Interactive Qualifying Projects (All Years) | Interactive Qualifying Projects

May 2014

An Augmented World: How wearable AR devices affect socialization

Evan Andrew Polekoff
Worcester Polytechnic Institute

Johnny Xavier Hernandez
Worcester Polytechnic Institute

Thomas Jeffery Watson
Worcester Polytechnic Institute

An Augmented World
How wearable AR devices affect socialization

An Interactive Qualifying Project
submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfilment of the requirements for the
degree of Bachelor of Science

by
Johnny Hernandez
Evan Polekoff
Thomas Watson

Date:
May 5, 2014

Report Submitted to:
Professor Robert Lindeman
Worcester Polytechnic Institute

Abstract:

While initial impressions of Augmented Reality devices, such as Google Glass, range from concerning to exciting, the studies detailed in this paper attempt to single out exactly what social effects occur when wearing them. One study used a survey to measure first impressions of a person wearing these devices while the other used a participant's stop distance. We concluded that orange Google Glass made participants approach closer than the control and that Google Glass in general has a positive impression on a person's intelligence, trustworthiness and ambitiousness and a negative effect on their perceived generosity. Most importantly, though,this paper lays a solid foundation for how future experiments with more resources can expand on our work to reach more substantial conclusions.

Table of Contents

List of Figures

1. Introduction

Augmented Reality devices have the potential to change society as we know it. This is not a naive statement based on our interest in the technology, but a fact of perception. We are the culmination of the things we experience and our reactions to them. So, when we add an augmentation layer to our visual and spatial perception, we inevitably change. The purpose of this IQP is to analyze one aspect of that potential change, the socialization factor. For thousands of years, our society has focused on verbal communication as the method of information transfer. It is only recently that we have developed rapid, non-verbal methods of communication. With decreased emphasis on personal, face-to-face interaction, society could be losing touch with our communication skills, such as visual and facial cues or tone recognition. We are interested in discovering if augmented reality will make a person more or less interested in interacting with others. We will investigate first impressions, social distance, and how the technology affects socialization with others.

When cell phones were first introduced, they revolutionized the way we communicate with each other. We could now connect with friends that were far away at any time. But there was a cost associated with this advancement. Nowadays people commonly become immersed in smartphone technology, constantly plugged in to the virtual stream of information available at a moment's notice. Some argue that personal connections and interactions have been permanently altered by this technology [2]. The same thing can be said of personal music devices. People can now walk through life to their own soundtrack, completely oblivious to the social potential of the people around them. We will look into whether or not there are similar concerns to be had with augmented reality systems.

2. Background

In order to better focus the study, we decided to research multiple areas of social psychology to see what we could reasonably deduce about the device and its impact on social interactions. By dividing social psychology into the four areas of Social Distance, First Impressions, Trust and Attributions, we were able to find similar research from which we could learn which approaches worked and did not work in the past. These past experiments often used cell phones or technology that was new at the time but has now pervaded our everyday lives. By expanding on what we found and applying it to the newer technology of Google Glass, we were able to structure our study properly to produce meaningful results. Listed in this section is the summary of our research on Google Glass and each of the areas of social psychology that we decided on which to focus.

2.1. Google Glass

Google Glass is a wearable Augmented Reality (AR) device, which from a semantics point of view, anything that changes the way one perceives their environment could be considered Augmented Reality. Google Glass augments a person's vision by providing users with a small projected screen in their field of view. The device can be connected to a phone to gain access to any applications that the user has installed, including games, utilities, and social networking applications. When connected to a phone, Google Glass can also receive notifications when the wearer receives an email, text or other compatible notifications.

The side of the device is a touchpad that serves as the main interface with the device. Users swipe down to close out of apps, swipe sideways to select options and apps, and tap to make a selection. Two-finger swipes and taps are also supported within the apps if developers want to use this functionality. Touching the device is not the only method of controlling the device. Google Glass has speech recognition technology that listens for voice commands. Users can send texts and emails, perform searches on Google and launch apps completely hands free using these commands.

Also embedded in the device is a front facing 5 megapixel camera, allowing the wearer to take pictures and videos at 720p. When connected to the internet, the camera can be used to stream video online. Because the camera is one of the more important features of the device, there is a small silver button on the top of the frame for even quicker access.

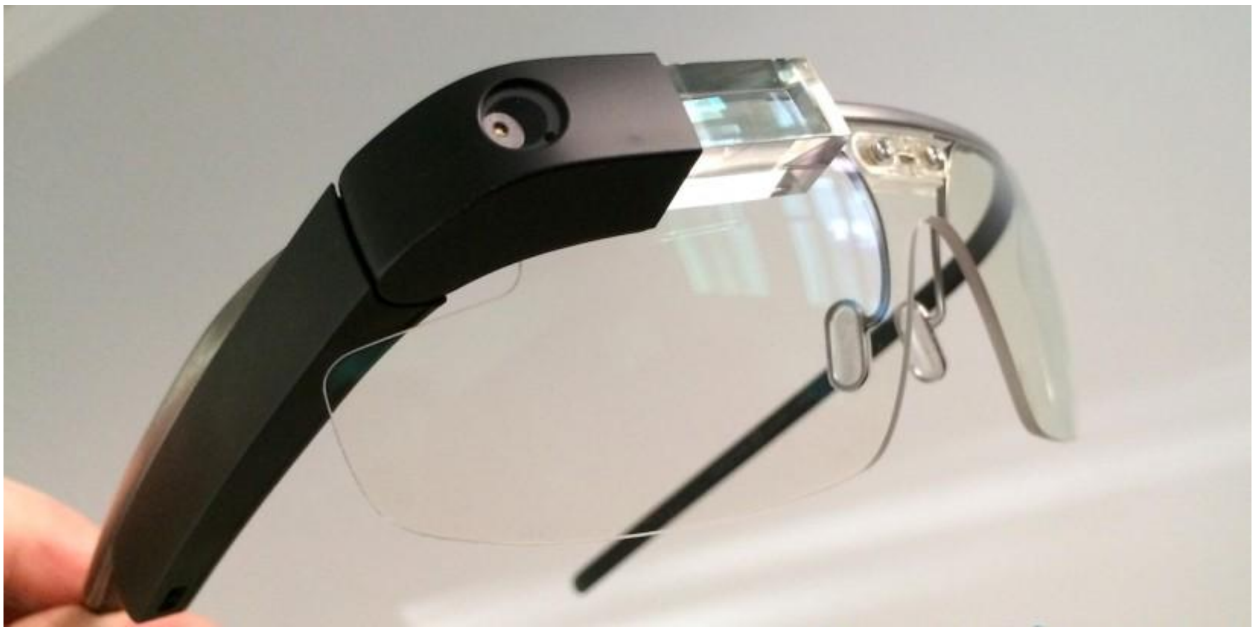

Figure 1: A picture of black Google Glass [http://www.slashgear.com/google-glass-up-for-sale-to-the-public-once-more-24326407/]

2.2. Social Distance

Social Distance is defined as "the extent to which individuals or groups are removed from or excluded from participating in one another's lives" by Dictionary.com. The concept extends from physical distance and can be attributed to emotional distance created by differences in race, social class, sexuality, or trust. First we will visit the three most common methods of measuring social distance. Then we will look at previous tests that were performed and results that were published that could relate to our work.

The simplest test used to measure social distance is known as the Social Distance Scale (SDS) [8]. This survey can measure someone's feelings of social distance using a specially designed scale known as the Bogardus Social Distance Scale. The survey asks demographic questions using the Bogardus Scale, and the answers can be used to determine social distance based on how the questions are designed.

Another method of testing social distance is the stop-distance experiment [20]. There are two common forms of this experiment. The first has a tester standing at the end of a long runway. The subject is asked to approach until they feel comfortable. The distance away from the tester at which they stop is indicative of their comfort level. This utilizes the concept of the "personal buffer zone," which is a physical bubble zone around a person where if something crosses into this buffer zone, the subject begins to feel uncomfortable. To get a better feeling of the size and shape of the buffer zone, the participant will often be asked to approach in different ways. For example, the person can walk forwards towards the tester or sideways at varying degrees. This will create a map around the person that can be used to measure their comfort level with someone. A second version of the experiment has the subject stand still and the experimenter approach from varying angles, instructing the subject to tell them to stop when they begin to feel uncomfortable [1].

In researching previous experiments that are applicable to AR devices, we looked into simpler versions of augmented reality as a jumping off point. Our first class of device under consideration was the portable audio device [8]. When researchers did the stop-distance experiment with people wearing headphones, they found that when listening to music, people held a larger buffer zone than they did under control conditions. Another experiment looked at lighting conditions and their effect on a person's personal buffer zone [1]. They had the experimenters approach the participants under different levels of illumination, and found that buffer zones were significantly larger when the lighting was dim. These findings suggest that personal space requirements increase when we have impairments in our vision and hearing. AR devices can cause similar losses in perception, and so these kinds of experiments could be useful for our work.

Once a comfortable distance has been established between two interacting individuals, direct interaction between individuals can begin. In this stage of social interaction, judgments of the participants are made, and explicit preconceived notions about another's personality traits, such as likeability or hostility, begin to form. Most of these judgments are made in the first seconds of a glance at another's face, forming the first impression of another individual. All of this could be influenced by the presence of AR devices.

2.3. First Impressions

First impressions define the judgments one makes when meeting with other people for the first time. This instinctive behavior to begin creating preconceptions of other individuals is due to an underlying behavioral trait that leads someone to judge another as being dangerous or suitable to interact with. Impression formation occurs daily beginning from the first instant where eye contact with others takes place, unconsciously making judgments that will alter how one will perceive future information of the second individual. The speed of these first impressions results in impressions based on facial features. Specific sets of facial features tend to skew judgments of a particular person based on criteria that is common in most people.

Preconceived notions generated from first impressions are of noteworthy study as they have been proven to be correlated to a trend of belief perseverance. Belief perseverance is a phenomenon that explains a trend of social impressions affecting judgment of other individuals as the first impression begins to affect how one perceives information of another party [21]. The effects of belief perseverance make the observation of first impressions important, as first impressions generally can dictate how comfortable one can feel in the presence of another. This may even lead to self-defeating behaviors or trust issues with whom one interacts. Several researchers have measured the existence and effects of first impressions through various uses of "stimulus materials (trait adjectives, paragraphs, direct observations of behavior) and judgment tasks (attribution of personal characteristics, general evaluative ratings, behavior predictions)" [27]. The lasting effect of first impressions has brought up a great amount of speculation, though three primary explanations have been agreed upon by most researchers. The first being a general regression of attention, where the solidification of ideas formed from the first impression stem from someone being less attentive as interactions flow, known as the attention decrement interpretation [27]. The second explanation states that one usually assumes that their first interpretation of another party is the more correct assertion, and will begin to assume incoming information is less reliable [27]. The third explanation, the biased assimilation interpretation, is using what was drawn from a first impression to judge incoming information of other individuals in a manner to keep it consistent with their respective first impression [27].

First impressions can be drawn from various sources, be it from the overall posture and composure of another person or unconsciously assuming traits from that person's facial features. The face is by far one of the most common sources of first impressions, and coincidentally forms the strongest first impression judgments. There is quite clearly a lot that can be taken in from a single glance at a face and while this analysis dates back to the 1700s, even today certain facial features are being attributed characteristics and judged very differently from others. For example, there is a study that demonstrates people attributing baby-faced adults as having "larger eyes, thinner eyebrows, and a rounder face" and correspondingly judging them in the experiment as "less socially autonomous and more naïve than their more mature faced peers" [31]. Another study found that a large number of participants would declare certain faces as threatening over others consistently [3]. This leads to the conclusion that there are universal characteristics of the human face that lead one to infer something specific about a person's facial features and that this assumption may easily be held by other individuals. The forming of these impressions is not only deeply rooted in those making judgments at the first glance of another's face, but several others will share this same judgment.

While a first impression is going to set the initial deeply rooted judgment of several people, the final aspect of an impression that must be observed is the amount of time it takes to reach this judgment. Multiple experiments have been performed that conclude several different time intervals for different judgments being made, such as attempting to gauge an individual's intelligence or aggressiveness. One experiment, focusing on the acquisition of impressions of neutral faces and their perceived levels of threat, found that mostly the same faces would trigger a perception of a threatening face in as quickly as 39 milliseconds [3]. This is not true for all judgments, as the quick perception of threat is due to human instinct to avoid threats quickly and deal with them accordingly. Other studies have shown there is a slight increase in time intervals to measure consistent impressions on other traits, such as intelligence or competence. Non-instinctive trait perceptions formed from first impressions have been studied to become consistent from individual to individual in as little as 100 milliseconds [30]. This same experiment also measures the confidence of participants in their first impressions. While the peak of correlation between participant answers with each other is at the 100 milliseconds mark, confidence levels rise up until the 1000 milliseconds mark. The confidence level within a first impression is definitely of note, as while it may not change the initial first impression, a greater belief in their first impression may solidify the initial base assumption, leading to a greater effect of the initial perception of the subject.

While facial features play a big role in determining the first impression, modifying one's face with the addition of an AR device could become a new element for an individual to take into consideration when forming a first impression. Using pre-existing information based on plain face first impression analysis, it would be of noteworthy interest to measure the effect of an AR device on the first impressions of an individual, as first impressions may dictate one's overall view of the wearer and all incoming information or judgments made on the wearer.

2.4 Trust

People often base the decisions they make and their behavior on more than a surface level impression of the other. The level of trust that one has in another is a major decision making factor. When a person trusts someone else, it means they believe that the other will do the right thing. Trust is something one person gives to another [12]. The definition of the "right thing" is determined by the person giving the trust, which can vary greatly between people. The amount of faith put into another person is dependent both on how trusting that person is and on any biases that person may have towards the other.

It is important to draw a distinction between trust and trustworthiness. Trustworthiness is also perceived by the person giving the trust, but is much more surface level. First impressions play a big role in how one measures the trustworthiness of another and everyone could have a different trustworthiness level for the same person. Most common survey questions used to measure trust ask participants how much they trust someone else. These types of questions are actually considered to be measuring trustworthiness [13]. Someone's own trustworthiness can be measured by how trusting they are of others. The level of trust one has in another varies over time, building up with positive interactions and decreasing with negative ones. The questions that should be asked in a survey attempting to measure trust should look at past experiences with people and ask how that person has previously trusted others [13]. These types of surveys work well when paired with experiments that can concretely measure trust as a byproduct of the interactions between participants. The survey questions are used to gather the participants' thoughts on the interactions and compare the results with the gathered data.

The demographic of the sample used for measuring trust plays a part in the results. Past studies have shown that people in wealthy societies tend to be more trusting than those in less wealthy ones. Wealthy people within societies are also more trusting than less wealthy people, even if they are in a generally wealthy society [14]. Other than wealth, factors affecting trust also include race, culture, morals, place of origin or essentially any differentiating factor. People tend to trust those more similar to themselves [4]. On top of this, general trust perceptions in the United States have been decreasing over time, potentially due to the increase in diversity.

