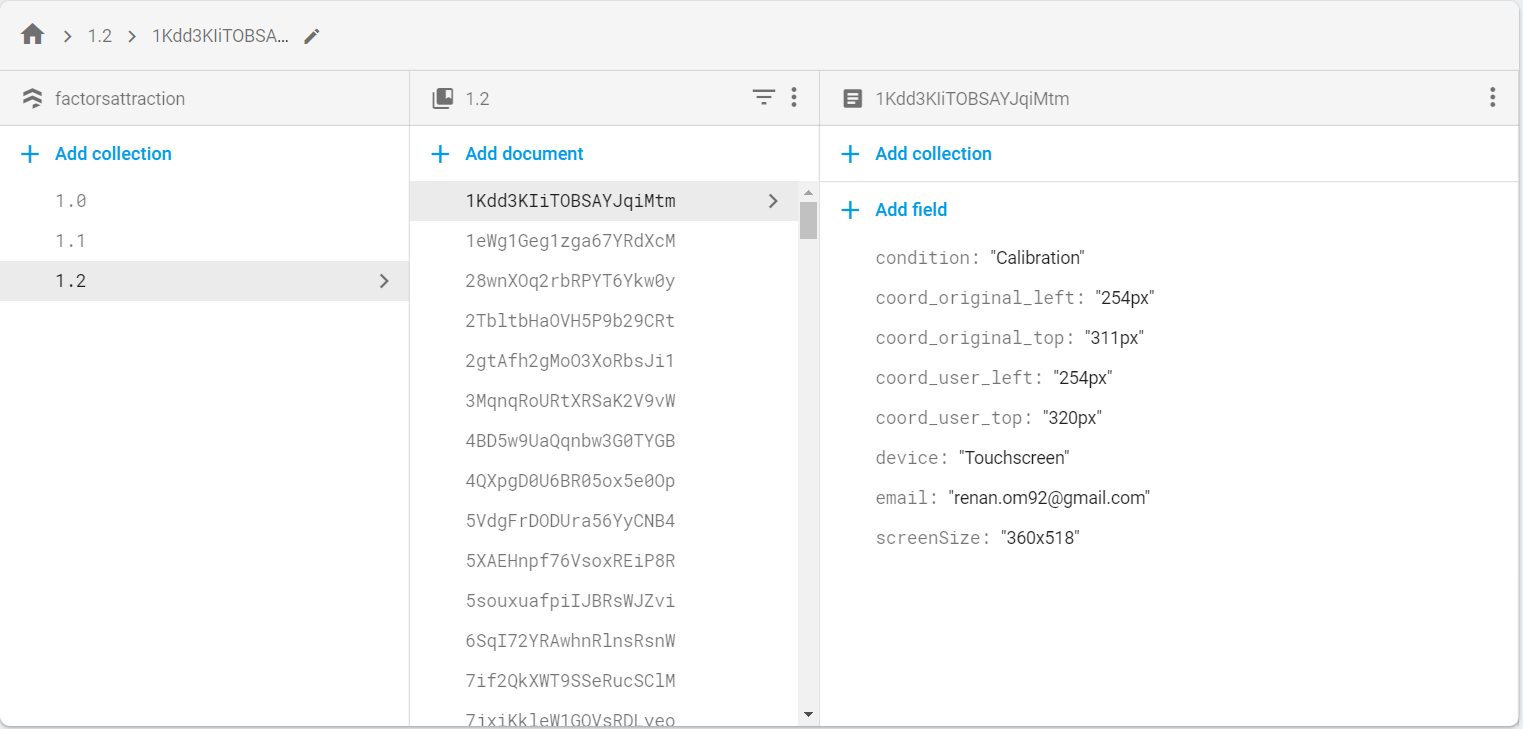
With all these qualities listed above, Angular was chosen as the framework used for constructing the website of the experiment. The version utilized was Angular 5, which has released date on November 1st of 2017. [insert references]

### 3.1.2 Firebase hosting and database structure

Firebase is a multifunctionality platform owned by Google which provides a large range of products, including hosting and database integration to web applications which were necessary for the deployment of this experiment. [insert reference <https://firebase.google.com/>] Firebase was the chosen platform because it offers a free plan with the following specifications, which covers all the needs of the project: hosting the web application, 100 simultaneous connections, saving the data on a database with a limit of 1 *Gigabyte* per month, extracting the data from the database and custom domain. For the reasons listed above, the website was then hosted on Firebase with the domain name: factorsattraction.firebaseapp.com. And all the data collected from the experiment was saved on the Cloud Firestore Database accordingly with the data structure supported: Collection, Document, Data as shown on the Figure X.

Figure 2 - Data structure of cloud firestore. Source: https://firebase.google.com/docs/firestore/images/structure-data.png

The experiment was released in different phases, as it will be explained on the section [4. Collecting data phases](#_4._Collecting_data), and it each phase was saved as Collections using the semantic versioning approach [insert reference <https://blog.codeship.com/best-practices-when-versioning-a-release/>]. Each interaction of the user (also called as one round) it was saved as a Document with ID generated randomly by Firebase, and all the data referent to the round was saved inside the document. Each document on the final phase of the project had the following structure of data: *condition*, *coord\_original\_left*, *coord\_original\_top*, *coord\_user\_left*, *coord\_user\_top*, *device*, *email*, *screenSize*. The database structure can be seen on the Figure X.

Figure 3 - Data structure example from the experiment. Source: Elaborathed by the author.

The meaning of each field value is better explained on the Table 1 below:

|  |  |  |
| --- | --- | --- |
| Field | Description | Type of data and exemplification |
| Collection | Container of the document. Used for control of version. | Numbers on the format:  1.1  2.0 |
| Document | Unit of storage, represents one interaction of the user playing the experiment. It’s an unique sequence of 20 characters composed of letters or numbers generated randomly by Firebase. | Sequence of 20 letters or digits:  1Kdd3KIiTOBSAYJqiMtm  4BD5w9UaQqnbw3G0TYGB  4QXpgD0U6BR05ox5e0Op |
| Condition | Represents the possible conditions in which the experiment can be played. | One of the following values:  Calibration  Face  SpiralLeft  SpiralCenter  ClickHereBottomLeft  ClickHereTopRight |
| coord\_original\_left | Represents the horizontal position on pixels which the cross was displayed at the beginning of the experiment. Where the 0 value is the left of the screen. | Number followed by pixel:  746px |
| coord\_original\_top | Represents the vertical position on pixels which the cross was displayed at the beginning of the experiment. Where the 0 value is the top of the screen. | Number followed by pixel:  520px |
| coord\_user\_left | Represents the horizontal position on pixels which the cross was answered by the user. Where the 0 value is the left of the screen. | Number followed by pixel:  777px |
| coord\_user\_top | Represents the vertical position on pixels which the cross was answered by the user.. Where the 0 value is the top of the screen. | Number followed by pixel:  12px |
| device | Contains which type of input device was used by the user. | One of the following values:  Touchscreen  Mouse |
| email | Contains the e-mail of the participant, used with the purposes of controlling the number of total interactions versus total of users. | Any possible value containing the e-mail format:  aaaaaa@aaaa.aaa |
| screenSize | Stores the value of the screen size of the device in which the participant used for starting the experiment. | Number x Number:  1366x626 |

Table 1 - Detailed explanation of data variables

## 3.2 Scenarios for testing

## 

When playing the game, the user can experience seven unique and different scenarios, they are: calibration, face, spiral centralized, spiral on the top right, spiral on the top left, “click here” button on the top right and “click here” button on the bottom left of the screen. Those scenarios will be detailed and explained in the following chapters. The scenario on which the user will play the game is randomly generated before the round starts (on the screen containing the instructions of the experiment in case of being the first round or at the end of the experiment for any rounds after that).

Even though the scenarios are randomly assigned, it was created a weighted value of appearance for each of them, with the purpose to have more appearance of a desired scenario. In the tested design, the weights were the selected ones: 33.3% for face scenario, 33.3% for the spiral scenario, 22.2% for the “click here” button and 11.1% for the calibration scenario. The calibration scenario has a lower weight when compared to the other scenarios because in the initial phases of the experiment it was collected only calibration data, which caused the abundance of calibration data, making unnecessary expose this scenario with the same frequency in the final version of the experiment. The “click here” button has a slightly lower weight due to the fact that it has only two variances, in contrast with the spiral that have three.