When conducting research in trust, the phrasing of the questions asked and how they ask about in-group trust versus out-group trust is also an important factor. In-group trust is how a person trusts others they have either known for a while or who live close by. Out-group trust is how a person would trust someone they met for the first time or someone from a different culture. A large number of past studies' survey questions asked participants questions in regards to "most people". The radius in which "most people" lie and whether that radius is small enough to be classified as in-group or not is ambiguous. This radius can be determined by a number of factors, such as cultural beliefs [10].

The amount of communication between people greatly increases the trust between them as well. Face to face communication has a more profound effect, but any kind of communication helps. Even in business communications over the phone, trust in a company has increased over companies that a customer has had no contact with [21]. This can tie in with the survey questions that ask about past actions rather than what a person would do. If someone interacted face-to-face with another in a lab setting, this would cause a difference in trust than if a hypothetical question had been asked.

There have been many studies in the past that have used trust-based games or activities in tandem with surveys to attempt to simulate real situations. The level of trust between participants can be measured by counting an amount of money exchanged, the number of points that each player receives or any other physical measure based on the type of game [7][11][13].

2.5. Attributions

The choices people make and the actions they perform are in part due to the past experiences of the individual. Because everyone is different, the reasons people perform certain actions and the actions themselves vary between individuals. It is near impossible to predict what the actions that a certain person will perform since their own upbringing is a mystery. Still, people find the need to make sense of these actions and situations, which is where attribution theory comes in. Fritz Heider (1958) defines attributions as ways for people to explain the actions of others after analyzing and observing their behavior [17].

There are two kinds of attributions: personal attributions and situational attributions. Personal attributions are based on the characteristics of the person performing the action, such as their personality, mood, physical characteristics and skills. Situational attributions are based on external factors, such as the events leading up to the current action, other people involved in the action and uncontrolled circumstances.

The correspondent inference theory says that people attempt to predict the reasons behind another's actions based on the actor's level of choice in the matter, how normal the behavior is and how beneficial the results are to the actor [17]. If an actor chose to perform one action over other possibilities, then that choice gives more information about the actor's background. For instance, if the only sweatshirt color available is bright pink and a man bought it, then the choice doesn't say much about him. However, if the man chose to buy a bright pink sweatshirt over a more commonly purchased color, observers will be more likely to attribute this behavior to personal factors. If a performed action is considered to be different from social norms, it reveals more about the person than if they were just following conventions. If everyone in a room was wearing the same clothing except one person, that outlier would receive attributions from everyone else because of his "abnormal" behavior. The way the outcomes of the actions affect the actor also affect what others infer from those actions. For instance, if the actor's job offers high pay, has a good location and is enjoyable, observers may not know the exact reason he chose his job. However, if the job is not enjoyable and is in a bad location, but it has a high salary, it is potentially more obvious why he chose that job [17].

Because people are biased towards others, their attributions about others are naturally going to be biased as well. The availability heuristic is the tendency of a person to make assumptions based on the events that they most commonly see or are most readily available to them [17]. This heuristic is a key influencer on a person's biases because it is entirely made up out of their own past experiences. If someone remembers more times where a man of a certain race has helped him rather than hurt him, he will more likely attribute the behavior of someone of the same race to a good cause. When this is paired with the false-consensus effect, which states that people tend to overestimate the degree to which the public perception of an issue lines up with their own beliefs, people have no problems making negative attributions with no evidence to back them up.

The key point of attribution theory is that people will attribute the behavior of others differently than they would attribute their own behavior. This is called the fundamental attribution error [17]. In a study on the intentionality of being disruptive while using a cell phone in public, 20.7% of participants attributed the other's behavior to being caused by boredom and 16.1% attributed the behavior to being intentionally rude or to promote self-importance. Of these same participants, only 12% would attribute their own use of cell phones to being bored and 1.3% would attribute it to being intentionally rude [9]. This data shows that the inferred reasons for behavior is judged more harshly for others than for one's self.

Tying in to the fundamental attribution error, people tend to pick self-serving biases in order to boost their own self-esteem. Using the cell phone study, perhaps the people who attributed cell phone use to promoting self-importance were jealous because they did not have cell phones of their own. This way, they are able to make themselves feel better about not having cell phones by attributing negative causes to their use. This self-serving bias can also be combined with the false-consensus effect to make the attributor feel as though others agree with his self-centered belief. Amy Mezulis (2004) analyzed 266 studies and found that self-serving biases are prevalent in every culture except for some Asian ones. As a result, a self-serving bias is something that needs to be considered and accounted for [17].

2.6. Personal Experience with Google Glass

Because the students (the authors) performing the study had access to two pairs of Google Glass for the entirety of the study, part of becoming accustomed to the technology involved wearing it on a daily basis. Included here are a few of the more interesting stories and observations from the authors when we wore the devices in public:

"I once wore the black Google Glass to a large show on campus. It was so packed, it was standing room only by the time I got there, so I was standing in an open area in the back. When I looked to my right, there was a group of about 5 Asian students talking to each other, looking and pointing at me. I can't confirm it, but I assume it was because of the Google Glass, seeing as how this never happened to me before."

"The first time I wore Google Glass outside, I couldn't help but laugh whenever anyone looked at me. I'm not sure why; it's not as though I felt silly for wearing them or that I felt like a celebrity. It wore off after the first day, though. I guess I just really liked the technology."

"The entire time I wore Google Glass to all of my meetings, only one person asked me if he could try it on. When he did, everyone else in the room spoke up and asked as well, even the ones who had seen me wear it for days before hand."

"I was wearing the orange Google Glass in daylight. It was around 4:00, so there were students changing classes, but it wasn't so busy that I wouldn't be noticed. I was walking through a 4-way intersection of walkways in the middle of campus when I noticed someone ahead of me looking at my Google Glass. As I continued to walk, I saw his head follow me, even though he was walking the opposite direction. Since he was no longer watching where he was going and was focused so intently on my Glass, he walked directly into a girl in the middle of the intersection."

"The first club meeting I wore Google Glass to, I sat next to two club members having a conversation. They brought up a point I wanted to add on to, so I spoke with my contribution. They looked at me because I was talking and the girl sitting next to me cut me off and said "hey, you have a thing on your face" in a surprised reaction. By this point, I was past the point where I was actively aware of my Google Glass, so it was getting annoying that I couldn't even contribute to a conversation without someone bringing it up."

"Once when I approached the intersection between my off-campus housing and the main campus, there were already a line of cars at each of the four ends of the intersection. Normally I would have to wait for up to a few minutes before someone would let me cross, but this time because I had Glass on, all the drivers immediately stopped and let me pass, some pointing to their copilots and what not."

"Once I was having a conversation with a professor where I was standing behind him as he worked on his computer. During our conversation he happened to glance back and noticed that I had Glass on, and although he did not pause in conversation he did immediately look away and smile. After we had finished talking, he asked me if I had been recording him the entire time, and only after I told him that I had not did he look back at me."

3. Methodology

Although we researched four areas of social psychology, we decided to use only Social Distance and First Impressions in our final study due to time constraints and limited resources. In the following sections, the methodologies for the studies we performed are listed along with our results.

3.1. First Impressions

The original proposal for the first impressions experiment, being mostly derived from an experiment performed by Janine Willis and Alexander Todorov in 2006, underwent several changes since its proposal. Originally, the experiment would have used full frontal head-shot photographs with neutral expressions with which a participant would then answer dichotomous judgment questions to determine the possible effects the device may have on certain first impression judgments. The viewing of a photograph would have been held at a constant fixed time, within 1500 milliseconds, to ensure the first impression is the base of the participant's decision. After each judgment question, the participant would have been asked how confident they are in their answer on a 7-point scale (1 being least confident and 7 being most confident). Response times would have been recorded, as we believed that the amount of time on any given question should be about 2,000 milliseconds to retain data for the first impression. The number of models for photograph stimuli would have been roughly twenty models, ideally both male and female in equal numbers.

The study performed in Willis and Todorov's research was performed with all participants being guided by an experimenter in the same room with a total of 245 participants, all of which were being compensated. Unfortunately, due to time constraints and limited resources, we were unable to gather such a large number of participants. For this reason, the study performed was slightly altered to be friendlier for an online user base and to reduce the size of the experiment to adequately construct and deploy the experiment within our given time frame while still obtaining a large response base.

The current experiment design used in this study utilized only four models within the WPI community, two of which were male and two female, under the criteria that there be no facial hair, minimal blemishes, or other facial obstructions, such as piercings or jewelry, to remove possible distractions. Participants were also required to wear neutral colored clothing.

Also, all of the selected models were Caucasian so as to remove race as a factor in participant responses. All photographs were taken in Tech Suite #212 in the WPI George C. Gordon Library, to maintain a consistently lit solid white background. Models were instructed to maintain a neutral expression, to remove emotions possibly affecting responses. Each model took a total of six photographs; all being frontal head shots, wearing different augmented reality devices and a single plain photograph. The devices used were headphones, sunglasses, glasses, an orange pair of Google Glass, and a black pair of Google Glass. An example of one of the images can be seen in Figure 2.

Figure 2: An example image that was shown in the survey of a student wearing orange Google Glass

Participants were informed in the beginning of the study that the survey is about first impressions and to make decisions quickly. After the introductory page explaining the experiment to the participant, the user proceeded to the next page containing the first photograph centered on the screen. The amount of time the participant could look at the photograph was not fixed, and participants would decide on their own when they wished to continue to the next page. The fixed time element was removed from the study as it could only be enforced when the study was performed in person with an experimenter. Various factors could possibly affect what the participant sees in that time frame, such as a faulty or slow internet connection. According to the results in Willis and Todorov's experiment, the amount of time viewing the photograph will not change the first impression, provided that the participant observes the photograph for at least 500 milliseconds, and the only variable they observed to change in different time intervals was the confidence. For this reason, the confidence levels accompanying the judgment traits were removed, and the judgment traits themselves were rated on the seven-point scale. Each participant would then repeat the process of viewing a photograph and agreeing or disagreeing with judgment statements until four photographs, one of each model, was rated. The total time to complete this study for each participant was also recorded along with their responses to ensure a good set of data and adequate first impressions were obtained.

The images shown to each participant were pseudo-randomly chosen from the pool of photographs. Six different sets of photographs were arranged so that a different device was shown, with the exception of the different colored Google Glass, and no single model was shown more than once. The images displayed can be seen in Figure 3. The order within these sets was also randomized to avoid possible skews in data depending on answers changing as users possibly become more aware of what was being studied as the experiment continues.

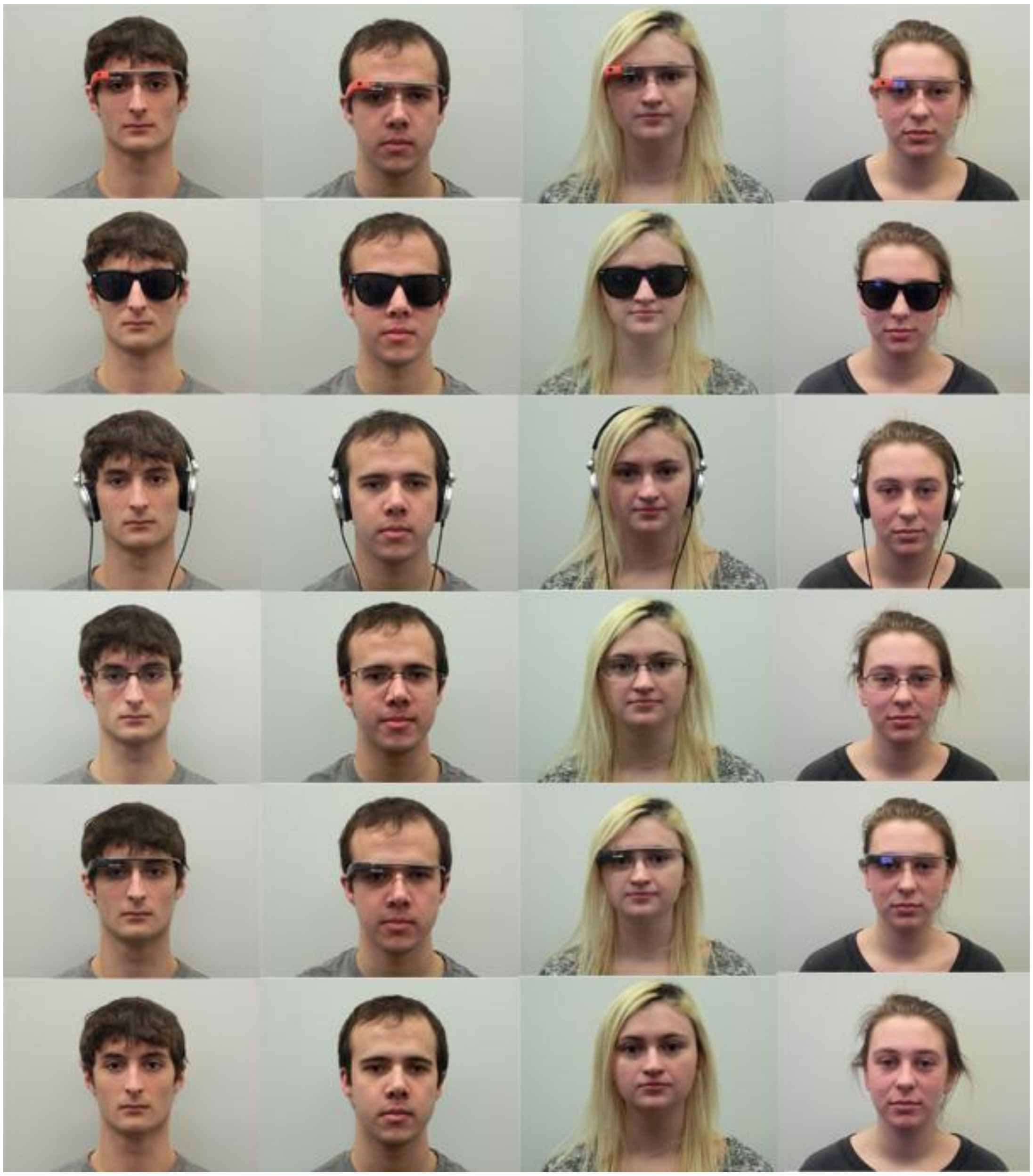

Figure 3: The images that were displayed in the survey, arranged in different orders for each participant

The study was deployed to two separate populations, the WPI community and online through the Amazon Mechanical Turk service. Amazon Mechanical Turk users were paid $0.20 for completing the survey. The survey for the students of WPI had to be adjusted for the scenario where a participant recognized one of the models by adding a question asking about the participant's familiarity with the model after the viewing of the photograph and before displaying the judgment statements. If the user answered that they knew the model, the trait judgments were skipped and the participant moved on to the next model. A total of 160 participants from Amazon Mechanical Turk completed the survey, and a total of 129 participants from the WPI community completed the survey with 86 partial responses recorded.

3.2. Social Distance

For this study, we performed a variation of the standard stop-distance approach test as outlined in the background. This allowed us to test a combination of first impressions, trust, and attribution properties associated with wearable AR devices. Using the concept of personal buffer zones, it was possible to measure a subject's comfort level with another person or thing by the distance at which they stop when asked to approach it. We took this physical stop distance as a measure of comfort level with the augmented reality devices. The physical set-up we used is shown in Figure 4.