### 3.2.1 Calibration

The calibration scenario was the first stage of the software development and contains the core of the experiment, and it also has the purpose of being a calibration for the following scenarios, meaning that the data collected with this scenario can be used as our control scenario to further comparisons, this is possible because the calibration scenario do not have any figure or image that could have a role of being a factor of attraction.

The calibration scenario consists of a white background at the start page and feedback page with a black fill-out screen between them. The start page contains the original position of the cross and it is displayed for one second. Then automatically it is displayed the fill-out screen for half second and then the feedback page can be visualized until the user inputs some value for the cross.

The original position where the cross can be displayed at the beginning of the experiment is generated randomly but it follows some conditions and one parameter. The conditions that it must follows are: not being too close to the edges of the screen, nor being in the center. The cross cannot be in the center due to the appearance of the button “play again” on the center of the results page, which could lead the button to cover some results.

The parameter for calculation of the position of the cross is the screen size of the device of the user. Due to the reason that each user can have different devices (mobiles, personal computers, tablets) with different screen resolutions, it is necessary to capture the screen size on which the user is playing and then based on that, generate the original position of the cross.

Based on that two information, the original position of the cross is a random value in pixels between 5% of the screen size until 45% or 55% until 95% (for both height and width independently). For example, if the screen resolution is 2000 × 1000 pixels (width × height), the possible values for the width would be from 100px until 900px, and from 1.100px until 1.900px and for the height would be from 50px until 450px and from 550px until 950px.

### 3.2.2 Face

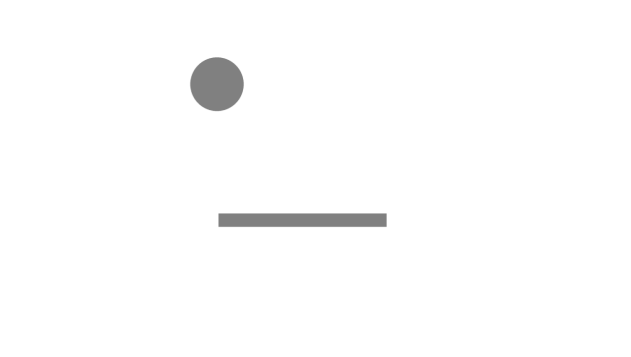
The face scenario has the same features of the calibration but with the addition of a background image in the start page and feedback page, this image (figure X) has the intention of recalling the shape of a face with a missing eye. The purpose of this scenario is to try out the hypothesis that the missing eye could be a factor of attraction. Due to the reason

Figure 4 - Face shaped image used on the scenario

The face shaped image is always centralized on the screen and has the feature of missing the right eye. The size of the image is adjustable and relative to the size of screen, which means that the figure does not have a fixed value in pixels, so our analysis of the feedback of the cross need to be based on percentage of the screen size of each user, so in this way we are able to examine if the feedback of the user is located in an approximate position of the missing eye of the shaped face figure. With that said, the center of the missing eye position figure (a hypothesized factor of attraction) is approximately on the coordinates 55% from the left and 27% from the top screen sizes.

Due to the position of the face and more specifically the missing eye being centered, I guided the generation of the random cross position to be more centered also. While the conditions of the cross position on the calibration scenario is being between 5% and 45% or 55% and 95%, in the scenario of the face, it was defined as between 30% and 49% or 51% and 70%. With these limits, the occurrence of the cross position on the center will be higher than in the calibration phase. At the moment of the design, it was believed that was worth generating values between 45% and 49% and 51% to 55%, meaning that the cross position could be further covered by the “Play again” button, affecting the usability and user experience quality but would increase the total of valuable samples for data analysis. The image size of the face shaped image will depend on the screen size and will be better explained in the section Data Calculation.