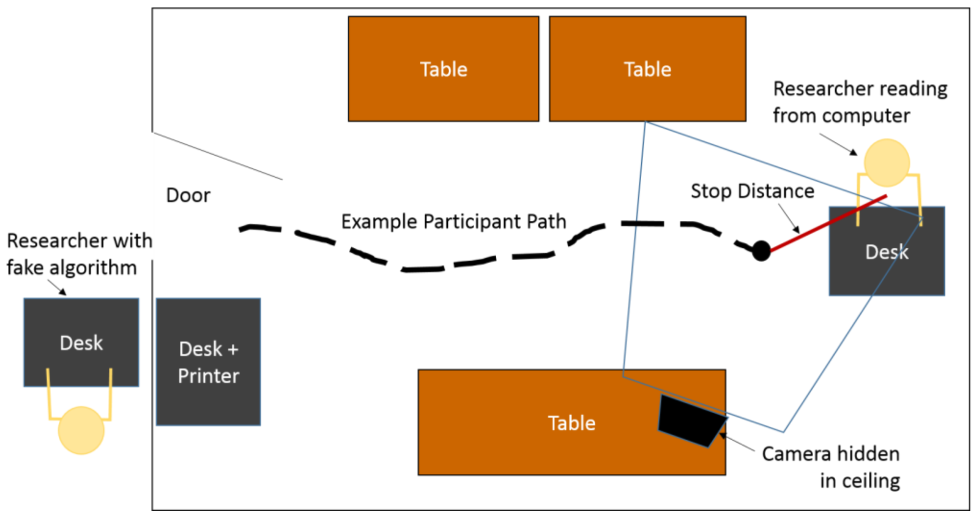

Figure 4: A sketch of the layout of the room used for the stop distance study.

During the study, participants approached the experimenter and were told to stop at a comfortable distance. The experimenter was wearing one of 4 devices: black Google Glass, orange Google Glass, headphones or nothing. The run where the experimenter was not wearing an extraneous device was considered the control. By using the same experimenter for all experiments, we eliminated the traits of the experimenter as a possible variable in the participants' stop distances, allowing us access to a less divided participant pool.

The measuring for the stop distance was done after each run by looking at the recordings from a hidden camera. There was tape placed on the floors in a chaotic way to the point where it was unnoticeable to participants but very useful from the camera's view. After the participants were recorded, we could lay lines over the tape marks and floor tiles visible from the camera view to give us a grid for analyzing the stopping distance, as shown in Figure 5. At the start of each new participant's research session, the experimenter climbed up to the rafters and turned the camera on, limiting the recording for each session. The camera was screwed into a tripod that was duct taped to a bar running across the ceiling. The duct tape held the tripod in place so that the camera was in relatively the same position each time.

Figure 5: A screenshot from one of the videos marked up to make measuring easier

To keep the subject from fixating on the device and also to keep from attracting only people who were over interested in the technology, we developed a fake IQP study which we called Food Fortune. We advertised that we could guess a subject's favorite food using a special computer algorithm. A very simple shell script was also created to further trick participants. The script would provide experimenters with their next prompt and take in input to further convince participants that it was working.

To get volunteers for the study, we had two methods. The most successful was using a resource at WPI called Sona Systems. Using Sona, a study was created in the web based system with study information and scheduling functionality built in. WPI students filling their social science requirement must get a certain number of credits from participating in student projects like this one. The majority of our volunteers ended up coming from this pool, which was very convenient and our most representative sample. Our other method of getting volunteers was by sending emails to the student body. Due to our low budget, this method proved less effective as we were asking busy students to volunteer their time for no reward. To manage registration of volunteers, we used a website called Slottr in conjunction with Sona, to support open registration.

When a subject arrived, we read them a brief overview of the food fortune experiment, and told them that there was another participant in the other room who they needed to ask a series of questions, although the participant in the other room was actually our experimenter. We told them they had to approach the subject until they thought they were at a comfortable distance to have a conversation. Whenever any skepticism, doubts or confusion with the way the algorithm worked arose, we pushed the question off by telling participants that we would only be able to answer their question at the end of the study. The experimenter explaining the study also had to tell participants that they had to approach the individual because otherwise they would not get the information that they needed. Participants were told that people tend to hold back information if they have conversations at uncomfortable distances or have to raise their voice to divulge personal information. This was just a ploy to get participants to approach the experimenter to a comfortable stop distance without explicitly telling them what the experiment was about. After the explanations were done, the participant would ask a series of four questions to the researcher within the other room, leaving the room and re-approaching each time. After they had done all four walks, we would bring them back in and debrief them, telling them the true nature of the study and what we were looking for.

4. Results

After performing the experiments outlined above, we collected and analyzed data for each type of test. We have consolidated and laid out our results in the sections below.

4.1 First Impressions

The results of the first impression experiment have been divided into two separate data sets. As a result of the different methods of online deployment for the survey, data representing two populations, the Amazon Mechanical Turk population and the WPI population, were recorded. As these populations differ in cultural norms and values, both data sets were analyzed separately and any agreements between the data sets could possibly link to a global perception, but will require further study and analysis to conclude on.

The Amazon Mechanical Turk responses for all four of the models and devices have each received a mean of 26.75 responses with a median of 27 and a standard deviation of 0.98907. This signifies a fairly even number of responses for each model and their associated judgment questions along with the device being tested.

The descriptive statistics for each of the statements can be found in appendices B and C. Figures 6 through 12 show the mean ratings for each judgment statement, along with all of the ratings corresponding to each device across all models. This data is from the Amazon Mechanical Turk community. Using the means for the plain frontal shots as the control group, the mean of responses for each device for each judgment statement was compared to the control. Running one tailed t-tests on each control and device pair for each judgment statement across all models reports whether the device has made an impact on the judgments of the Amazon Mechanical Turk population. Significance is being observed at both the 90% and 95% confidence levels.

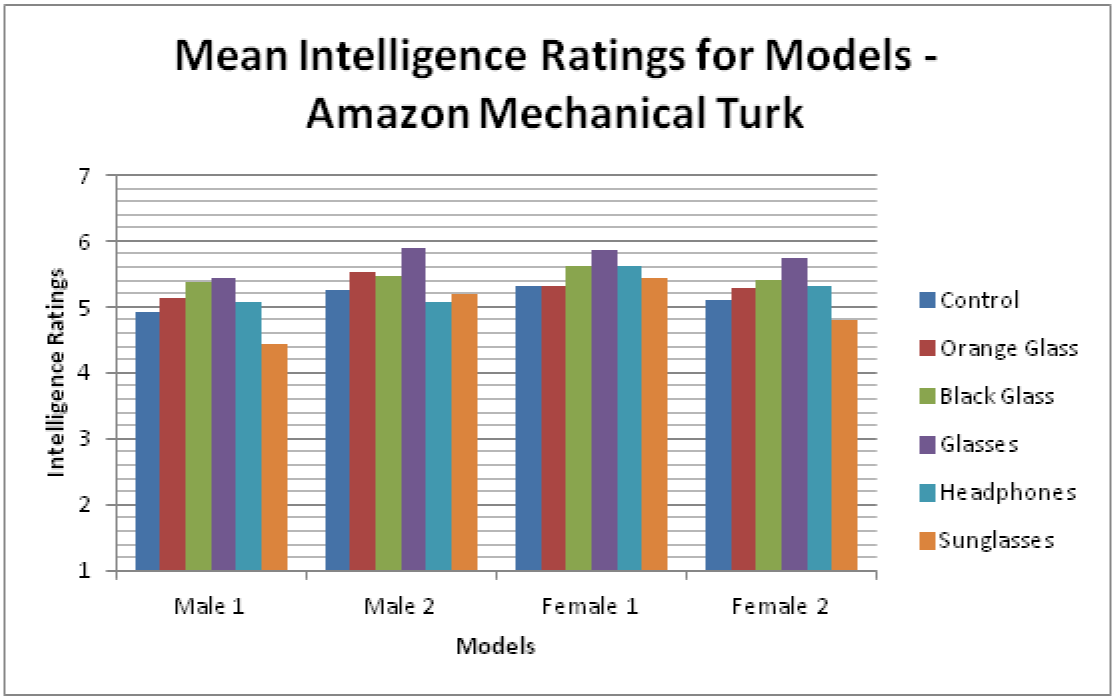

Figure 6: A graph showing the mean intelligence ratings from the survey results

For the judgment statements involving intelligence, as shown in Figure 6, statistical significance for glasses was found across all models, at a confidence level of 95%. Comparing the mean scores for the controls of all the models with their respective scores for the glasses demonstrates impressions of greater intelligence among wearers of glasses within this population. The black Google Glass did not test as favorably, but Male 1 and Female 1 means both received statistically significant mean increases within a 90% confidence level, which may cause the black model of the Google Glass to affect the impression of intelligence. Headphones and Sunglasses both had one statistically significant comparison in Female 1 and Male 1, respectively. Headphones increase intelligence scores and sunglasses decrease intelligence scores in these significant model scores. As only one of the three models obtained significance for both cases this data would require further testing before conclusive assumptions can be made. No significance could be found for the orange model of the Google Glass across all models.

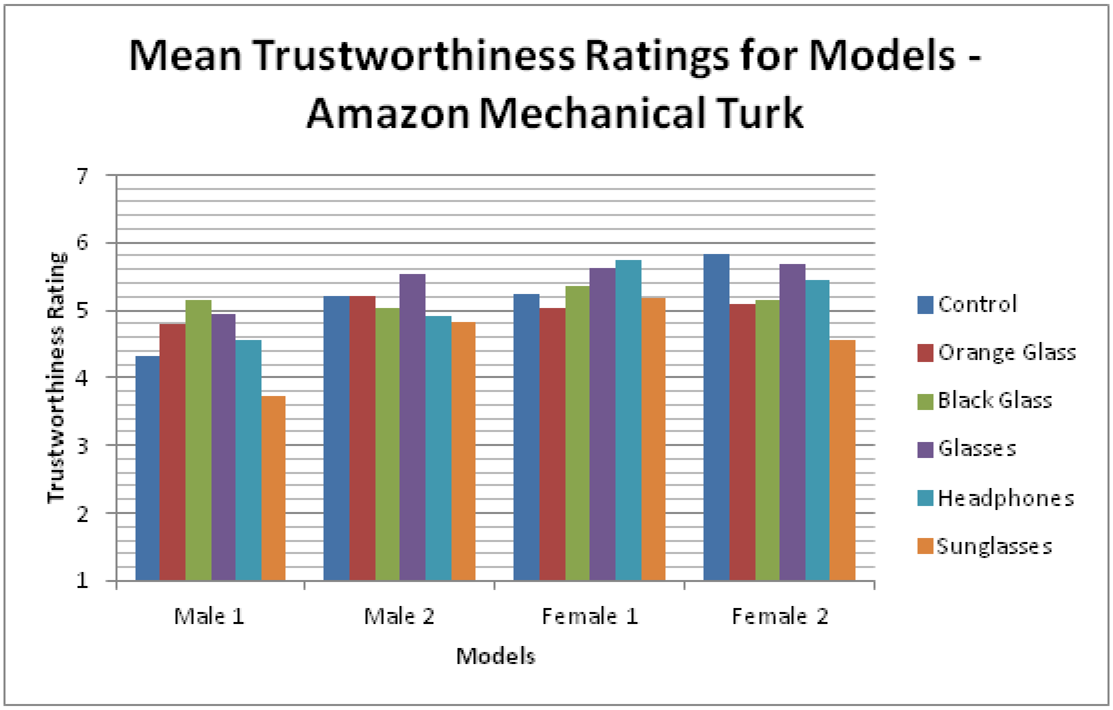

Figure 7: A graph showing the mean trustworthiness ratings from the survey results

Judgment statements involving trustworthiness, as shown inFigure 7, for both models of the Google Glass resulted in two significant comparisons within the 90% confidence level. While a statistically significant shift in trustworthiness was observed for both models of the Google Glass, whether the effect is negative or positive cannot be concluded from the significant data provided by Male 1 and Female 2 for both Google Glasses. The changes observed in one model, specifically Male 1, appear to place more trust in the Orange Glass and less trust for the Black Glass, but the opposite is observed in Female 2's data. Sunglasses also appear to have two statistically significant comparisons in the 95% confidence level, in Male 1 and Female 2's results. Observing the means for the data appears to indicate the use of sunglasses makes one appear less trustworthy for this population. Both glasses and headphones received only one significant point of reference from the four models, Male 1 and Female 1 respectively, both appearing to increase the trustworthiness of an individual, but as these observations were not found in more than a single model perhaps more research should be done to conclude on this.

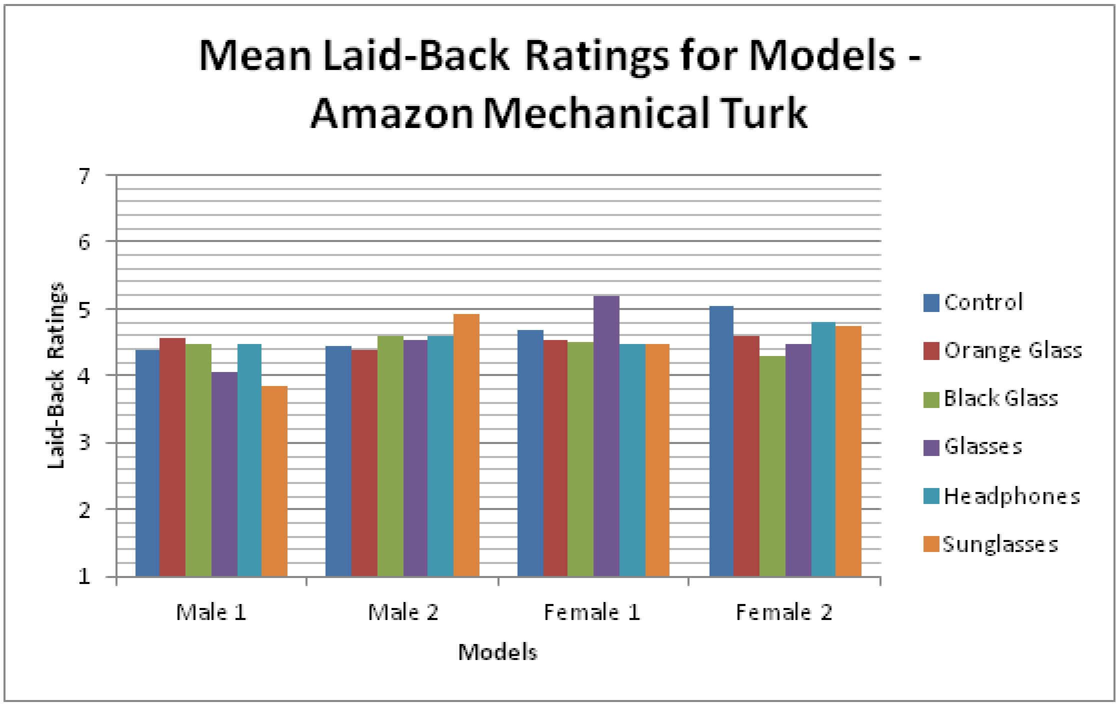

Figure 8: A graph showing the mean "laid-back" ratings from the survey results

Judgment statements covering the trait of relaxedness or the appearance of being laid-back, as shown inFigure 8, found both male models control to sunglasses means within a 90% confidence level. Though the ratings for one male are positive and the other is negative, sunglasses appear to have an observable effect on the impressions of relaxedness for males, though the same significance could not be said for females. Statistical significance within a 90% confidence level for both models of Google Glass and glasses was observed when comparing the means of Female 2, each one having a negative effect on the outlook of relaxedness. Black Google Glass appeared to have the strongest significance among the three, being the only one significant at a 95% confidence level, but as none of the other models demonstrated this significance, we will not deem this reliable at this point.