### 3.2.3 Spiral

The spiral scenario like the previous one is an extension of the calibration, but it has the addition of showing a spiral on the fill-out screen for 0.5 seconds, this spiral (figure X) has three different variances, based on the location where it can appear (top-left, centralized or top-right) on the screen. This scenario has the purpose of to try out if the center of the spiral could be a factor of attraction. The image measurements (width and height) of the spiral will always be half of the width screen. Due to the reason



Figure X

### 3.2.4 Button

The button scenario is also based on the calibration, but it has the addition of a button on the fill-out screen which is necessary to be clicked to proceed with the experiment, this feature is different from all other scenarios, in which the fill-out screen is showed for 0.5 seconds then automatically the feedback screen is shown. The button (figure X) has two different variances, based on the location (bottom-left and top-right) on the screen. The possible factor of attraction tested on this scenario would be the center of button itself. The button have measures of 87x32 px. Due to the reason



Figure X

# 4. Data collection

The website started gathering data on the day 28th of January of 2019 and is available until the present day, but at 05th of Apil of 2019 it was taken a snapshot of the database. This snapshot contains the data that will be used for analysis in this thesis. The web-based experiment was announced over the social media profile of the author of the thesis with the purpose to attract users of different locations. A total of XXXX interactions were registered during the collecting period, in which XXXX were performed by unique users, meaning that the participants group of our experiment consisted of a total of XXXX people, in which those participants could play a role in both the control scenario as also in the factors of attraction hypothesized scenarios.

## 4.1 Calibration phase

The first deployment of the website, which followed the software development stages of the website itself, only contained the calibration scenario and it was published online on 28th of January of 2019. The data collected was saved into the Firsetore collection as version “1.0”. This first deployment did not collect the information regarding the device of the user, and it lasted for one day and collected 53 interactions. The data will be used or not?

On the 29th of January, it was deployed a new version, saved as version “1.1” which contained the device of the user, this version also only had the calibration scenario. The duration period of the “1.1” version was approximately four (4) full days, lasted until 1st of February and collected 313 interactions.

The first two versions of the experiment collected 366 interactions of the calibration scenario, scenario which will be used as our control scenario and baseline for further comparisons. The calibration scenario will still be present in the next versions, but with the appearance rate reduced as explained in the section XXXX.

## 4.2 Testing phase

On the day 2nd of February of 2019, the website was deployed in a testing phase, as known as version “1.2”, where introduced plenty of different scenarios with the purpose to test their usability and encounter possible misleading features, these usability tests were perfomed by myself and Christophe Heintz. Some scenarios (e.g. different face shaped image, white background fill-out) were not approved and will not be taken into consideration, meaning that they were not selected to the next phase and consequently will not be in the data analysis. The scenarios that were approved can be found elicited on the chapter XXXX and will be used on the data analysis. This phase lasted until 5th of February of 2019, a total amount of four full days and collected 49 interactions of the calibration scenario, 27 of the scenario “click here” button top right, 24 of the “click here” button bottom left, 8 of the scenario spiral left, 1 of spiral center and 7 of spiral right.

## 4.3 Attractor condition phase

This attractor condition phase is the last collecting data phase of the experiment, and the one that contain all the scenarios designed at the beginning of the experiment, namely Calibration, Face Shaped, Spiral and Click Here Button, with all the possible variations. It started on 6th of February of 2019 and lasted until XXXX of April of 2019, totalizing an amount of XXX days. This phase collected 49 interactions of the calibration scenario, 27 of the scenario “click here” button top right, 24 of the “click here” button bottom left, 8 of the scenario spiral left, 1 of spiral center and 7 of spiral right.

During this phase, it was found a bug in the scenario Spiral on the top right of the screen (the spiral was not showed at the screen, consequently the fill-out screen was totally black and did not contain the hypothesized factor of attraction), and all the data related to this specific scenario will not be taken into analysis.

# 5. Data analysis

The data analysis can be separated in different sections, first and foremost, it was necessary to extract the data from the Cloud Firestore Database, where all the data was saved. After that, it was used the programming language Python to manipulate all the following activities: load the data, clean it, perform all the necessary calculations and plot the results. The code and the data file are available at the git repository from the author (RenanOm92/factorsAttractionFirebase, 2019) under the folder “Data extraction and analysis”.

## 5.1 Data extraction and cleaning

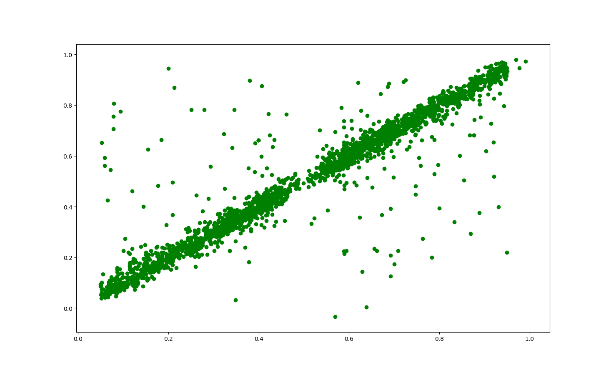
To perform the data extraction from Cloud Firestore, it was used the external package available at Node.js package manager (mostly known as npm) called “node-firestore-import-export”. Please visit their website (node-firestore-import-export, 2019) for a complete guidance and explanation of all functionalities from the package, on this thesis it will be only explained the basic tasks performed to extract data.

To install this package, is necessary to have Node.js installed and then execute on the Command Prompt the following command: *npm install node-firestore-import-export*. After the package being successfully installed, is necessary to get the credentials from the Cloud Firestore to be able to extract the data. On the Firebase Console, click on Project Settings (the gear icon button on top left of the page), navigate to Service Accounts and then click Generate New Private Key. There will be downloaded a JSON file with all the credentials necessary for the data extraction. With the package installed and with credentials downloaded, it is possible to perform the data extraction, to do so, the following command should be executed: *firestore-export --accountCredentials path/to/credentials/file.json –backupFile factorsAttraction.json*. It is necessary to adjust the path accordingly with the path of where it was downloaded the credentials file. As consequence, there will be created a JSON file called factorsAttraction.json containing all the data from the experiment.

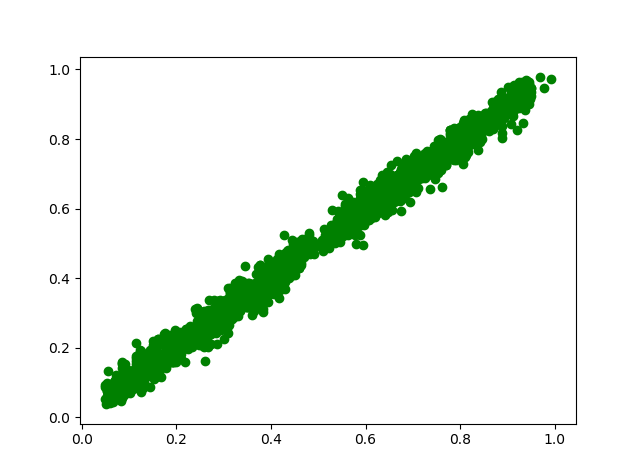
Using python and the packages JSON and Pandas, the data was imported from the file factorsAttraction.json to a DataFrame, DataFrame is a 2-dimensional labeled data structure perfect for manipulating and performing the data analysis. For more information about how it was performed the action of transforming the JSON file into a DataFrame, check the file called dataReading.py on (RenanOm92/factorsAttractionFirebase, 2019).

The data cleaning process consisted on two steps, the first one is transforming the data from the database in usable data, for example: the field “screeSize” was broken down in two columns: width and height, another example is the value left and top are transformed in X and Y of the cartesian coordinate system, and another one is the transformation of pixels in proportion of the screen size. The second step of the data cleaning consists on removing data from the DataFrame which will not be valuable to the analysis, this data is called noisy data. Noisy data on this experiment can be considered answers from the users where the feedback of the position is far away from the original position. This mistake could be a result of lack of attention from the user while playing, or even mistakes due to learning the game and another possibility is that those mistakes could be malicious answers with the purpose to damage the experiment.

On the Figure XXX is possible to visualize all the data collected on the experiment, each point represents the width or height values of an interaction, where the X-axis is the original position of the cross, and the Y-axis is the feedback from the user. The scale of the values is from 0 to 1, this value means the proportion of the position of the cross regarding the screen size. So, for example one screen resolution of 1000x1000px (width x height), where the original position of the cross is located at 500x200px, and the feedback of the user at 580x190px, the original position would be represented as 0.5x0.2, the feedback as 0.58x0.19. Then one point would be for the width (0.5,0.58) and another point would be for the height (0.2,0.19). If the feedback of the user is somewhat close to the original position, then when the point is plotted, it should be located on the diagonal, on the Figure XXX is possible to visualize the biggest distribution of points are located on the diagonal. (create new picture, with the Axis and description)



Based on this distribution, all the data that has more than 10% of discrepancy from the original width position to the user feedback width position will be discarded. The same rule is applied for the height values. After applying this rule, the data looked like the shown in the Figure XXX. From a total of XXXX interactions, it remained a total of XXXX valid interactions.