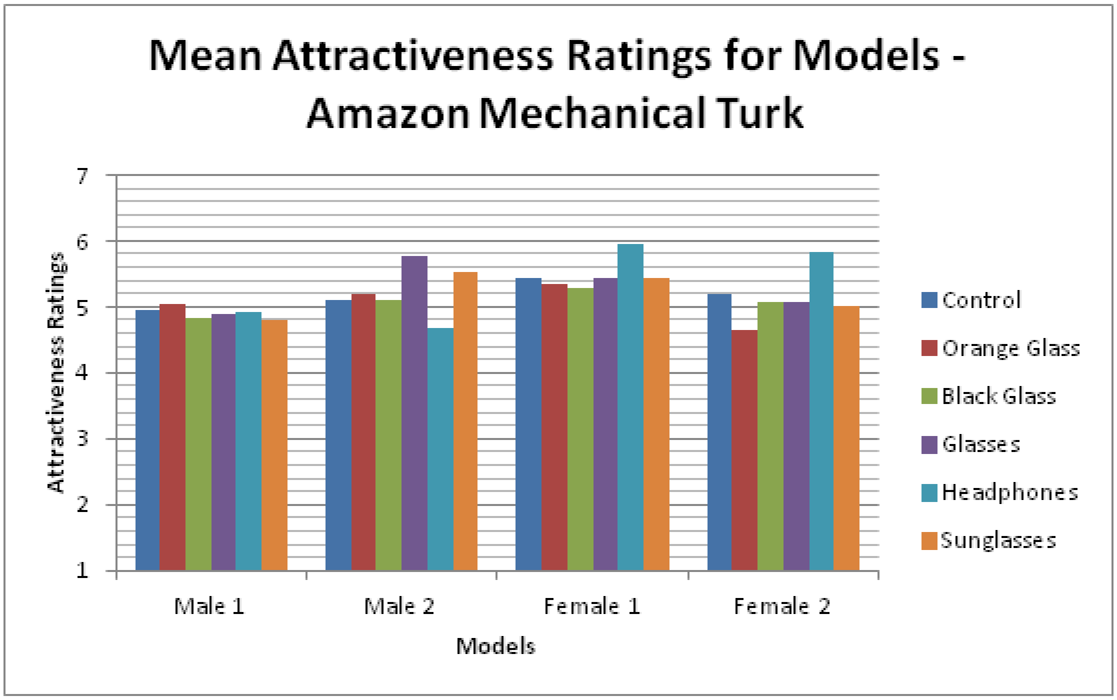

Figure 9: A graph showing the mean attractiveness ratings from the survey results

Judgment statements on attractiveness, as shown in Figure 9, yielded both female models with statistically significant comparisons for headphones at the 90% confidence level. It would appear from the data that headphone use possibly increases impressions of attractiveness for females, though this was not observed in males. Orange Google Glass and glasses both had a statistically significant comparison on one model, Female 2 and Male 2 respectively, both decreasing the attractiveness of the models, though this was only observed in a single model out of the four possible. No statistically significant comparisons were made for the black model of Google Glass or sunglasses.

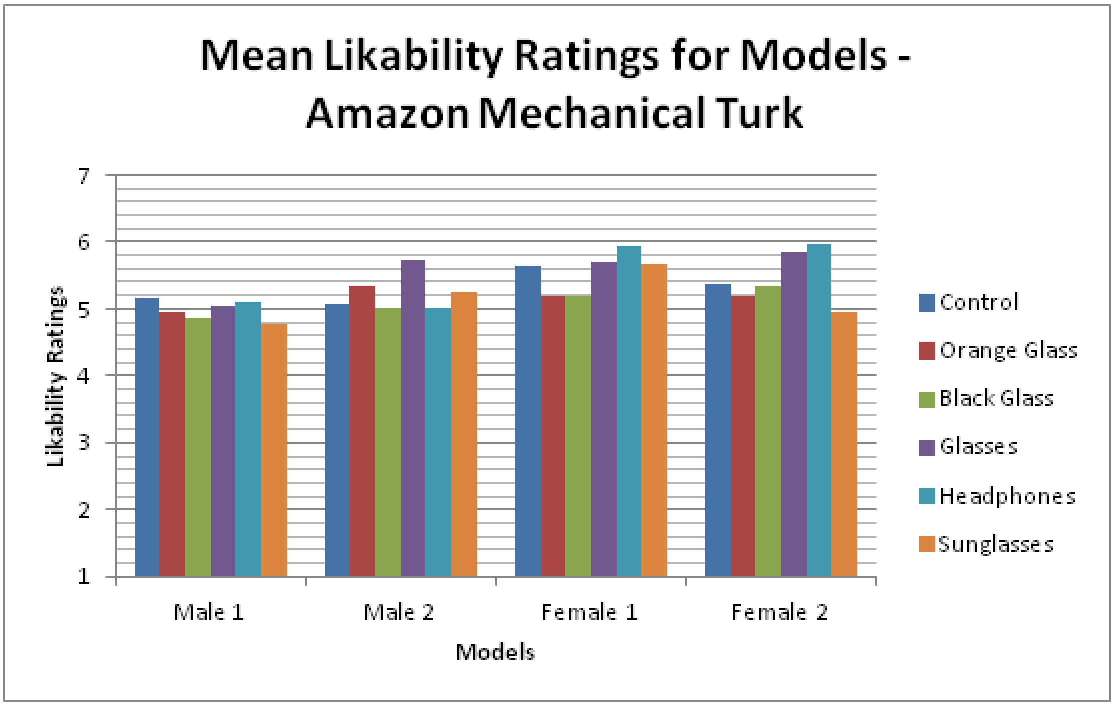

Figure 10: A graph showing the mean likability ratings from the survey results

Judgment statements on likability, as shown inFigure 10, draw a parallel between the means of glasses and their respective controls with a confidence level of 95% within Male 2 and Female 2. According to the comparisons between the means, it appears the impression of the likability of an individual increases with the use of glasses. Both models of the Google Glass found significance in Female 1's mean comparisons, both decreasing the model's likability at a confidence level of 90%. Headphones and glasses were also found to be significant in Female 2's mean comparisons, but at a 95% confidence level. Headphones increased this model's likability while sunglasses had decreased the likability.

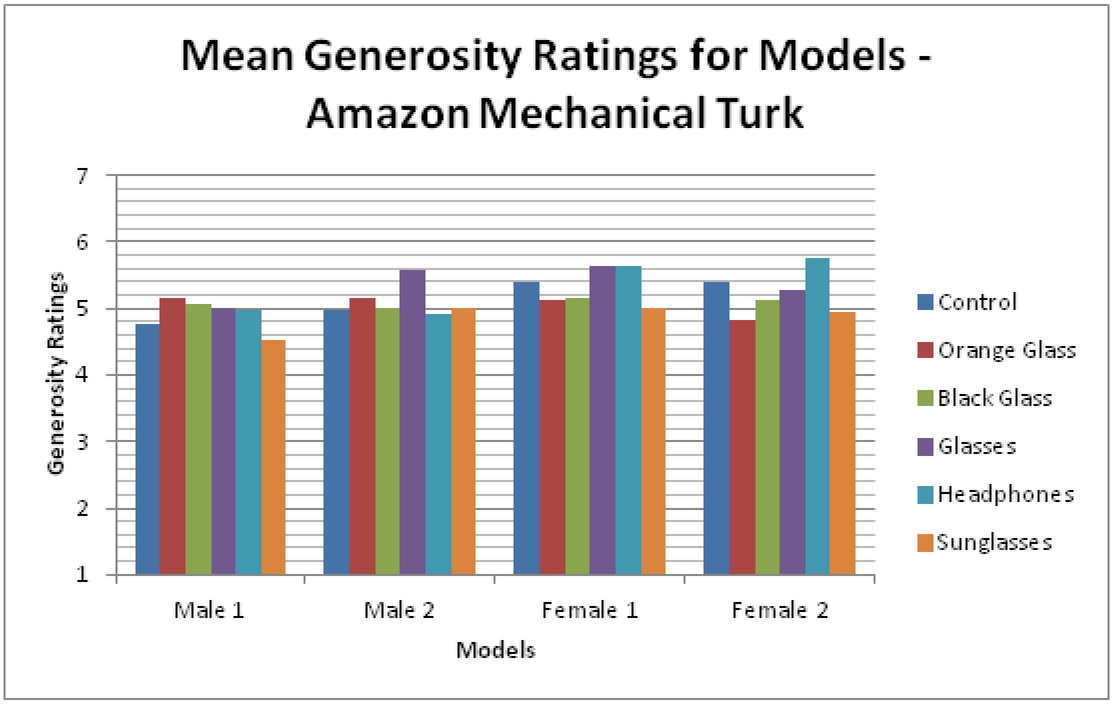

Figure 11: A graph showing the mean generosity ratings from the survey results

Judgment statements on the generosity of the models, as shown inFigure 11, only yielded significance in one model for the orange Google Glass, glasses, headphones, and sunglasses. The orange Google Glass, with a confidence level within 95% in Female 2's mean comparison, decreased this model's generosity. Glasses also having a 95% confidence, but within Male 2's comparisons, increased the impression of generosity. Both headphones and sunglasses found significance at a 90% confidence level within Female 2's mean comparison, headphones positively affecting generosity while sunglasses negatively affecting generosity.

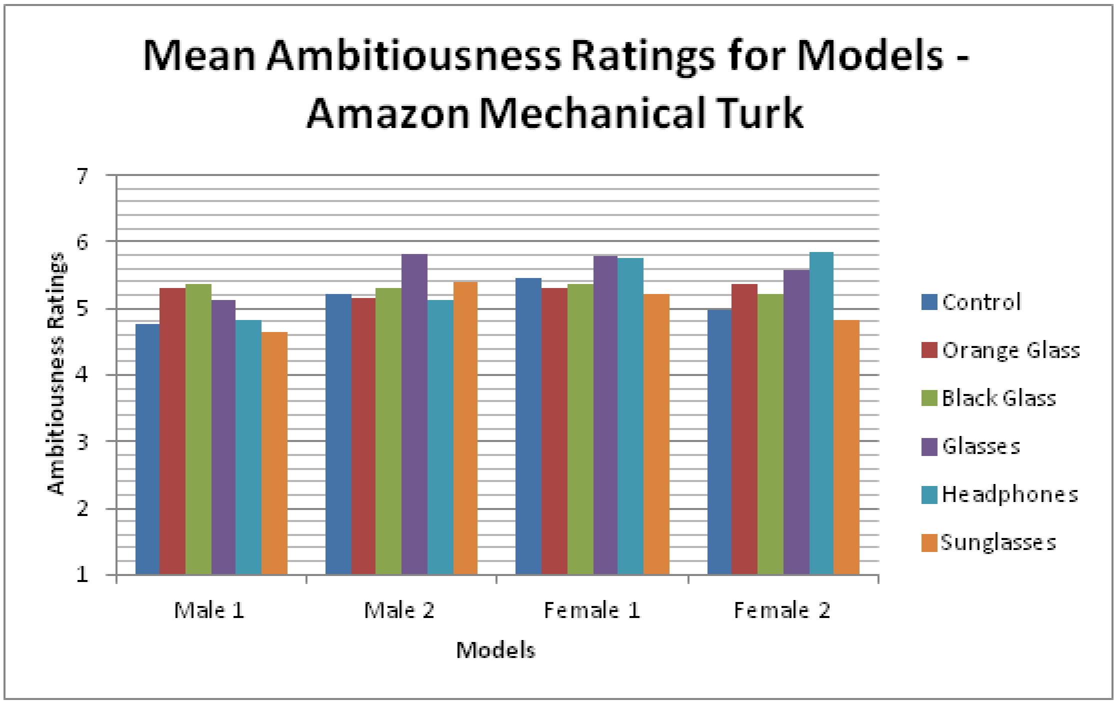

Figure 12: A graph showing the mean ambitiousness ratings from the survey results

Judgment statements on the ambitiousness of the models, as shown inFigure 12, havea 95% significance for Male 2 and Female 2 for glasses. The mean comparison seems to demonstrate that glasses increase the perceptions of ambitiousness in the wearers of glasses. This same observation was found for the orange model of Google Glass, but with a confidence level of 90% in Male 1 and Female 2. To some extent, this can be branched to the black model of Google Glass, as a 95% confidence interval for only Male 1's data was found, though as this observation was found in one model, the observation is not as strong as the orange Google Glass. Headphones also had a strong increase in ambitiousness within Female 2's mean comparisons, with a confidence level of 95%.

The same method of analysis for the Amazon Mechanical Turk population will be performed on the WPI data in that significance within a 90% or 95% confidence level on the changing means between control and a particular device will be observed. Two or more models experiencing this same significance results in a more conclusive impact on the first impression, while having less may indicate further study is required or does not correlate.

The WPI community responses for all four models and their devices reached a mean of 22.21 responses with a median of 22 and a standard deviation of 4.93. The number of responses for the WPI community is lower than that of the Amazon Mechanical Turk community as no incentives were provided to participants, and fewer methods of publicizing the study. The standard deviation also signifies an imbalance of responses between the models, unfortunately due to an error within the creation of the experiment that will be discussed later. The calculations for the one-tailed t tests as described above should account for these imbalances by requiring a greater change within data points to observe significance.

Figures 13 through 19 are graphs containing the mean ratings for each judgment statement, containing the same format as the graphs above for the Amazon Mechanical Turk community. The following data is representative of the WPI community.