## 5.2 Data calculations

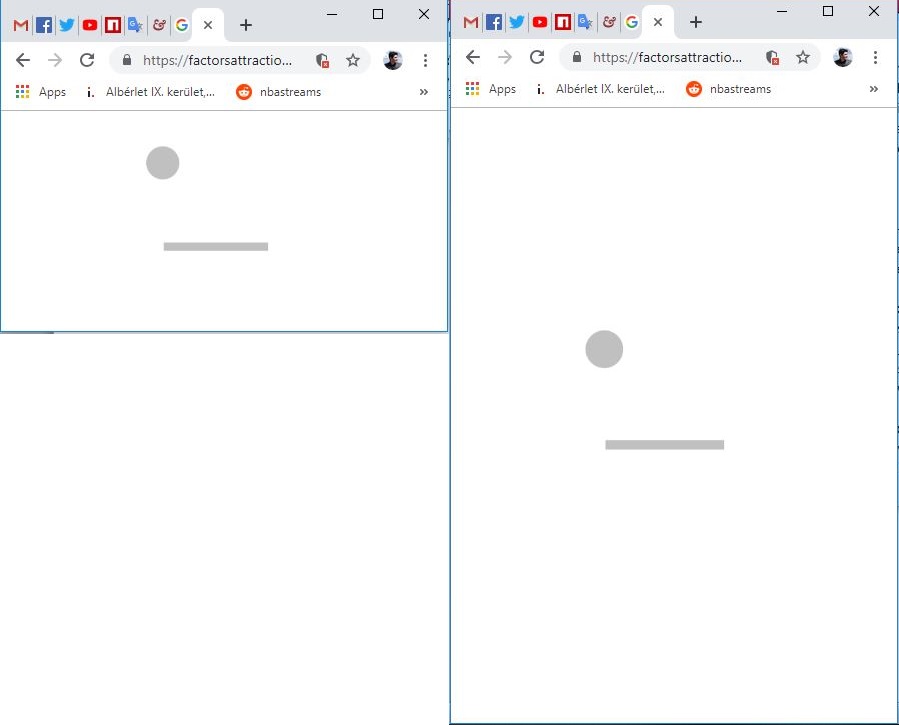
With the data cleaned, it was possible to perform all the operations related to the analysis. The first measurement that was explored is the distance between the original position of the cross (a point in the cartesian coordinate system, where the width is the X-axis and height is the Y-axis) versus the feedback of the user. All the coordinates here are also expressed in the scale 0 to 1 relative to the screen size. The formula used to calculate the distance between two points is: (x2 – x1)2 + (y2 – y1)2.

The goal to calculate the distance between the original position and the user feedback is to validate if the calibration scenario has an average points distance lower than from the other scenarios supposed to contain attraction factors. In the calibration scenario, there is no variables to disrupt or effect the feedback of the user, so the average distance on this scenario should be considered the baseline error based on the lack of precision. In our hypothesized scenarios with factors of attraction, the tendency of the feedback of the user should get closer to the factor itself, resulting in an average longer distance from the original to the user position when compared to the average distance in the calibration. It was calculated the average distance from the points original position and user feedback of all scenarios separately, and the results will be presented in the next section.

Another calculation which has extreme importance is the location of the factor of attraction in the cartesian coordinate system. This is a necessary step because some of the images used, as the face shaped image and spiral, have their dimension calculated relative to the screen size, meaning that the image size relies on the device from the user, the “click here” button image has a fixed size of 87x32 px. The face shaped image size depends on the aspect ratio of the screen size, the original image size has 1280 x 720 px, an 16:9 aspect ratio, or 1.777 ratio, and it will always keep this ratio (width versus height), so it is necessary to calculate if the screen size have a bigger aspect ratio (common on personal computers in our data), meaning that the image size will be based on the height. Or if the aspect ratio is smaller (frequent on cellphones/mobile devices in our experiment) and the image size will be based on the width screen.

The original spiral image has a dimension of 720 x 720 px, a ratio of 1 and will always have the image size based on the width screen, more precisely half of the width. So, if the screen resolution is 350 x 700 px, the image size will be 175 x 175 px. This rule applies for all spiral’s scenarios (left, center and right).

Not only the image size relies on the screen resolution as we have seen before, but the image disposition always depends on the screen resolution. The figure XXX exemplifies the disposition of the face shaped image in two different screen resolutions with same width, which gives the same image size, but with different screen heights, affecting where the possible factor of attraction would be. In the left image, the missing eye is closer to the top border of the screen, so the Y value is way higher than the Y value in the image on the right, where the missing eye is almost centered, having an Y value closer to 0.5 (50% of the height screen). Because of this behavior, it is necessary to calculate each interaction in an individual level and compare the original cross position and user cross position to the factor attraction point for each interaction in terms of the screen measurements (width and height).



Cortar o cabeçalho dessa fiura

The formula for the face shaped image that gives the point of attraction of the missing eye relative to the screen size is defined as: . This formula should be applied for both axis, X (width) and Y (height). Where the *screen size* refers to the width or height of the screen in pixels, the *image size* it is the width or height of the face shaped image in pixels, and *the location of the factor of attraction in the image* refers to the missing eye in percentage in reference to the image, having a fixed value of 0.55 in the X axis and 0.73 in the Y axis. The result of this calculation is the relative position from 0 to 1 of the factors of attraction in terms of the screen size. With this value is possible to compare if the user feedback cross position was closer or farther to the factor of attraction compared to the original position.

For the spiral scenarios, the possible factor of attraction to be analyzed is the center of the spiral, which is exactly at the center of the image for both height and width. As the spirals image always has half of the width screen size, the factor of attraction on the X-axis for the left spiral has value 0.25 (meaning is at 25% of the screen width) and the center spiral has 0.5 (50% of the width size, exactly centered). For the Y-axis it is necessary to execute some calculations, for example the left spiral has a vertical alignment always at the top of the screen, so the formula for calculation the factor of attraction is: .. For the center spiral the vertical alignment is centered, just like the horizontal alignment, so the formula is: . Both formulas give the factor of attraction position in the Y-axis relative to the screen height in the scale from 0 to 1 to their respective scenarios.

For the “click here” button scenarios, the image size has a fixed width size of 87 pixels and height of 32 pixels. It will be taken in consideration the factor of attraction as the center of the button for both axes, so the image size of 87x32 px should be multiplied by 0.5. The disposition of the left bottom button starts at 20% of the screen width and height in the cartesian coordinate system, and the right top button start from 80% from the cartesian coordinate system. Based on this information, the formula used to calculate the left bottom button for both axis separately, is: . The right top button values should also be calculated separately for the width and height and has the following formula: .

The calculations explained on this chapter produced all the necessary information that it was still missing from the data collection for proceeding with the analysis: Relative position of the factor of attraction for every interaction based on the scenario for the X and Y axis, and the average distance between the two points (original cross position and user feedback cross position) for every scenario. That information, summed up with the relative position of the original cross position in the X and Y axis and the relative position of the user feedback cross position in the X and Y axis, enabled this study to analyze the possible impact of the hypothesized images in the users’ feedback.

With all the data in hands, it is necessary to define a threshold value where it should be investigated the radius of action of the factor of attraction, meaning, how distant should be the cross position from the factor of attraction so it could still suffer the attraction by it. On this question, a dilemma arises: taking in consideration a small radius of action, which explores less data but analyzes the points which will be more affected by the attraction factor or a bigger radius of action that covers more data but at the same time may lose the effect of the factor of attraction itself.