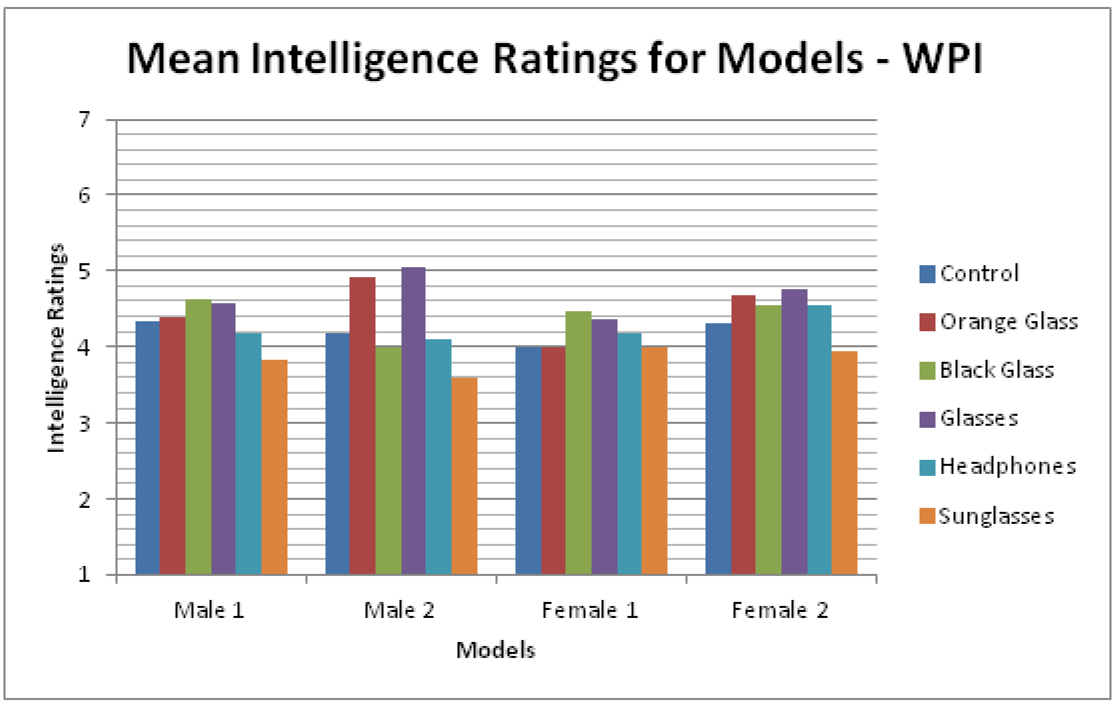

Figure 13: A graph showing the mean intelligence ratings from the survey results from the WPI population

For judgment statements involving intelligence, as shown in Figure 13, sunglasses were found to detract from the intelligence ratings of both male models at a 95% confidence level. Both female models did not receive consistent positive or negative results, nor did they have significant results which may indicate the impression judgments only affect males. Glasses also received two significant values in Male 2 and Female 2, though one of these values was within

90% confidence, both models received a noticeable increase in the intelligence rating. Both the black and orange Google Glass had received statistically significant results in at least one model, Male 2 and Female 1 respectively, both with an observed increase in intelligence. Nothing statistically significant was observed within the comparisons of control and headphones data.

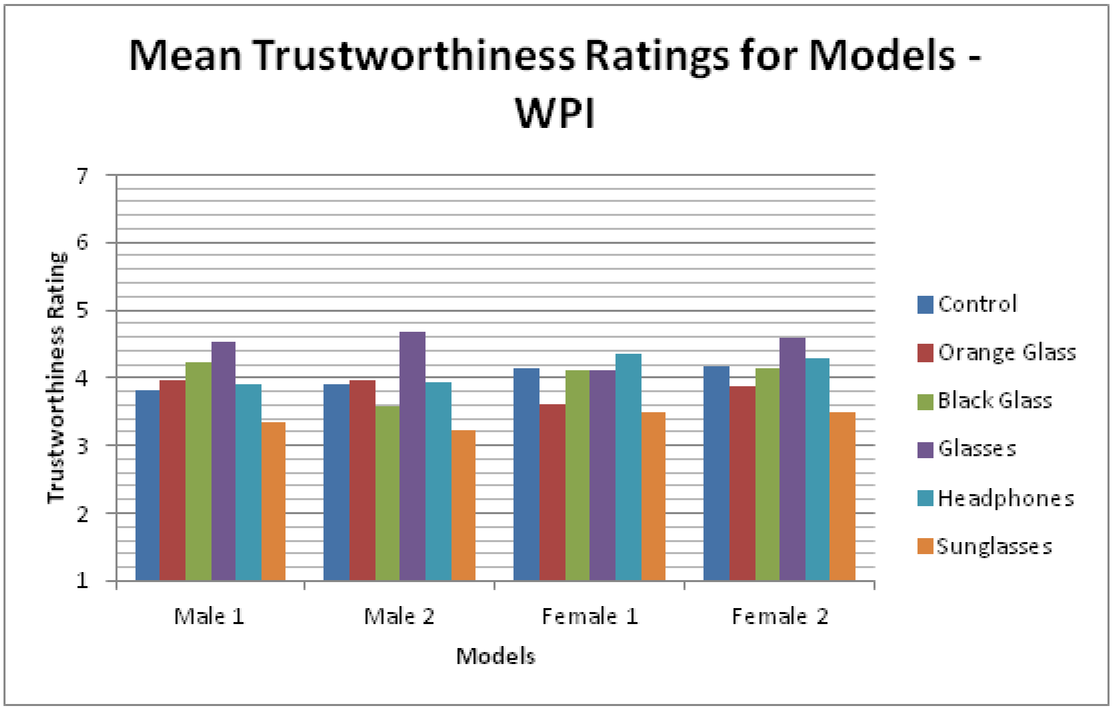

Figure 14: A graph showing the mean trustworthiness ratings from the survey results from the WPI population

For judgment statements involving trustworthiness, as shown in Figure 14, sunglasses were found to substantially diminish trustworthiness ratings across all four models at the 95% confidence level, strongly indicating that obstructing the face with a pair of sunglasses will make one appear less trustworthy. Glasses on the other hand, had a substantial increase in trustworthiness within a 95% confidence level in both male models and 90% confidence level within a Female 2. This could indicate a first impression against males with glasses will be more affected by glasses than females. Both models of Google Glass received only one statistically significant model in the 95% confidence level, Female 1 for the Orange Glass and Male 1 for the Black Glass, but the orange model had decreased trustworthiness while the black model had increased it. Nothing statistically significant was observed for headphones within this judgment.

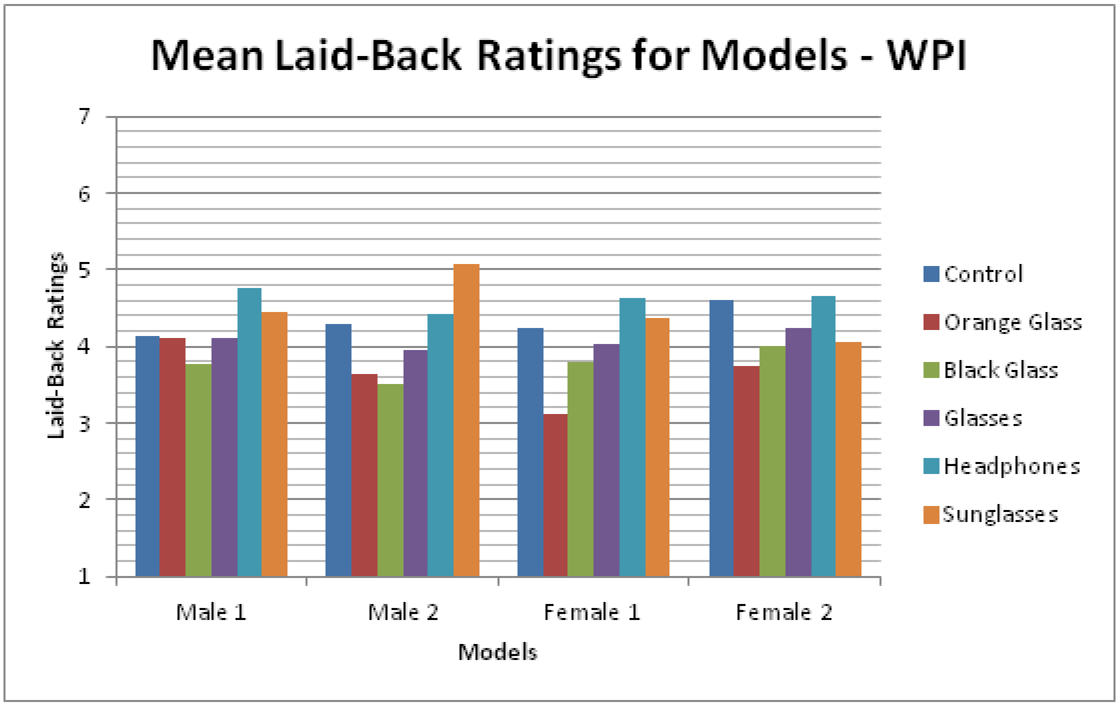

Figure 15: A graph showing the mean "laid-back" ratings from the survey results from the WPI population

For judgment statements involving relaxedness, or the trait of being laid-back, as shown in Figure 15, both Google Glass' were found to be statistically significant within a 95% confidence level in all of the models except for Male 1. Both models of Glass had decreased the perception of relaxedness within all of the models found significant, possibly indicating the wearers of Google Glass may not appear laid back to others. For headphones, Male 1 and Female 1 were found to be statistically significant within the 90% confidence level, both of which were found to have an increase in this judgment. Sunglasses were also found statistically significant within the 90% confidence level within Male 2 and Female 2, though the scores for this device were more erratic. This may indicate the effect sunglasses have on an individual's first impression may depend on how the device compliments the wearer's facial features. No significance was found among any of the models with respect to glasses on this judgment.

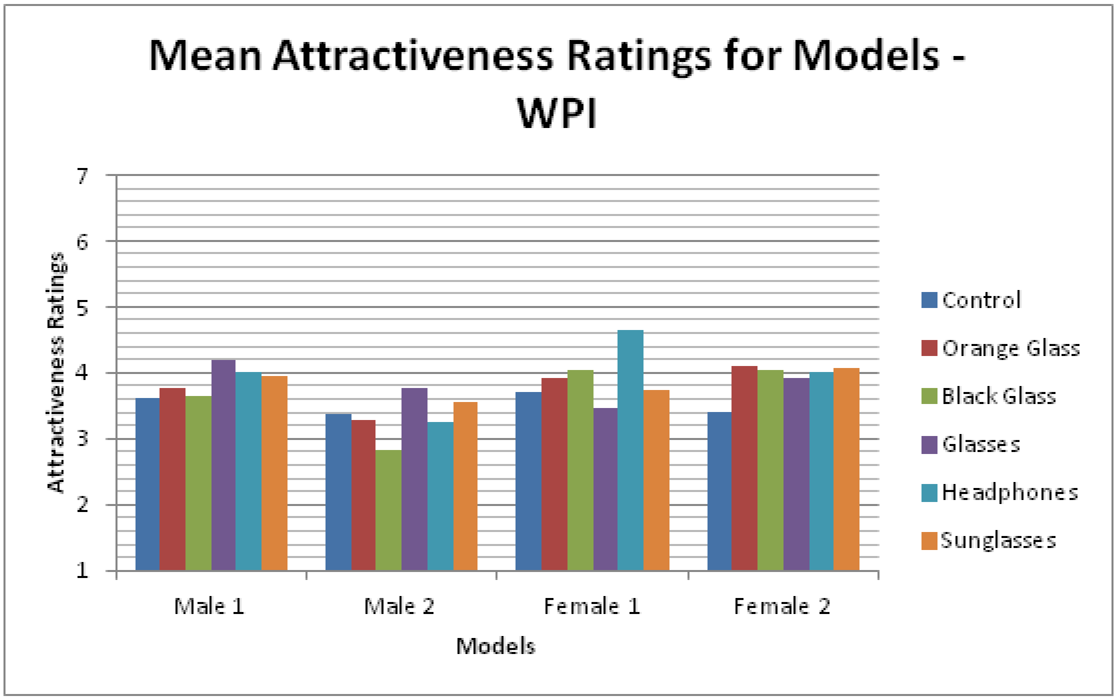

Figure 16: A graph showing the mean attractiveness ratings from the survey results from the WPI population

For judgment statements involving attractiveness, as shown in Figure 16, significance among all models except Female 1 was observed for the data on glasses within a 90% confidence interval. Within the WPI community, glasses appeared to increase the attractiveness of these three models. Two statistically significant changes within both female models in the headphones data was found within a 95% confidence level. Both models had received an increase in their attractiveness scores, which may indicate headphones increase attractiveness for females. Black Google Glass had two statistically significant data points within the 90% confidence level in Male 2 and Female 2, but both models had varying results. The male had a decrease in attractiveness, while the female model had increased attractiveness. Only one statistically significant data point was acquired for the orange Google Glass within the 95% confidence level on the same female as the black Glass, once again increasing the attractiveness. This could possibly mean Google Glass may increase attractiveness among females within this community, though more testing would be needed for more conclusive data. Sunglasses had only received one statistically significant change within the 95% confidence level found in Female 2, increasing attractiveness though significant changes within the other models were not found.

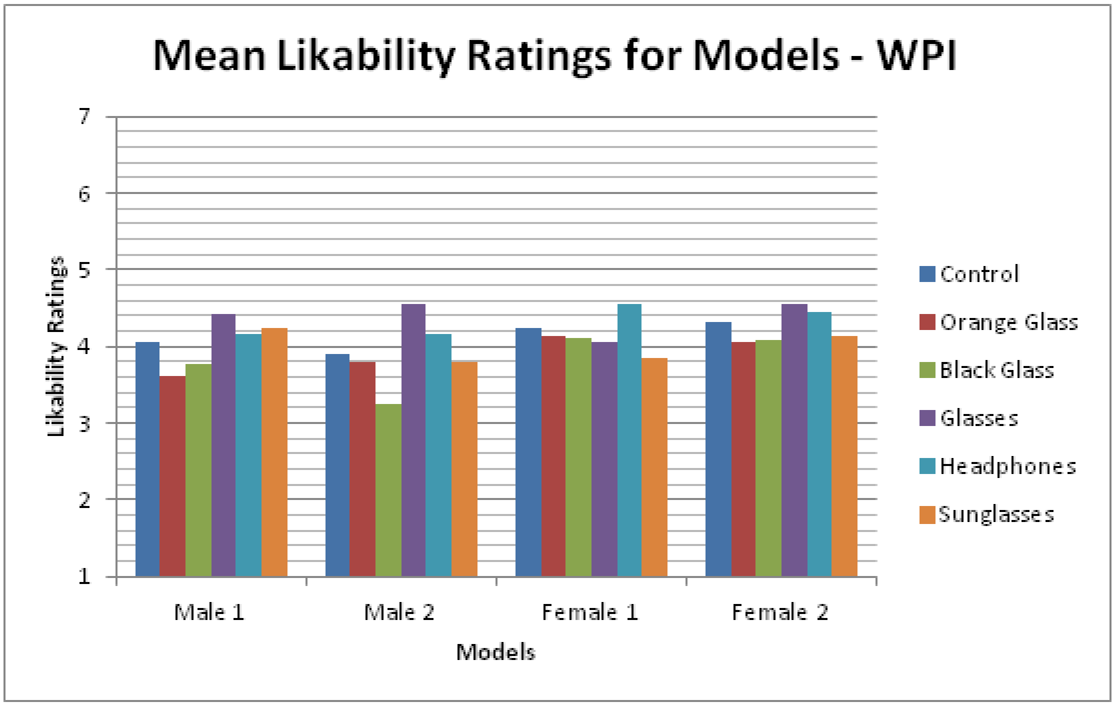

Figure 17: A graph showing the mean likability ratings from the survey results from the WPI population

For judgment traits involving likability, as shown in Figure 17, the data for glasses had observed two statistically significant changes within both male models at a 90% confidence level. Both data points increased the likability of the models, indicating the impressions of likability increases among males when wearing glasses. One statistically significant value, at the 95% confidence level in Male 2, was found for the black Google Glass while none were found for the orange Google Glass, making conclusions for the device difficult to ascertain. Only one statistically significant value at the 95% confidence level was found within sunglasses in Female 1, potentially decreasing users of sunglasses. Once again, as only one statistically significant value was found among the four, additional tests would be necessary for a stronger conclusion.

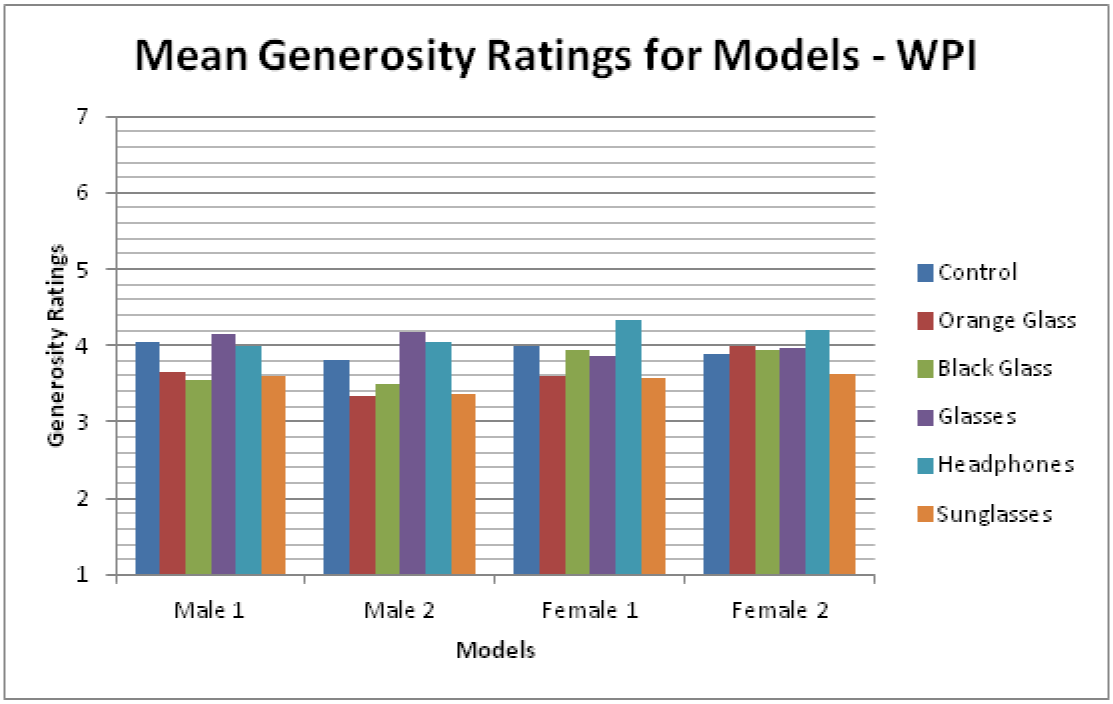

Figure 18: A graph showing the mean generosity ratings from the survey results from the WPI population

For judgment traits involving generosity, as shown in Figure 18, Male 2, Female 1, and Female 2 had statistically significant data points for sunglasses at the 95% confidence level. All three models had received a decreased score with the introduction of sunglasses, meaning the first impressions for generosity are negatively affected when wearing sunglasses. The orange Google Glass had received a similar effect within Male 1, Male 2, and Female 1, but within a 90% confidence level, and only Male 1l for the black Google Glass at a 95% confidence level. Male 1, Male 2, and Female 1 with statistically significant data points at a 90% confidence level for headphones were observed, but with a positive effect to their corresponding generosity scores. Only one statistically significant comparison was found for glasses in Male 2’s data.

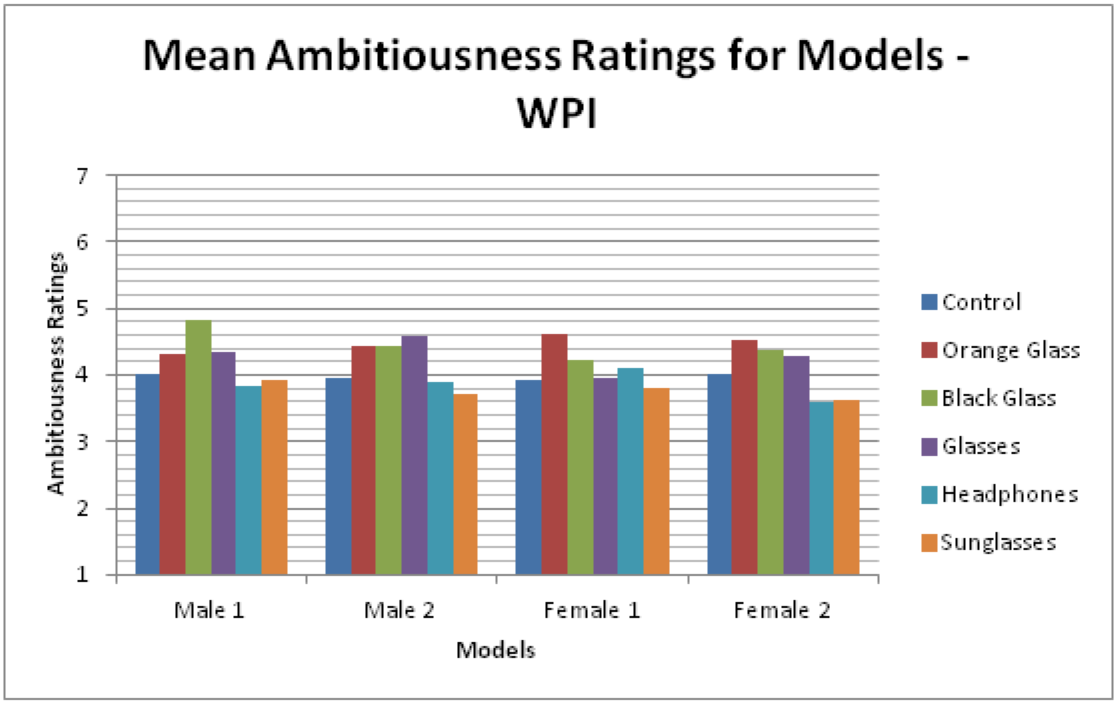

Figure 19: A graph showing the mean ambitiousness ratings from the survey results from the WPI population

For judgment traits involving ambitiousness, as shown in Figure 19, all models except Male 1 with statistically significant data points for the orange Google Glass were observed at 95% confidence level, all models receiving an increase in their ambitiousness scores. The same effect was observed on the black Google Glass model, though only Male 1 was found within 95% statistical significance with the other models coming only close to the 90% confidence level. Glasses had also received only one statistically significant model comparison in Male 2, increasing ambitiousness, though as this was not observed with significance among other models, more tests would be needed to affirm this effect holds true. Headphones had received only one statistically significant comparison at the 90% confidence level within Female 2, and sunglasses found no statistically significant data for this judgment trait.

4.2. Social Distance

The results for each of the participants who signed up for the Social Distance part of our study have been compiled in Appendix A. The total number of participants who signed up was 25 and this pool was split into four subsections, one for each device. Because of this split, the number of participants in each section can be very small, and the subsections get even smaller when gender is taken into account.

Figure20 and Figure 21 show the graphs of the distances that the participants stopped from the experimenter for each of the four walks and for the total average stop distance.

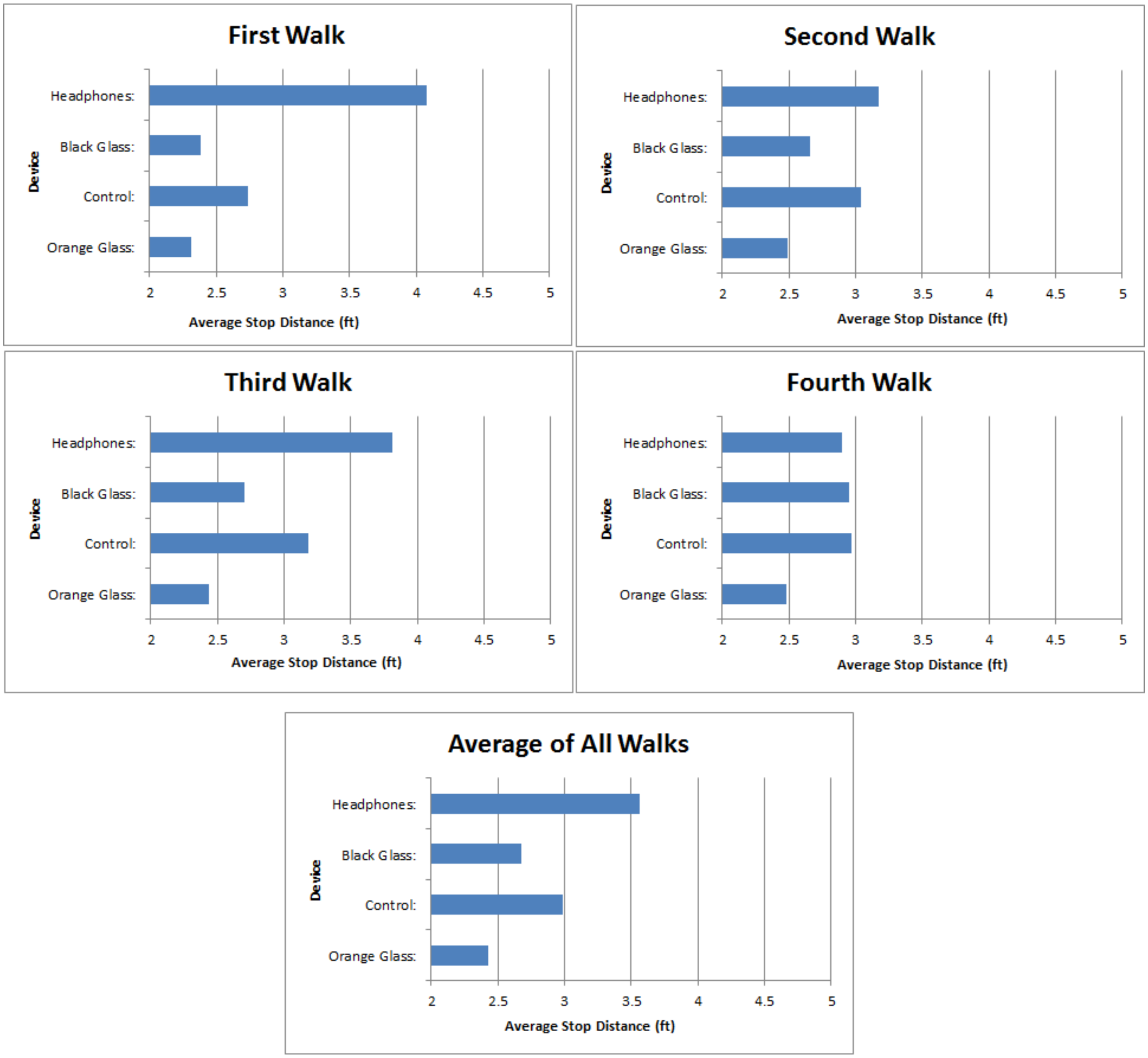

Figure 20: Graphs showing the average stop distance for each device during each walk.

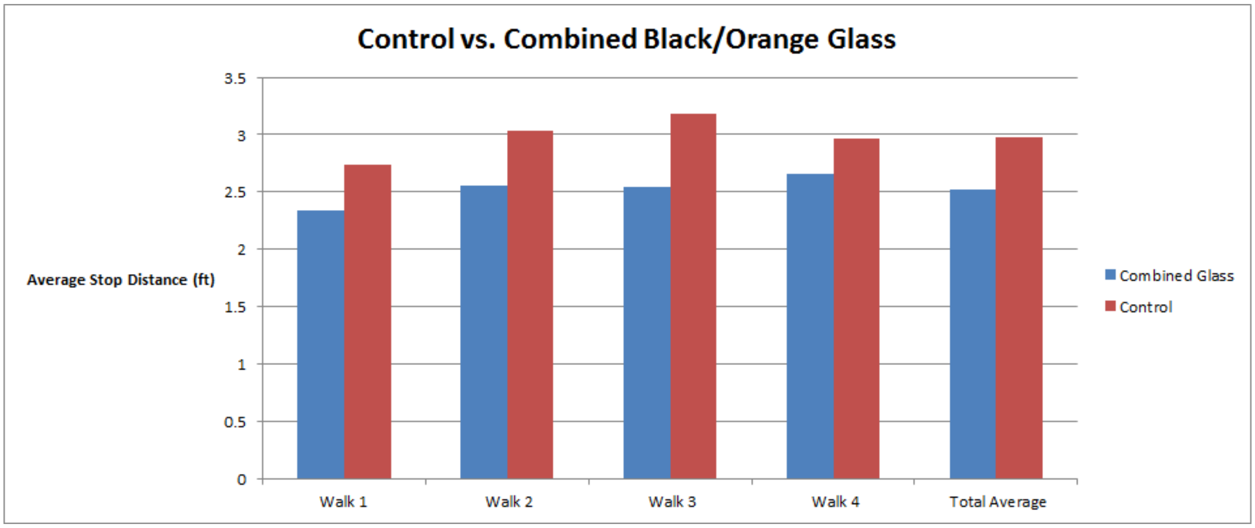

Figure 21: Graph showing the average stop distance when the participant pool for black and orange Google Glass were combined

From a glance, it seems that wearing black Google Glass and orange Google Glass tended to make participants stop closer than the control while wearing headphones made participants stop farther away. The combined averages of the stop distances when wearing black or orange Google Glass are lower than the distance for the control for each walk. The averages are still very close and the participant pool was relatively small, so we ran T-Tests on the data to determine its strength. The confidence level that we decided to aim for was 95% confidence in order to draw conclusions from our data.

Walk 1:	Average: Walk 1	Median: Walk 1	STDEV: Walk 1	Samples:	T-Test Lookup:	**Confidence:**
Orange Glass:	2.31	2.11	0.617	8	0.9982863906	**82.165%**
Control:	2.74	2.46	0.977	7		
Black Glass:	2.38	2.48	0.318	4	0.9022933908	**78.329%**
Headphones:	4.08	4.63	1.613	5	-1.651304206	**91.299%**

Figure 22: The descriptive statistics and T-Test results for the first walk in the stop distance experiment

Total:	Total Average:	Total Median:	Total STDEV:	Samples:	T-Test Lookup:	Confidence:
Orange Glass:	2.43	2.16	0.715	8	1.160783599	**85.509%**
Control:	2.98	2.55	1.067	7		
Black Glass:	2.67	2.58	0.481	4	0.6584689488	**72.137%**
Headphones:	3.56	3.56	1.390	3	-0.6492116417	**70.860%**

Figure 23: The descriptive statistics and T-Test results for the average of all of the walks in the stop distance experiment

Figure22 and Figure 23 show the results from the T-Test in bold. The number displayed in the confidence column shows how confident we are that the differences in stop distances between each device and the control are significant. These are the numbers that we compared to our confidence level threshold when drawing our conclusions.