This dilemma can be solved on the data collection process, a wide and extensive data collection which generates plenty of data, would enable a study to concentrate only on the question of how far the factor of attraction extends it radius of influence. Unfortunately, this experiment had no funding and all the data collected was over social media from acquaintances of the author, this experiment collected XXXX unique interactions, counting all the scenarios and calibration, and out of these XXXX, only XXXX were valuable after the data cleaning. From those XXXX, if we take in consideration a radius of action of 15% of the total screen size from the factor of attraction as the center (a diameter that covers 30% of the total size of the screen) it would leave only XXXX unique interactions to all factors of attraction scenarios. Because of this, it will be also analyzed a radius of action of 25% (diameter of 50%), meaning that half of screen size for height and width will be taken in consideration, on this case, there will be XXXX unique interactions for all the factors of attraction scenarios.

## 5.3 Results

### 5.3.1 Calibration

Say the distance

### 5.3.2 Face

Say the distance

Show image of closer farther for 15 and 25%

### 5.3.3 Spirals

Say the distance

Show image of closer farther for 15 and 25%

### 5.3.4 Buttons

Say the distance

Show image of closer farther for 15 and 25%

Show boxplot 15 and 25%

Compare with ANOVA

Null results – all the groups shouldn’t have difference. How we prove that the click button have difference? F-score and p value

Part III: Conclusion

# 6. Discussion and future work

## 6.1 Research objectives: Summary of findings

## 6.2 Limitations

## 6.3 Future research

You do not specify much what would be different if the input was a truly 'social' one or the one that is given by the experimenter.

One possibility is that the input is given by a participant rather than the experimenter ... but what does it change for us?

The other possibilities are multiple: they include different aspects of social interactions, such as eye contact, providing communicative cues, etc. The effect of these features is not what we study at the moment. But it is not featured or clearly analysed in most of the studies on cultural chain. At this stage, I think it is good to abstract away from these aspects. But you could include a few thoughts about how to include them again.

Second question, I would like to sort of understand and explain why we are testing these specific scenarios (spiral, face and click here button), there is any the reason behind the button or spiral would be possible factors of attraction? I think those information would be valuable to be explained, and with your info and feedback it would be great :D

Yep. We could improve on that one. At this point, we are just shooting in the dark in order to document potential effects of the 'context' of transmission. We could definitively try to have conditions that are actually informed by the psychology of visual perception and memory.

I'll try to get some relevant information on my side.

ANALYSIS: USER\_LEFT - ORIGINAL % SCREEN SIZE - POSITIVE MEANS: MORE LEFT

NEGATIVE MEAN MORE RIGHT

USER\_TOP - ORIGINAL % SCREEN SIZE - POSITIVE MEANS MORE TO TOP

NEGATIVE MEAN MORE BOTTOM

LOOK SPECIFICALLY EACH QUADRANT

# References

*Merriam-Webster Dictionary*. (2018, March 11). Retrieved from Culture: https://www.merriam-webster.com/dictionary/culture

*node-firestore-import-export*. (2019, 01). Retrieved 03 30, 2019, from npm: https://www.npmjs.com/package/node-firestore-import-export

*RenanOm92/factorsAttractionFirebase*. (2019, 03 30). Retrieved 03 30, 2019, from GitHub: https://github.com/RenanOm92/factorsAttractionFirebase

Bisin, A., & Verdier, T. (2008). Cultural Transmission. In S. N. Durlauf, & L. E. Blume, *The New Palgrave Dictionary of Economics* (pp. 1225-1229). London: Palgrave Macmillan.

Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture 7*, 89-111.

Henrich, J., & Boyd, R. (2002). On Modeling Cognition and Culture. *Journal of Cognition and Culture 2.2*, 87-112.

Hershberger, S. L. (2014). Cultural Transmission. *Wiley StatsRef: Statistics Reference Online*, doi:10.1002/9781118445112.stat06740.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509), 3489-3501.

Mesoudi, A., Whiten, A., & Laland, K. N. (2004). Is human cultural evolution Darwinian? *Evolution*, 1-11.

Reips, U.-D. (2002). Standards for Internet-Based Experimenting. *Experimental Psychology, 49 (4)*, 243-256.

Sperber. (1996).

Verschuren, P., & Doorewaard, H. (2010). *Designing a Research Project.* The Hague: Eleven International Publishing.

Reference list

* Five Misunderstandings about Cultural Evolution (2002) Joseph Henrich, Robert Boyd and Peter J. Richerson.
* Sperber 1996? Find this, cited on introduction

Standards for Internet-Based Experimenting, Ulf-Dietrich Reips 2002