Walk 1:	Average: Walk 1	Median: Walk 1	STDEV: Walk 1	Samples:	T-Test Lookup:	Confidence:
Combined Glass:	2.34	2.21	0.507	12	1.014602054	**82.527%**
Control:	2.74	2.46	0.977	7		

Figure 24: The descriptive statistics and T-Test results for the first walk in the stop distance experiment when data for the orange and black Google Glass was combined

Total:	Total Average:	Total Median:	Total STDEV:	Samples:	T-Test Lookup:	Confidence:
Combined Glass:	2.52	2.37	0.641	12	1.034191474	**82.954%**
Control:	2.98	2.55	1.067	7		

Figure 25: The descriptive statistics and T-Test results for the average of all walks in the stop distance experiment when data for the orange and black Google Glass was combined

The same tests were applied in Figures 24 and 25. In these figures, the descriptive statistics shown are for the combine data of both black Google Glass and orange Google Glass.

By combining the data, we were able to get a data set and make generalizations about the device, regardless of its color.

After performing the analysis, we discovered that none of the results reached our threshold of 95% confidence. Despite this, the less significant results that we discovered are as follows:

- 91% confident that headphones cause people to approach less that the control on their first walk.
- 78% confident that black Google Glass causes people to approach more than the control on their first walk.
- 72% confident that black Google Glass causes people to approach more than the control in the average of all walks.
- 70% confident that headphones cause people to approach less than the control in the average of all walks.
- 82% confident that either pair of Google Glass causes people to approach more than the control in both their first walk and as an average of all of their walks.

5. Discussion

The design of the experiment being based on a previous experiment done in person has led to a few flaws within our first impressions experiment. Our unfamiliarity with online recruiting tools and survey software also led to some of these shortcomings. It is our hope that by listing these oversights, any work or studies performed deriving methodologies from this experiment will be able to avoid repeating these mistakes.

5.1. First Impressions

Designing an online based experiment from an experiment previously performed in person where several environmental variables can be controlled with the use of an experimenter leading participants through the experiment was the first flaw in the design of this experiment. Ideally, a fixed amount of time when displaying the photographs would have been able to control the factor of confidence, and the study would then be able to test whether a device may impact the confidence in certain judgments. Another issue with time lies in that a participant decides how long they wish to view the photo, which if viewed in the 100ms range may cause a faulty first impression, or if viewed for far too long may affect the first impression by having participants rethink their original judgments. Willis and Todorov's experiment along with Bar, Neta, and Linz's experiment proved first impressions within 100ms to 1700ms are similar in judgment, but participants may decide to view the photo for longer than that period of time. The speed of the answers may also affect the participant's judgments, as Willis and Todorov emphasized quick results and eliminated responses significantly above the average response time. The survey software used to create the experiment was able to capture the total amount of time a participant would take to complete the survey, but individual blocks were not recorded nor could the time be accurately correlated with a participant who is thinking about his answers as opposed to a participant on a slower connection.

Amazon Mechanical Turk (AMT) is a rather cheap method of obtaining valuable survey data, as the rating system and certain constraints on participant approval allow for users of the service to find subjects that consistently provide good data. The first attempt at obtaining data used Amazon's Masters Qualification program, which allowed only Amazon Master level contributors to perform the study. While this ensures good results, accounts that have been granted this qualification are a very miniscule fraction of the AMT population, and the requirements needed to be granted a Masters Qualification are not known. Using this program, the number of participants performing the experiment was low with an estimated time of two to three weeks to fulfill a 150 participant quota. To ensure quicker results, we opted towards using the standard service, but setting the qualifications to perform the experiment so that users who have not completed at least 5,000 HITs and do not have an approval rating of more than 98% would be unable to participate in the experiment. While setting these constraints ensured quicker yet reliable results, unfortunately a large majority of the users who completed the survey were accounts from India. The standard AMT constraints service does allow for region specific blocking, but as we were unfamiliar with the user base and workings of AMT, the online survey data was not representative of the global population as planned, but is now representative of the population in India.

The experiment was designed through Qualtrics, making an extensive use of blocks as a photo with a page break followed by the question of participant model familiarity and then the seven judgment questions. These blocks were then pseudo randomly organized into six different sets, which a randomizer within the survey flow would pick one set to display to a participant while evenly presenting the sets. This arrangement worked well for the AMT experiment, as there each AMT worker was expected to complete the entire survey as it was determined to be unlikely for a worker to recognize one of the students. This was not the case for the experiment when performed for the WPI population. The students were asked whether they knew the model in the photograph and if a participant answered yes, the judgment questions were skipped. Since the six sets were not randomly generated by Qualtrics, any blocks skipped within the set would still appear as the participant answering the entire set, leaving data for some of the models underrepresented when compared to others. Another issue raised through the use of Qualtrics that possibly led to an uneven distribution of respondents in both versions of the survey was due to partial survey results being recorded by Qualtrics. The default option for unfinished surveys on Qualtrics is to close and record unfinished results. As authentication on the AMT experiment gave participants a confirmation code to later receive compensation, unfinished surveys were not an issue within that population. On the other hand an overwhelming number of WPI students, roughly 86 of the 215 surveys, had dropped out of the survey or submitted partial data. It is possible that when displaying a set to a participant who had dropped out, the counter responsible for evenly presenting the blocks was not decremented.

The last few mistakes within the experiment lie within the availability of adequate models for the photographs, and a few mistakes when planning the photographs. Initially, a total of ten models were to be photographed. Of those ten, four would be selected for being ideal representations of a face without facial obstructions as described in the methodology section. Unfortunately, due to an unforeseen blizzard and haphazard rescheduling, a small portion of the original models could attend and the remaining models necessary to have an even gender distribution had to be found within the University Library at that moment. Some of these models did not completely fit the criteria. Having to find willing models on a short notice also caused one participant to have a neutral colored shirt, but with a slight pattern that could possibly be distracting to participants performing the experiment, though it may also be out of sight to participants. Another possible effect on the data may be one of the model's hairs getting slightly messier as devices were being switched, though this was dismissed early in the experiment as being inconsequential.

5.2. Social Distance

The social distance experiment required live test subjects in a controlled testing environment. This limited our participant pool by availability and also came with several challenges that you would expect to see in human interaction experiments. We discuss our results and the data analysis below.

5.2.1. Strength of Data

After performing some analysis of the data in the form of One-Tailed T Tests, we found that the confidence in our data was not very high. Even with a 90% confidence level threshold, the only conclusion we could draw was that the first approach to an experimenter wearing headphones had a larger stop distance than that of the control. With an 80% confidence threshold, we can conclude that orange Google Glass causes people to stop closer than the control does for both the first walk and the average of all walks. Even though the combined average for the two Google Glass colors was lower in the graphs, the strength of the data was too weak to conclude anything significant.

The T Test results of the data split between male and female participants was significantly less conclusive. Some of the samples only had a single participant after splitting the data up this much. As a result, we could not look at how gender played into the stop distances of the participants.

One important detail about the participant pool is that it was made up entirely of students at Worcester Polytechnic Institute. Even within this population, our sample was not extremely representative, as the study was only advertised to students in psychology classes and Computer Science, Interactive Media and Game Development and Robotics Engineering majors.

5.2.2. Challenges

One of the biggest challenges we faced was getting people to approach the experimenter. Originally we had the experimenter facing the door and about twenty feet away from the entrance. When the subject came in, they immediately saw the researcher standing attentively, and so they had no reason to approach them. We tried several other variations to see how we could force the subject to approach, and the way that worked the best was to have the researcher's attention be focused on something else, like a laptop. The next iteration involved the researcher sitting at a desk which was facing the wall, with their back to the door the participant would enter in. This proved unwieldy, as the participant would try and get the researcher's attention by tapping them on the shoulder or yelling from the door. We determined that it would be necessary to both instruct the participant that they had to approach in addition to giving them a clear avenue for approach. Even when the desk was turned to be at a 45 degree angle with the door, participants would consistently approach parallel to the laptop screen. Our final design involved the researcher at a 90 degree angle to the door, facing the laptop screen. We instructed the participant to approach until they were comfortable, and the researcher only changed their focus to the participant after they were hailed, physically or verbally.

One important factor in the approach distance we had not considered was the location of the participants' belongings. Because we ran the study in an academic building during the day, most of the participants' had backpacks with them. Without explicitly telling them what to do with their belongings, participants handled this in varying ways. One participant left her bag in the room with the experimenter and her approach path was skewed as she tended to stay closer to her bag for her walks. Other participants didn't want to set their bags down and walked all four times with their bags on their backs. This may have skewed the data as well, dissuading the participant from walking farther even if there was a more comfortable stop distance. The best way to solve this issue is to explicitly tell participants set their bags down at the door before starting.

A recurring issue we faced was with impatient participants. Because most of our population came from volunteers on a social science participant pool that gave class credit to volunteers, a lot of these volunteers were anxious to get out with as little work as possible simply to finish their requirement. On a number of occasions the experimenter would explain the entire study and the participant would nod the entire time, but when it was time to start they would stare back at the experimenter in blank confusion. This could also factor into the stop distance but it is impossible to say for sure.

Another challenge we faced involved the camera used for taking measurements. The batteries on the camera would die after only a few participants. The number of batteries used was not a problem, but on two separate occasions the batteries died in the middle of an important run. This caused us to lose some of our already limited supply of student volunteers. The fact that we had to climb up to the rafters and take the camera down to add the batteries also meant that the camera was not in the same position every time. There were a few runs where the view on the camera was a little skewed, making it harder to see and count the floor tiles properly. The tripod holding the camera to the rafters also slowly fell as the runs went on due to the weakening duct tape. Given better equipment, the images of the participants could have been much more consistent.

6. Conclusion

The conclusions we can draw from the collected data have varying levels of confidence. The two studies each had their own problems due to time constraints, limited resources and a few mistakes made along the way. After twenty one weeks, we have honed two methodologies that could be run again in the future over a larger, more representative sample to produce better results. While we do have a few solid preliminary results, we hope this paper will lay a solid groundwork for how social situations are affected by the presence of an augmented reality device.

With our results from the survey, we were able to draw a few conclusions. Within our Amazon Mechanical Turk population, the data showed that black Google Glass makes people in general seem more intelligent while orange Google Glass makes people seem more ambitious. Both colors of Google Glass make men seem more trustworthy and makes women seem less trustworthy. The population of students at Worcester Polytechnic Institute mostly had different results. The data showed that orange Google Glass made people seem more ambitious and less generous. Both colors of Google Glass made people seem less laid back. The confidence levels of the conclusions drawn from the other questions on the survey were not high enough to meet our threshold, meaning future experiments may be able to get more meaningful results if a larger participant pool is used.

While the data that was collected in the social distance study may not have a lot of strength to back it up, the averages do make it seem that wearing Google Glass causes others to have a closer stop distance. The only conclusion that met our stretch confidence level was that orange Google Glass causes people to approach the experiment more closely. Our biggest problem in this study was getting a large, representative sample for our participant pool. Future iterations should be able to reach more concrete conclusions with a larger sample.

Other than gaining access to a larger sample size, there are other ways to improve the results from the studies. The quality of the camera used for the social distance study was sub par and the fact that there was no battery indicator meant the camera would die during subjects' runs. This caused us to lose data from a few participants. Using a room solely for the experiment is also important so that materials and furniture is not moved and rearranged. For the online survey portion, it is important to restrict the location of the study so that participants don’t all come from the same country in order to get a more representative population.

The results that were calculated and the groundwork of the study are important in predicting how this new technology will affect social interactions in the world. Google Glass is slowly becoming more readily available for consumers while other devices, such as smart watches, are growing in popularity. While our preliminary data shows a heightened interest in the devices, more studies are always helpful in tracking the social effect of these devices in an evolving world.

7. References

[1] L. Adams and D. Zuckerman. The effect of lighting conditions on personal space requirements. *J. Gen. Psychol. 118(4),* pp. 335-340. 1991.

[2] K. Aoki and E. J. Downes.An analysis of young people's use of and attitudes toward cell phones.*Telematics Inf. 20(4),* pp. 349-364. 2003. . DOI:http://dx.doi.org/10.1016/S0736-5853(03)00018-2.

[3] M. Bar, M. Neta and H. Linz. Very first impressions.*Emotion 6(2),* pp. 269. 2006.

[4] C. Bjørnskov. On the determinants of honesty perceptions in the united states. *Rationality and Society 24(3),* pp. 257-294. 2012. . DOI: 10.1177/1043463112453546.

[5] K. Bodine and F. Gemperle.Effects of functionality on perceived comfort of wearables.Presented at Proceedings of the Seventh IEEE International Symposium on Wearable Computers (ISWC'03). 2003, .

[6] D. B. Bugental, C. Johnston, M. New and J. Silvester.Measuring parental attributions: Conceptual and methodological issues.*Journal of Family Psychology 12(4),* pp. 459-480. 1998. . DOI: 10.1037/0893-3200.12.4.459.

[7] C. F. Camerer and E. Fehr, "Measuring Social Norms and Preferences using Experimental Games: A Guide for Social Scientists," *MacArthur Foundation Anthropology,* 2002.

[8] R. M. Crane .Social distance and loneliness as they relate to headphones used with portable audio technology. 2005.

[9] K. Cumiskey, "Surprisingly, nobody tried to caution her: Intentionality and the role of social responsibility in the public use of mobile phones." *Mobile Communications: Re-Negotiation of the Social Sphere,* pp. 225-225-236, 2005.

[10] J. Delhey, K. Newton and C. Welzel. How general is trust in "Most people"? solving the radius of trust problem. *American Sociological Review 76(5),* pp. 786-807. 2011. . DOI: 10.1177/0003122411420817.

[11] M. Dufwenberg and U. Gneezy.Measuring beliefs in an experimental lost wallet game.*Games Econ. Behav. 30(2),* pp. 163-182. 2000. . DOI:http://dx.doi.org/10.1006/game.1999.0715.

[12] B. W. Fitzpatrick and B. Collins-Sussman ,*Team Geek.* O'Reilly Media, 2012.

[13] E. L. Glaeser, D. I. Laibson, J. A. Scheinkman and C. L. Soutter.Measuring trust.*The Quarterly Journal of Economics 115(3),* pp. 811-846. 2000. . DOI: 10.1162/003355300554926.

[14] T. Hamamura. Social class predicts generalized trust but only in wealthy societies. *Journal of Cross-Cultural Psychology 43(3),* pp. 498-509. 2012. . DOI: 10.1177/0022022111399649.

[15] K. N. Hampton and N. Gupta. Community and social interaction in the wireless city: Wi-fi use in public and semi-public spaces. *New Media & Society 10(6),* pp. 831-850. 2008.

[16] L. Humphreys. Cellphones in public: Social interactions in a wireless era. *New Media & Society 7(6),* pp. 810-833. 2005.

[17] S. Kassin, S. Fein and H. R. Markus, *Social Psychology.* Belmont, CA 94002-3098: WADSWOTH CENGAGE Learning, 2011.

[18] G. Kortuem and Z. Segall. Wearable communities: Augmenting social networks with wearable computers. *Pervasive Computing, IEEE 2(1),* pp. 71-78. 2003.

[19] P. M. Leonardi. Problematizing "new media": Culturally based perceptions of cell phones, computers, and the internet among united stateslatinos. *Critical Studies in Media Communication 20(2),* pp. 160-179. 2003. Available:http://dx.doi.org/10.1080/07393180302778. DOI: 10.1080/07393180302778.

[20] D. M. Lloyd, A. Coates, J. Knopp, S. Oram and S. Rowbotham. Don't stand so close to me: The effect of auditory input on interpersonal space. *Perception 38(4),* pp. 617. 2009.

[21] R. E. Nisbett and L. Ross.*Human Inference: Strategies and Shortcomings of Social Judgment* 1980.

[22] Z. O'Donoghue. "Friend me": The impacts of technology on human interaction.

[23] R. Overå. Networks, distance, and trust: Telecommunications development and changing trading practices in ghana. *World Dev. 34(7),* pp. 1301-1315. 2006. . DOI:http://dx.doi.org/10.1016/j.worlddev.2005.11.015.

[24] T. Pierce. Social anxiety and technology: Face-to-face communication versus technological communication among teens. *Comput. Hum. Behav. 25(6),* pp. 1367-1372. 2009.

[25] S. Porter, L. England, M. Juodis, L. ten Brinke and K. Wilson. Is the face a window to the soul? investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement 40(3),* pp. 171. 2008.

[26] J. Rico and S. Brewster. Usable gestures for mobile interfaces: Evaluating social acceptability. Presented at Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010, .

[27] P. E. Tetlock. Accountability and the perseverance of first impressions.*Soc. Psychol. Q.* pp. 285-292. 1983.

[28] R. Wei and L. Leung. Blurring public and private behaviors in public space: Policy challenges in the use and improper use of the cell phone. *Telematics Inf. 16(1),* pp. 11-26. 1999.

[29] M. Williams, O. Jones and C. Fleuriot. Wearable computing and the geographies of urban childhood: Working with children to explore the potential of new technology. Presented at Proceedings of the 2003 Conference on Interaction Design and Children. 2003, .

[30] J. Willis and A. Todorov. First impressions making up your mind after a 100-ms exposure to a face.*Psychological Science 17(7),* pp. 592-598. 2006.

[31] L. A. Zebrowitz and J. M. Montepare.Impressions of babyfaced individuals across the life span.*Dev. Psychol. 28(6),* pp. 1143. 1992.

Appendix A. Raw Data for Stop-Distance Test

Averaged Data:	Participant Number:	Walk 1:	Walk 2:	Walk 3:	Walk 4:
Orange Glass:	101	2.79	3.28	3.69	4.18
Control:	102	2.83	3.86	4.75	4.24
Black Glass:	103	2.76	2.79	3.20	3.24
Headphones:	104	4.63	3.75	null	null
Orange Glass:	105	2.21	2.35	2.56	2.35
Control:	106	2.96	4.70	4.78	3.95
Control:	107	2.13	2.21	2.03	2.10
Black Glass:	108	2.19	2.29	2.00	2.39
Headphones:	109	1.87	1.83	1.83	1.58
Orange Glass:	110	3.58	4.21	3.35	3.04
Orange Glass:	111	1.84	1.87	1.87	1.84
Black Glass:	112	2.48	2.71	2.63	3.27
Headphones:	113	3.01	3.09	3.48	3.50
Orange Glass:	114	2.00	2.00	2.00	2.00
Orange Glass:	115	1.82	2.42	2.13	2.42
Orange Glass:	116	2.44	2.16	2.16	2.16
Orange Glass:	117	1.83	1.61	1.78	1.85

Headphones:	118	5.04	4.03	3.79	3.62
Control:	119	2.00	1.70	1.71	2.07
Black Glass:	120	1.94	2.12	2.33	2.51
Control:	121	2.03	1.91	2.58	2.22
Headphones:	122	5.86	null	6.16	null
Control:	123	2.46	2.54	2.57	2.50
Control:	124	4.78	4.34	3.85	3.71
Black Glass:	125	2.52	3.37	3.37	3.37

Appendix B. Descriptive Statistics from the Amazon Mechanical Turk Community

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness	#Respondents
Male 1 - Control	Mean	4.92	4.31	4.38	4.96	5.15	4.77	4.77	28
	Median	5	4.5	4.5	5	5.5	5	5	
	Std. Dev	1.23	1.258	1.525	1.455	1.377	1.243	1.032	
Male 1 - Orange Glass	...	5.14	4.79	4.57	5.04	4.96	5.14	5.29	27
	...	5	5	6	5	5	5	5.5	
	...	1.177	1.315	1.317	1.138	1.17	1.079	1.084	
Male 1 - Black Glass		5.39	5.14	4.46	4.82	4.86	5.07	5.36	28
		6	5	5	5.5	5	5.5	6	
		1.066	1.145	1.503	1.701	1.533	1.514	1.193	
Male 1 - Glasses		5.44	4.93	4.07	4.89	5.04	5	5.11	27
		5	5	4	5	5	5	5	
		0.892	1.107	1.072	1.155	1.16	1.144	1.013	
Male 1 - Headphones		5.07	4.56	4.48	4.93	5.11	4.96	4.81	27
		5	5	4	5	5	5	5	
		1.035	1.251	1.122	1.385	1.013	1.018	1.001	
Male 1 - Sunglasses		4.44	3.72	3.84	4.8	4.76	4.52	4.64	25
		5	4	4	5	5	5	5	
		1.121	1.339	0.987	1.384	1.422	1.295	1.319	

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness	#Respondents
Male 2 - Control	Mean	5.26	5.22	4.44	5.11	5.07	4.96	5.22	27
	Median	5	5	4	5	5	5	5	
	Std. Dev	1.023	1.188	1.188	1.396	1.207	1.16	1.013	
Male 2 - Orange Glass	...	5.52	5.22	4.37	5.19	5.33	5.15	5.15	27
	...	6	6	5	6	6	5	6	
	...	1.312	1.423	1.573	1.469	1.301	1.406	1.379	
Male 2 - Black Glass		5.48	5.04	4.59	5.11	5	5	5.3	27
		6	5	5	5	5	5	5	
		1.122	1.16	0.971	1.155	1.144	1.24	1.068	
Male 2 - Glasses		5.88	5.52	4.52	5.76	5.72	5.56	5.8	25
		6	6	5	6	6	6	6	
		1.092	1.262	1.122	1.128	1.061	1.261	1.354	
Male 2 - Headphones		5.08	4.92	4.58	4.69	5	4.92	5.12	26
		6	5	5	5	5	5	5	
		1.383	1.294	1.391	1.289	1.131	1.23	1.211	
Male 2 - Sunglasses		5.18	4.82	4.93	5.54	5.25	5	5.39	28
		5.5	5	5	6	6	5	6	
		1.307	1.442	1.359	1.401	1.206	1.361	1.066	

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness	#Respondents
Female 1 - Control	Mean	5.32	5.24	4.68	5.44	5.64	5.4	5.44	25
	Median	5	5	5	6	6	6	6	
	Std. Dev	0.748	1.3	1.406	1.158	1.186	1.258	1.294	
Female 1 - Orange Glass	...	5.3	5.04	4.52	5.33	5.19	5.11	5.3	27
	...	5	5	5	6	5	5	5	
	...	1.031	1.126	1.156	1.387	1.111	0.847	1.031	
Female 1 - Black Glass		5.62	5.35	4.5	5.27	5.19	5.15	5.35	26
		6	5	5	5	5	5	6	
		0.852	0.892	1.449	1.151	1.096	1.156	1.231	
Female 1 - Glasses		5.86	5.61	5.18	5.43	5.71	5.64	5.79	28
		6	6	6	6	6	6	6	
		0.891	0.916	1.492	1.23	1.117	0.951	0.995	
Female 1 - Headphones		5.63	5.74	4.48	5.96	5.93	5.63	5.74	27
		6	6	5	6	6	6	6	
		0.926	1.023	1.578	1.372	0.781	1.006	0.984	
Female 1 - Sunglasses		5.44	5.19	4.48	5.44	5.67	5	5.22	27
		5	5	5	6	6	5	5	
		1.086	1.21	1.221	1.625	1	1.144	1.05	

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness		#Respondents
Female 2 - Control	Mean	5.11	5.82	5.04	5.18	5.36	5.39	4.96		28
	Median	5	6	5	5	5	5.5	5		
	Std. Dev	1.257	1.124	1.427	1.278	0.911	1.133	1.261		
Female 2 - Orange Glass	...	5.27	5.08	4.58	4.65	5.19	4.81	5.35		26
	...	5	5	5	5	5	5	5		
	...	1.041	1.055	1.137	1.263	1.096	0.981	0.797		
Female 2 - Black Glass		5.41	5.15	4.3	5.07	5.33	5.11	5.22		27
		5	5	4	5	5	5	6		
		1.01	0.949	1.564	1.299	1.109	1.281	1.219		
Female 2 - Glasses		5.74	5.67	4.48	5.07	5.85	5.26	5.56		27
		6	6	5	5	6	6	6		
		1.196	0.961	1.626	1.639	0.989	1.289	1.34		
Female 2 - Headphones		5.32	5.44	4.8	5.84	5.96	5.76	5.84		25
		6	6	5	6	6	6	6		
		1.108	1.325	1.291	0.987	0.889	0.926	1.068		
Female 2 - Sunglasses		4.81	4.56	4.74	5	4.96	4.93	4.81		27
		5	5	5	5	5	5	5		
		1.145	1.05	1.13	1.074	0.808	1.107	1.178		

Appendix C. Descriptive Statistics from the WPI Community

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness		#Respondents
Male 1 - Control	Mean	4.33	3.81	4.14	3.62	4.05	4.05	4		21
	Median	4	4	4	4	4	4	4		
	Std. Dev	0.913	0.814	1.195	1.244	1.071	0.74	1		
Male 1 - Orange Glass	...	4.38	3.96	4.12	3.77	3.62	3.65	4.31		26
	...	4	4	4	4	4	4	4		
	...	1.602	1.371	1.423	1.505	1.388	1.263	1.35		
Male 1 - Black Glass		4.62	4.23	3.77	3.65	3.77	3.54	4.81		26
		4.5	4	4	4	4	4	5		
		0.804	0.587	1.032	1.164	1.142	1.14	0.981		
Male 1 - Glasses		4.57	4.52	4.1	4.19	4.43	4.14	4.33		21
		4	4	4	4	4	4	4		
		0.926	0.814	0.768	0.814	0.746	0.655	0.796		
Male 1 - Headphones		4.17	3.92	4.75	4	4.17	4	3.83		12
		4	4	5	4	4	4	4		
		0.389	0.515	0.965	0.603	0.835	0	0.389		
Male 1 - Sunglasses		3.84	3.36	4.44	3.96	4.24	3.6	3.92		25
		4	4	4	4	4	4	4		
		0.746	0.995	1.193	1.098	1.012	0.913	1.038		

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness		#Respondents
Male 2 - Control	Mean	4.19	3.9	4.29	3.38	3.9	3.81	3.95		21
	Median	4	4	4	4	4	4	4		
	Std. Dev	0.873	0.7	1.231	1.117	0.944	0.68	1.161		
Male 2 - Orange Glass	...	4.92	3.96	3.65	3.27	3.81	3.35	4.42		26
	...	5	4	4	4	4	4	4		
	...	0.796	0.774	1.018	1.041	1.021	0.936	0.857		
Male 2 - Black Glass		4	3.58	3.5	2.83	3.25	3.5	4.42		12
		4	4	3.5	3	3	4	4.5		
		1.206	0.996	1.087	1.03	0.866	1	1.165		
Male 2 - Glasses		5.05	4.68	3.95	3.77	4.55	4.18	4.59		22
		5	4.5	4	4	4	4	4		
		0.999	0.995	1.046	0.685	0.671	0.664	0.854		
Male 2 - Headphones		4.11	3.95	4.42	3.26	4.16	4.05	3.89		19
		4	4	4	4	4	4	4		
		0.994	0.524	1.017	1.098	0.501	0.405	0.459		
Male 2 - Sunglasses		3.6	3.24	5.08	3.56	3.8	3.36	3.72		25
		4	3	5	4	4	4	4		
		1.19	1.128	1.352	1.325	1.384	1.036	1.275		

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness		#Respondents
Female 2 - Control	Mean	4.32	4.16	4.6	3.4	4.32	3.88	4		25
	Median	4	4	4	4	5	4	4		
	Std. Dev	1.249	1.214	1.041	1.19	1.108	0.971	1.225		
Female 2 - Orange Glass	...	4.68	3.89	3.74	4.11	4.05	4	4.53		19
	...	4	4	4	4	4	4	4		
	...	0.946	0.994	1.046	1.049	0.705	0.471	0.697		
Female 2 - Black Glass		4.55	4.14	4	4.05	4.09	3.95	4.36		22
		4.5	4	4	4	4	4	4		
		0.739	0.56	0.816	1.174	0.811	0.653	1.002		
Female 2 - Glasses		4.76	4.59	4.24	3.93	4.55	3.97	4.28		29
		5	4	4	4	5	4	4		
		0.83	0.825	0.83	1.334	1.088	0.823	0.882		
Female 2 - Headphones		4.56	4.3	4.67	4	4.44	4.19	3.59		27
		4	4	4	4	4	4	4		
		0.847	0.953	1.038	1.074	0.801	0.681	0.971		
Female 2 - Sunglasses		3.94	3.5	4.06	4.06	4.13	3.63	3.63		16
		4	4	4	4	4	4	4		
		0.998	0.894	1.34	1.124	1.025	0.885	0.957		

Model		Intelligence	Trustworthiness	Laid-Back	Attractiveness	Likability	Generosity	Ambitiousness		#Respondents
Female 1 - Control	Mean	4	4.14	4.25	3.71	4.25	4	3.93		28
	Median	4	4	4	4	4	4	4		
	Std. Dev	0.903	0.803	0.645	1.15	0.752	0.77	0.858		
Female 1 - Orange Glass	...	4	3.6	3.13	3.93	4.13	3.6	4.6		15
	...	4	4	3	4	4	4	5		
	...	1.069	0.91	0.915	1.335	1.187	0.91	1.056		
Female 1 - Black Glass		4.47	4.11	3.79	4.05	4.11	3.95	4.21		19
		4	4	4	4	4	4	4		
		0.964	0.658	0.787	1.177	0.459	0.524	0.787		
Female 1 - Glasses		4.37	4.11	4.04	3.48	4.07	3.85	3.96		27
		4	4	4	4	4	4	4		
		1.418	1.219	1.224	1.341	1.328	0.949	1.4		
Female 1 - Headphones		4.18	4.36	4.64	4.64	4.55	4.32	4.09		22
		4	4	4	5	4	4	4		
		0.795	0.727	0.902	1.177	0.963	0.716	0.426		
Female 1 - Sunglasses		4	3.5	4.36	3.75	3.86	3.57	3.79		28
		4	4	4	4	4	4	4		
		0.943	0.694	1.162	1.005	0.891	0.836	0.738		